SACRO CUORE DON CALABRIA

I R C C S

ECI CANCER CENTRE

SCUOLA DI METODOLOGIA DELLA RICERCA CLINICA

2025 - 11ª EDIZIONE

MODULI SPECIALISTICI - S3

VENERDÌ 11 - SABATO 12 APRILE 2025

NEGRAR DI VALPOLICELLA (VR)

Centro Formazione IRCCS "Sacro Cuore - Don Calabria"

Classificazione degli studi osservazionali: descrittivi *Vs* analitici

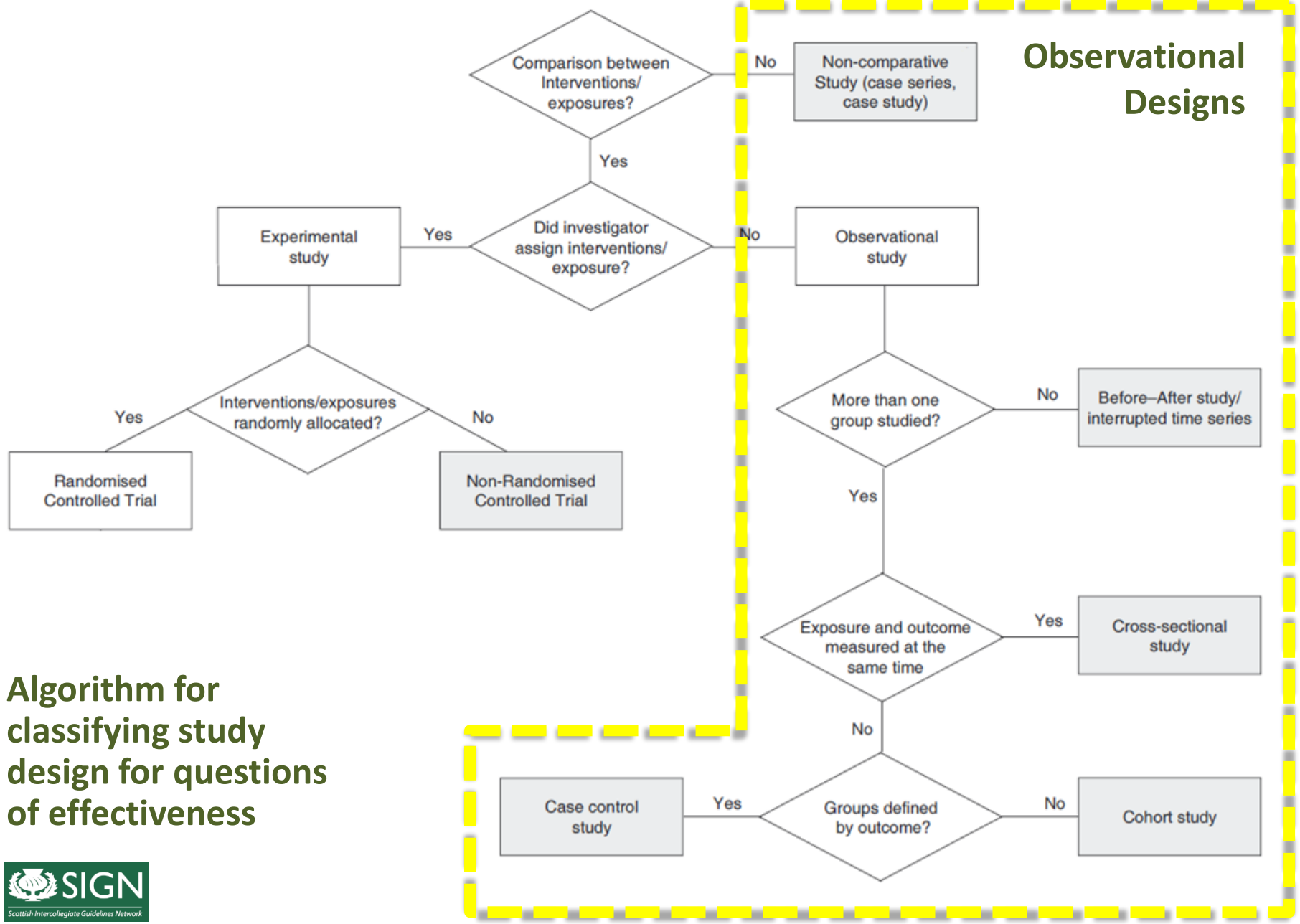
(G.L. Pappagallo)

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

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

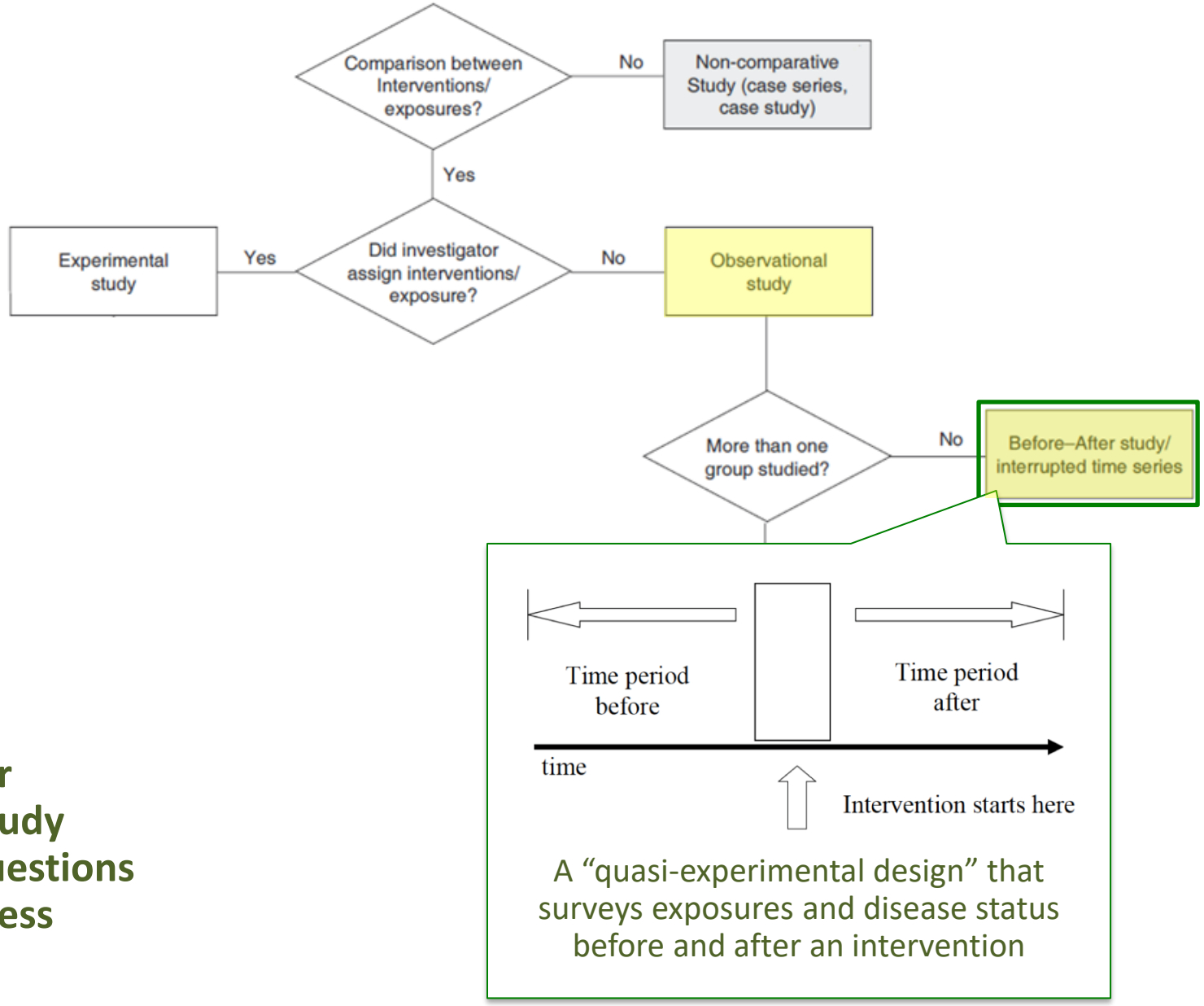
Identify the incidence of a particular outcome over time

Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions



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

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

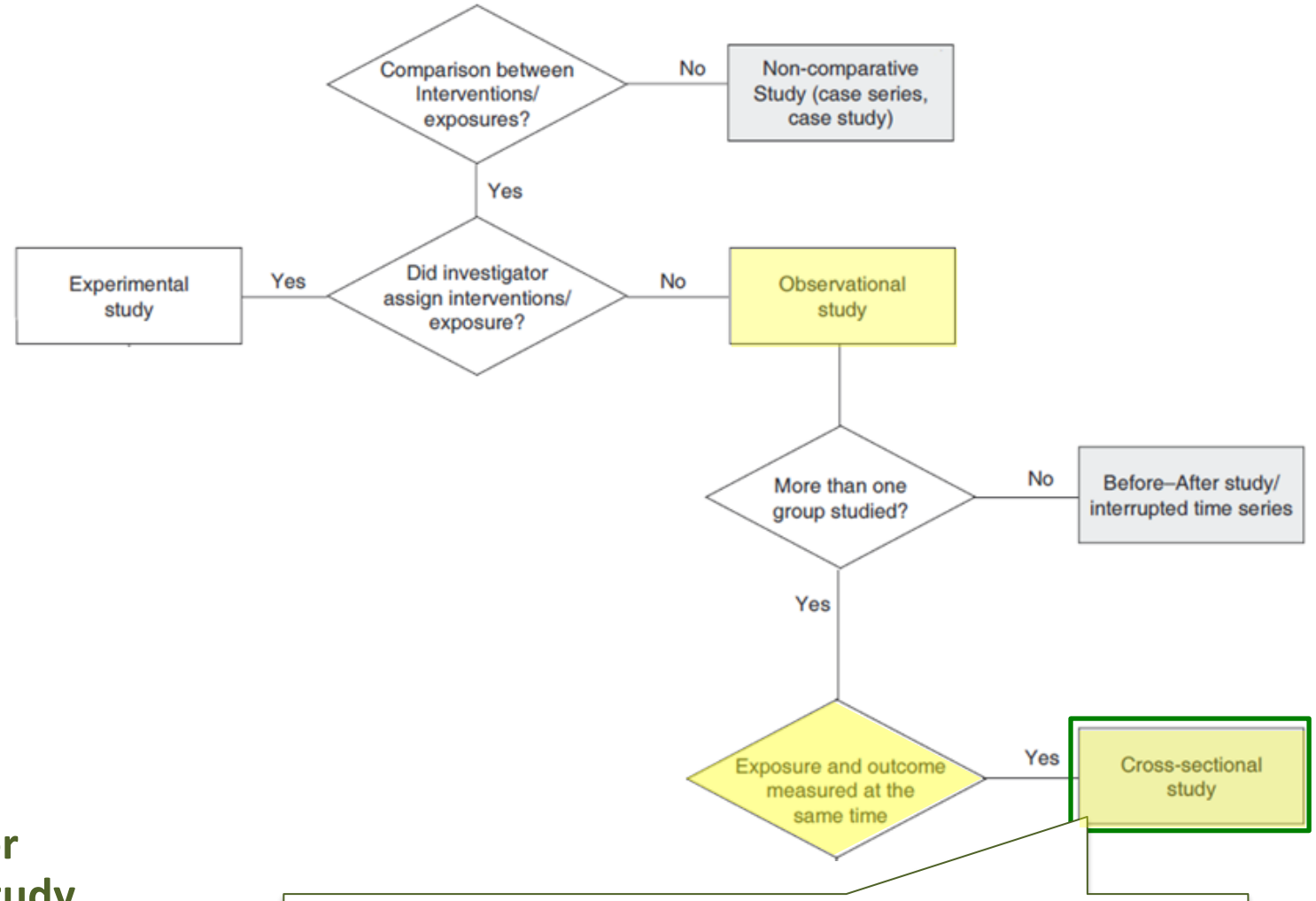
Non-comparative case series

Report outcomes of patients who received a specific intervention

Comparative case series

Compare outcomes between patients who received different interventions

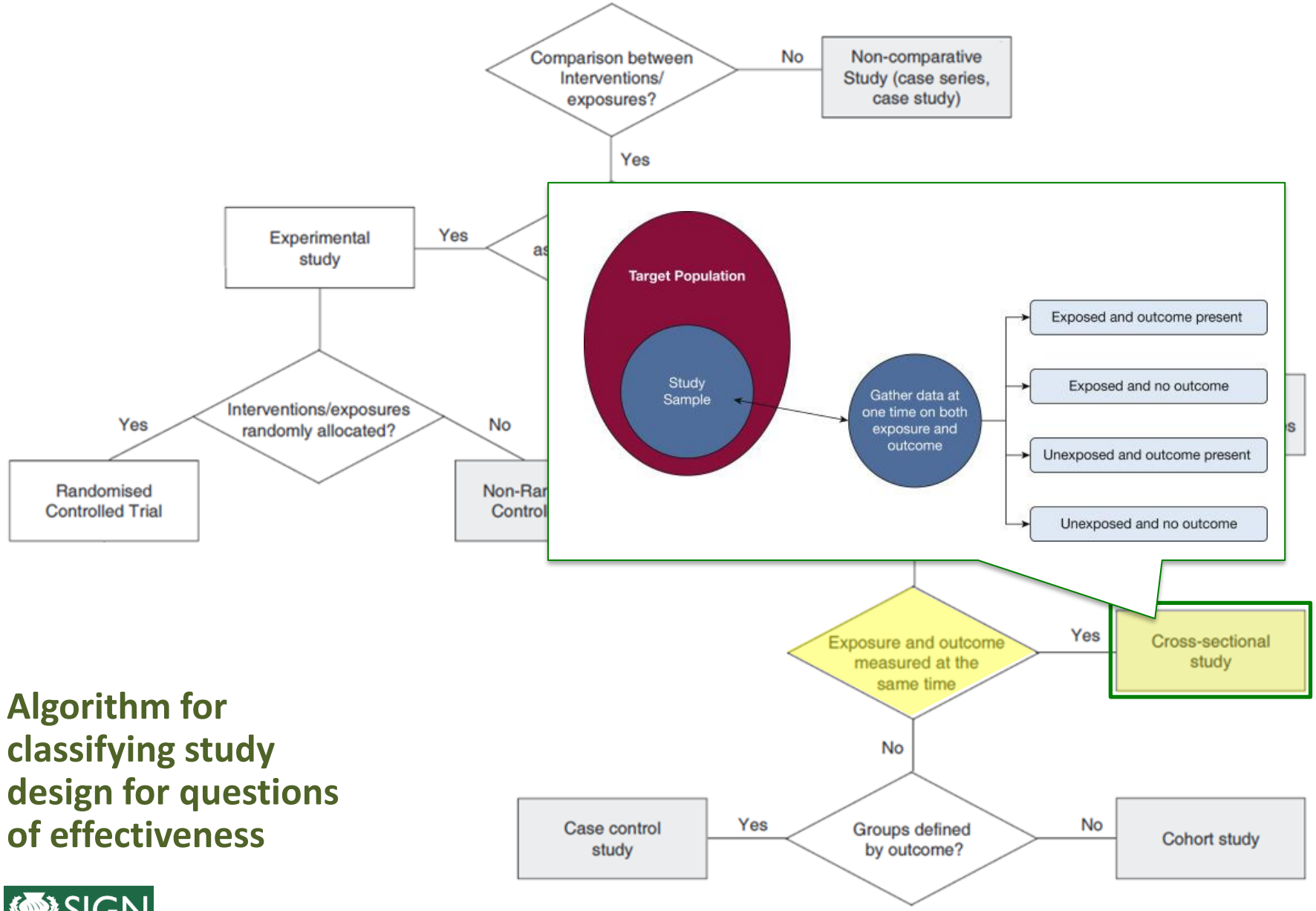
Algorithm for classifying study design for questions of effectiveness



Subjects selected irrespective of the presence or absence of the characteristics of interest. 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

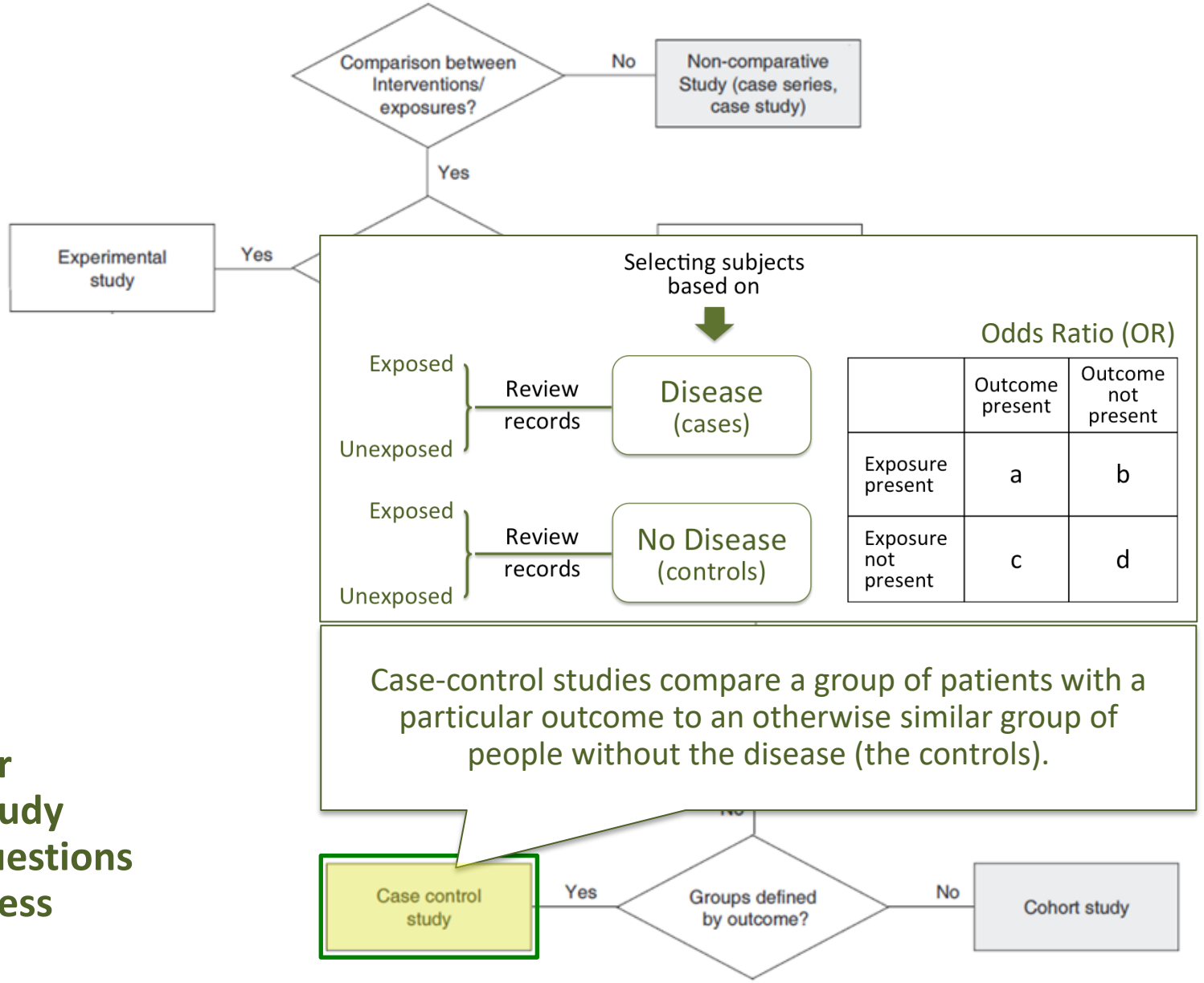
Non-comparative case series

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

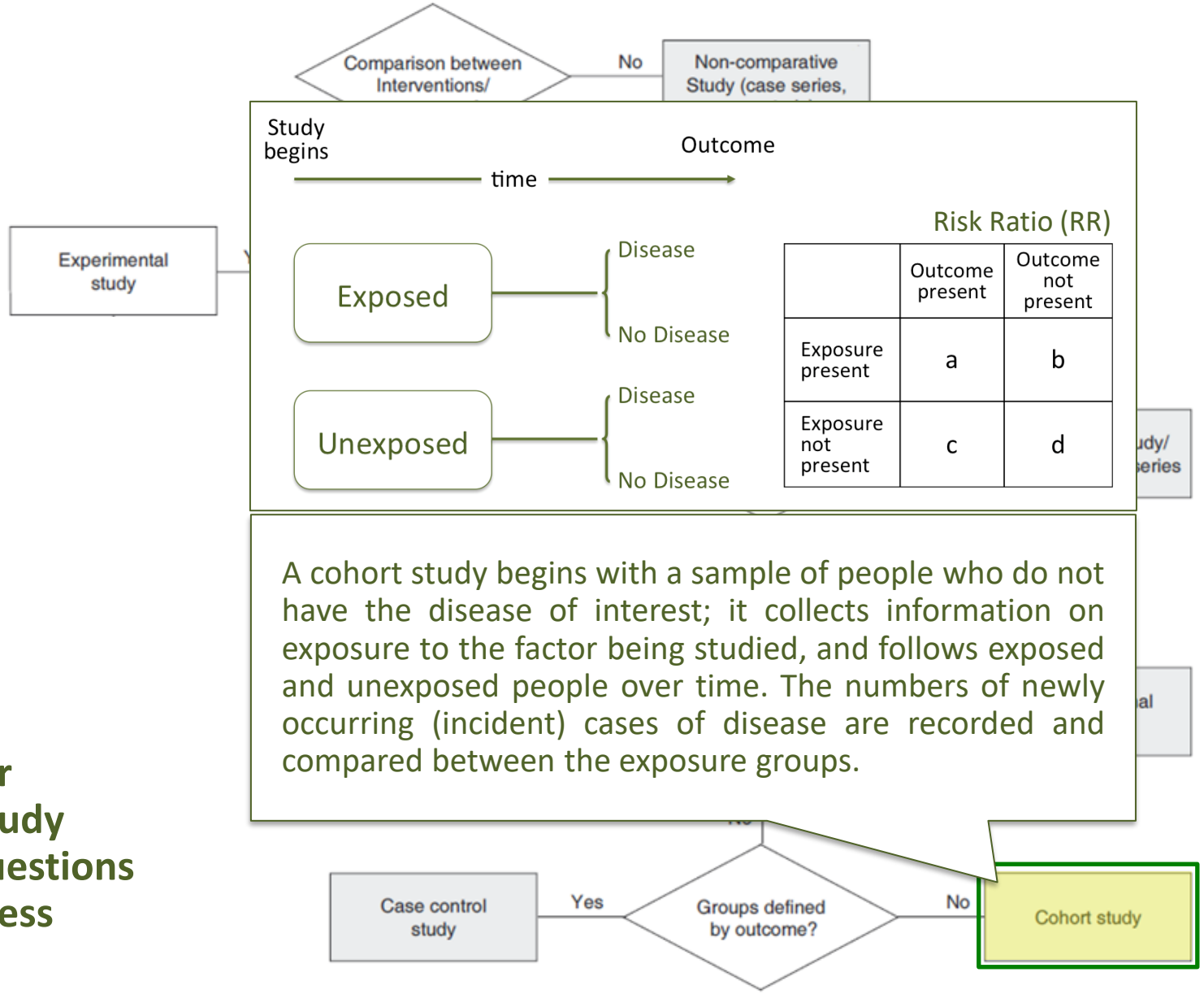
Non-comparative case series

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



SIGN
Scottish Intercollegiate Guidelines Network

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?

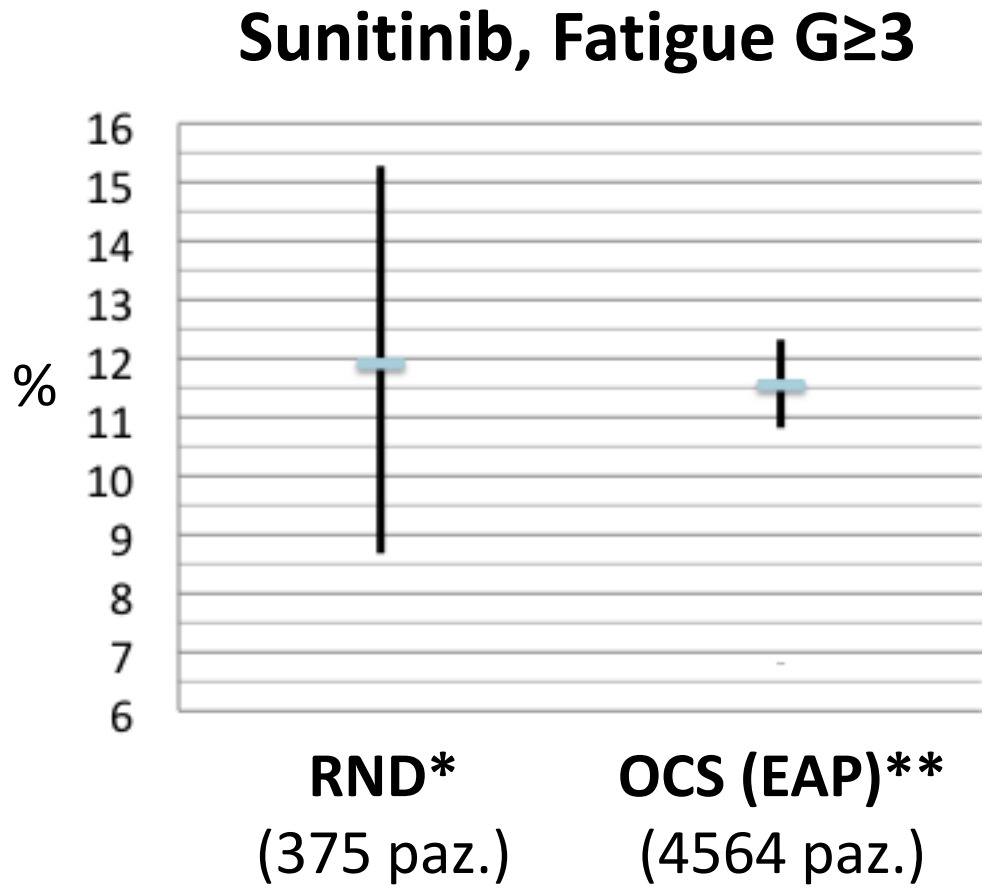
An OCS is an analysis of a group of individuals who have specific features in common and who are followed over a defined period of time.

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]

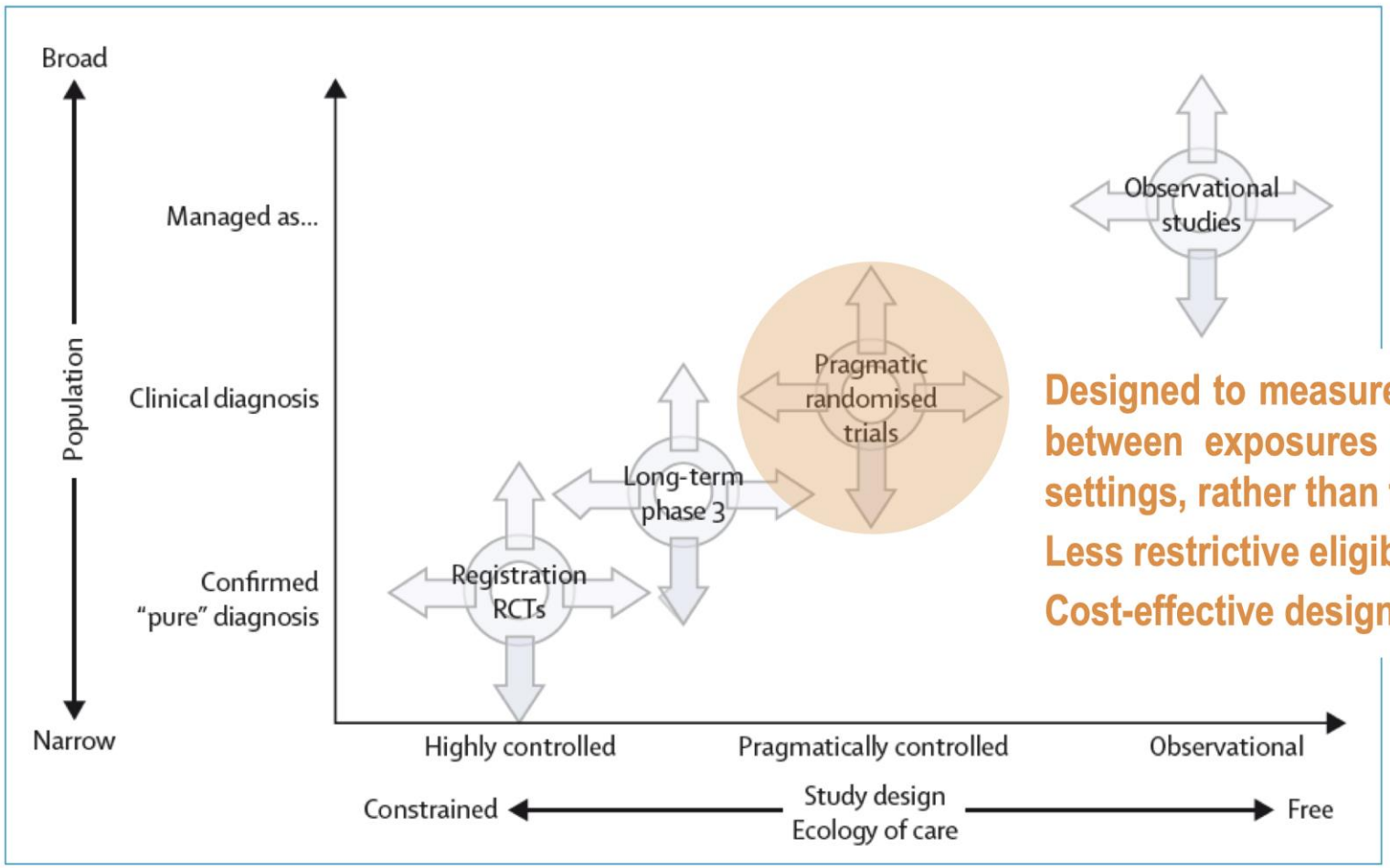
To some, the best evidence may be viewed as research that minimizes bias to the greatest extent possible, while others may prioritize research that is deemed most pertinent to clinical practice.

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

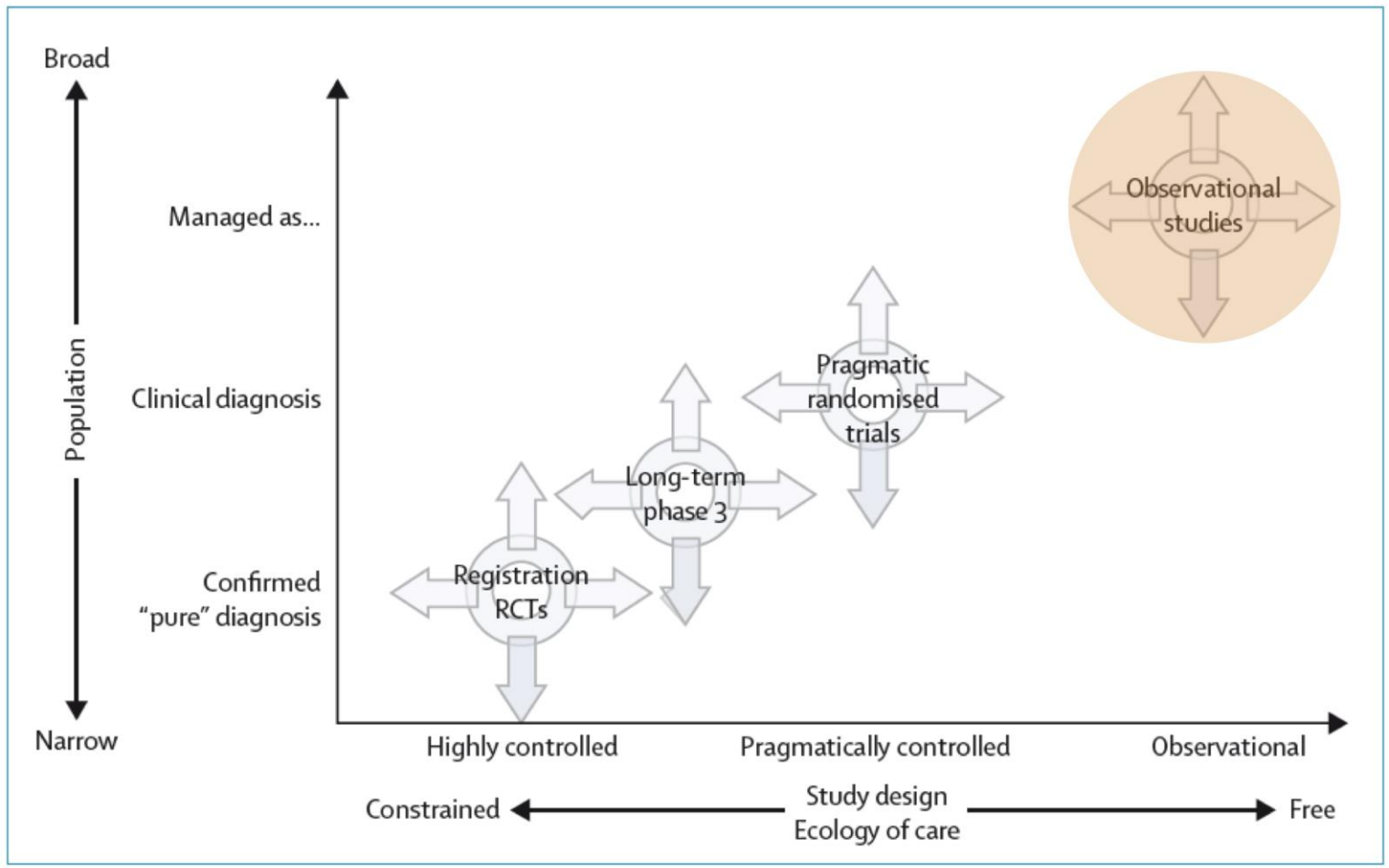
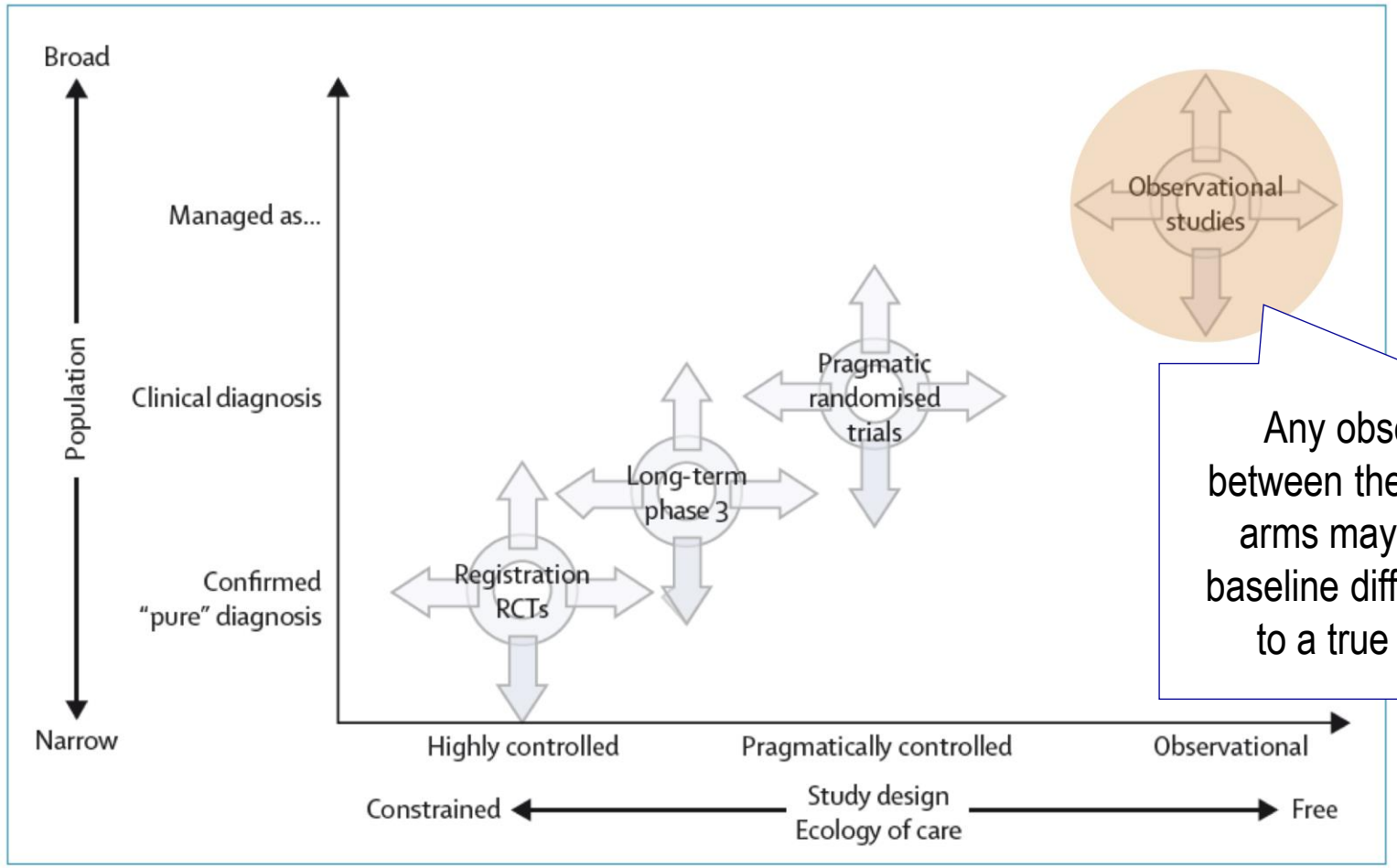


Figure 1: A conceptual framework for therapeutic research

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.

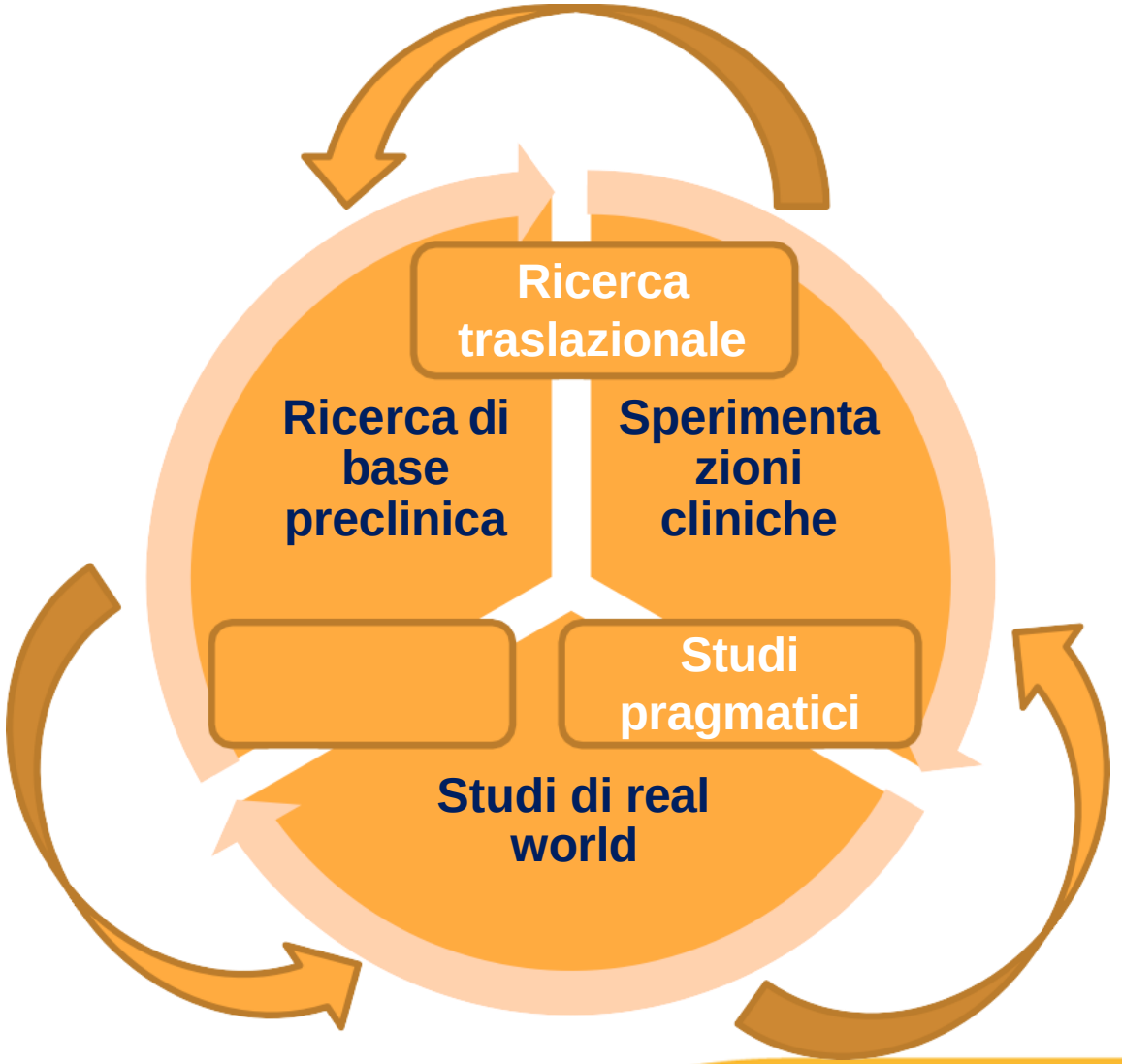
Potential Solutions:

Careful Study Design
Appropriate Statistical Methods

Figure 1: A conceptual framework for therapeutic research

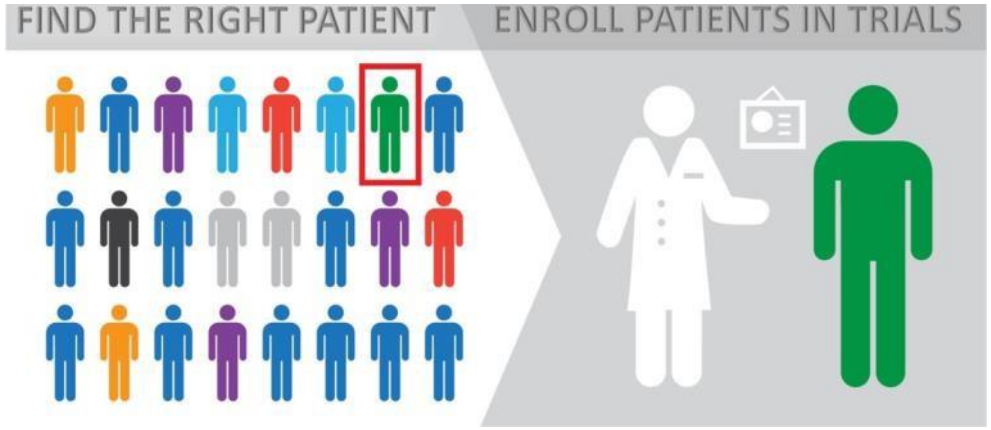
Studi di *real world:*
punti di forza/debolezza,
metodologie di pianificazione,
qualità metodologica,
conduzione e analisi

(O. Nanni)


Ricerca traslazionale
Ricerca di base preclinica
Sperimentazioni cliniche
Studi pragmatici
Studi di real world

Major PAIN POINTS in clinical trials

PATIENTS RECRUITMENT



Less than 5% of patients with cancer enroll in clinical trials and 1 in 5 trials are stopped for **poor accrual**!

Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, Torres A, Holtzen H, Coverdill SE, Williamson MP, Chau Q, Rhee K, Vinegra M. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. JCO Clin Cancer Inform. 2020 Jan;4:50-59. doi: 10.1200/CCI.19.00079. PMID: 31977254.

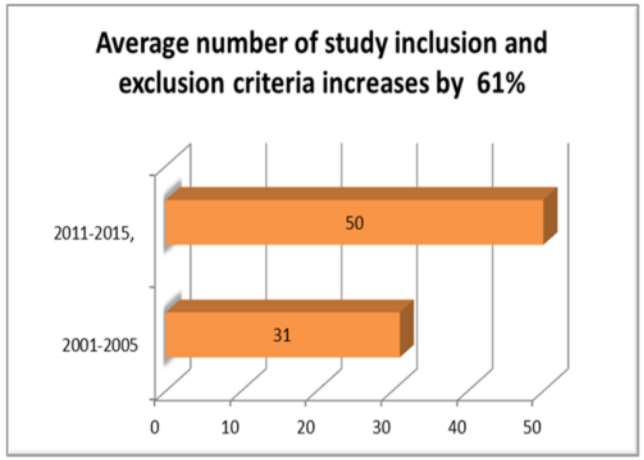
DATA COLLECTION

70% to 90% of eCRF data is duplicated in EHRs eSource systems, that account for 70 to 80 hours of **manual re-transcription effort**!

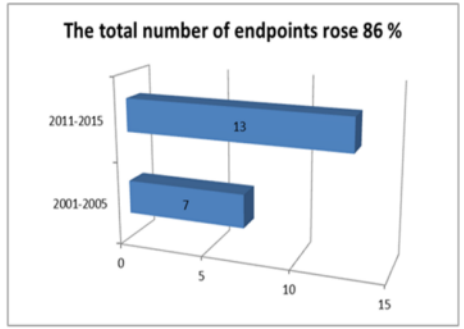
Sundgren, M., PhD., Ammour, Nadir,M.B.A., D.M.D., Hydes, D., Kalra, D., & Yeatman, R. (2021). Innovations in data capture transforming trial delivery. Applied Clinical Trials, 30(7), 16-20. Retrieved from https://www.proquest.com/scholarly-journals/innovations-data-capture-transforming-trial/docview/2821706597/se-2

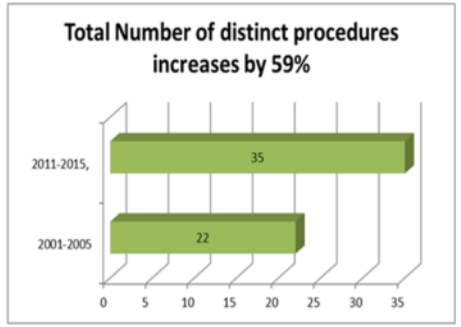
Major PAIN POINTS studi clinici: COMPLESSITÀ

Courtesy of Nadir AMMOUR, Sanofi R&D



- Sponsors' focus in targeting
  - Unmet medical needs
  - Narrowly defined and stratified patient populations.





AUMENTO COMPLESSITÀ
delle sperimentazioni cliniche

BIOBUSINESS BRIEFS

TRIAL WATCH

Trends in clinical trial design complexity

Kenneth A. Getz is at the Tufts Center for the Study of Drug Development, Tufts University, 75 Kneeland Street, Boston, Massachusetts 02111, USA.
Rafael A. Campo is at Medidata Solutions, 350 Hudson Street, New York, New York 10014, USA.
Correspondence to K.A.G.
kenneth.getz@tufts.edu

doi:10.1038/nrd.2017.65

NATURE REVIEWS | DRUG DISCOVERY VOLUME 16 | MAY 2017 | 307

ABOUTOpen Clinical Research

AboutOpen | 2022; 9: 42-44
ISSN 2465-2628 | DOI: 10.33393/ao.2022.2437

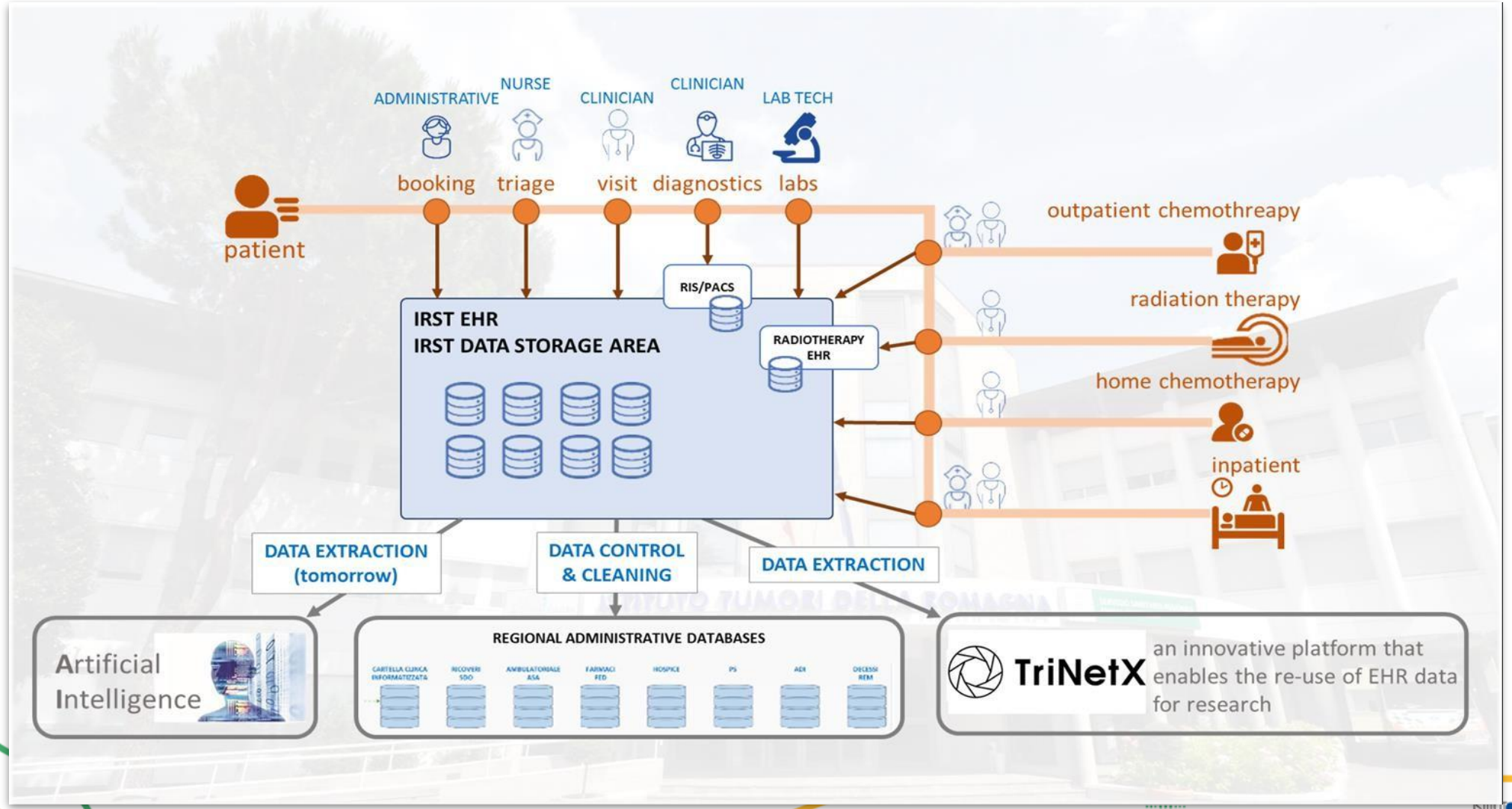
EDITORIAL

New trends in clinical trials—between complexity and the need for renewal

Celeste Cagnazzo

S.C. Oncoematologia Pediatrica – AOU Città della Salute e della Scienza Presidio Infantile Regina Margherita, Turin - Italy

Raccolta dei dati in IRST - Cartella Clinica Elettronica



TriNetX LIVE™ Networks

Define, explore, and analyze cohorts from hundreds of millions of de-identified patients around the globe

STUDY DESIGN AND FEASIBILITY

Ensuring that proposed criteria, comparators, endpoints, and other protocol specifications will allow for rapid enrollment and diverse trial participants

TRIAL SITE IDENTIFICATION AND OUTREACH

Connecting users with trial offices at healthcare organizations (HCOs) around the world along with counts of qualifying patients

REAL-WORLD EVIDENCE AND INSIGHTS

Revealing today's unmet needs, burden of disease, and comparative effectiveness and safety of approved treatments

Networks that span the globe

Cohorts on TriNetX LIVE™ are always drawn from a network: a set of HCOs united by geography, care specialty, or research mission. Our regional networks, described below, are the largest and most essential to trial operation and evidence generation.

GLOBAL
21 Countries
167 HCOs
204M Patients

USA
1 Country (United States)
81 HCOs
152M Patients

EMEA
14 Countries
60 HCOs
38M Patients

USA + APAC
6 Countries
91 HCOs
156M Patients

LATAM
2 Countries
18 HCOs
9M Patients

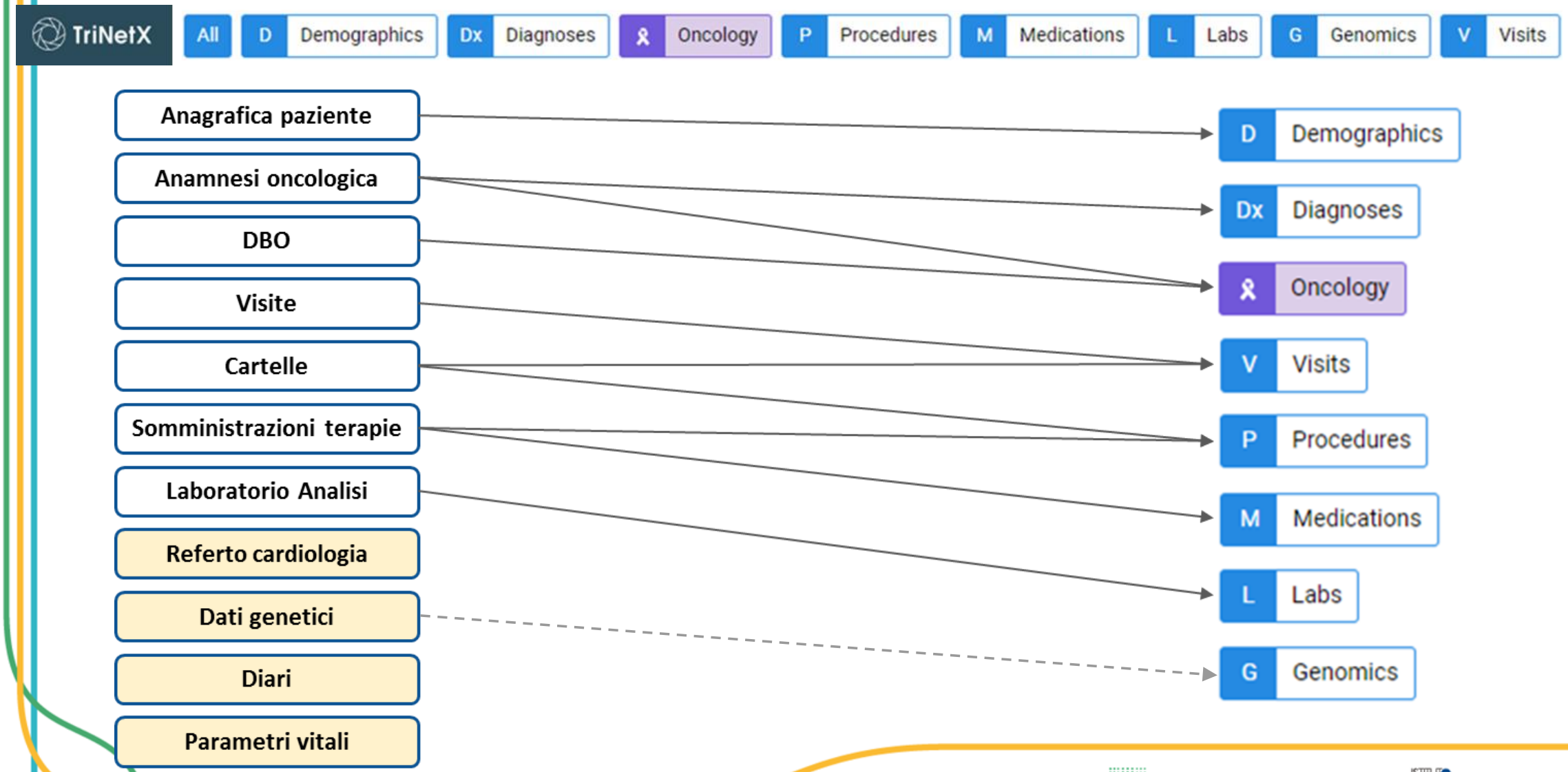
JAPAN
1 Country (Japan)
460 HCOs
43M Patients

OUR NETWORK IN ITALY

- **ASST Brianza**, Vimercate (MB)
- **ASST Lodi,** Lodi (LO)
- **ASST Grande Ospedale Metropolitano Niguarda,** Milano
- **ASST Spedali Civili Brescia**, Brescia (BS)
- **Azienda Ospedaliero-Universitaria di Ferrara**, Cona, Ferrara (FE)
- **Fondazione IRCCS San Gerardo dei Tintori,** Monza (MB)
- **IRCSS Istituto Dermatopatico Italiano (IDI),** Roma
- **IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori"**, Meldola (FC)
- **IRCCS Ospedale Galeazzi - Sant'Ambrogio**, Milano
- **IRCCS Ospedale San Raffaele**, Milano
- **IRCCS Policlinico San Donato**, San Donato Milanese (MI)
- **IRCCS Istituti Clinici Scientifici Maugeri**, Pavia
- **Azienda Ospedaliera Universitaria "SS Antonio e Biagio e C. Arrigo" Alessandria (AOU AL)**, Alessandria

TriNetX

Raccolta dei dati in IRST - Cartella Clinica Elettronica



What is TriNetX Connect?

The communication and workflow application within the TriNetX platform, supporting sponsors and study sites with efficient global engagement and evidence-based *identification* of study sites

TriNetX

RECLUTAMENTO: criticità

original reports

Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center

J. Thaddeus Beck, MD[1]; Melissa Rammage, PharmD[2]; Gretchen P. Jackson, MD, PhD[2]; Anita M. Preininger, PhD[2]; Irene Dankwa-Mullan, MD[2]; M. Christopher Roebuck, PhD[3]; Adam Torres, RN[4]; Helen Holtzen, RN[1]; Sadie E. Coverdill, MIM[2]; M. Paul Williamson, MSc[5]; Quincy Chau, PhD[5]; Kyu Rhee, MD[2]; and Michael Vinegra, MBA[5]

TABLE A3. Time for Eligibility Determination (n = 90 Patients With Cancer and Hematology Patients)

Variable	Time	Patients With Breast Cancer Identified	Trial 1[a]	Trial 2[b]	Trial 3[c]
Manual screening against 3 trials	1 hour 50 minutes	18			
WCTM-assisted screening against 3 trials	24 minutes	18			
WCTM sensitivity, %			100	100	100
WCTM specificity, %			76.5	94.1	76.5
WCTM PPV, %			20.0	50.0	20.0
WCTM NPV, %			100.0	100.0	100.0
Accuracy, % WCTM-manual agreement			77.8	94.4	77.8

Abbreviations: NPV, negative predictive value; PPV, positive predictive value; WCTM, Watson for Clinical Trial Matching.
[a]ClinicalTrials.gov identifier: NCT02437318.
[b]ClinicalTrials.gov identifier: NCT01923168.
[c]ClinicalTrials.gov identifier: NCT01633060.

Beck JT, Rammage M, Jackson GP, Preininger AM, Dankwa-Mullan I, Roebuck MC, Torres A, Holtzen H, Coverdill SE, Williamson MP, Chau Q, Rhee K, Vinegra M. Artificial Intelligence Tool for Optimizing Eligibility Screening for Clinical Trials in a Large Community Cancer Center. JCO Clin Cancer Inform. 2020 Jan;4:50-59. doi: 10.1200/CCI.19.00079. PMID: 31977254.

DOI:10.1093/jnci/dju229
First published online September 4, 2014

COMMENTARY

Adult Cancer Clinical Trials That Fail to Complete: An Epidemic?

Kristian D. Stensland, Russell B. McBride, Asma Latif, Juan Wisnivesky, Ryan Hendricks, Nitin Roper, Paolo Boffetta, Simon J. Hall, William K. Oh, Matthew D. Galsky

Manuscript received November 30, 2013; revised June 9, 2014; accepted June 15, 2014.

Correspondence to: Matthew D. Galsky, MD, Mount Sinai School of Medicine, Tisch Cancer Institute, 1 Gustave L Levy Place, New York, NY 10029 (e-mail: matthew.galsky@mssm.edu).

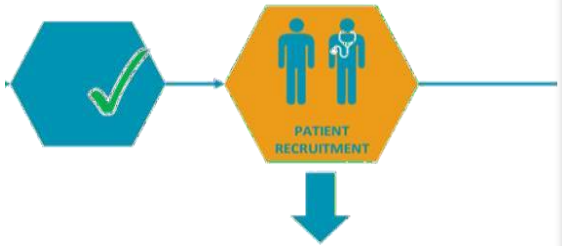
Reason for termination	Number of trials	Percent of terminated trials
Poor accrual	362	38.7%
Logistics	192	20.5%
Cancellation of trial by sponsor	99	10.6%
Inadequate budget	52	5.6%
Unavailability of study agent	23	2.4%
Departure of Principal Investigator from institution	18	1.9%
Toxicity/efficacy	169	18.1%
Poor results at interim analysis	92	9.9%
Toxicity/adverse events	77	8.2%
Other	212	22.7%
None given	92	9.9%
Other	87	9.3%
No longer necessary	33	3.5%

Figure 1. Reasons for adult cancer clinical trials failing to complete. Breakdown of the reasons for failure to complete among 935 adult interventional cancer clinical trials.

Stensland KD, McBride RB, Latif A, Wisnivesky J, Hendricks R, Roper N, Boffetta P, Hall SJ, Oh WK, Galsky MD. Adult cancer clinical trials that fail to complete: an epidemic? J Natl Cancer Inst. 2014 Sep 4;106(9):dju229. doi: 10.1093/jnci/dju229. PMID: 25190726.

RECLUTAMENTO e CARTELLA CLINICA ELETTRONICA

Utilizzando i dati della cartella clinica è possibile gestire il processo di reclutamento dei pazienti?



Speeding up recruitment by making EHR data **searchable for investigators** and establishing a **unified communication path** between sponsors and sites.

The «STRONG» use case

NIH U.S. National Library of Medicine
ClinicalTrials.gov

AstraZeneca

An Open-Label, Multi-Centre, Study to Assess the Safety of Fixed-Dose Durvalumab + Tremelimumab Combination Therapy or Durvalumab Monotherapy in Advanced Solid Malignancies. (STRONG)

STRONG[a]
Science and innovation in immuno-oncology

First Submitted Date ICMJE	March 7, 2017
First Posted Date ICMJE	March 21, 2017
Last Update Posted Date	February 1, 2018
Actual Start Date ICMJE	April 17, 2017
Recruitment Information	
Recruitment Status ICMJE	Recruiting
Estimated Enrollment ICMJE	1200
Estimated Completion Date	March 26, 2023
Estimated Primary Completion Date	March 26, 2023 (Final data collection date for primary outcome measure)

September 20, 2017
ETHICS COMMITTEE APPROVAL

November 3, 2017
InSite PARTECIPATION REQUEST

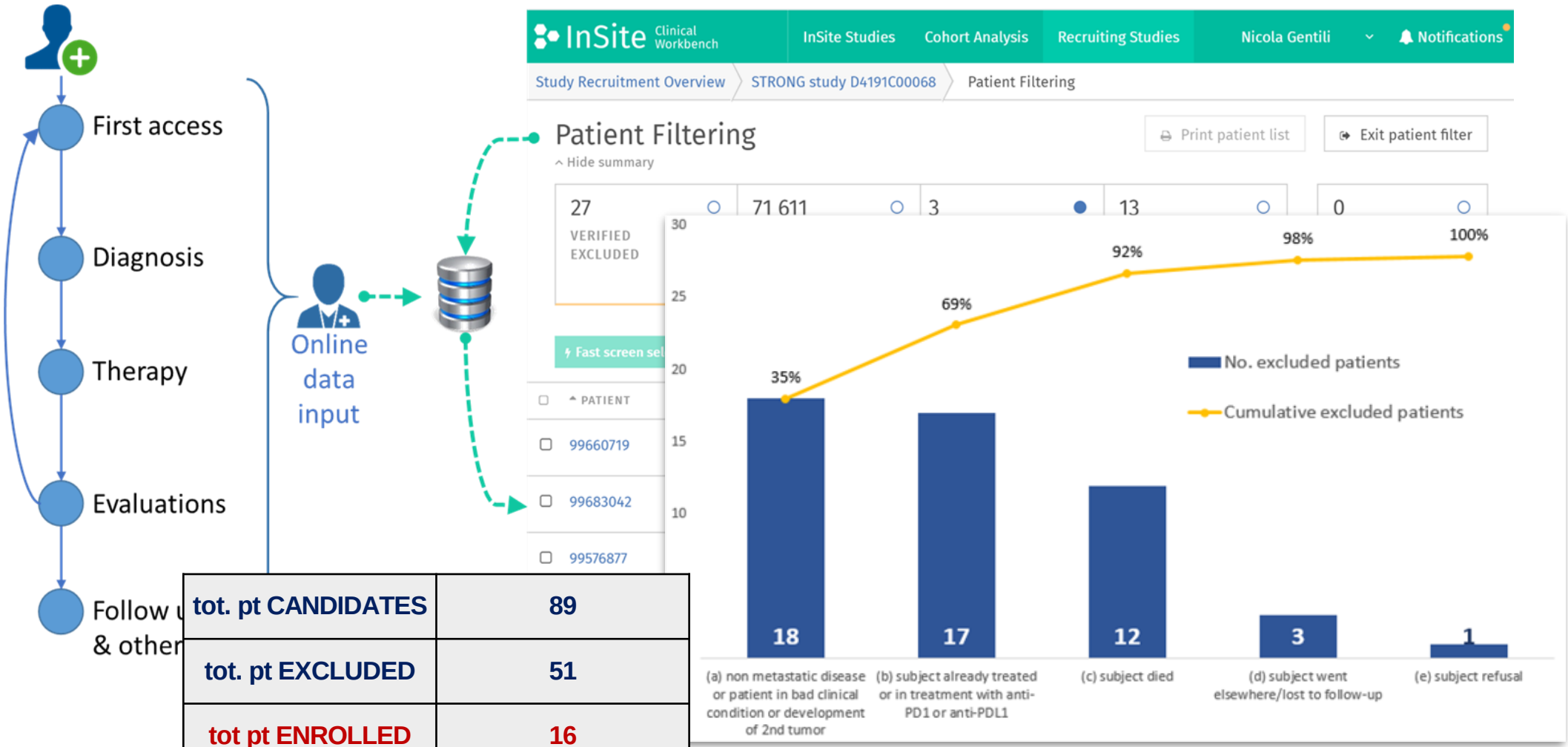
December 14, 2017
SITE INITIATION VISIT

December 22, 2017
InSite RECRUITMENT START

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

RECLUTAMENTO e CARTELLA CLINICA ELETTRONICA



tot. pt CANDIDATES	89
tot. pt EXCLUDED	51
tot pt ENROLLED	16
tot. pt VERIFIED	22

RACCOLTA DATI e EHR: TRANSFAIR study

È possibile trasferire i dati dalla cartella clinica elettronica alle CRF?

Facilitating **EHR data extraction** for applications used during trial execution (e.g. prefilling of CRFs and of SAE reports).

Ammour, Nadir, et al. "TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection." BMJ health & care informatics 30.1 (2023).

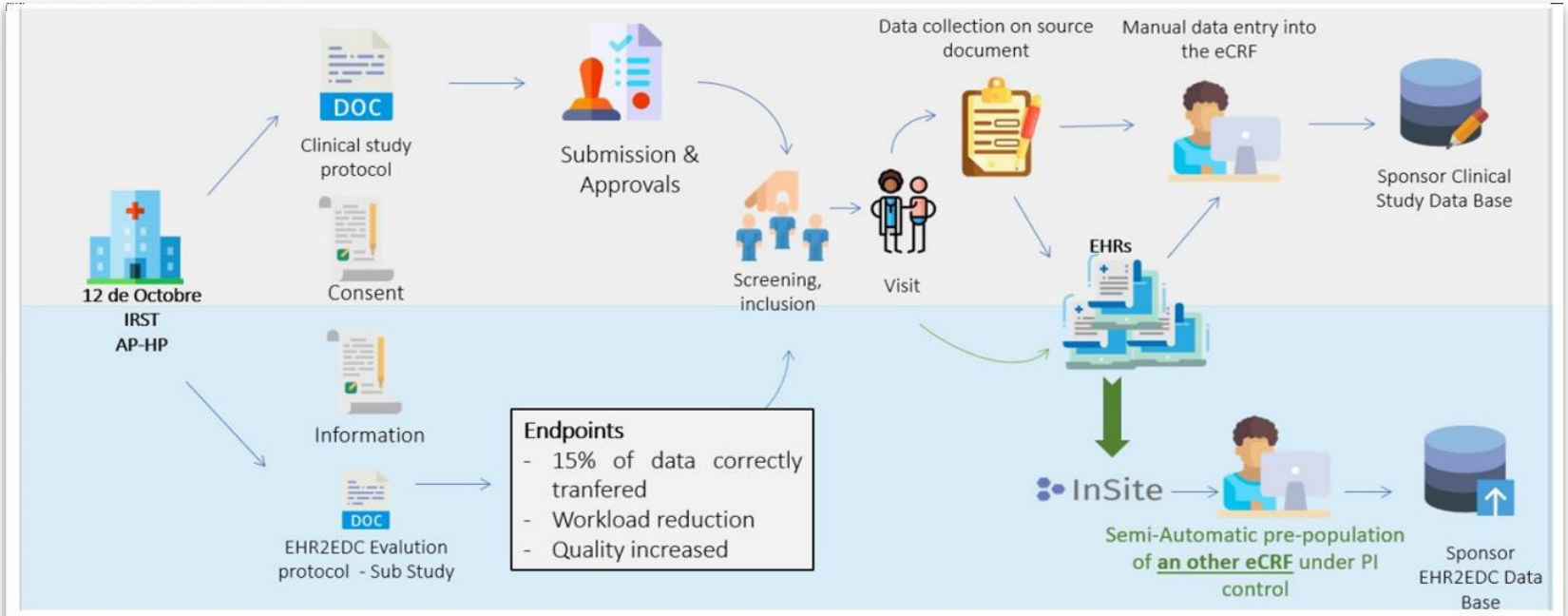


Figure 1 General organisation of the TransFAIR study. AP-HP, Assistance Publique-Hôpitaux de Paris; EHR2EDC, Electronic Health Records to Electronic Data Capture; PI, principal investigator ; eCRF, electronic Case Report Form.

- Six protocols from three pharma companies (AstraZeneca, Janssen, Sanofi)
- Four categories of health data (demographics, vital signs, laboratory and concomitant medication)
- Hospitals from three countries: AP-HP (France), 12 Octubre (Spain), IRST (Italy)
- 44 patient consents collected
- >11000 data points automatically transferred

TRANSFAIR study: RESULTS

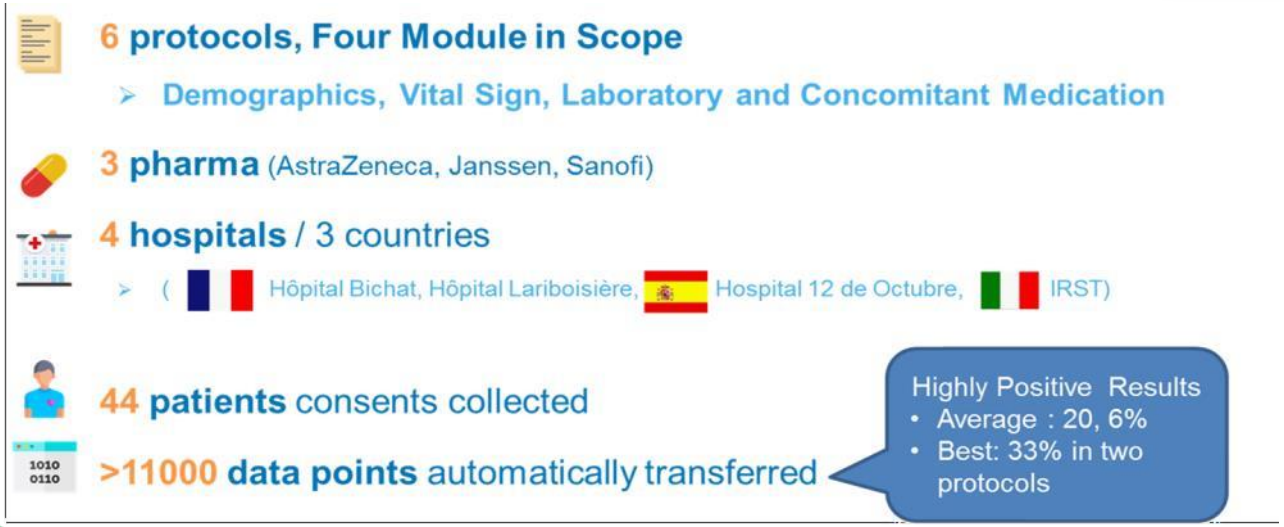
CORRETTAMENTE trasferito circa il 20% dei dati!

Open access Original research

BMJ Health & Care Informatics

TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection

Nadir Ammour,[1] Nicolas Griffon,[2,3] Juliette Djadi-Prat ,[4] Gilles Chatellier,[5,6] Martine Lewi,[7] Marija Todorovic,[7] Augustin Gómez de la Cámara,[8] Maria Teresa García Morales,[8] Sara Testoni,[9] Oriana Nanni,[9] Christoph Schindler,[10] Mats Sundgren,[11] Almenia Garvey,[12] Tomothy Victor,[12] Manon Cariou ,[1] Christel Daniel[2,3]



Source: EHR2EDC Project - Courtesy of Nadir AMMOUR, Sanofi R&D

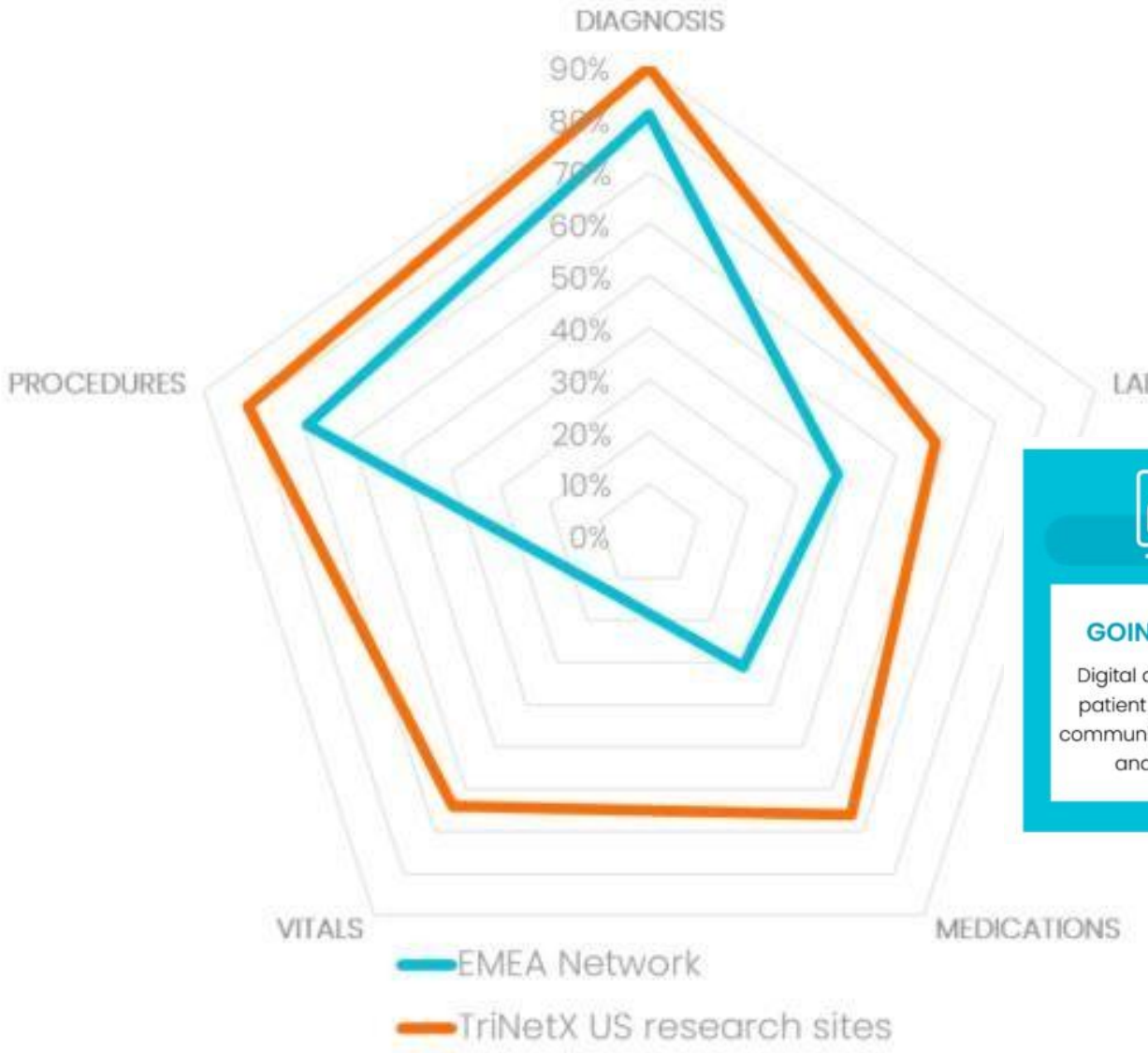
Table 4 Proportion of data collected and not collected for the four domains in the TransFAIR study scope

Data domain	% of data correctly transferred	% of missing data
Demographics	34.2	65.8
Laboratories	40.6	59.4
Vital signs	40.9	59.1
Concomitant medications	7.7	92.3

Ammour, Nadir, et al. "TransFAIR study: a European multicentre experimental comparison of EHR2EDC technology to the usual manual method for eCRF data collection." BMJ health & care informatics 30.1 (2023).

SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI



GOING DIGITAL

Digital approaches to patient engagement, community engagement and outreach

AI/ML

Artificial Intelligence (AI)/ Machine Learning (ML) for protocol optimization

PREDICTIONS

Using ML for patient ID and predicting timelines

NLP

Natural Language Processing (NLP) for chart review

Associazione Farmaceutici Industria
Società Scientifica

Prima esperienza: ABSTRACT ESGO2023

Young Ivestigator Session
03. Endometrial cancer

PDF

#434 Association between endometriosis and endometrial cancer: a real world evidence study FREE

Alberto Farolfi [1], Sara Testoni [1], Francesca Rusconi [2], Nicola Gentili [1], Ilaria Massa [1], Valentina Danesi [1], Amelia Altavilla [1], Maria Concetta Cursano [1], Salvatore Luca Burgio [1], Gema Hernandez Ibarburu [2] and Ugo De Giorgi [1]

Abstract

Introduction/Background Endometriosis is a benign pathological condition characterized by the ectopic presence of endometrial tissue. Whether endometriosis predisposes the pathogenesis of endometrial cancer (EC) is still debated. This study uses realworld data (RWD) from the network of TriNetX healthcare organization (HCO) networks in the US (TNX-US) and EMEA (TNX-EMEA) to analyze the impact of endometriosis as a risk factor for the development of EC.

Methodology Using TriNetX Platform, we defined a cohort of 284,287 patients with endometriosis and at least 6 months of follow up at the HCO, 254,726 from TNX-US and 29,561 TNX-EMEA. Propensity score matching between these cohorts and the female control cohorts in each regional network was used to remove the possible confounding effects of age, body mass index (BMI), previous diagnosis of pelvic inflammatory disease, breast cancer, other cancer of female genital organs or genetic susceptibility to cancer. Hazard ratio (HR) was used to compare the incidence of EC between the matched cohorts. Kaplan Meier analysis was used to compare the overall survival (OS) of EC patients with previous endometriosis vs those without endometriosis patients after propensity score matching. The time window of observation in both analyses was 10 years.

Results Patients with endometriosis diagnosis had a higher risk of developing EC in both TNX-US (2,151/237,034 vs 620/238,837, HR 3.49, 95% CI 3.19–3.82) and TNX-EMEA (319/28,241 vs 41/28,282, HR 7.58, 95% CI 5.48–10.50). The OS of EC patients with endometriosis was demonstrated to be significantly better than those without endometriosis: the 10-year OS probability was 78.37% vs 62.41% (p<0.01) and 73.11% vs 49.61% (p<0.01), in TNX-US and TNX-EMEA, respectively.

Conclusion Our RWD supports the association between endometriosis and an increased risk of developing EC. Endometriosis-associated tumors appear to have a better prognosis.

Disclosures U.D.G. has received advisory board or consultant fees from Merck Sharp & Dohme, Bristol My-ers Squibb, Janssen, Astellas, Sanofi, Bayer, Pfizer, Ipsen, Novartis, and Pharmamar and institutional research grants from Astrazeneca, Sanofi, and Roche. A.F. has received personal honoraria for lectures from Astrazeneca, GSK-Tesaro, Clovis, and advisory board from Jannsen, Astrazeneca, GSK-Tesaro. The other authors declare no conflict of interest.

TriNetX contributed in the collection and analyses of the data, but had no role in interpretation of data, in the writing of the manuscript, or in the decision to present the results.

https://doi.org/10.1136/ijgc-2023-ESGO.55

IRCCS Istituto Romagnolo per lo Studio dei Tumori "Dino ...
4.249 follower
5m •

La relazione tra #endometriosi e il rischio di sviluppare un #tumore all'#utero è stato il tema centrale dello studio presentato dal dott. Alberto Farolfi, oncologo di IRST "Dino Amadori" Irccs della SC #Oncologia #Clinica e #Sperimentale in #Terapie #innovative ed alte dosi (diretta dal dott. Ugo De Giorgi), nel corso del congresso European Society of Gynaecological Oncology 2023 che si è tenuto nei giorni scorsi a #Instabul.

Una presentazione che ha ricevuto il primo premio come miglior studio della sessione #Young #Investigator #Esgo.

ESGO European Society of Gynaecological Oncology
5.162 follower
5m • Modificato •
+ Segui

Congratulations to all speakers of the Young Investigator Session during the #ESGO2023 Congress!

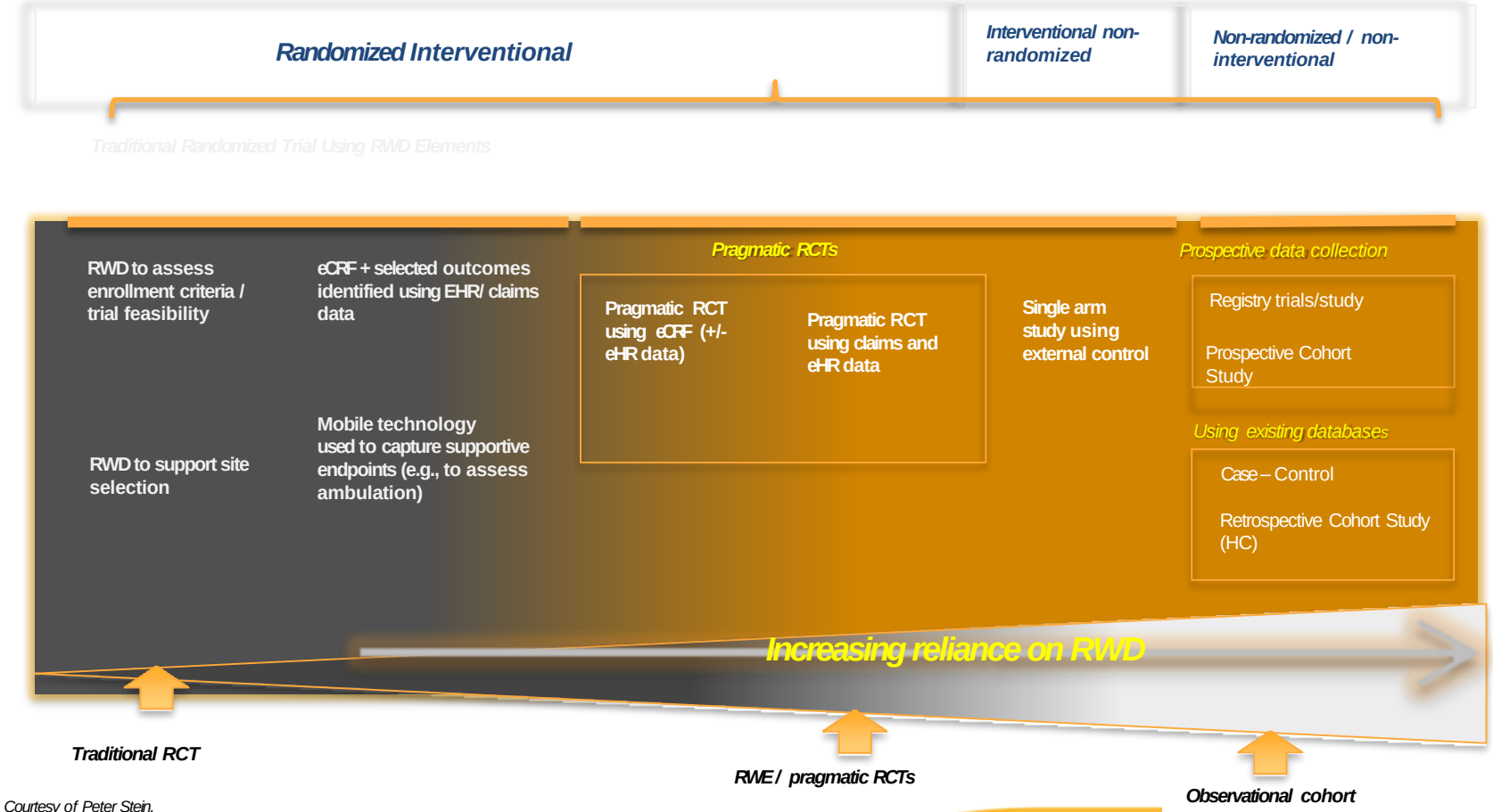
The 1st prize was awarded to dr Alberto Farolfi for the study entitled "Association between endometriosis and endometrial cancer: a real world evidence study
Abstract: https://buff.ly/3Pwb86A

Strategic goals and core recommendations - Human medicines[1]

1. Catalysing the integration of science and technology in medicines' development	**Support developments in precision medicine, biomarkers and 'omics**
	Support translation of advanced therapy medicinal products (ATMPs) into patient treatments
	Promote and invest in the PRIME scheme
	Facilitate the implementation of novel manufacturing technologies
	Create an integrated evaluation pathway for the assessment of medical devices, in vitro diagnostics and borderline products
	Develop understanding of, and regulatory response to, nanotechnology and new materials in pharmaceuticals
	Diversify and integrate the provision of regulatory advice along the development continuum
2. Driving collaborative evidence generation – improving the scientific quality of evaluations	Leverage non-clinical models and 3Rs principles[¥]
	Foster innovation in clinical trials
	Develop the regulatory framework for emerging clinical data generation
	Expand benefit-risk assessment and communication
	Invest in special populations initiatives
	Optimise capabilities in modelling, simulation and extrapolation
	Exploit digital technology and artificial intelligence in decision making
3. Advancing patient-centred access to medicines in partnership with healthcare systems	**Contribute to HTA's preparedness and downstream decision making for innovative medicines**
	Bridge from evaluation to access through collaboration with payers
	Reinforce patient relevance in evidence generation
	Promote use of high-quality real-world data (RWD) in decision-making
	Develop network competence and specialist collaborations to engage with big data
	Deliver improved product information in electronic format (ePI)[¥]
	Promote the availability and support uptake of biosimilars in healthcare systems
	Further develop external engagement and communications to promote trust and confidence in the EU regulatory system
4. Addressing emerging health threats and availability/ therapeutic challenges	Implement EMA's health threats plan, ring-fence resources and refine preparedness approaches
	Continue to support development of new antibacterial agents and their alternatives[¥]
	Promote global cooperation to anticipate and address supply problems[¥]
	Support innovative approaches to the development, approval and post-authorisation monitoring of vaccines
	Support the development and implementation of a repurposing framework
5. Enabling and leveraging research and innovation in regulatory science	**Develop network-led partnerships with academic/research centres to undertake research in strategic areas of regulatory science**
	Leverage collaborations between academia and network scientists to address rapidly emerging regulatory science research questions
	Identify and enable access to the best expertise across Europe and internationally
	Disseminate and exchange knowledge, expertise and innovation across the network and to its stakeholders

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

Different Challenges and Opportunities for Each Approach



Courtesy of Peter Stein, OND

Definitions

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

  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 and potential benefits or risks of a medical product derived from analysis of RWD.**

  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).

REAL WORLD STUDIES

INTERVENTIONAL STUDIES

- Pragmatic clinical trials

OBSERVATIONAL STUDIES: *PROSPECTIVE*

- Traditional cohort studies
- Patient surveys
- Disease registries

OBSERVATIONAL STUDIES: RETROSPECTIVE

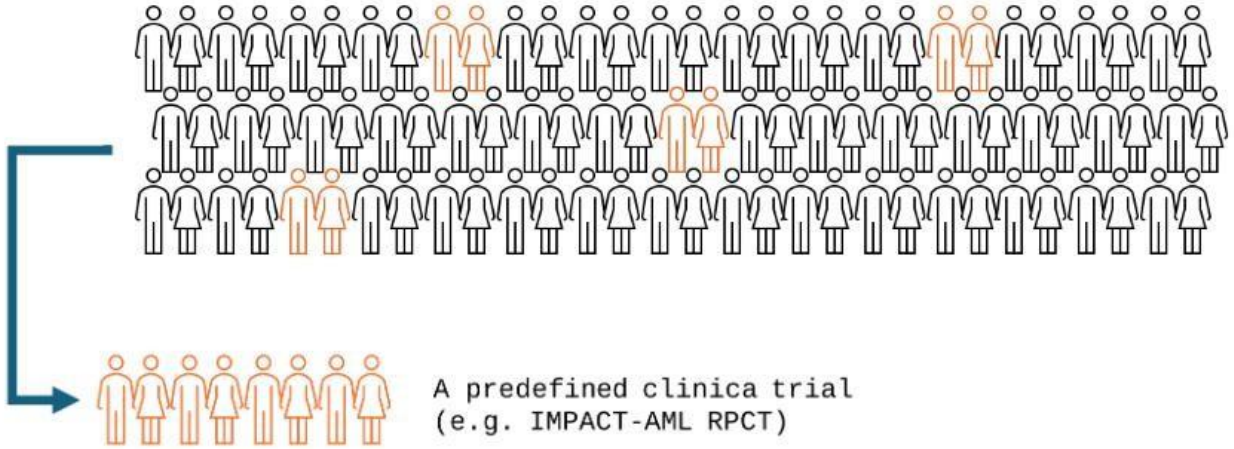
- Electronic medical records
- Medical claims data
- Birth or death registries
- Surveillance databases

WP2

STREAM: Concept & ambition

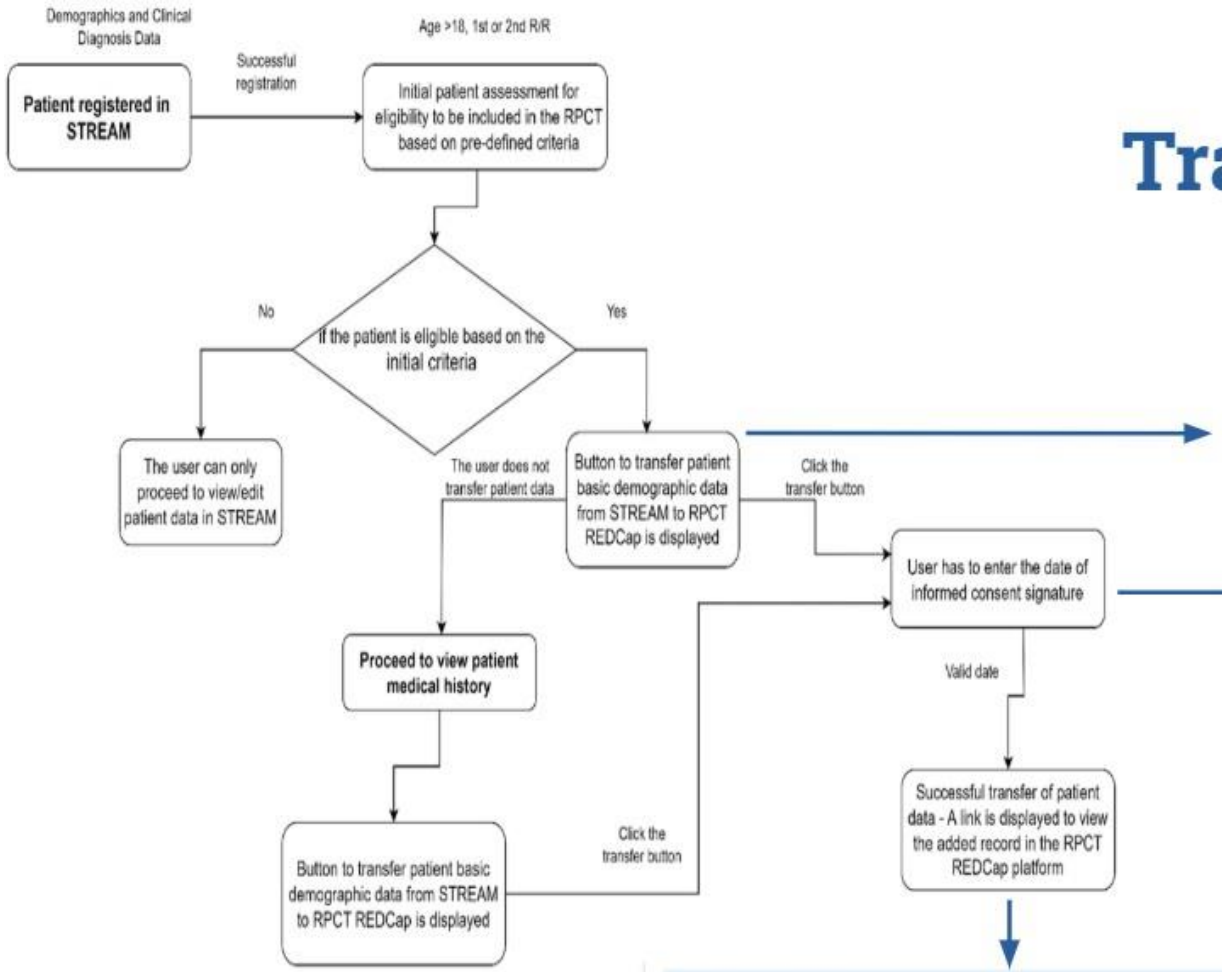
maSTer framework for RElapsed or REfractory Acute Myeloid leukemia

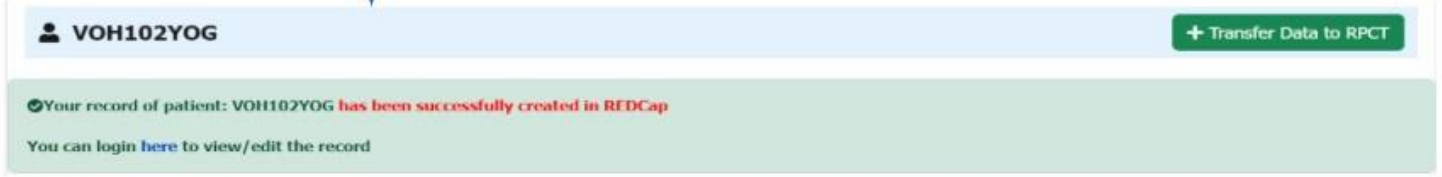
R/R AML population



Funded by the European Union

Transfer Patient Data to IMPACT-AML RPCT





Funded by the European Union

IMPACT- AML RPCT, A Randomized Pragmatic Clinical Trial for Relapsed/Refractory AML

What challenges?

Project funded **Horizon Mission Cancer program**

- 1st or 2nd R/R AML
- ECOG<4, M/F>18 yrs
- Candidate to low and high intensity treatments according to clinical judgement
- No alternative therapies

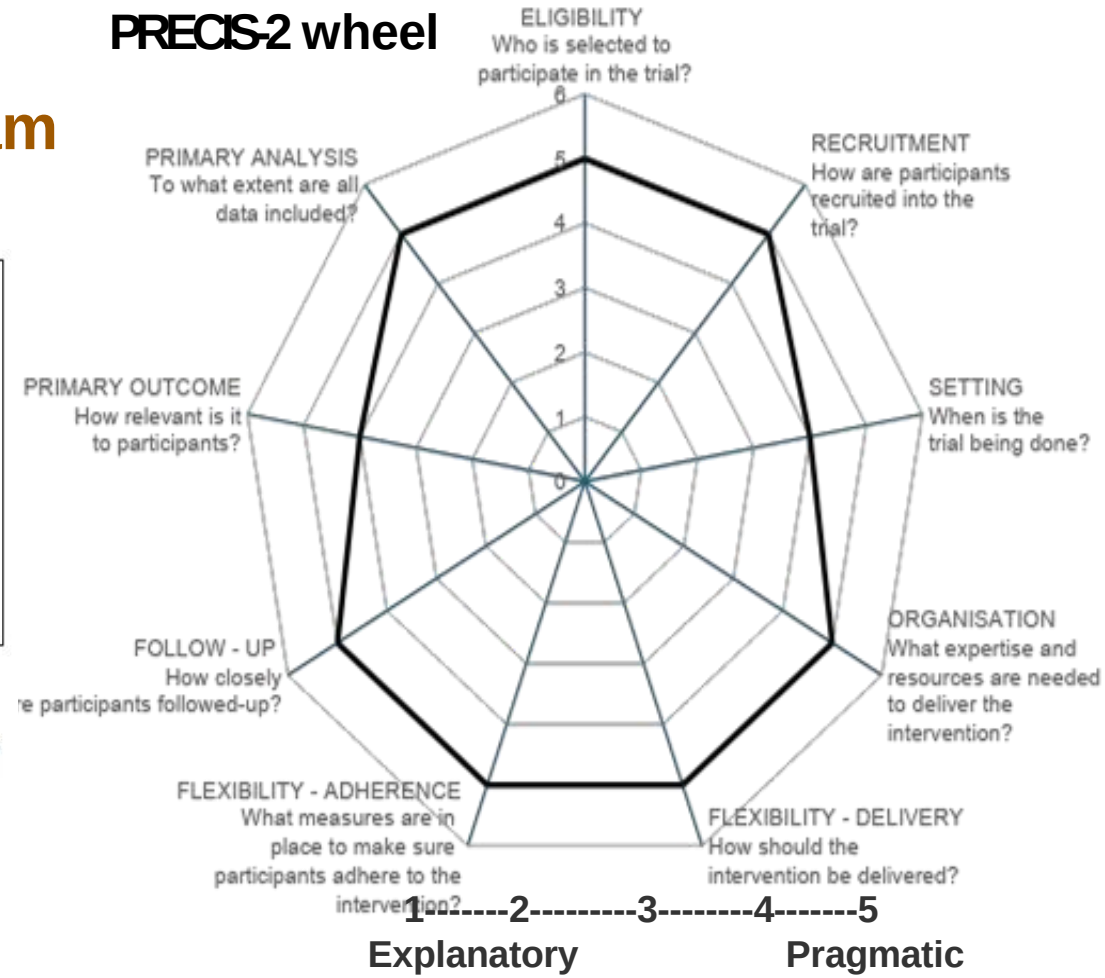
R 1:1

- High Intensity CHT[1] (1 or 2 courses)
- Low Intensive Treatment (max 4 months)

HSCT whenever possible
Or
Treatment according to physician decision/observation

- 339 patients (2 yrs enrollment; 1 yr Follow up)
- Randomisation 1:1 to receive intensive chemotherapy or low intensity treatment as per country/site clinical practice.
- The **low intensity or intensive regimen** must be selected by the physician before randomization among the options **available according to clinical practice at country/clinical site level**, as per consolidated use or clinical practice

PRECIS-2 wheel



1-------2---------3--------4-------5

Explanatory **Pragmatic**


AIOM e ASCO: insieme per più qualità nel disegno dei trial

A cura di Alessio Malta, Luciano De Fiore

3 Aprile 2023

Esiste ancora un gap fra studi registrativi e real world evidence», è la riflessione di **Francesco Perrone, Presidente eletto AIOM**. «*Servono studi di sequenza terapeutica, di confronto testa a testa e adattivi, in grado cioè di aggiornarsi con l'evoluzione degli scenari diagnostici e terapeutici. E gli endpoint a cui fare riferimento devono essere solidi, includendo sopravvivenza, qualità di vita e tossicità».* Dati del genere avrebbero ricadute positive anche dal punto di vista regolatorio.

A questo scopo, prosegue Perrone, accanto ai grandi studi registrativi ed ai trial promossi meritoriamente dall'industria, «*serve una ricerca indipendente più forte, promossa dal Servizio sanitario nazionale, capace di rispondere a questi bisogni. Non dovremmo considerare impossibile l'obiettivo di tornare a quel quasi 30% di studi indipendenti di dieci anni fa, quando ormai oggi, in Italia, solo un quinto degli studi su nuovi farmaci è indipendente*».

Leading Edge
Commentary

Cell

The Master Observational Trial: A New Class of Master Protocol to Advance Precision Medicine

Dane Dickson,[1,2,*] Jennifer Johnson,[3] Raymond Bergan,[1] Rebecca Owens,[2] Vivek Subbiah,[4] and Razelle Kurzrock[5]
[1]Knight Cancer Institute, Oregon Health and Science University, Portland, OR, USA
[2]Taproot Health, Salt Lake City, UT, USA
[3]Sidney Kimmel Cancer Center, Thomas Jefferson University, Philadelphia, PA, USA
[4]University of Texas MD Anderson Cancer Center, Houston, TX, USA
[5]Moores Cancer Center, University of California at San Diego, San Diego, CA, USA
*Correspondence: dane.dickson@taprootco.com
https://doi.org/10.1016/j.cell.2019.12.009

This commentary introduces a new clinical trial construct, the Master Observational Trial (MOT), which hybridizes the power of molecularly based master interventional protocols with the breadth of real-world data. The MOT provides a clinical venue to allow molecular medicine to rapidly advance, answers questions that traditional interventional trials generally do not address, and seamlessly integrates with interventional trials in both diagnostic and therapeutic arenas. The result is a more comprehensive data collection ecosystem in precision medicine.

Studio ODHIN Real World Data e dati genetici in rete

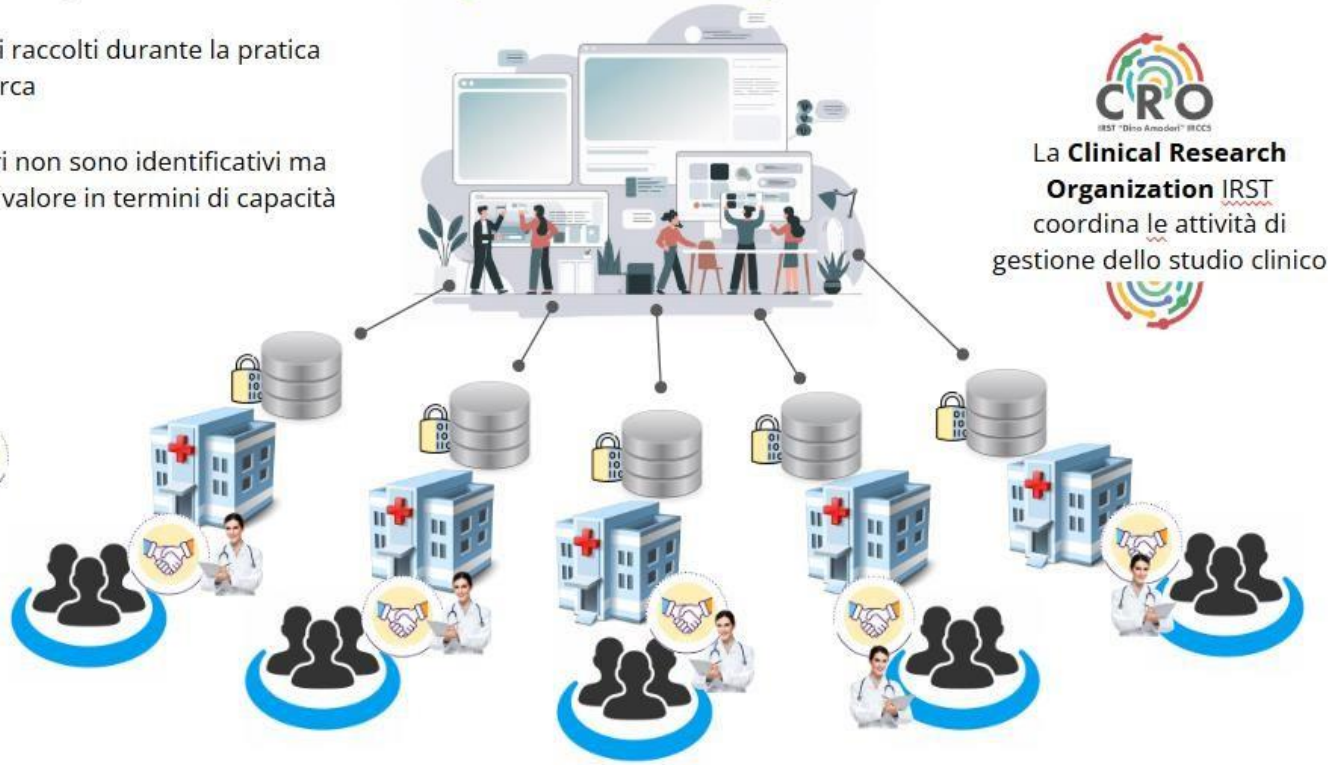
Progetto proposto da IRST e promosso alla rete degli IRCCS Oncologici

I **Real World Data** sono dati raccolti durante la pratica clinica e "riutilizzati" per ricerca

I **dati genetici** necessari non sono identificativi ma aggiungono un enorme valore in termini di capacità di analisi

I pazienti affidano all'Istituto i propri dati che, in base alla normativa sulla privacy, trattano i le informazioni in modo sicuro, adottando tutte le misure necessarie per proteggerli e condividerne l'uso nel progetto.

Ogni Istituto stringe un rapporto fiduciario con i propri pazienti con l'obiettivo di garantire un uso dei dati etico una collaborazione di ricerca tra i ricercatori della rete Alleanza Contro il Cancro.

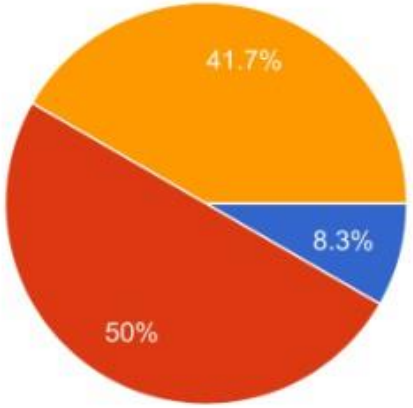


La **Clinical Research Organization** IRST coordina le attività di gestione dello studio clinico

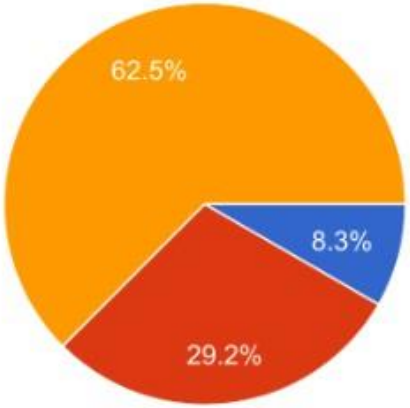
Maturità digitale

Questionario per la valutazione preliminare di fattibilità per lo studio Odhin

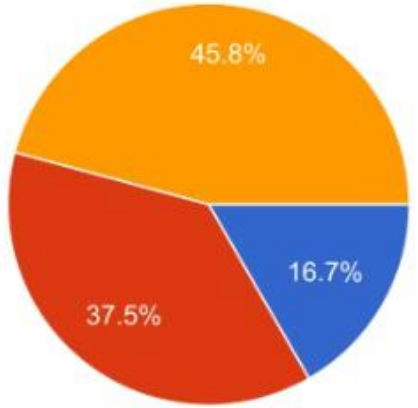
Dati anamnesi oncologiche



Dati terapie somministrate



Analisi genetiche NGS



La survey rileva una elevata disponibilità di dati da parte degli Istituti (anamnesi oncologica 91,7%; terapie somministrate 91,7%, analisi genetiche 83,3%), tuttavia circa la metà delle informazioni risultano essere riportate in campi non strutturati (testo libero).

- No
- Si, ma solo informazioni riportate in testo libero
- Si, in forma strutturata e codificata

The Oncologist 2020;25:e746–e752 www.TheOncologist.com

The Oncologist®

Commentary

Real-World Evidence in Oncology: Opportunities and Limitations

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

[a]Department of Oncology, University of Turin; Ordine Mauriziano Hospital, Torino, Italy; [b]Clinical Trial Unit, National Cancer Institute, IRCCS Fondazione Pascale, Napoli, Italy; [c]Department of Surgery, Oncology and Gastroenterology, University of Padova and Oncologia Medica 2, Istituto Oncologico Veneto, I.R.C.C.S., Padova, Italy

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Real-world evidence • Clinical trials • Cancer treatments

SHARP

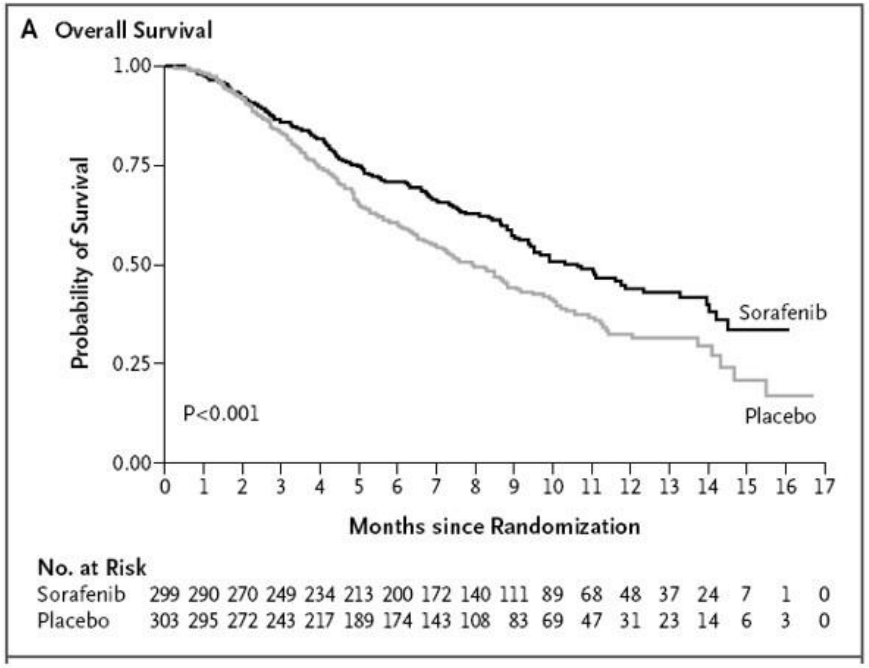
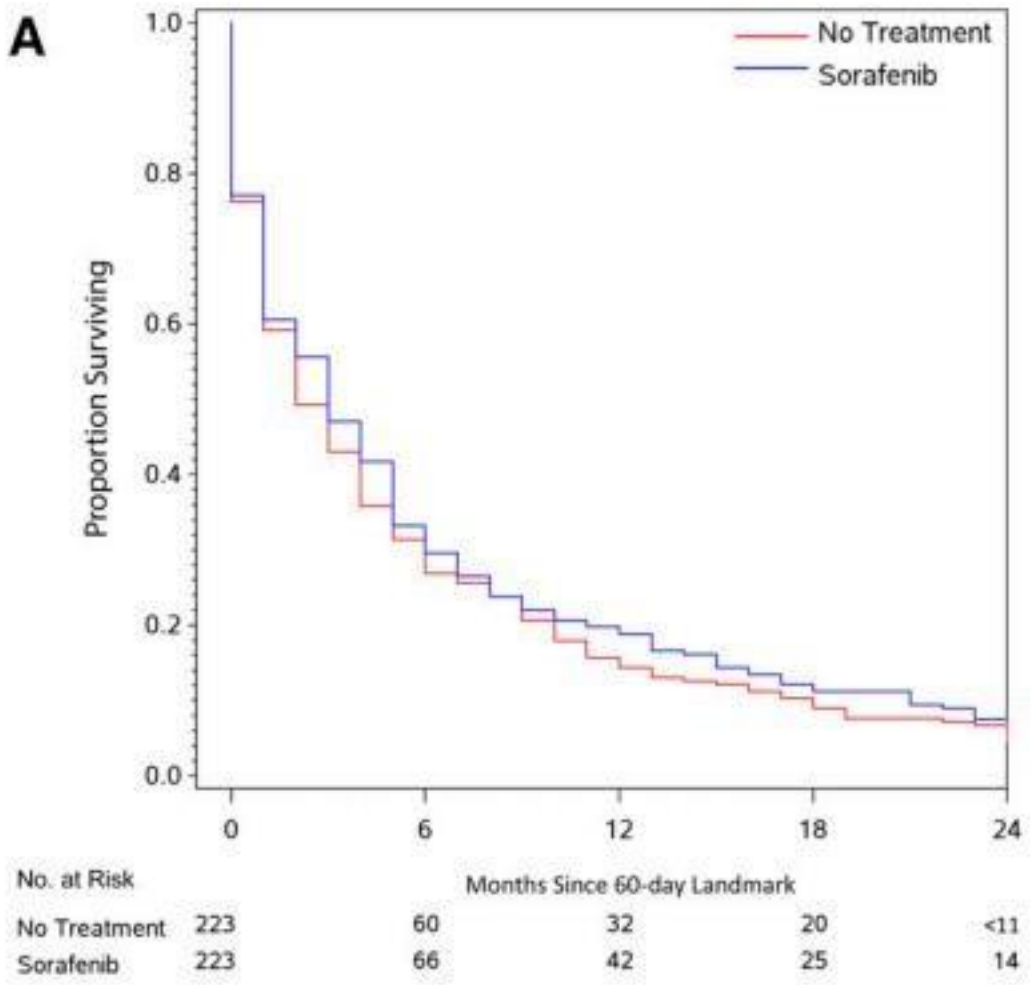


Table 2. Summary of Efficacy Measures.*

Outcome	Sorafenib (N=299)	Placebo (N=303)	Hazard Ratio (95% CI)	P Value
Overall survival (mo)			0.69 (0.55–0.87)	<0.001
Median	10.7	7.9		
95% CI	9.4–13.3	6.8–9.1		
1-yr survival rate (%)	44	33		0.009
Time to symptomatic progression (mo)†			1.08 (0.88–1.31)	0.77
Median	4.1	4.9		
95% CI	3.5–4.8	4.2–6.3		
Time to radiologic progression (mo)			0.58 (0.45–0.74)	<0.001
Median	5.5	2.8		
95% CI	4.1–6.9	2.7–3.9		
Level of response (%)‡				
Complete	0	0		NA
Partial	2	1		0.05
Stable disease	71	67		0.17
Disease-control rate (%)§	43	32		0.002



The findings of a median survival of o
beneficiaries with HCC prescribed so
the questionable value of sorafenib in
Patients should be cautioned that
randomized trials, their life expectanc
of sorafenib is likely to be quite small.
Given that sorafenib causes consider
symptom palliation, supportive care s
alternative to sorafenib, particularly fo
performance status or advanced cirrh

Sanoff HK . TheOncologist 201

ESMO GROW
GUIDANCE FOR REPORTING ONCOLOGY
REAL-WORLD EVIDENCE

ESMO-GROW Checklist for Authors and Reviewers

This checklist integrates all ESMO-GROW recommendation and could be used by authors and reviewers when assessing the reporting standards of a real-world evidence study in Oncology.

For the the ESMO-GROW checklist, the following criteria are considered:
"Yes, fully reported" – The recommendation is adequately considered.
"Yes, partially reported" – The recommendation is considered, but some important details are missing.
"Not reported" – The recommendation is applicable for the case, but it was not considered.
"Not applicable" – The recommendation is not applicable for this study.

Name of Author/Reviewer: **Date:**

Title of Manuscript or Identifier:

Recommendations	Yes, fully reported	Yes, partially reported	Not reported	Not applicable
1. Title				
1.1: Concisely include relevant key terms referring to the study type, study population, objectives, data sources and outcomes, depending on the study. Consider including the terms 'real-world' or 'observational'	◯	◯	◯	◯
2. Introduction				
2.1: Explain the scientific rationale for the research question(s), providing concise background information on previous core evidence from systematic reviews, meta-analyses, clinical trials and/or real-world evidence studies	◯	◯	◯	◯
2.2: Identify the gaps in evidence and explain why and how they can be suitably addressed by real-world evidence research. Specify the new evidence that is expected from the current study	◯	◯	◯	◯
2.3: Briefly introduce the aim(s) of the study	◯	◯	◯	◯
3. Methods				
Study objective(s), design, data sources and variables				
3.1: Provide the study research question(s) including a description of the patients or the object under study and the target outcome(s)	◯	◯	◯	◯
3.2: Provide the study objective(s) and consider classifying the type of research as descriptive and/or analytical (explanatory or predictive)	◯	◯	◯	◯
3.3: Provide relevant information to describe and classify the study design used to address the research question	◯	◯	◯	◯
3.4: Give a clear definition of the eligibility criteria used to select the patients or objects under study, particularly regarding cancer-related aspects	◯	◯	◯	◯
3.5: Report the specific type and purpose of real-world data source(s) used, providing a detailed description and the reason(s) why the source was considered appropriate for the study objectives	◯	◯	◯	◯
3.6: When multiple real-world data sources are used, provide details on interoperability, including identification of duplicated cases or data linkage from separate databases	◯	◯	◯	◯
3.7: Provide details and timings of source and study data management. Consider specifying methods of raw data collection, updates and completeness, data extraction, cleaning and/or quality controls and validation	◯	◯	◯	◯
3.8: Provide core details on database and/or study registration, governance, ownership, metadata and full data accessibility in the main text or supplementary material	◯	◯	◯	◯
3.9: Identify the data source of each core variable, its definition, if the variable was derived or coded, and describe how the derivation or coding was conducted and validated.	◯	◯	◯	◯
3.10: Specify the time points of core variables in relation to the cancer disease trajectory	◯	◯	◯	◯
3.11: Provide a complete list of core variables included in the study. Variables can be grouped as baseline characteristics, exposure and outcomes or endpoints	◯	◯	◯	◯
3.12: For biomarker-related studies, provide details on biomarker description, timing, and methods of assessment and analytical validation	◯	◯	◯	◯
Statistical analysis and artificial intelligence methods				
3.13: Summarise the main aspects of the statistical analysis	◯	◯	◯	◯
3.14: When applicable, provide details on the pre-planned sample size requirements and power of the study	◯	◯	◯	◯
3.15: Specify the pre-planned strategies to identify and mitigate the main sources of bias	◯	◯	◯	◯
3.16: Clearly distinguish prespecified from *post hoc* analyses, especially for subgroup analyses	◯	◯	◯	◯
3.17: Provide information on internal and external validity, as well as any sensitivity analyses	◯	◯	◯	◯
3.18: For analytical studies, the full version of the statistical analysis plan should be provided in the supplementary material, including a brief explanation of any amendments	◯	◯	◯	◯
3.19: When applicable, specify which machine learning, deep learning or alternative artificial intelligence method has been used	◯	◯	◯	◯
3.20: When reporting real-world data analysis with artificial intelligence (e.g. machine learning and deep learning) algorithms, include comprehensive aspects on data pre-processing techniques, feature engineering strategies and model development	◯	◯	◯	◯
3.21: Address the artificial intelligence model explainability and interpretability, and present the plan for integration into clinical practice, if applicable	◯	◯	◯	◯
3.22: When applicable, briefly describe the multidisciplinary team required for the study and explain how these needs were met	◯	◯	◯	◯

ESMO GROW
GUIDANCE FOR REPORTING ONCOLOGY
REAL-WORLD EVIDENCE

ESMO-GROW Checklist for Authors and Reviewers

Recommendations	Yes, fully reported	Yes, partially reported	Not reported	Not applicable
4. Results				
4.1: Provide number of cases excluded or nonparticipating and reasons at each stage of sample selection, as well as numbers lost to follow-up. Compare the cases excluded with those included in the analyses. Illustrate this with a flowchart	◯	◯	◯	◯
4.2: Describe the baseline characteristics of the cases included (e.g. clinico-demographic and tumour characteristics). The baseline characteristics of different groups under analysis should be compared, if applicable	◯	◯	◯	◯
4.3: Report the results of the primary analysis of study outcomes. Briefly describe the results of exploratory analyses if relevant (prespecified and/or *post hoc*). Provide details of how readers can access the full results	◯	◯	◯	◯
5. Discussion and conclusions				
Discussion				
5.1: Summarise the core results that address the primary research question(s) and objectively discuss the data in relation to the best available evidence on the topic. Avoid a convenient selection of literature to support a point	◯	◯	◯	◯
5.2: Discuss the strengths and limitations of the current study, including the main biases, how the strategies applied contributed to bias avoidance or mitigation, and, if applicable, in which direction the authors estimate that residual bias may influence the core results of the study	◯	◯	◯	◯
5.3: Discuss the generalisability of the study results and their potential implications for clinical practice, health policies or public health and for the generation of hypotheses for future research	◯	◯	◯	◯
Conclusions				
5.4: Provide a balanced summary of core results relating to the primary research question and the main implications for clinical practice, health policies and/or public health. Suggest further research considering the remaining unmet needs and limitations from the reported study	◯	◯	◯	◯
6. Final considerations				
6.1: Specify all relevant study sponsorship(s) as well as direct and/or indirect or in-kind funding	◯	◯	◯	◯
6.2: Specify all relevant acknowledgements, author disclosures, individual contributions and other final considerations as per journal regulations	◯	◯	◯	◯

Notes:

To access full manuscripts and for citations, please consider the following references and links:

- Castelo-Branco L et al. "ESMO Guidance for Reporting Oncology real-World evidence (GROW)". Ann Oncol 2023; 34: 10.1016/j.annonc.2023.10.001
- Castelo-Branco L et al. "ESMO Guidance for Reporting Oncology real-World evidence (GROW)". ESMO Real World Data & Digital Oncol 2023; 1: 10.1016/j.esmorw.2023.10.001

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

protocollo di studio ricerca osservazionale

completezza e accuratezza:

- Impiego di un appropriato disegno dello studio e di una adeguata dimensione del campione;
- Accertamento della presenza di eventuali bias e messa in atto delle strategie per limitare il loro effetto;
- Valutazione dei possibili fattori confondenti e stabilire eventuali possibilità di mitigazione;
- Tenere conto della possibile presenza dei dati mancanti e del loro impatto sui risultati;
- Redazione ‘a priori’ di un piano per l’analisi dei dati consistente rispetto alle linee guida della letteratura;
- Budget

Aesth Plast Surg
https://doi.org/10.1007/s00266-023-03661-z

The Impact of Adjuvant Radiotherapy on Immediate Prepectoral Implant-Based Breast Reconstruction

Edvin Ostapenko[1,2] · Larissa Nixdorf[1] · Yelena Devyatko[1] · Ruth Exner[1] · Kerstin Wimmer[1] · Florian Fitzal[1,3]

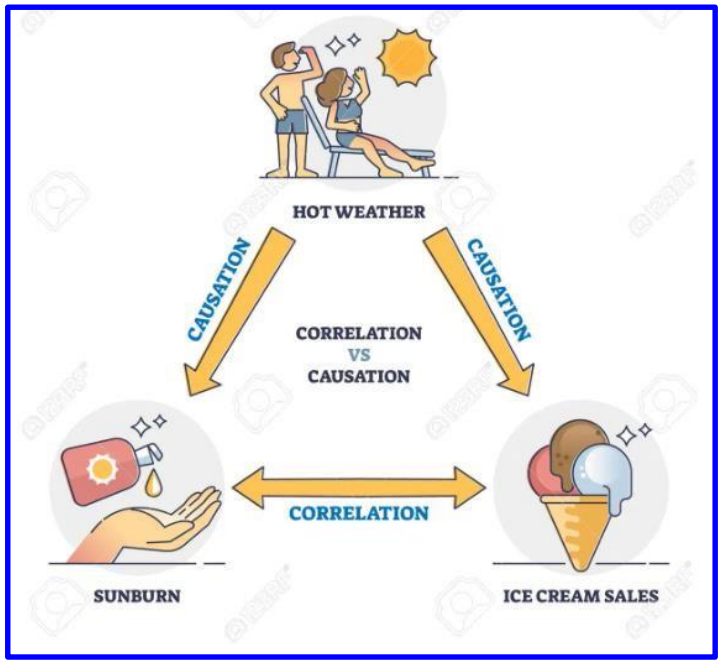
Statistical analysis

Data analysis was performed using IBM SPSS Version 24 (IMB Corp., Armonk, NY, USA). For the descriptive analysis of continuous variables, we calculated standard indexes, such as the mean, median, standard deviation, minimum, maximum, range and quartiles. We examined the association between complications and clinical predictors by using logistic regression. Numerical variables among the groups were analyzed and compared using either with independent-samples t test or Mann–Whitney U test depending on the normality. P value < 0.05 was considered to indicate statistical significance.

Analisi statistica

- Descrizione dei metodi di analisi statistica
- Eventuali analisi ad interim
- Giustificazione della numerosità campionaria
- Popolazione da analizzare
- Trattamento dati mancanti

Correlazione e Causalità

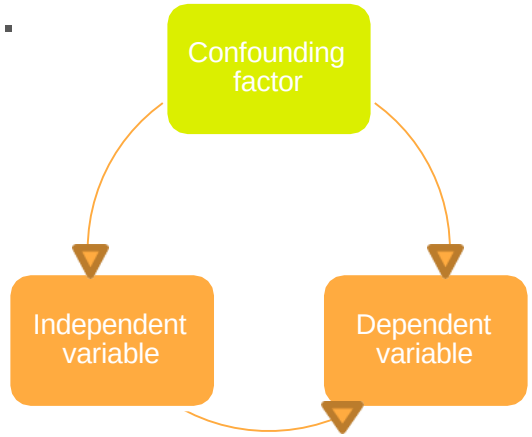


I 9 criteri di Bradford Hill

1. **Forza** dell'associazione tra esposizione e malattia (forte aumento dell'incidenza negli esposti rispetto all'incidenza nei non esposti)
2. **Consistenza** dell'associazione (studi condotti in luoghi e tempi diversi e da ricercatori diversi riportano gli stessi risultati)
3. **Specificità** (relazione caratteristica tra una specifica esposizione e una specifica malattia)
4. **Temporalità** (la causa deve precedere l'effetto)
5. **Gradiente biologico** (l'incidenza della malattia aumenta con l'aumentare dell'intensità dell'esposizione)
6. **Plausibilità** (spiegazione plausibile del modo in cui l'esposizione determina la malattia)
7. **Coerenza** (il meccanismo con cui riteniamo che l'esposizione determini la malattia è coerente con le conoscenze biomediche consolidate)
8. **Evidenza sperimentale** (l'incidenza della malattia diminuisce con la riduzione dell'intensità dell'esposizione)
9. **Analogia** (l'associazione che osserviamo tra esposizione e malattia è analoga ad altre in cui il rapporto causale è già accettato)

Confondimento

Presenza di altre variabili associate all'esposizione e che hanno effetti sull'outcome di interesse.



- Restriction (exclusion)
- Stratification
- Multivariable analysis (ex. Propensity score)

Propensity score

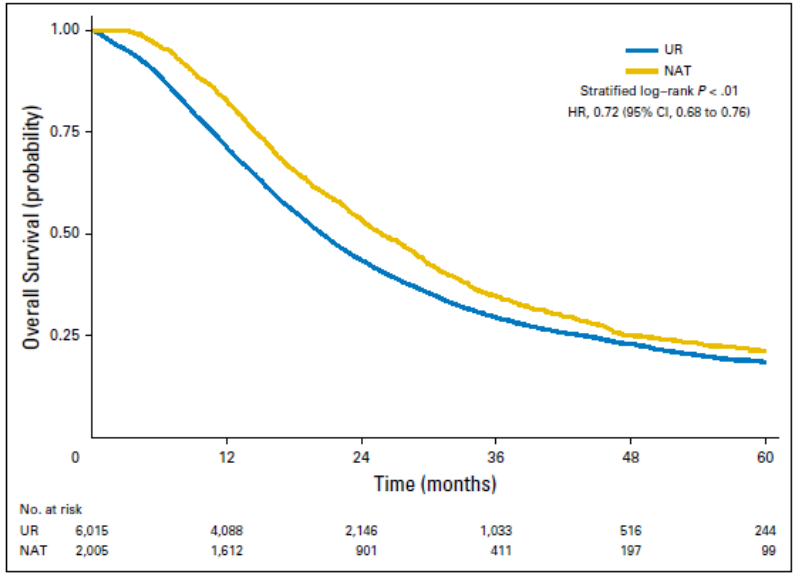
- “For confounding bias, the relevant question is: why did a patient receive one particular drug over any other?” (Haneuse S. Research. Med Care. 2016)
- In un contesto non randomizzato, si applica il PS per limitare lo squilibrio tra due o più gruppi nei fattori prognostici nella stima dell’effetto del trattamento;
- Score calcolato per ciascun soggetto in studio, rappresenta la probabilità di essere trattato, date le covariate di baseline osservate;
- Applicazione di matching, weighting, adjusting e relativa diagnostica;
- Analisi dei risultati dello studio;

Example: main results

- Confrontati con il gruppo NAT, i pazienti che eseguono una chirurgia upfront hanno un pT più alto e più frequentemente hanno i linfonodi e i margini di resezione positivi;
- Gli indicatori di riammissione e mortalità a 30/60 giorni sono simili tra i due gruppi;
- OS del gruppo NAT più alta rispetto al gruppo UR (26 vs 21 months);
- Gli autori concludono che, vista la mancanza di trial randomizzati sul quesito clinico, hanno effettuato una analisi attraverso l'analisi del PS match al fine di replicare uno scenario realistico di due campioni trattati con due trattamenti diversi. I risultati mostrano come effettuare una terapia neoadiuvante seguita da chirurgia porti ad un aumento della sopravvivenza.

Table 2. Comparison of Surgical and Early Postoperative Outcomes in the Matched Data Set

Variable	UR (n = 6,105)	NAT (n = 2,005)	*P*
Pathologic T stage			< .01
pT0/Tis	22 (0)	47 (2)	
pT1	171 (3)	212 (11)	
pT2	667 (11)	291 (15)	
pT3	4,728 (79)	1,193 (60)	
pT4	427 (7)	262 (13)	
Positive lymph nodes	4,306 (73)	932 (48)	< .01
Positive resection margin	1,417 (24)	335 (17)	< .01
Length of stay, mean ± SD, days	11 ± 10	11 ± 10	.86
30-day unplanned readmission	502 (9)	147 (8)	.17
30-day mortality	175 (3)	50 (3)	.16
90-day mortality	343 (6)	125 (6)	.22



Five-year cardiovascular outcomes in patients with chronic myeloid leukemia treated with imatinib, dasatinib, or nilotinib: A cohort study using data from a large multinational collaborative network

Rafael Amorim Belo Nunes[1]*,
Precil Diego Miranda de Menezes Neves[2],
Leandro Menezes Alves da Costa[1], Philip Bachour[3],
Marcelo José de Carvalho Cantarelli[1],
Gustavo Bernardes de Figueiredo Oliveira[2] and Álvaro Avezum Jr.[2]

Background: Breakpoint cluster region-Abelson gene (BCR-ABL) tyrosine kinase inhibitors (TKIs) have revolutionized the treatment of patients with chronic myeloid leukemia (CML). However, concern has arisen about the cardiac safety profile of these drugs.

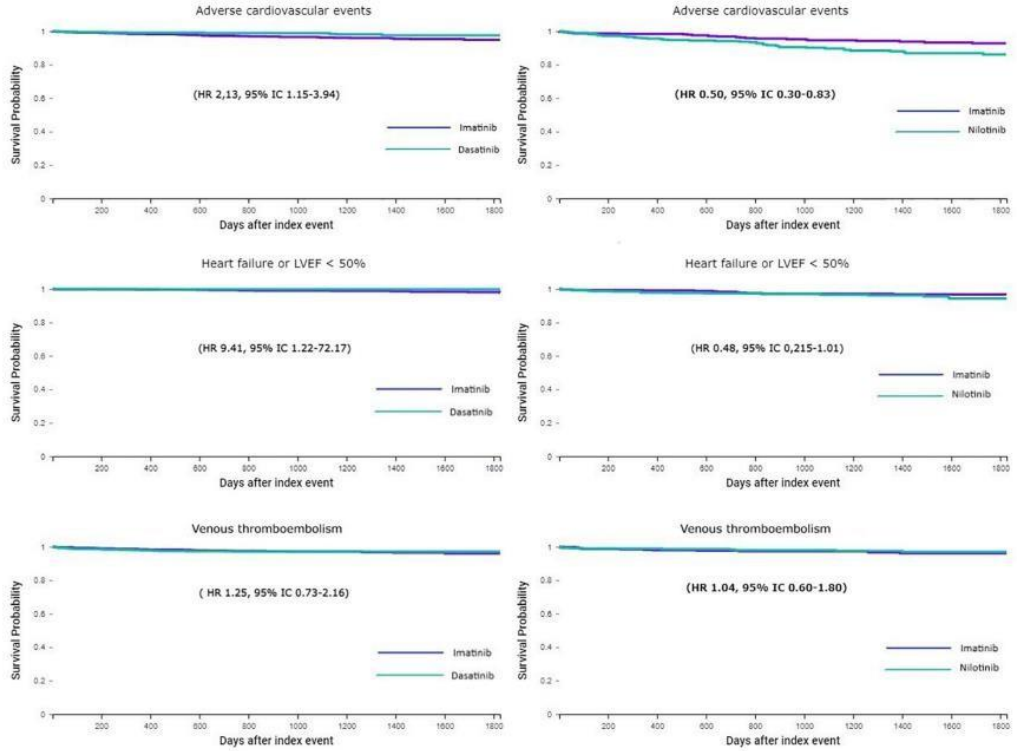
Objectives: This study aims to compare long-term risks of adverse cardiovascular and cerebrovascular events (ACE), heart failure or left ventricular ejection fraction (LVEF) < 50%, and venous thromboembolic events (VTE) in patients with CML treated with BCR-ABL TKIs, using data from a large multinational network.

Methods: Patients aged ≥ 18 years with CML treated with imatinib, dasatinib, or nilotinib without prior cardiovascular or cerebrovascular disease were included. We used propensity score matching to balance the cohorts. The 5-year cumulative incidences and hazard ratios were calculated.

Results: We identified 3,722 patients with CML under treatment with imatinib (n = 1,906), dasatinib (n = 1,269), and nilotinib (n = 547). Patients with imatinib compared to dasatinib showed a higher hazard ratio (HR) for ACE (HR 2,13, 95% CI 1.15–3.94, p = 0.016). Patients with imatinib presented a lower HR than nilotinib for ACE (HR 0.50, 95% CI 0.30–0.83, p = 0.0074). In relation to heart failure or LVEF < 50%, patients with imatinib had a higher HR than dasatinib (HR 9.41, 95% CI 1.22–72.17, p = 0.03), but no significant difference was observed between imatinib and nilotinib (HR 0.48, 95% CI 0.215–1.01, p = 0.064).

TABLE 1 Baseline characteristics of the cohort for the imatinib, dasatinib, and nilotinib groups.

	Imatinib	Dasatinib	*P*-value*	Nilotinib	*P*-value**
Cohort size, *n*	1906	1269		547	
Demographics					
Mean age (SD), years	55 (16)	47.7 (15)	<0.0001	53 (15)	0.0086
Sex					
Male, *n* (%)	1029 (54)	700 (55)	NS	261 (48)	0.0092
Female, *n* (%)	877 (46)	569 (45)	NS	285 (52)	0.0073
Race					
White, *n* (%)	1323 (69)	820 (65)	0.0281	362 (66)	NS
Black or African American, *n* (%)	215 (11)	185 (15)	0.0026	81 (15)	0.0230
Asian, *n* (%)	35 (2)	30 (2)	NS	10 (2)	NS
Unknown, *n* (%)	323 (17)	223 (18)	NS	97 (18)	NS
Comorbidities					



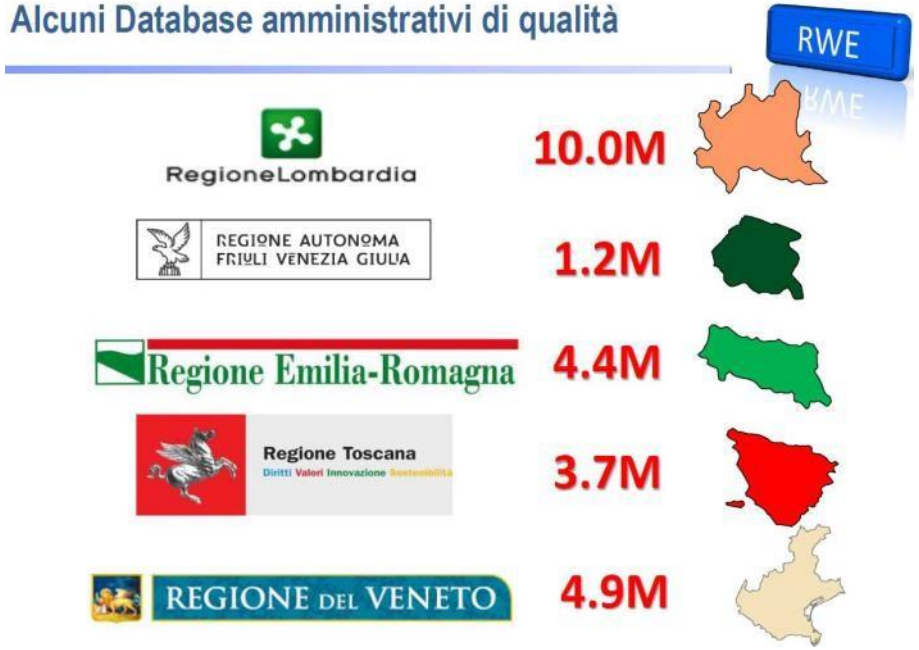
Metodi per l’analisi dei missing data

- Metodi “semplici”
  - Complete-case analysis
  - Last observation carried forward (LOCF)
  - Imputazione tramite media
  - Imputazione tramite regressione
- Metodi Complessi
  - Imputazione Multipla
  - Aggiustamento per variabili di baseline

RWE – BIG DATA

Il percorso di cura e la registrazione di informazioni cliniche

Alcuni Database amministrativi di qualità

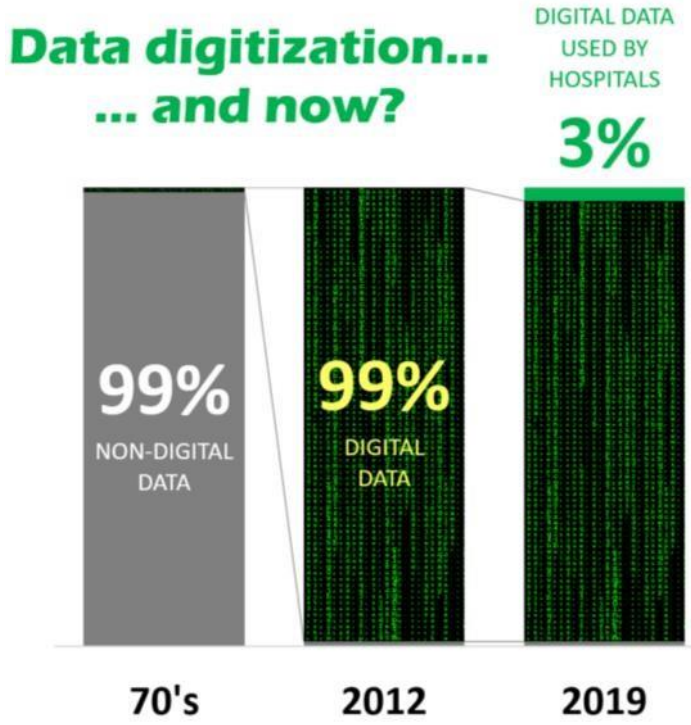


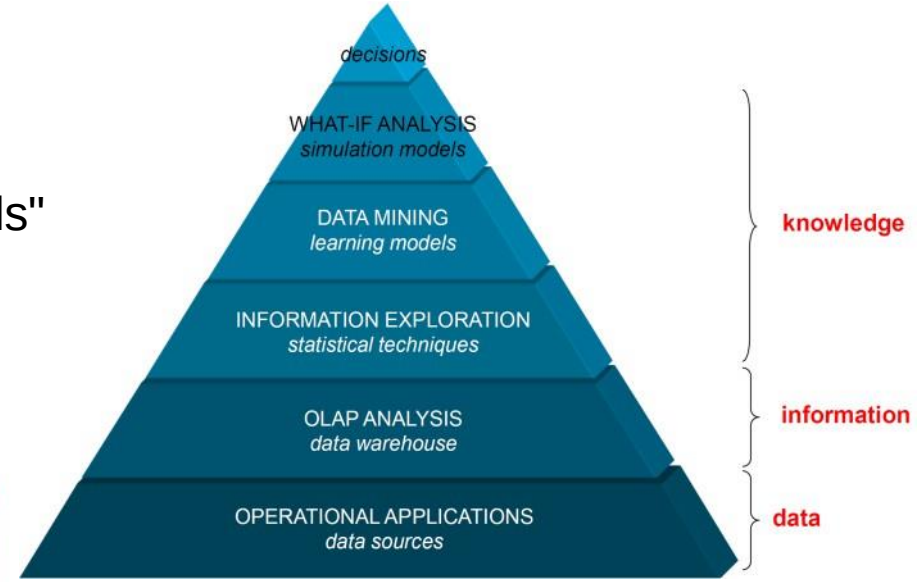
Immagini, ngs……

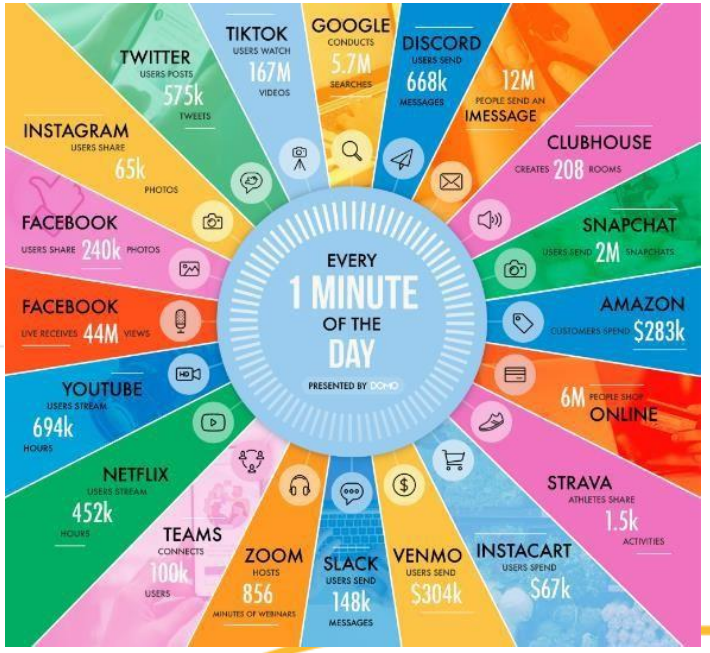
BIG DATA:

Noun: “We have *big data*"

Adjective: “We use *big data* tools"







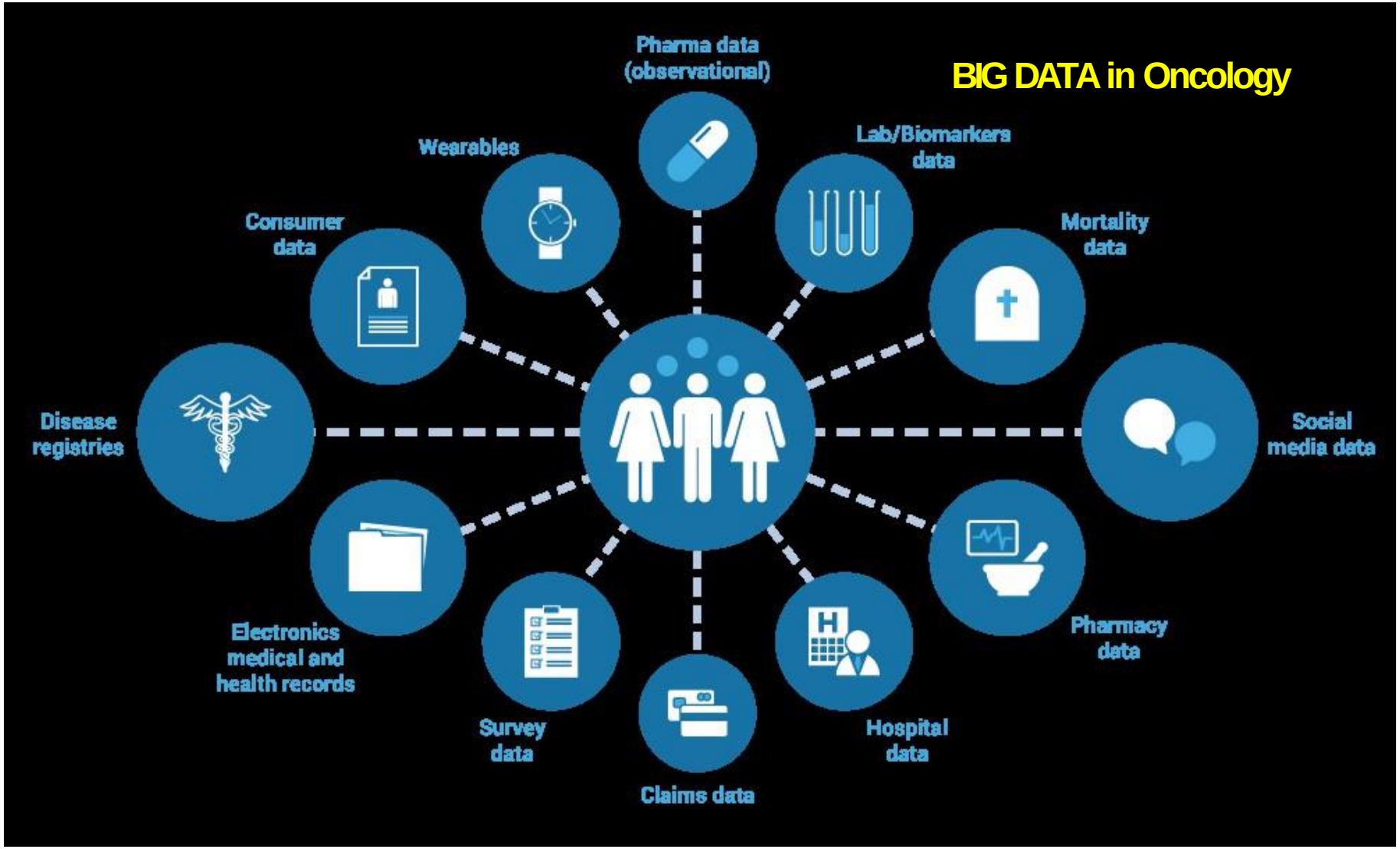
BIG DATA – The V’s

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


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

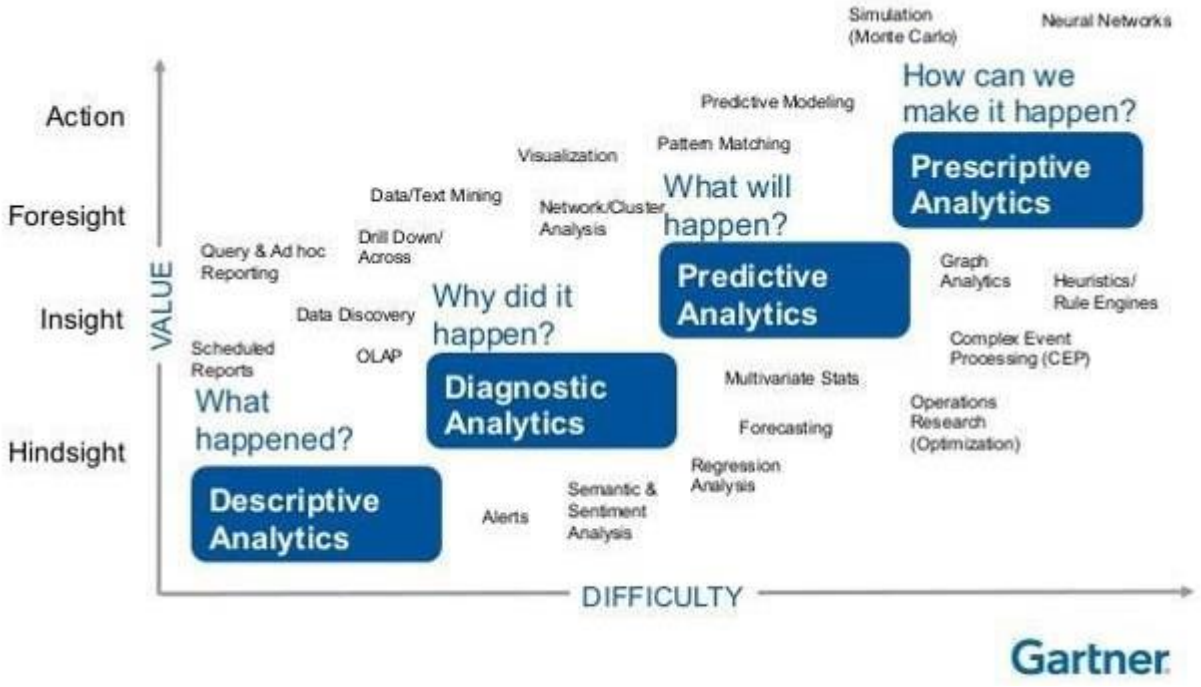
Analytics

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



«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:

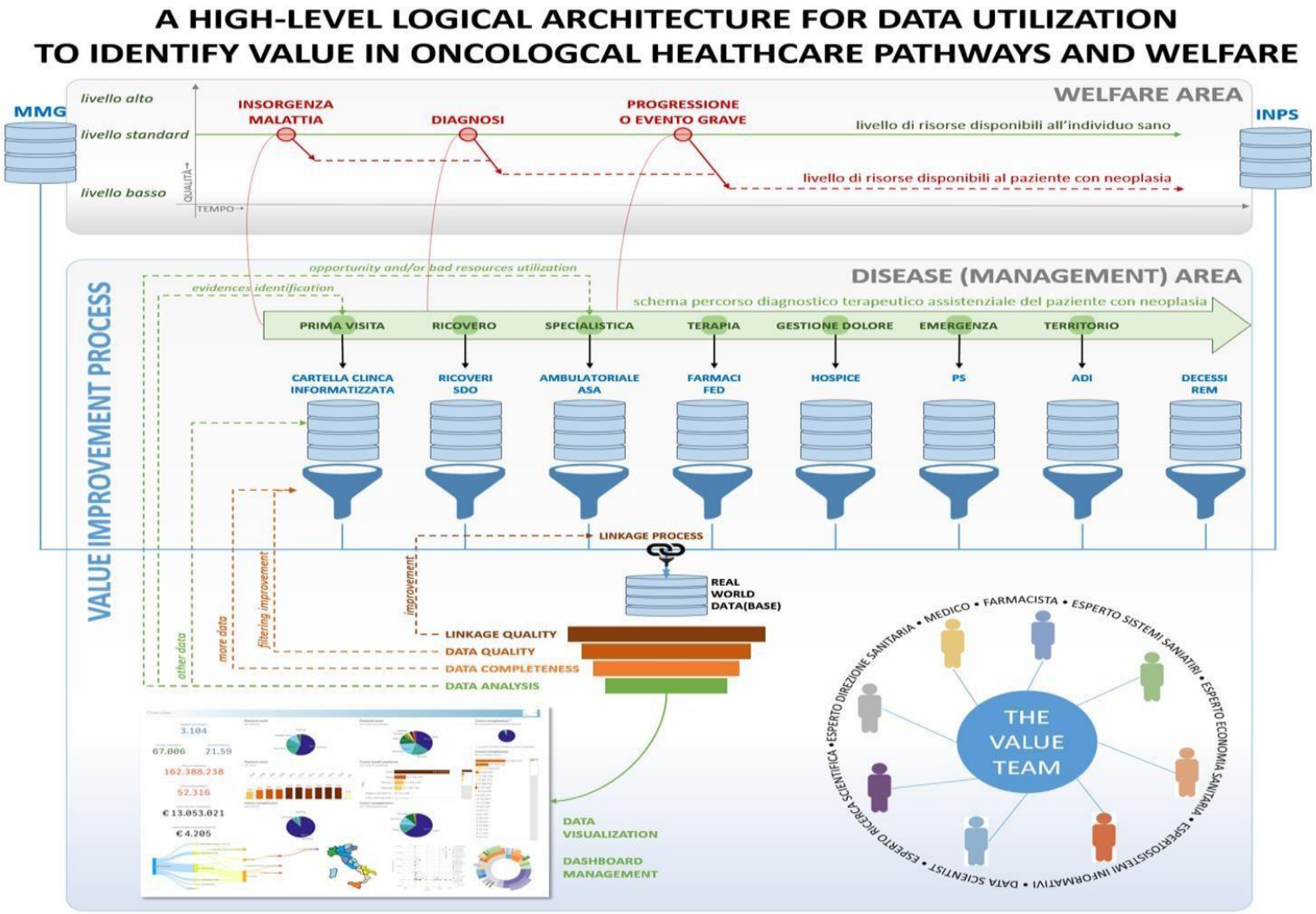
- Data Collection and Processing
- Programming
- Machine Learning
- Data Munging
- Data Visualization
- Applying Statistical Analysis

Data Scientist

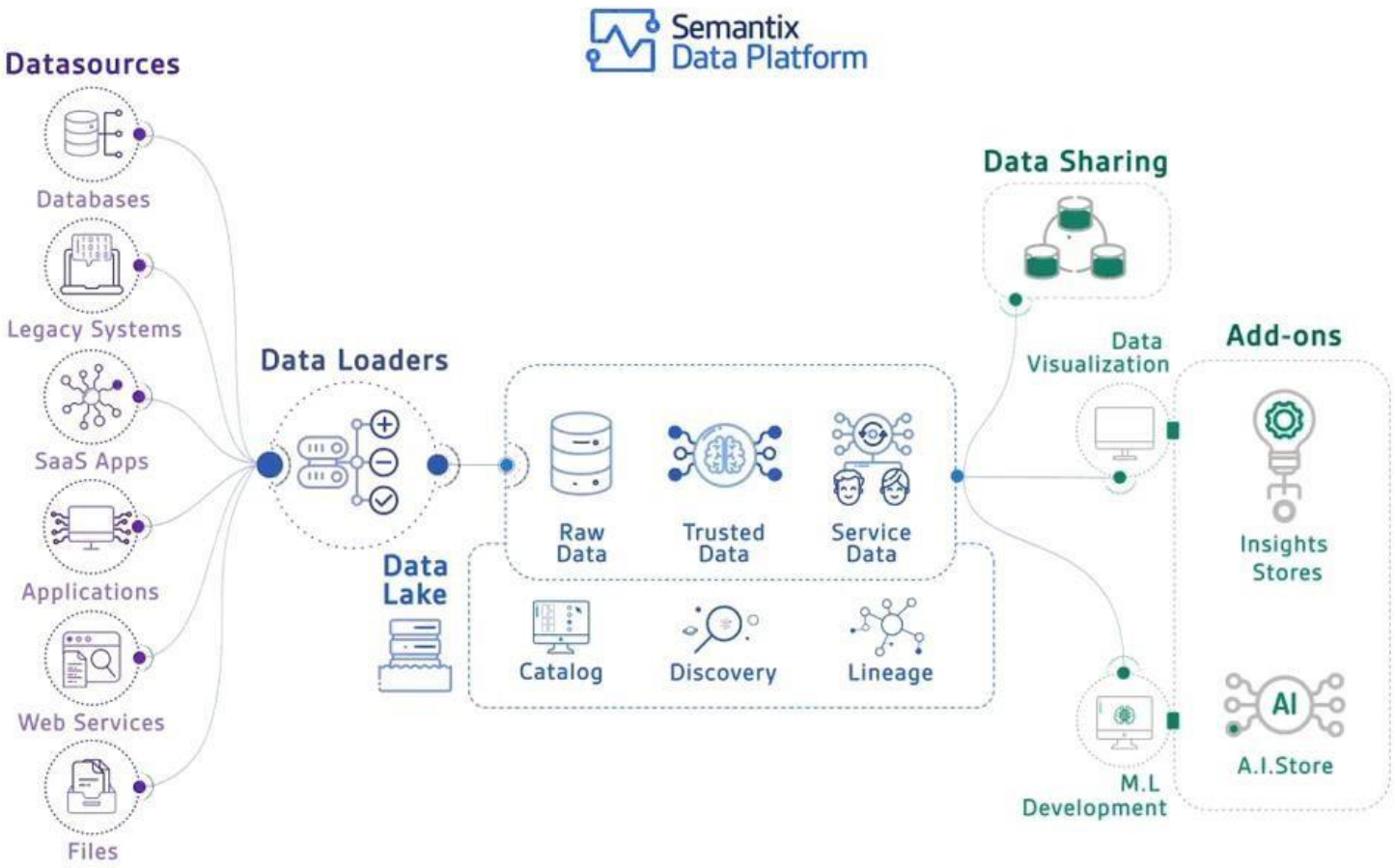
Data analysis once data volume and velocity reaches a level requiring sophisticated technical skills. Core tasks are:

- 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



DATA UNIT: STRUTTURA

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

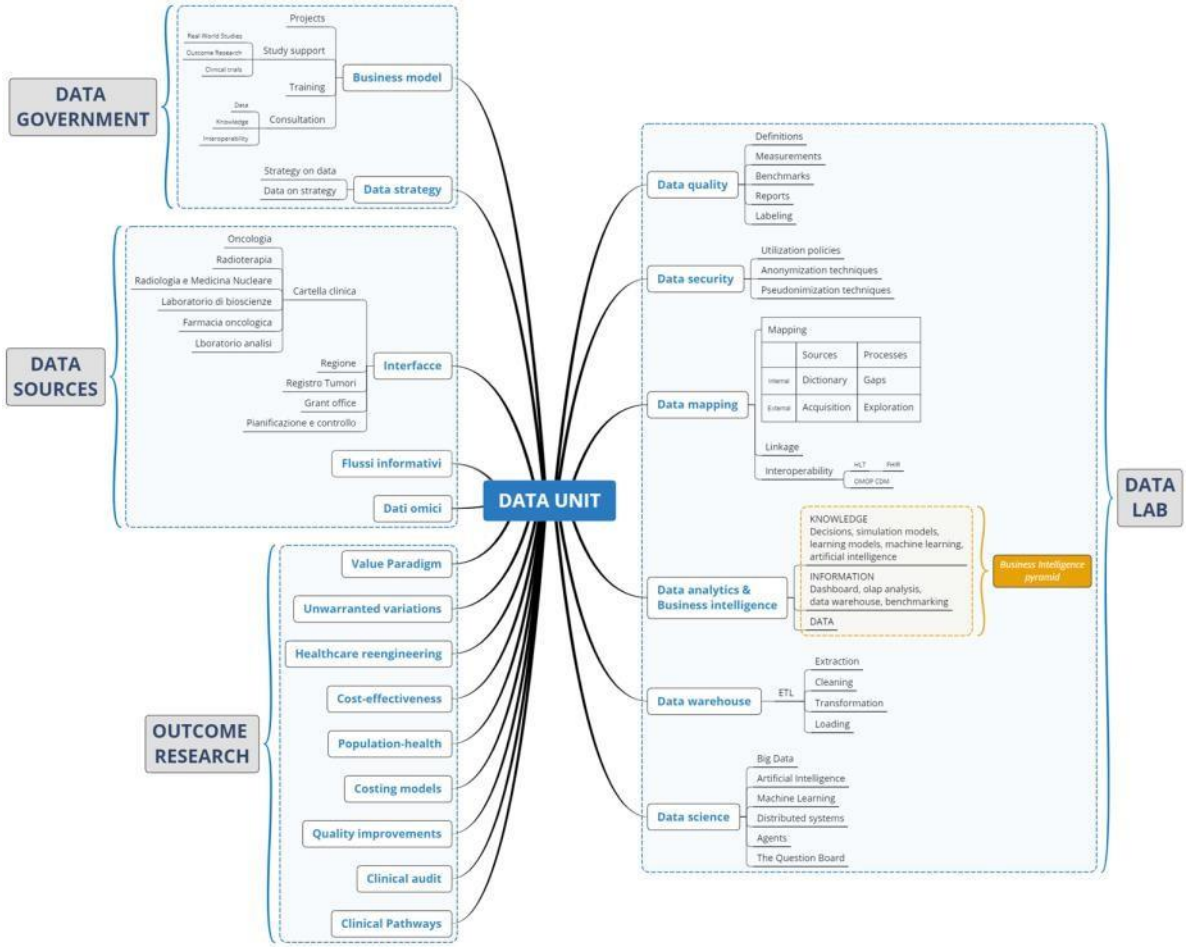
DATA GOVERNMENT

DATA SOURCES

DATA LAB

OUTCOME RESEARCH

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

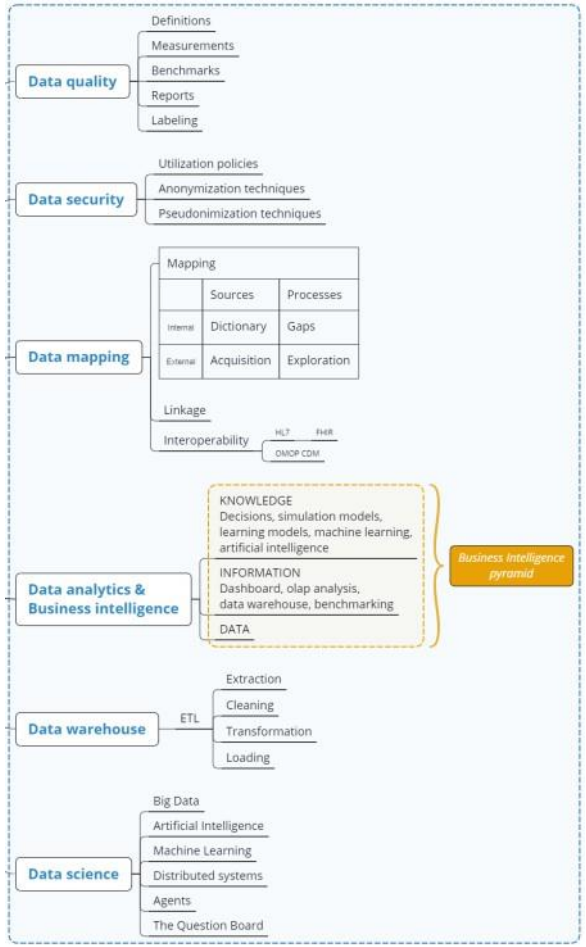


DATA LAB

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'



DATA QUALITY

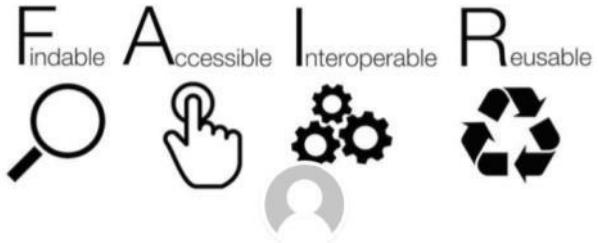
Data quality definition from ISO 9000:2015:

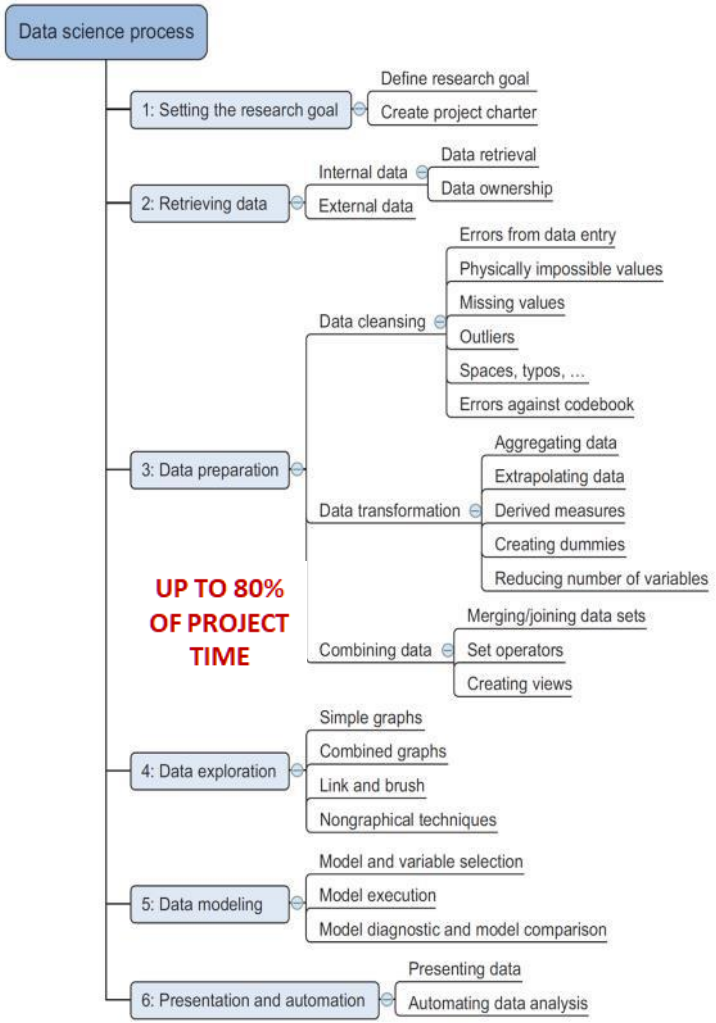
“**degree to which a set of characteristics of data fulfills requirements”**.

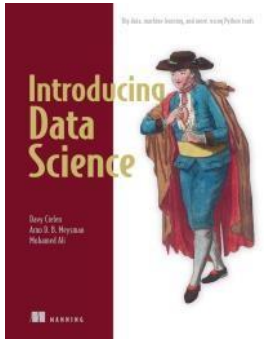
Requirements are defined as the need or expectation that is stated, generally implied or obligatory.

Some **Data quality** *suggestions*:

- Degree of excellency, consistency, timeliness, accuracy
- Ability of data to satisfy a given purpose



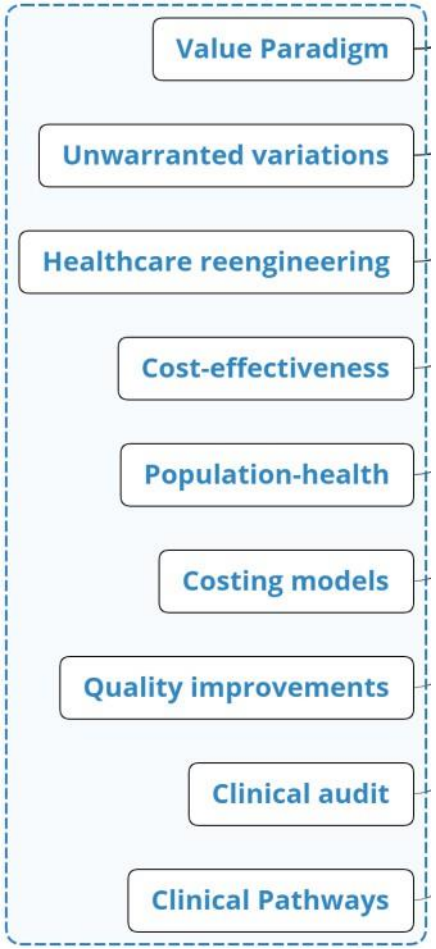




OUTCOME RESEARCH

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.

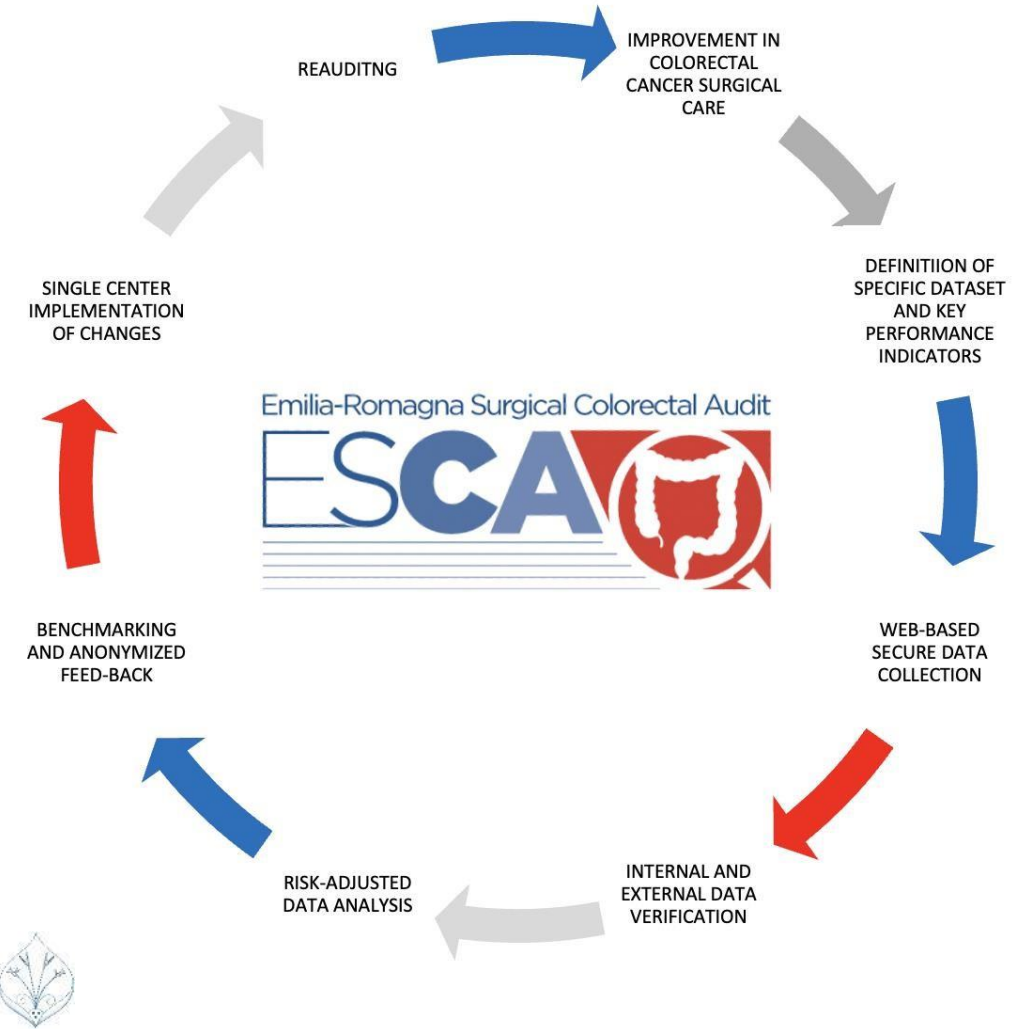


Emilia-Romagna Surgical Colorectal Audit (ESCA)

"*Improving outcomes in cancer patients*

throughout a collaborative

and systematic auditing activity"

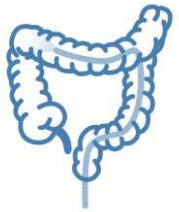


Emilia-Romagna Surgical Colorectal Audit

ESCA

1. PREOPERATIVE ASSESSMENT

2. DIAGNOSTIC WORK UP



3. SURGERY

4. POST-OPERATIVE COURSE

5. PATHOLOGY

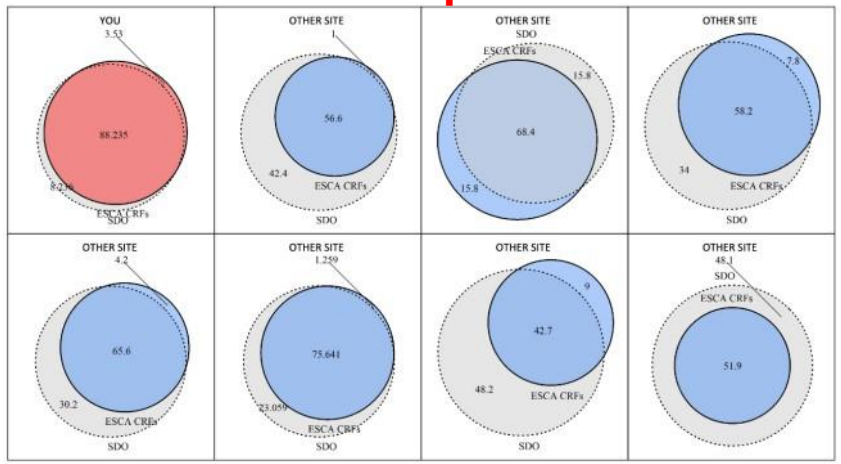
6. POST-OPERATIVE TREATMENT

7. FOLLOW-UP

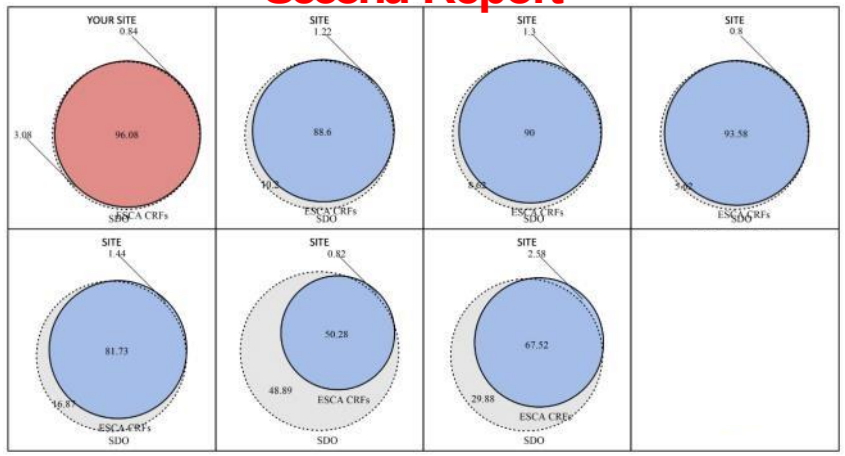
SERVIZIO SANITARIO REGIONALE EMILIA-ROMAGNA
Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori

ISTITUTO SCIENTIFICO ROMAGNOLO PER LO STUDIO E LA CURA DEI TUMORI

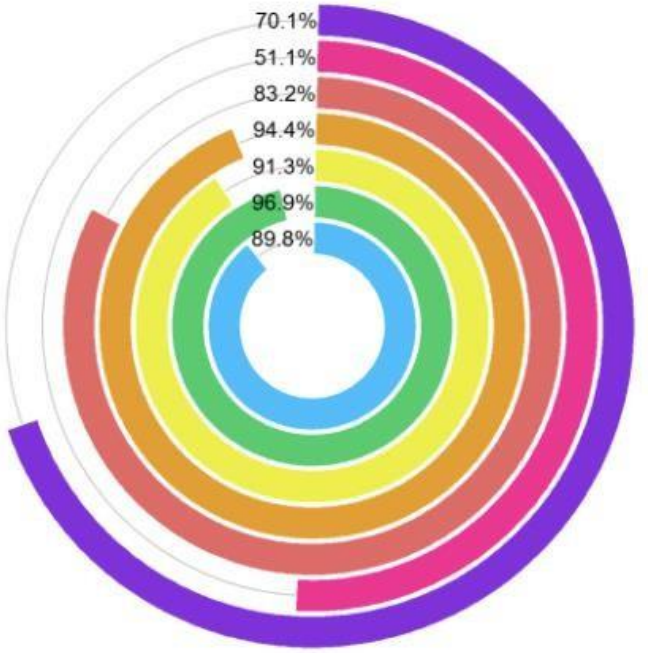
First Report



Second Report



Overall enrollment rate



74.6% → Feedback on enrollment → **82.8%**

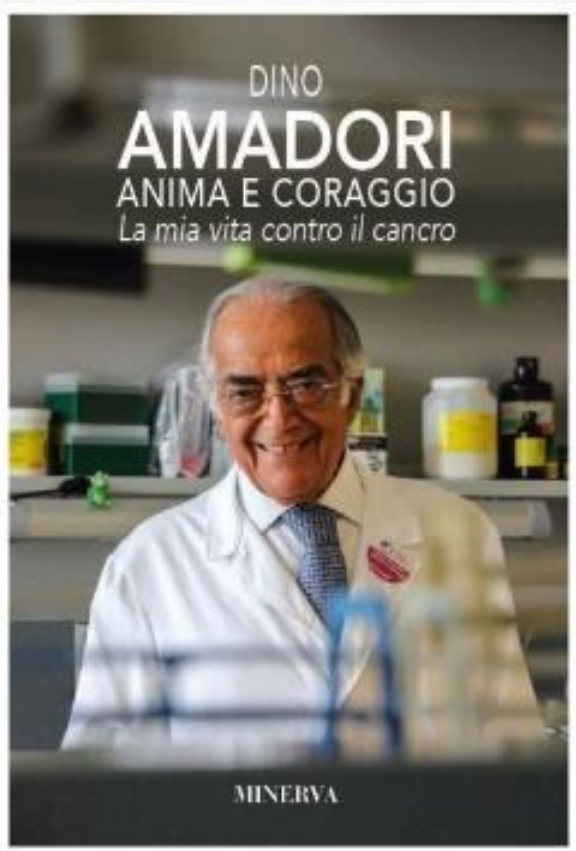
- **I quattro pilastri della Real World Evidence**

- **Real-World Data: infrastruttura e qualità**
- **Framework regolatorio**
- **Cultura, partnership e nuove competenze per garantire un nuovo ecosistema dati**
- **Generazione di Real-World Evidence di valore**

«…. L'oncologia è sicuramente la disciplina che più di altre ha cambiato il modo di fare medicina e di fare ricerca clinica nel mondo … [la multidisciplinarietà, la interdisciplinarietà, la capacità di lavorare in gruppo, lo sviluppo di relazioni a doppia direzione tra i gruppi di ricerca e soprattutto la metodologia della ricerca stessa].

Ha fatto giustizia dei «secondo me», dei «ricordo un caso che» dei «se ve la dico io è cosi» eccetera.

L'oncologia ha introdotto il metodo scientifico non solo negli studi clinici ma anche nella pratica clinica, per cui ogni scelta terapeutica deve avere buonissimi livelli di evidenza disponibili in quel momento, derivati dagli studi clinici controllati e condotti secondo rigorosi criteri etici e metodologici …»



Grazie dell'attenzione

oriana.nanni@irst.emr.it

L'impatto degli studi osservazionali nella pratica clinica

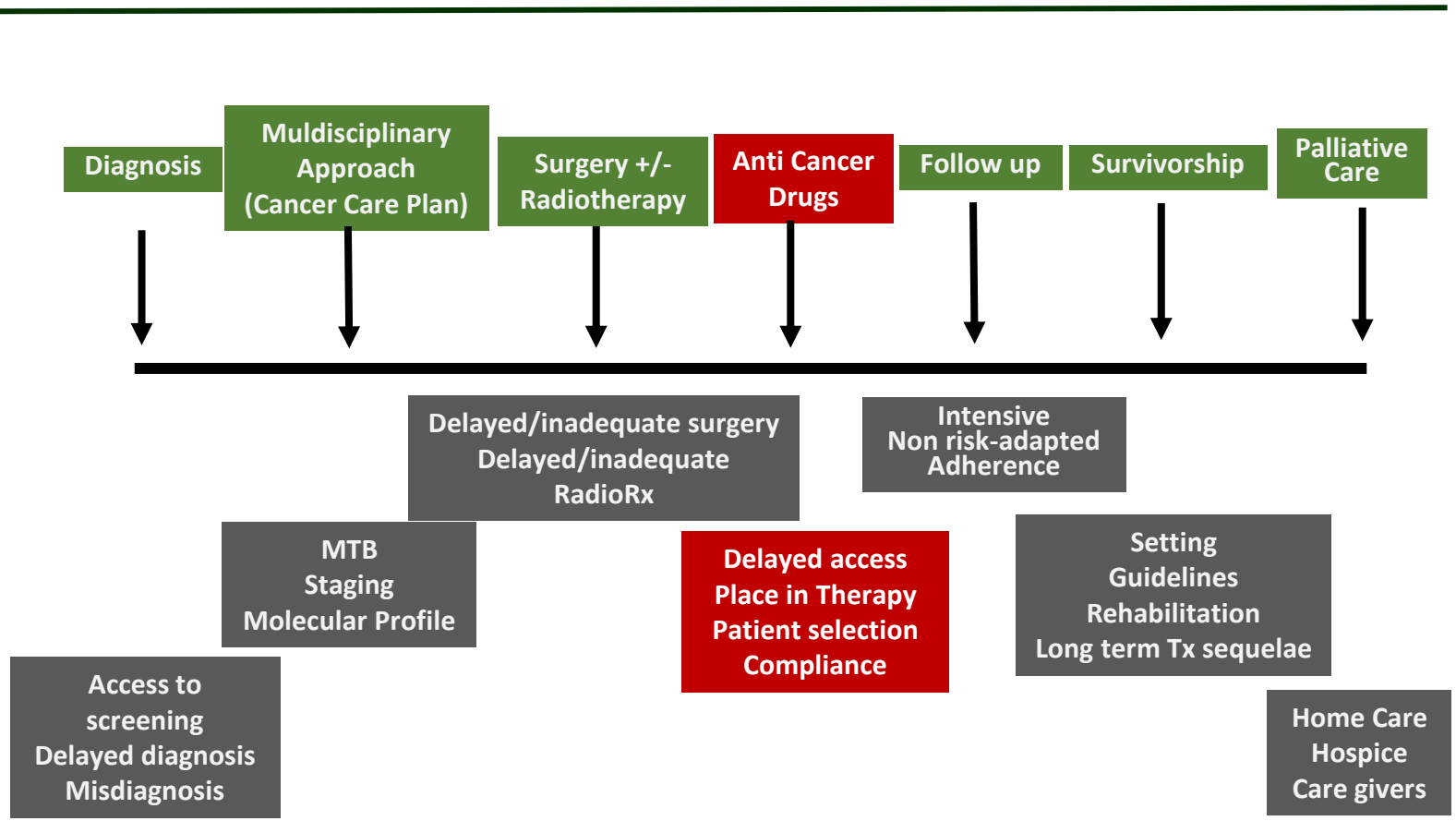
(P.F Conte)

Oncology at the Cross Roads

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

fondazione Periplo

Patients' Journey in Oncology

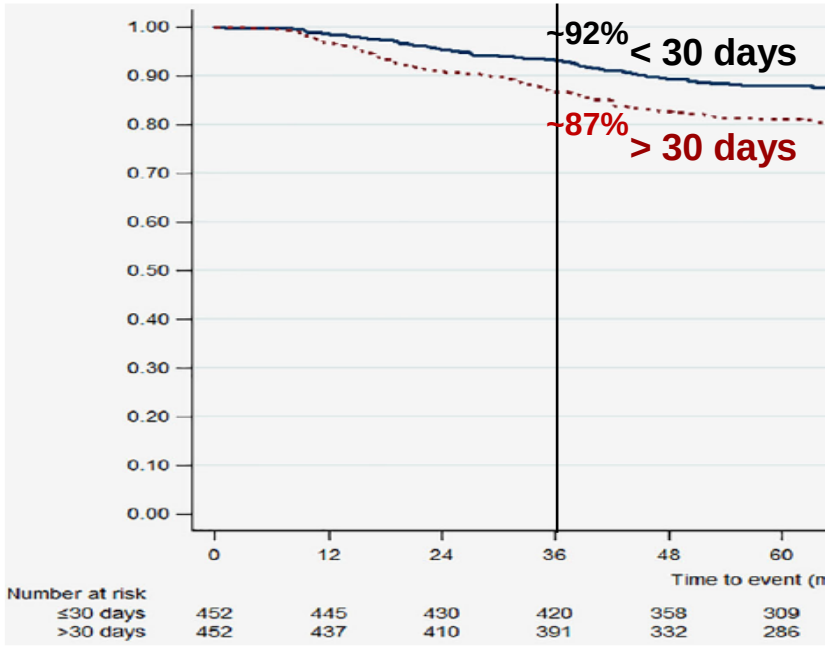


fondazione Periplo

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

Periplo Foundation
Outcome-COsts-CAncer Programme

Studio PERSEO (PERiplo-SEnO):

Confronto analitico per la valutazione e monitoraggio degli indicatori per la terapia adiuvante e neoadiuvante del tumore mammario ricavati da dati amministrativi elettronici da archivio sanitario centro-specifico

ISPRO
Istituto per lo studio, la prevenzione e la rete oncologica

ISTITUTO ROMAGNOLO PER LO STUDIO DEI TUMORI DINO AMADORI

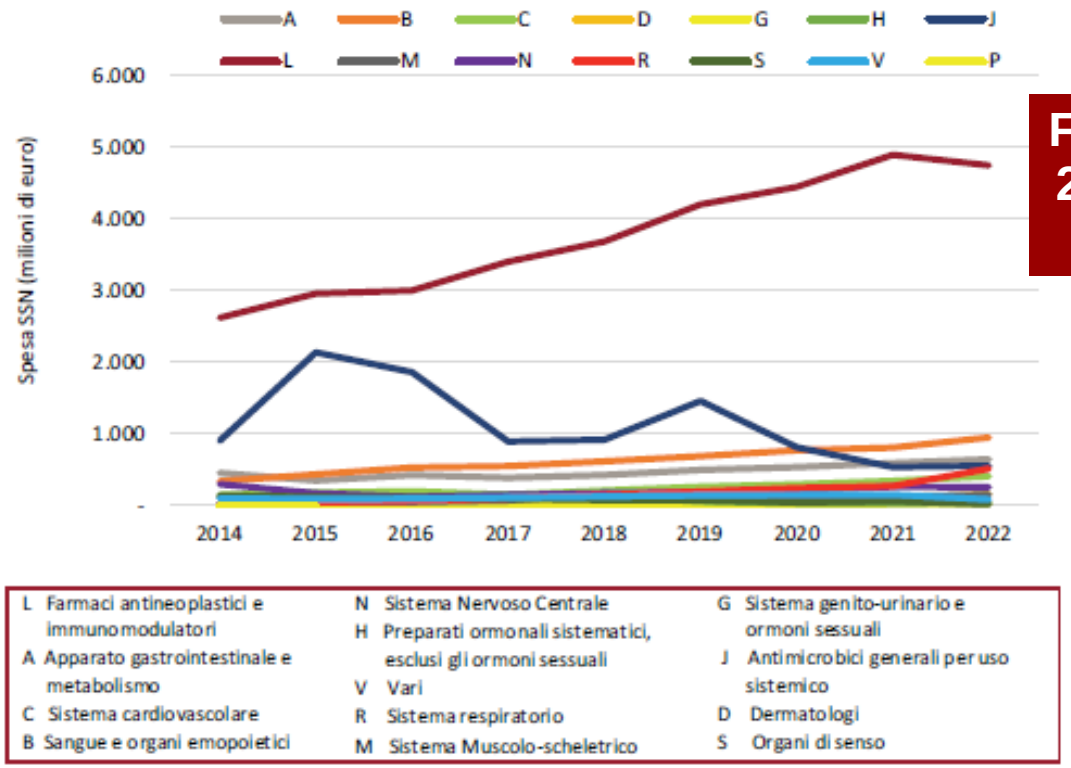
fondazione Periplo

Oncology at the Cross Roads

- **Diagnostic-therapeutic Pathways: markers and outcomes**
  - no new data from PNE or Regional Oncology Networks
  - Periplo Foundation, together with IRST and ISPRO, is carrying forward the PERSEO project (major barriers: funding, data protection officer)
- **Innovation is sustainable; way of funding the NHS is NOT**
- Evidence- based medicine: from efficacy to effectiveness
- Survivorship

fondazione Periplo

Spesa nuova entità terapeutiche



Farmaci Oncologici
2014 - 2022 +113 %
+ 9,9% annuo

L Farmaci antineoplastici e immunomodulatori	N Sistema Nervoso Centrale	G Sistema genito-urinario e ormoni sessuali
A Apparato gastrointestinale e metabolismo	H Preparati ormonali sistematici, esclusi gli ormoni sessuali	J Antimicrobici generali per uso sistemico
C Sistema cardiovascolare	V Vari	D Dermatologi
B Sangue e organi emopoietici	R Sistema respiratorio	S Organi di senso
	M Sistema Muscolo-scheletrico	

Periplo Foundation
Outcome-COsts-CAncer Programme

O-CO-CA Project

**(impact of diagnostic&therapeutic innovation
on health outcome, costs and quality of care in NSCLC patients)**

Studio osservazionale di coorte che prevede la raccolta di tutti i casi di NSCLC diagnosticati in Regione Veneto in tre anni differenti (2017,2019,2021) per valutare:

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

Sources of Data

Identificazione dei casi incidenti tra i residenti nelle ULSS Venete

Caratterizzazione dei casi: creazione del registro ad alta risoluzione

Calcolo degli indicatori e dei costi

Data links and interpretation:
a multidisciplinay team including physicians, pathologists, pharmacists and epidemiologists

SDO

Schede di morte

Cartelle cliniche

Referti esami radiologici

Schede dimissione ospedaliera (SDO)
Specialistica ambulatoriale (SPS)
Farmaceutica (DDF e CINECA)
Assistenza domiciliare integrata (ADI)
Certificati di morte
Hospice
Dispostivi medici
Pronto soccorso

cancers

MDPI

Article

NON-SMALL-CELL LUNG CANCER: Real-World Population-Based Cohorts' Study

Alessandra Buja [1,*], Massimo Rugge [2], Alberto Bortolami [3], Manuel Zorzi [4], Federico Rea [1], Anna Zanovello [1], Giovanna Scroccaro [3], Pierfranco Conte [5,6], Giulia Pasello [7,8], Valentina Guarneri [7,8], on behalf of Rete Oncologica Veneta and Periplo Foundation

Cancers, 2025

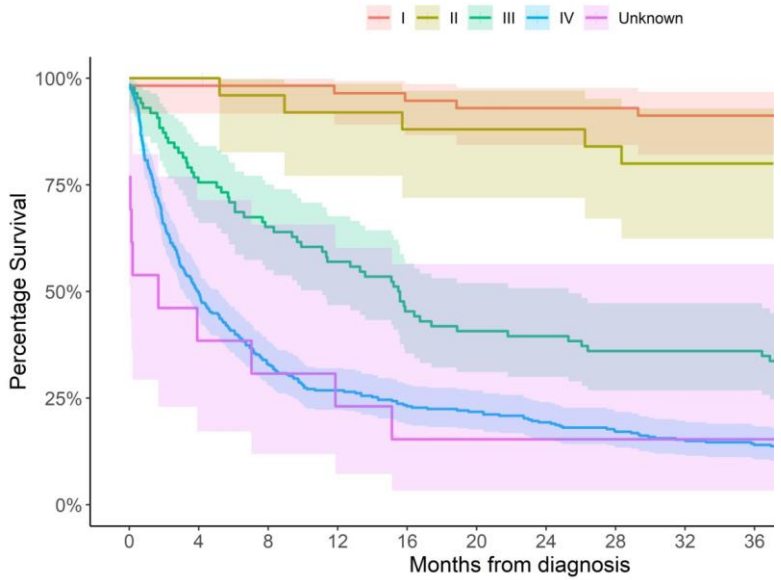
Non–Small-Cell Lung Cancer: un update Real-World Cost Consequence Analysis

Mean costs into three years from diagnosis by year and difference in costs								
	Total	Inpatient drugs	Outpatient services	Hospitalisation	Hospice	Emergency Dept visits	Outpatient drugs	Medical devices
2017	45590.46	20718.6	9161	11469.2	1221.97	623	1581.8	814.5
2019	47846.74	2226.6	10805	10968.7	603.13	683.4	1455.7	1065
Δ	2256.28	1547.96	1643	- 500.54	- 618.84	60.44	- 126.08	250.21

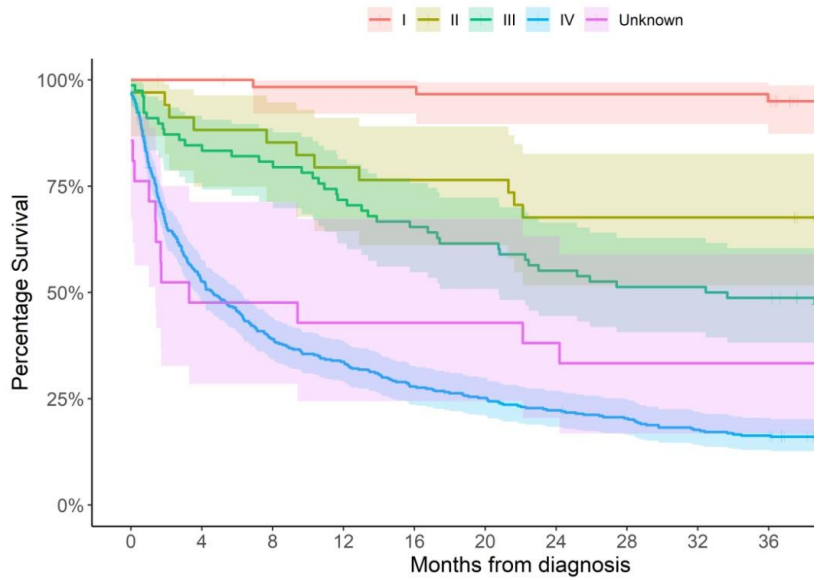
fondazione Periplo

OCOCA- Lung
NSCLC: OS by stage and year of diagnosis

2017



2019



Significant overall OS improvement in 2019 cohort: HR 0.84 (95%CI 0.72-0.98); p value 0.024.
Significant LCS-OS improvement in 2019 cohort for stage III disease: HR 0.61 (95% CI 0.41-0.91); p value 0.025

A Buja et al, Cancers 2025

Oncology at the Cross Roads

- **Diagnostic-therapeutic Pathways: markers and outcomes**
  - no new data from PNE or Regional Oncology Networks
  - Periplo Foundation, together with IRST and ISPRO, is carrying forward the PERSEO project (major barriers: funding, data protection officer)

- **Innovation is sustainable; way of funding the NHS is NOT**
  - health care costs of cancer patients and their outcome can be derived from multiple data bases
  - closed silos financing is inadequate to sustain NHS

- **Evidence- based medicine: from efficacy to effectiveness**

- **Survivorship**

fondazione Periplo

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

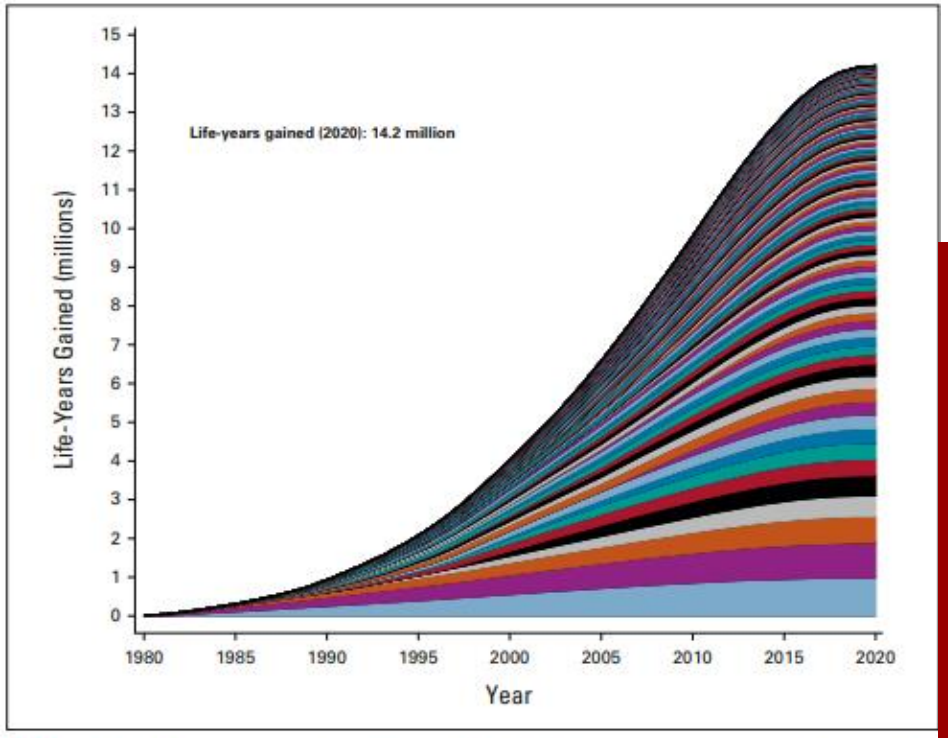
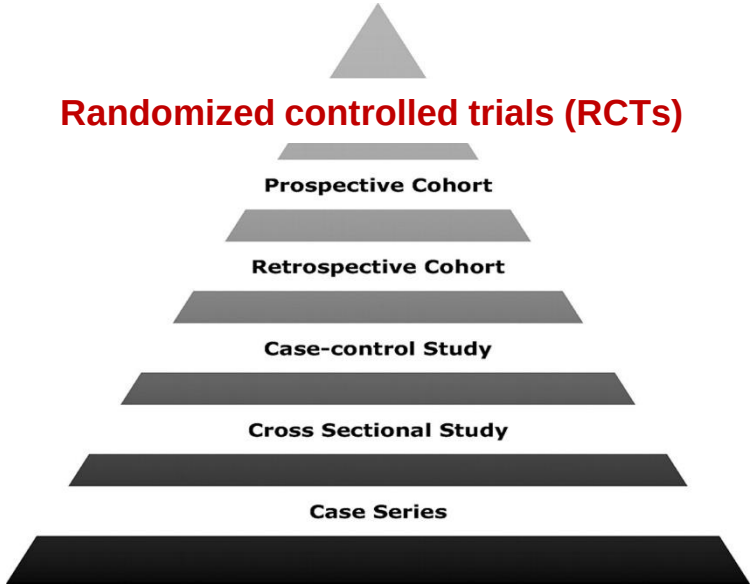


FIG 1. Cumulative life-years gained through 2020 by study. Each color-coded area represents cumulative life-years for 1 of 133 studies for which life-year gains were estimated.

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
- long term outcomes & toxicities rareley available
- include selected patients:
  - elderly, poor PS patients, comorbid patients are under-represented or excluded
  - differences in ethnic/racial composition
- budget impact estimates highly questionable

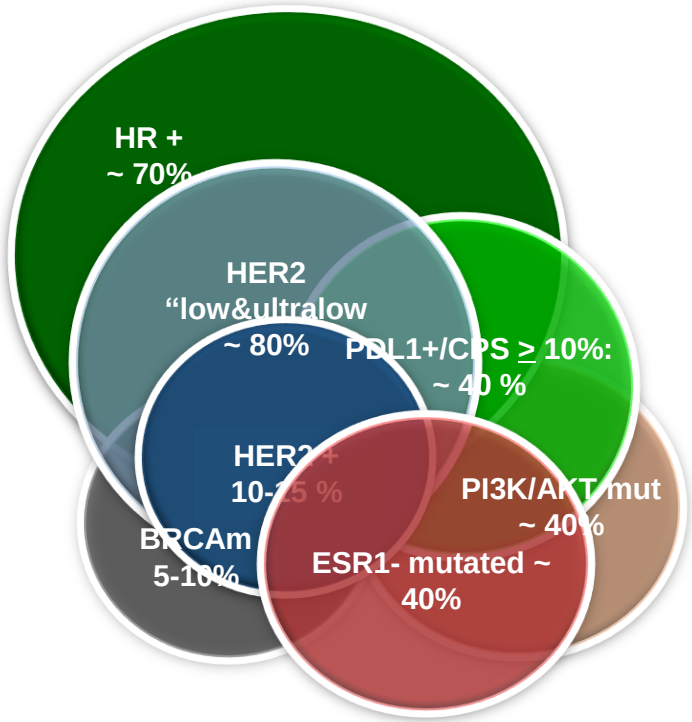
Unger JM et al, JCO 2023

SOUNDING BOARD

Obse
(i.e. I
Repl
false

ICI Cambridge Symposium 1985

Patients' Journey in Oncology – Breast Cancer Molecular Profile



+ GEPs
for HR+/HER2- early breast cancer

Qs to be addressed

- Who to test
- When to test
- How many times to test
- Where to test
- Why to test

fondazione Periplo

MINISTERO DELLA SALUTE

DECRETO 18 maggio 2021.

Modalità di riparto e requisiti di utilizzo del fondo per i test genomici ormonoresponsivo per il carcinoma mammario in stadio precoce.

BASSO RISCHIO	ALTO RISCHIO
Le seguenti 5 caratteristiche	Almeno 4 delle seguenti caratteristiche
G1 T1 (a-b)* Ki 67 <20% ER>80% N Negativo	G3 T3 T4 Ki 67>30% ER<30% N Positivo (>3 linfonodi non indicazione al *test*)
*In caso di T1a non è indicato l'accesso al *test* in presenza di almeno altri 2 parametri favorevoli	

Stadio I-IIIA, ER+/HER2-

Non indicazione per:

- ≥4 linfonodi positivi
- Basso o alto rischio
- Paziente non candidabile a CT
- Paziente che rifiuta CT

Utilizzo dei Tests Genomici in alcune Regioni			
Regione	**Previsti**	**2022**	**2023**
Veneto	915	61	398
Liguria	308	118	154
Toscana	654	261	322

fondazione Periplo

Aggiornamento criteri per l'accesso al test

Pubblicazione Marzo 2023

Se test urgente predittivo, invio da parte dell'oncologo, chirurgo o genetista.

Invio al test da parte dell'oncologo.

Consulenza genetica tradizionale.

Se test urgente predittivo, consulenza entro 1 settimana o invio da parte dell'oncologo o chirurgo.

Storia personale di:
Variante patogenetica nota in un gene predisponente in un familiare
Uomo con carcinoma mammario
Donna con carcinoma mammario e carcinoma ovarico
Donna con carcinoma mammario ≤40 anni
Donna con carcinoma mammario triplo negativo
Donna con carcinoma mammario bilaterale < 50 anni
Donna con carcinoma mammario in stadio iniziale a recettori ormonali positivi e ≥ 4 linfonodi positivi
Donna con carcinoma mammario a recettori ormonali positivi con precedente CT neoadiuvante, residuo di malattia e CPS/EG score≥3
Donna con carcinoma mammario metastatico recettori ormonali positivi/HER2-negativo già sottoposta a chemioterapia con antracicline/taxani e trattamento endocrino (qualora possibili), in progressione dopo inibitori di CDK 4/6 per la malattia avanzata.
Storia personale di carcinoma mammario 46-50 anni e familiarità di primo grado* per:
Carcinoma mammario <50 anni
Carcinoma ovarico non mucinoso o borderline a qualsiasi età
Carcinoma mammario bilaterale
Carcinoma mammario maschile
Carcinoma del pancreas
Carcinoma della prostata
Storia personale di carcinoma mammario >50 anni e familiarità per carcinoma mammario, ovarico, pancreatico in 2 o più parenti in primo grado* tra loro (di cui uno in primo grado con lei)

**Presenza di un familiare di primo grado (genitore, fratello/sorella, figlio/a) con le caratteristiche di malattia specificate. Per il lato paterno della famiglia, considerare anche familiari di secondo grado (nonna, zie).*

fondazione Periplo

A-BRAVE TRIAL: A PHASE III RANDOMIZED TRIAL WITH AVELUMAB IN EARLY TRIPLE NEGATIVE BREAST CANCER WITH RESIDUAL DISEASE AFTER NEOADJUVANT CHEMOTHERAPY OR AT HIGH RISK AFTER PRIMARY SURGERY AND ADJUVANT CHEMOTHERAPY

P Conte, MV Dieci, GC Bisagni, P Schmid, V Fotia, F Piacentini, M de Laurentiis, A Favaretto, S Tamberi, G Bianchi, C Zamagni, S Cinieri, D Corsi, L Del Mastro, A Ferro, A Gennari, M Mion, A Musolino, GL De Salvo, V Guarneri on behalf of A-BRAVE study team.

Medical Oncology 2, Istituto Oncologico Veneto IRCCS
DiSCOG-University of Padova, Italy

477 patients from 67 institutions
Accrual from june 2016 to october 2020
BRCA status known in 284 patients (59.5%)

MINISTERO DELLA SALUTE

DECRETO 30 maggio 2023.

Istituzione dei *Molecular tumor board* e individuazione dei centri specialistici per l'esecuzione dei test per la profilazione genomica estesa *Next generation sequencing* (NGS).

"Istituzione dei Molecular Tumor Board e individuazione dei centri specialistici per l'esecuzione dei test per la profilazione genomica estesa Next Generation Sequencing (NGS)"

2.ISTITUZIONE DEI MOLECULAR TUMOR BOARD NELL'AMBITO DELLE RETI ONCOLOGICHE REGIONALI

fondazione Periplo

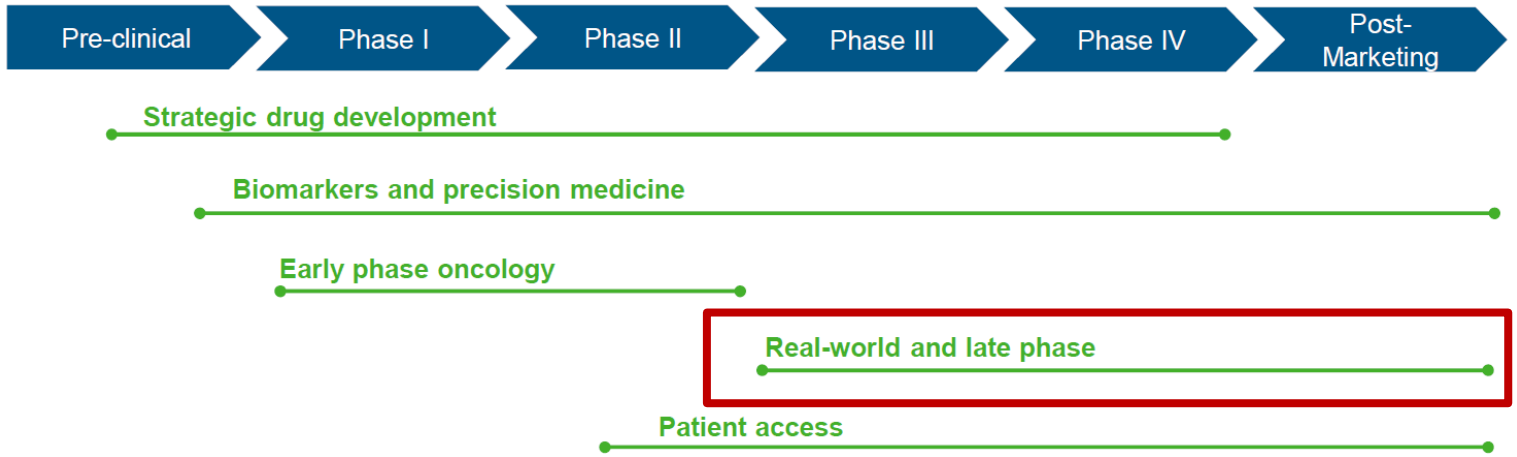
Innovative drugs and clinical research revolution

Trials with innovative drugs are conducted globally to assure rapid accrual and data acquisition

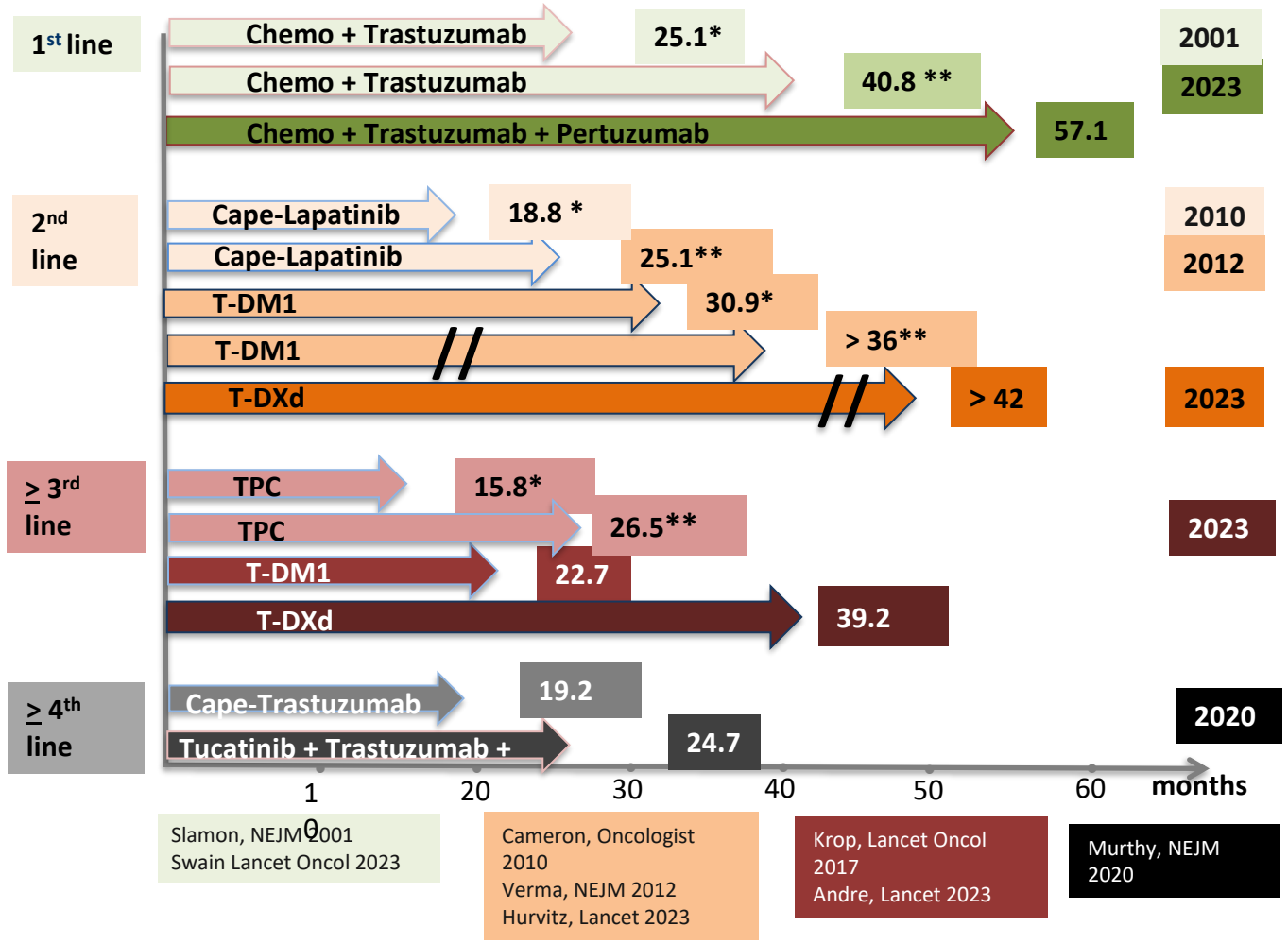
.

However:

- Impact on treatment sequence is unknown
- Global availability of effective therapies (beyond control arm) is unknown
- Reliability of molecular diagnostics outside central labs, is unknown
- Efficacy & tolerability in different ethnicities is largely unknown
- Multidisciplinary treatment of oligometastases or residual disease not considered
- Efficacy & tolerability of «maintenance» treatment not available



Advanced HER2+ BC: OS over time



OS (months) of HER2+ aBC over time

OS increases over time with the same treatment & line of therapy (* one/last line available; ** multiple lines available)

OS at ≥3rd line of therapy today longer than at 1st line in 2001

fondazione Periplo

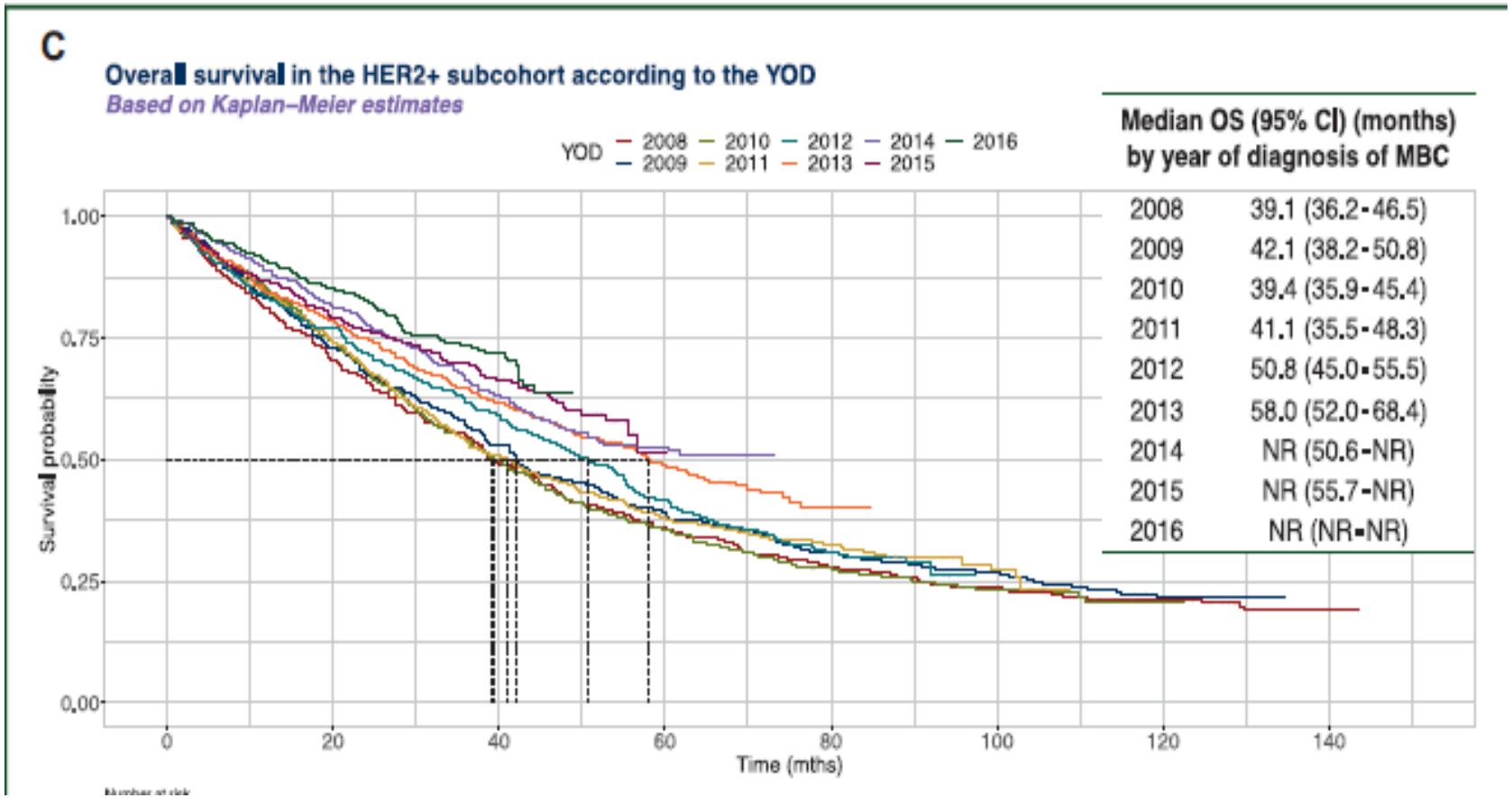
ESMO OPEN | CANCER HORIZONS

ORIGINAL RESEARCH

Evolution of overall survival and receipt of new therapies by subtype among 20 446 metastatic breast cancer patients in the 2008-2017 ESME cohort

T. Grinda[1], A. Antoine[2], W. Jacot[3], C. Blaye[4], P.-H. Cottu[5], V. Diéras[6], F. Dalenc[7], A. Gonçalves[8], M. Debled[3], A. Patsouris[9], M.-A. Mouret-Reynier[10], A. Mailliez[11], F. Clatot[12], C. Levy[13], J.-M. Ferrero[14], I. Desmoulins[15], L. Uwer[16], T. Petit[17], C. Jouannaud[18], M. Lacroix-Triki[19], E. Deluche[20], M. Robain[21], C. Courtinard[21,22,23], T. Bachelot[24], E. Brain[5], D. Pérol[2] & S. Delaloge[2*]

ESMO Open 2021



DESTINY-Breast09:
a phase III, randomized study of T-DXd with or without pertuzumab in the first-line treatment of HER2-positive metastatic breast cancer

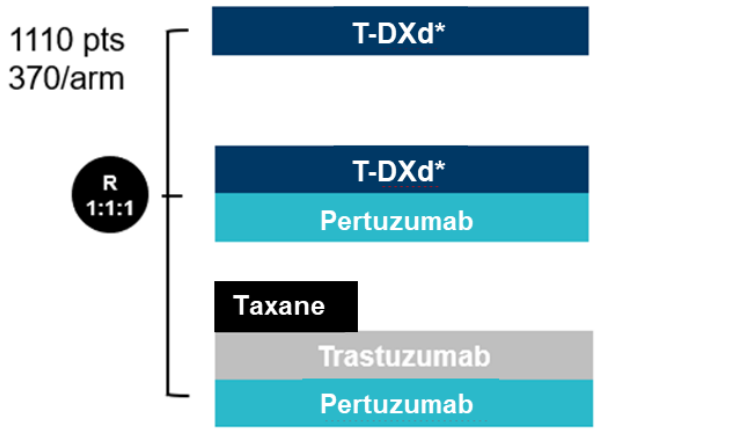
POPULATION

- HER2+ MBC
- DFI >12 months from last chemo or HER2-targeted therapy
- No prior systemic treatment for mBC

Stratification factors under consideration:

- *De novo* vs. recurrent (cap at 50% *de novo*)
- PIK3CA mut.(TBD)
- HR pos vs. neg

STUDY DESIGN



* If T-DXd is discontinued due to toxicity, patients may continue with trastuzumab if there is no contraindication.

ENDPOINTS

Primary:
- PFS (BICR)

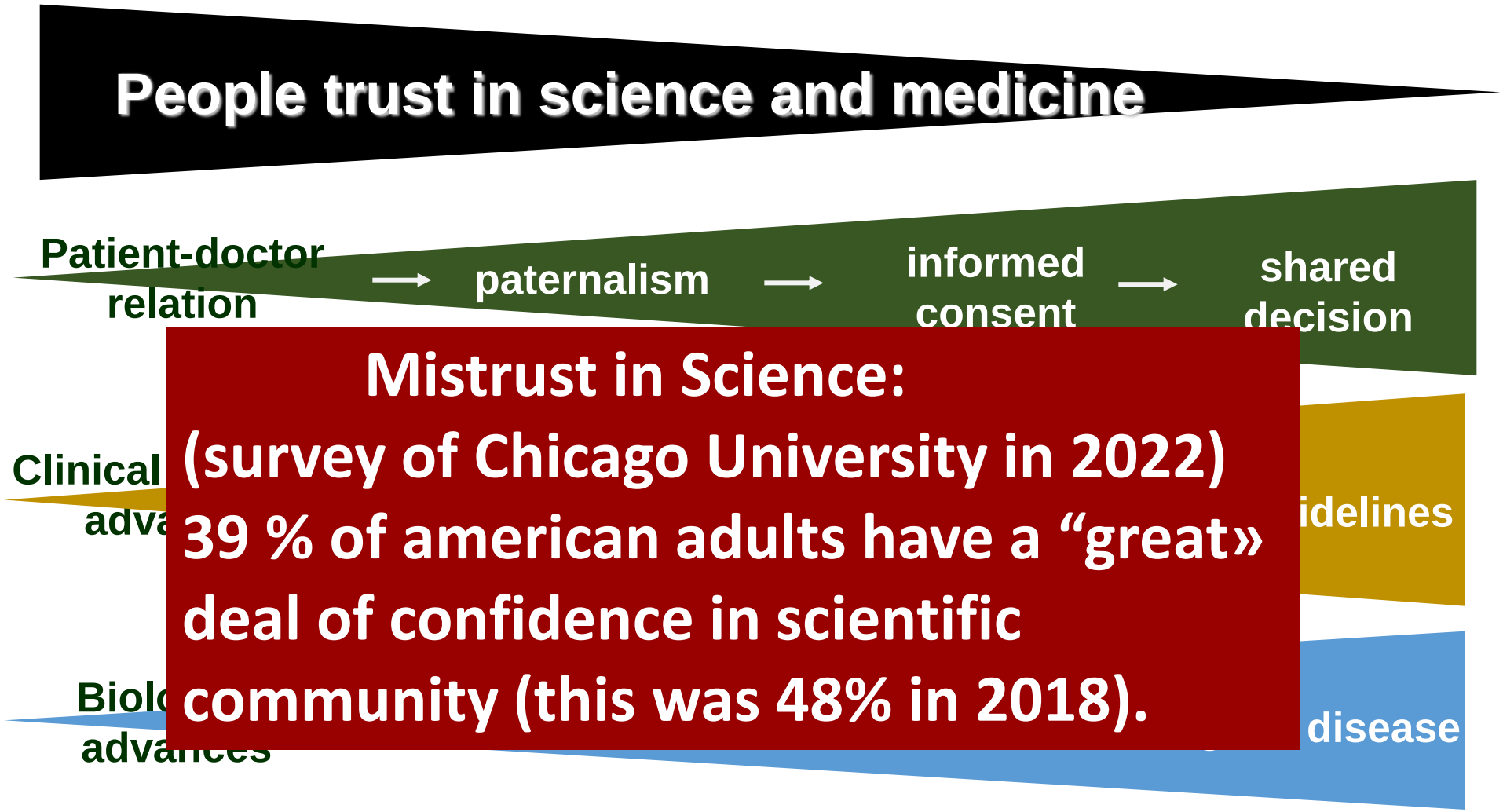
Secondary:
- OS
- ORR, DOR
- PFS2
- PRO
- Safety and tolerability

Exploratory:
- HRQoL
- PD biomarkers

If the trial will be positive:
Which salvage options are available in different world regions?
Access to trastuzumab as maintenance in case of toxicity is the same worldwide?
Which therapy is still effective after T-DXd?

fondazione Periplo

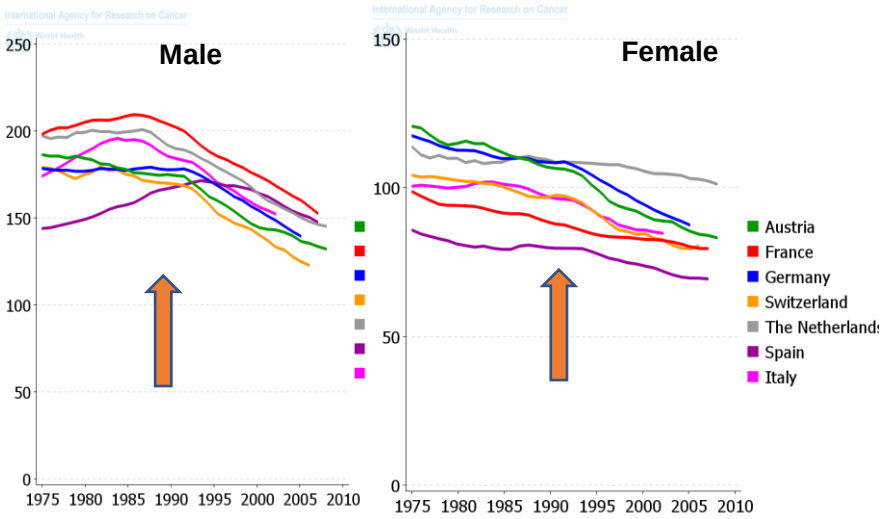
The paradox of the inverse relation between knowledge and trust



Mistrust in Science:
(survey of Chicago University in 2022)
39 % of american adults have a “great» deal of confidence in scientific community (this was 48% in 2018).

fondazione Periplo

Trends in mortality from cancer in Europe & Italy: age-standardised rate (W) per 100,000

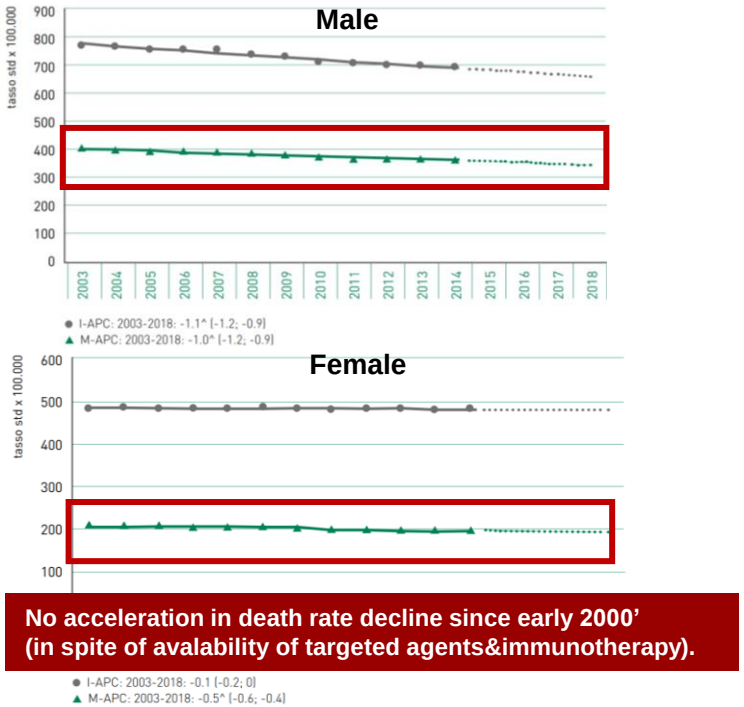


Death rates started to decline sharply in early (female) and late (male) 80'.

Main reasons for cancer death decline

1) Reduction in cigarette smoking
2) Healthy diet & refrigerator
3) Vaccination (HBV, HPV) & retroviral drugs
4) Screening & Early Diagnosis

www.who.int/gho



No acceleration in death rate decline since early 2000' (in spite of avalability of targeted agents&immunotherapy).

AIRTUM 2014-2018

fondazione Periplo

European Journal of Epidemiology
https://doi.org/10.1007/s10654-023-01039-8

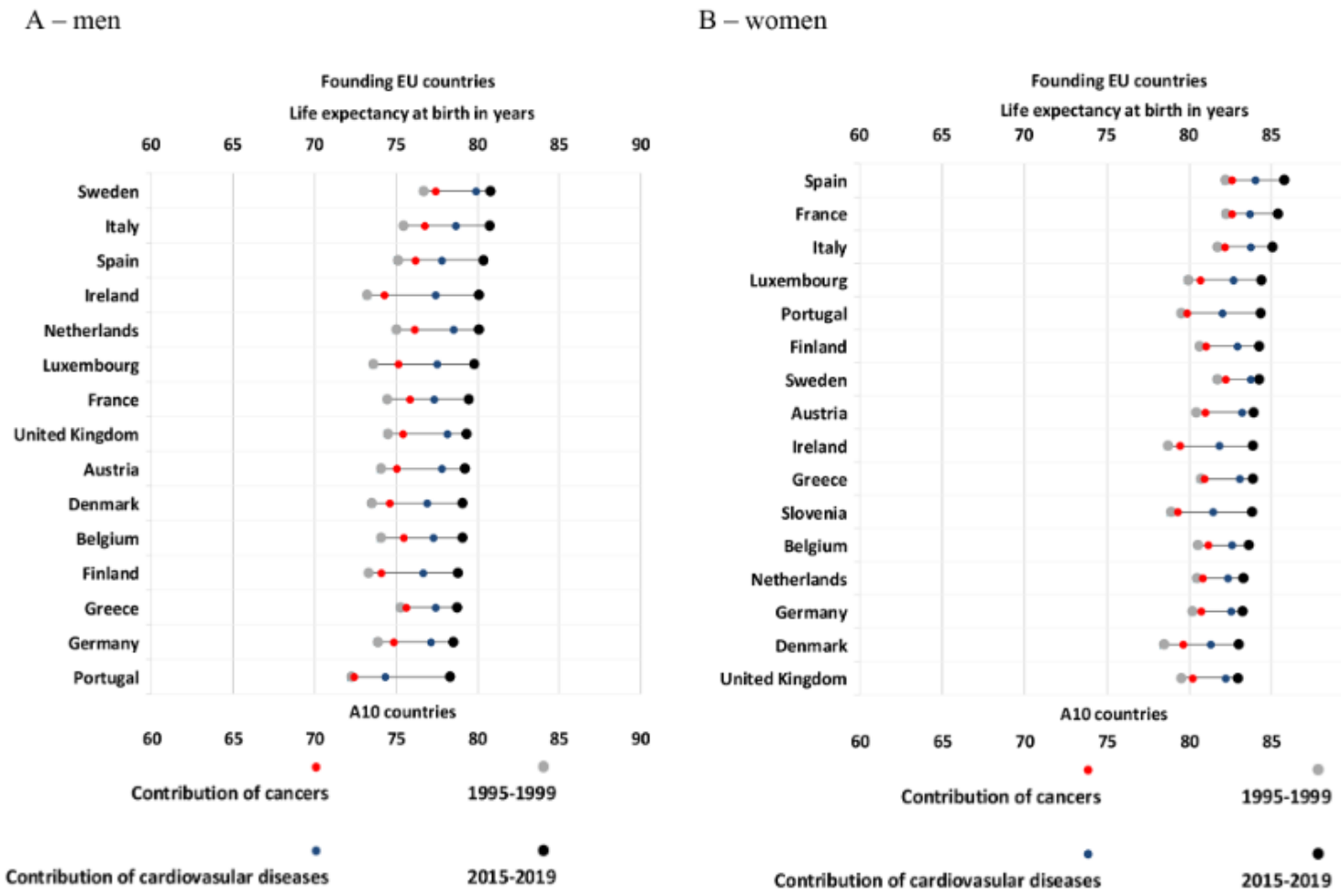
MORTALITY

Gains in life expectancy from decreasing cardiovascular disease and cancer mortality – an analysis of 28 european countries 1995–2019

András Wéber[1,2] · Mathieu Laversanne[1] · Péter Nagy[3,4,5] · István Kenessey[2,6] · Isabelle Soerjomataram[1] · Freddie Bray[1]

Received: 17 April 2023 / Accepted: 2 August 2023

Men:
2.26 years gained by CVD declines versus 1.07 years for cancer
Women:
1.81years gained by CVD declines versus 0.54 years for cancer



fondazione Periplo

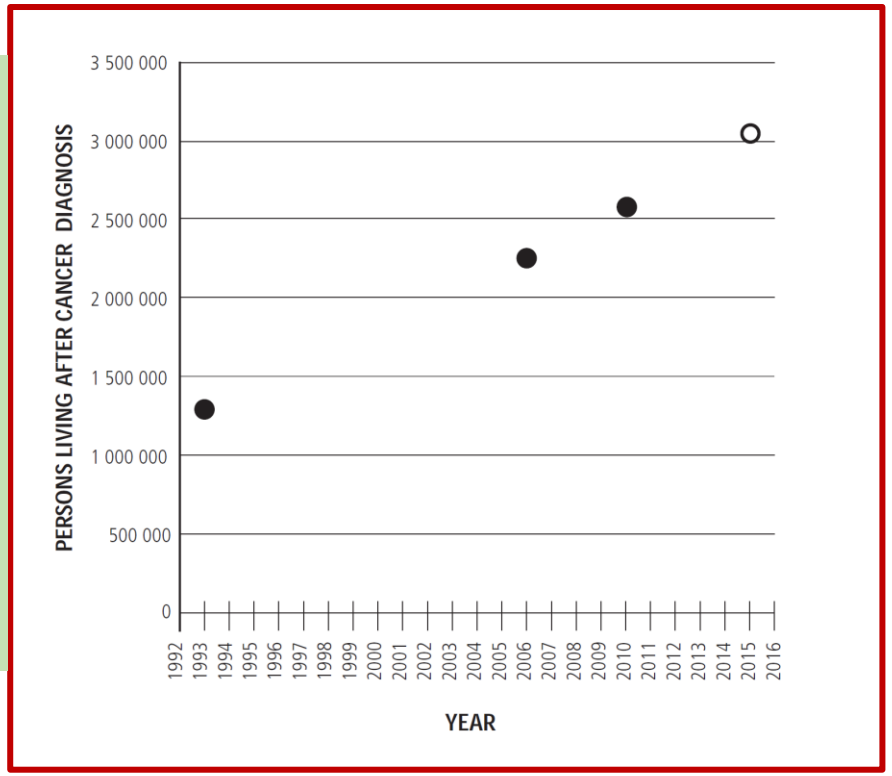
Oncology at the Cross Roads

- **Diagnostic-therapeutic Pathways: markers and outcomes**
  - no new data from PNE or Regional Oncology Networks
  - Periplo Foundation, together with IRST and ISPRO, is carrying forward the PERSEO project (major barriers: funding, data protection officer)
- **Innovation is sustainable; way of funding the NHS is NOT**
  - health care costs of cancer patiens and their outcome can be derived from multiple data bases
  - closed silos financing is inadequate to sustain NHS
- **Evidence- based medicine: from efficacy to effectiveness**
  - Research globalisation and molecular tumor profile make RWD mandatory
  - RWD are what people see & trust!!
- **Survivorship**

fondazione Periplo

Life after or with Cancer is increasingly common

- In Italy in 2019 almost **3.460.000 people** had a previous diagnosis of cancer
- **Estimated increase : 3%/year**
- **~ 25% cured** (similar death rates as general cancer-free population)
- **~ 75% living with cancer or still at risk of relapse**
- Patients living with cancer survive longer (chronic phase)



LEGGE 7 dicembre 2023, n. 193

Disposizioni per la prevenzione delle discriminazioni e la tutela dei diritti delle persone che sono state affette da malattie oncologiche. (23G00206)

(GU n.294 del 18-12-2023)

Law on «Right to be Forgotten» for Cancer patients eliminates discrimination for:

- access to work
- access to bank loan
- access a insurance policy
- request for children adoption

fondazione Periplo

Long-term Consequences of Cancer Therapies

Cancer therapies can be associated with important **short- and long-term sequelae**:

- **Second neoplasms**
- **Cardiotoxicity**
- **Endocrinopathies**
- **Infertility**
- **Sexual dysfunctions**
- **Peripheral & Central Neurotoxicities**
- **Psychiatric disorders**
- **Socio-relational issues**

→ Important impact on patients' **quality of life**

Rare & Long-term toxicities are an unrecognised clinical need

- In clinical trials patients are treated and observed for a limited period of time
- Drugs for rare molecular dysregulations are approved on the basis of phase I/II trials with very few patients and very short observation period
- More patients are cured or survive for years (impact on working ability largely unknown)
- Incidence of CNS metastases is increasing as well as survival of these patients

fondazione Periplo

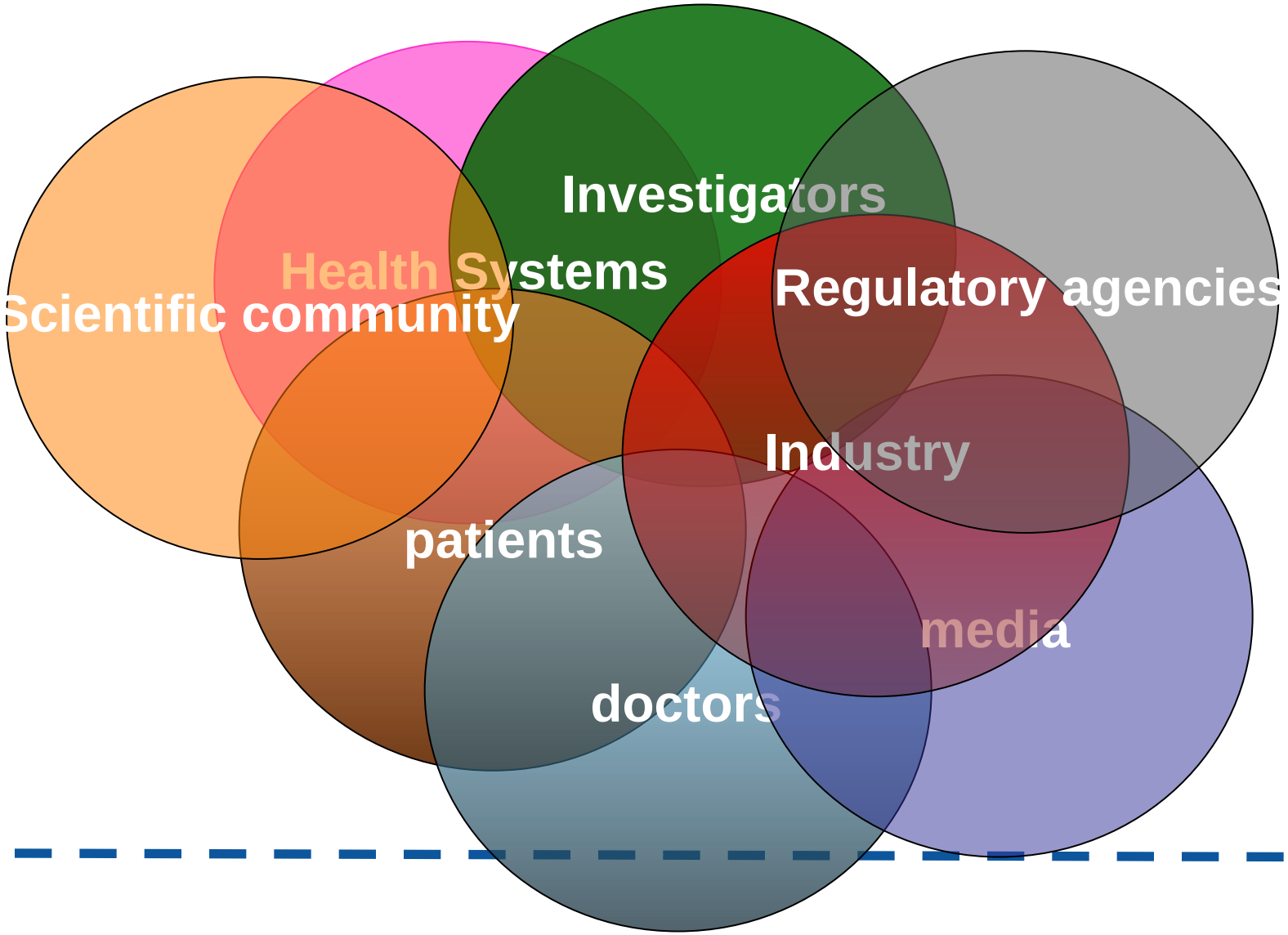
fondazione Periplo

San Camillo IRCCS: PROJECTS

- **PNRR COMBINE Project *(IRCSS San Camillo, Lido di Venezia and IRCSS Pascale, Napoli):***
  - Integration between clinical data and biological markers in patients with breast cancer and cognitive impairment.
  - Cognitive telerehabilitation in cancer patients
  - Shared virtual Biobank
- **WOV ‘Wellness in Oncology in Venice’ PROJECT:**
  - Development of neuropsychological tests for the assessment of cancer-related cognitive impairment.
  - Investigation of cognitive, functional, biological, and neural profile in patients treated with innovative anti-cancer therapies.
- **ONCOLOGY WELLNESS SERVICE:**
  - Oncological and psychological assessment
  - Individual in-depth packages (neurological, physiatric, nutritional, sexological, speech therapy)

,

Stakeholders in Biomedical Research



fondazione Periplo

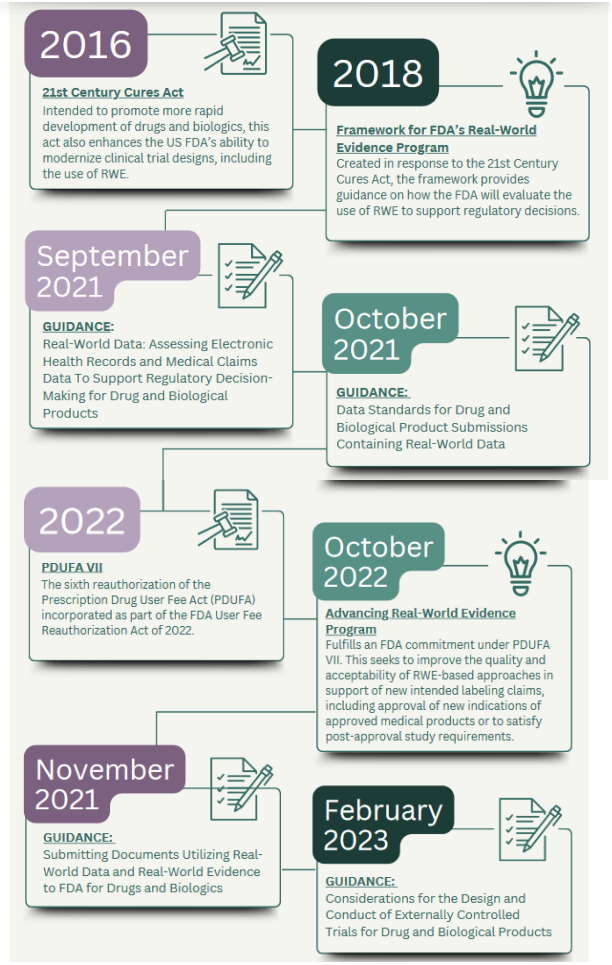
Milestones in the FDA's **Real-World Evidence** Activities

Real-world data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. Examples of RWD include data derived from electronic health records, medical claims data, data from product or disease registries, and data gathered from other sources (such as digital health technologies) that can inform on health status.

Real-world evidence (RWE) is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

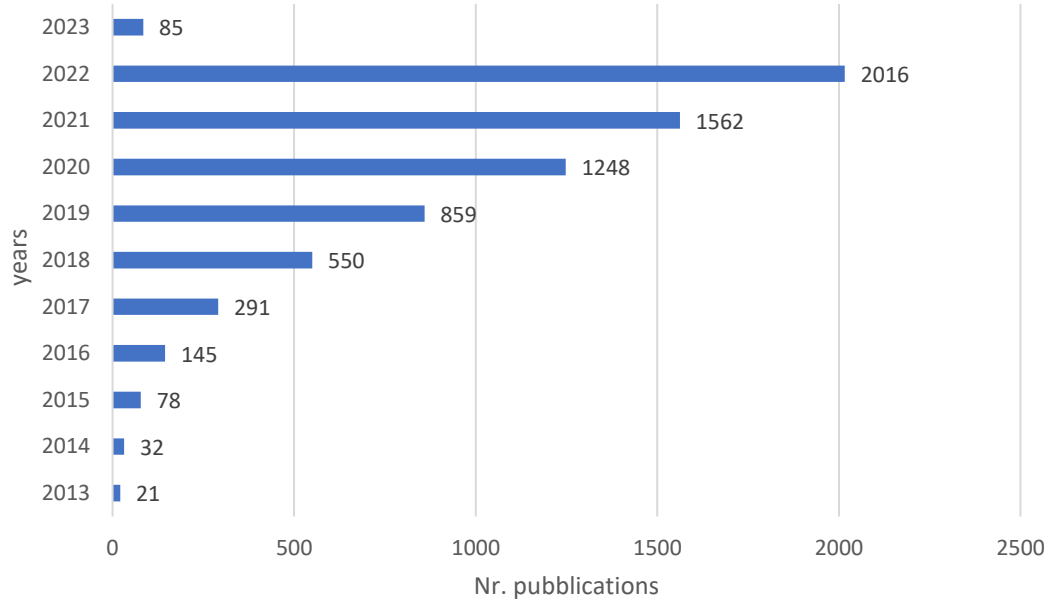
The Evidence Base®

https://www.evidencebaseonline.com/



fondazione Periplo

Real World Publications - Oncology



ESMO GROW

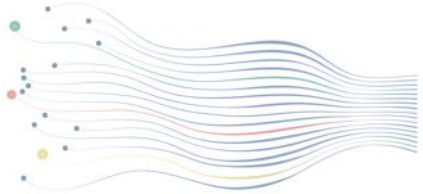
GUIDANCE FOR REPORTING ONCOLOGY
REAL-WORLD EVIDENCE

ESMO-GROW Checklist for Authors and Reviewers

This checklist integrates all ESMO-GROW recommendation and could be used by authors and reviewers when assessing the reporting standards of a real-world evidence study in Oncology.

ESMO GROW

GUIDANCE FOR REPORTING ONCOLOGY
REAL-WORLD EVIDENCE



ESMO-GROW flowchart for real-world evidence studies in oncology

- Castelo-Branco L et al. "ESMO Guidance for Reporting Oncology real-World evidence (GROW)". Ann Oncol 2023; 34: 10.1016/j.annonc.2023.10.001
- Castelo-Branco L et al. "ESMO Guidance for Reporting Oncology real-World evidence (GROW)". ESMO Real World Data & Digital Oncol 2023; 1: 10.1016/j.esmorw.2023.10.001

ESMO

fondazione Periplo

BARRIERS to good quality Real World Data

- Access to regional/national Data Bases
- Data Availability (i.e. molecular characteristics/mutational profiles)
- Set of Data (i.e. proportion of patients; TTF; emergency room access; attrition rate)
- Quality of Data (EMR, source of data and data verification)
- Data interpretation (comparative effectiveness)
- Funding

Editorial

The Italian health data system is broken

The population of Italy is projected to decrease by approximately 8% by 2050, falling from 59 million in 2022 to 54.4 million, due to increased ageing and a declining birth rate. By 2050, more than 35% of Italians will be older than 65 years, while children younger than 14 years will represent only 11.7% of the population.

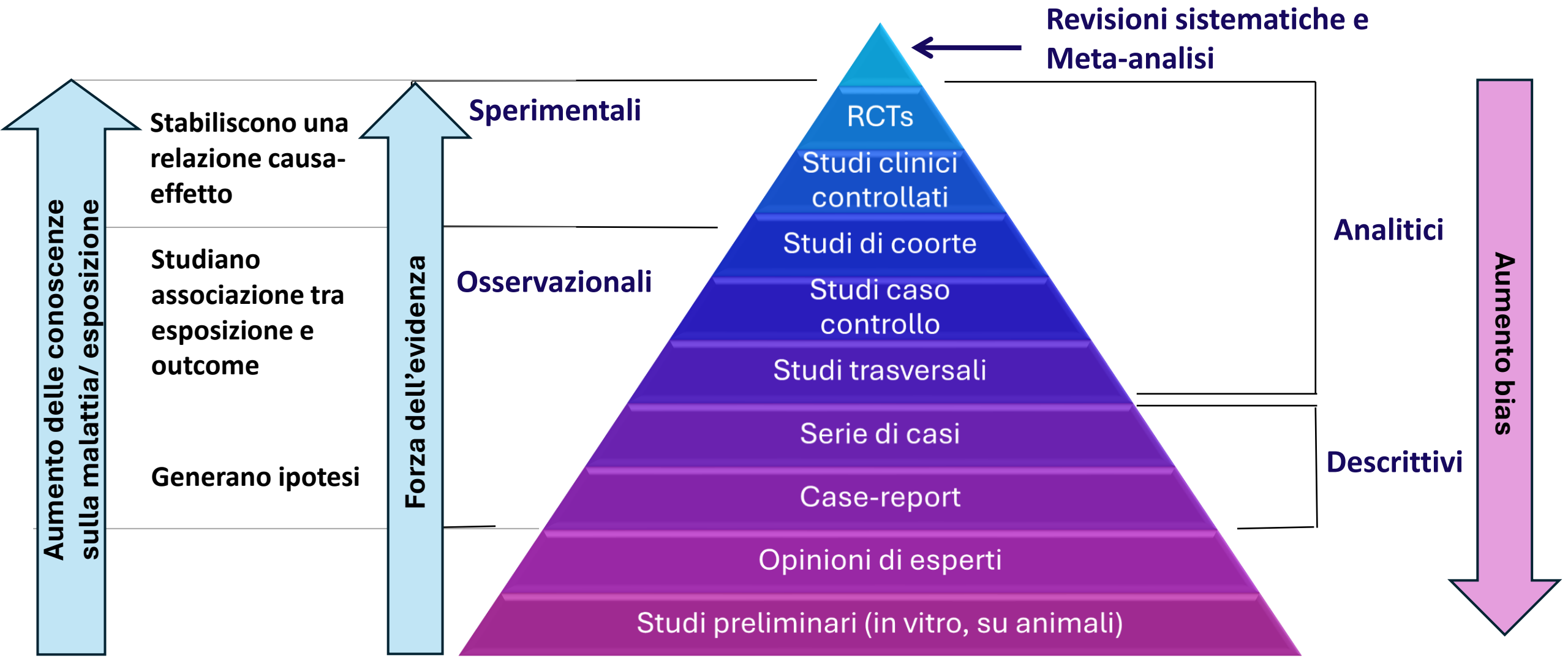
systems, hospitals in the north often cannot access patient records, resulting in repeated diagnostic tests and delayed care. This duplication inflates costs—interregional health-care mobility alone accounts for around €3.3 billion annually—and undermines patient outcomes.

The Lancet Regional Health - Europe
2025;48: 101206
https://doi.org/10.1016/j.lanepe.2024.101206

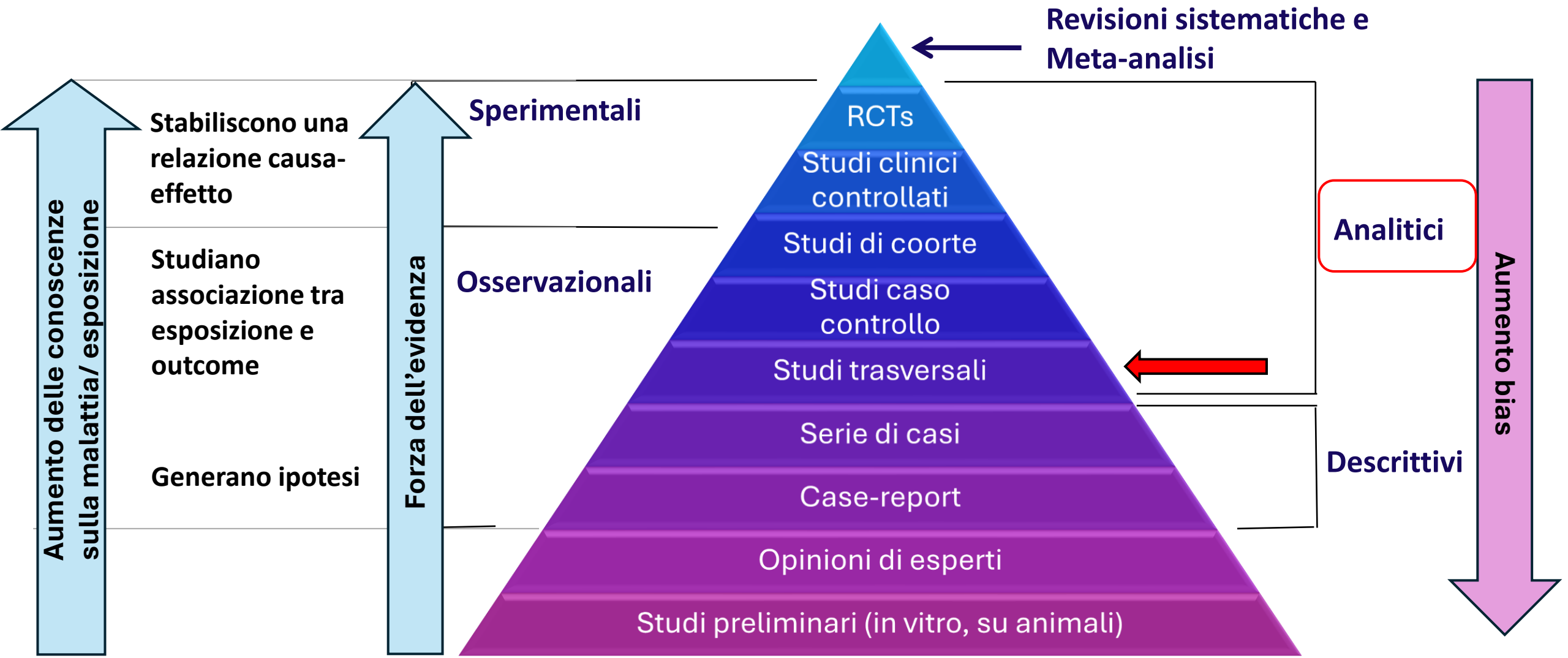
Studi trasversali: punti di forza/debolezza, metodologie di pianificazione, qualità metodologica, conduzione e analisi

(E. Rulli)

Gerarchia dei disegni di studio

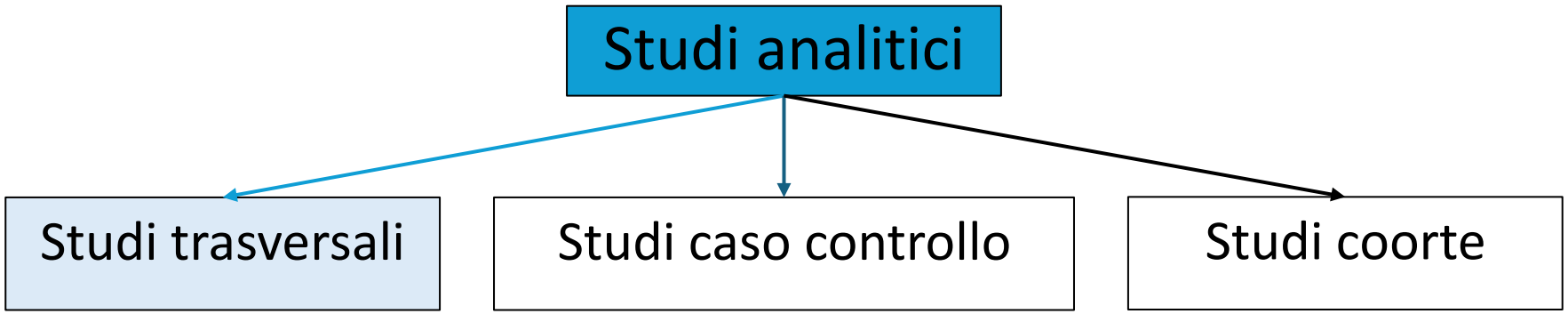


Gerarchia dei disegni di studio



Studi analitici

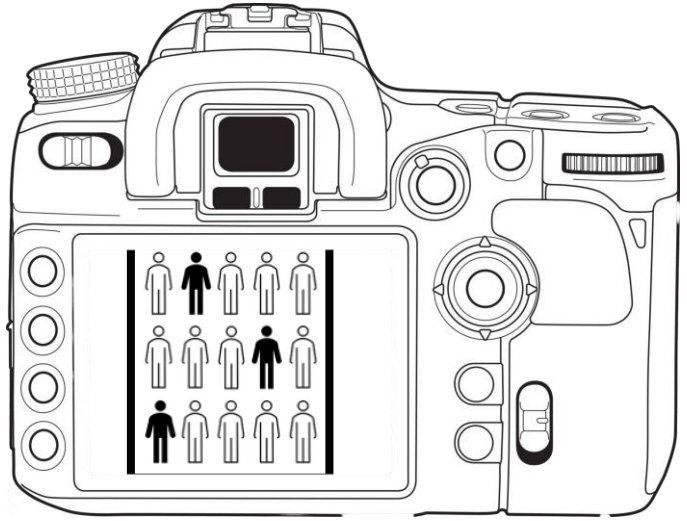
L’epidemiologia analitica si propone di **verificare la presenza di associazioni tra un’esposizione** (fattore di rischio o protezione) **e un evento outcome** (malattia)



Studi trasversali

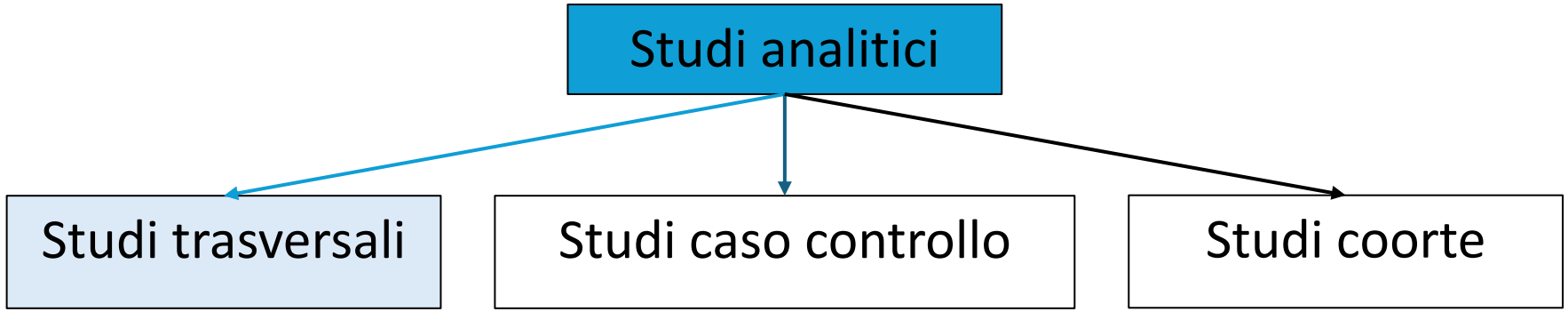
Gli studi trasversali (o ***cross-sectional***) esaminano una **popolazione definita** (o più spesso un suo campione) **in un preciso istante temporale**, in cui si rileva la presenza di uno o più eventi, quali ad esempio lo stato di malattia, l'esposizione ad un particolare fattore di rischio

Possono essere considerati come una «fotografia» istantanea di un gruppo di individui.



Individuo libero dall'esito di interesse

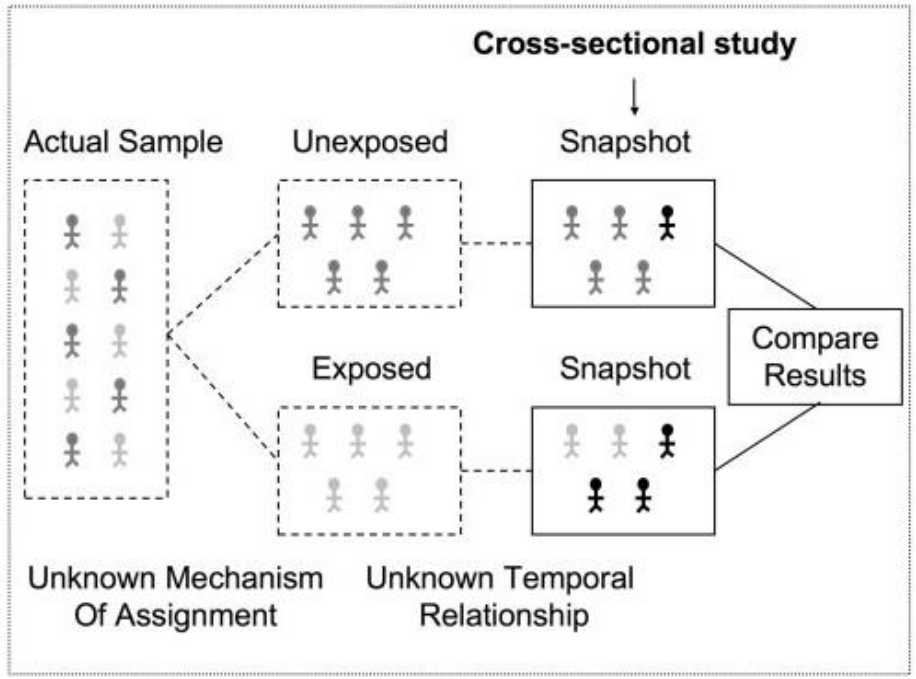
Individuo affetto dall'esito di interesse

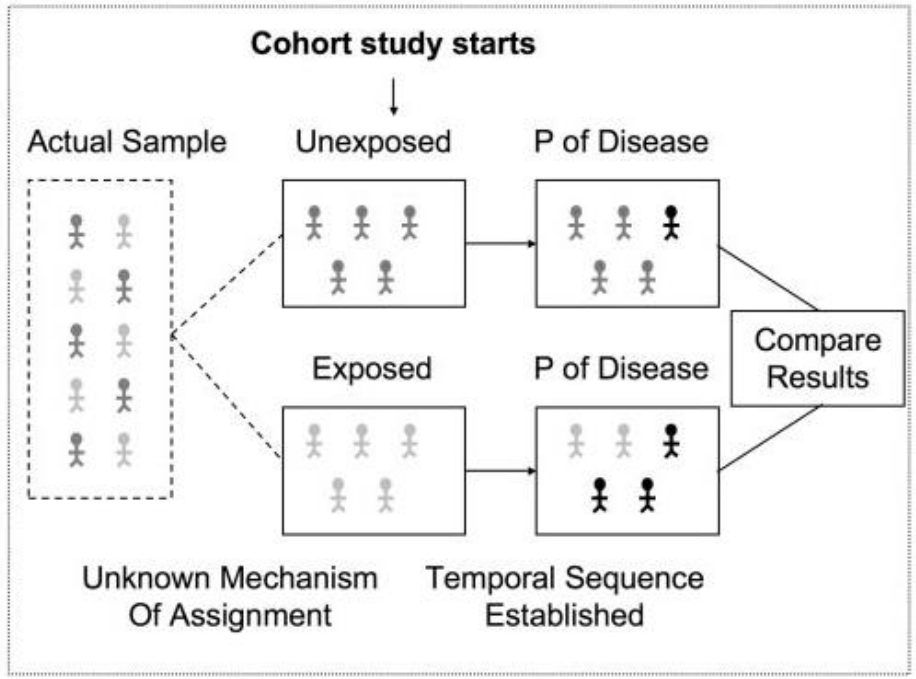


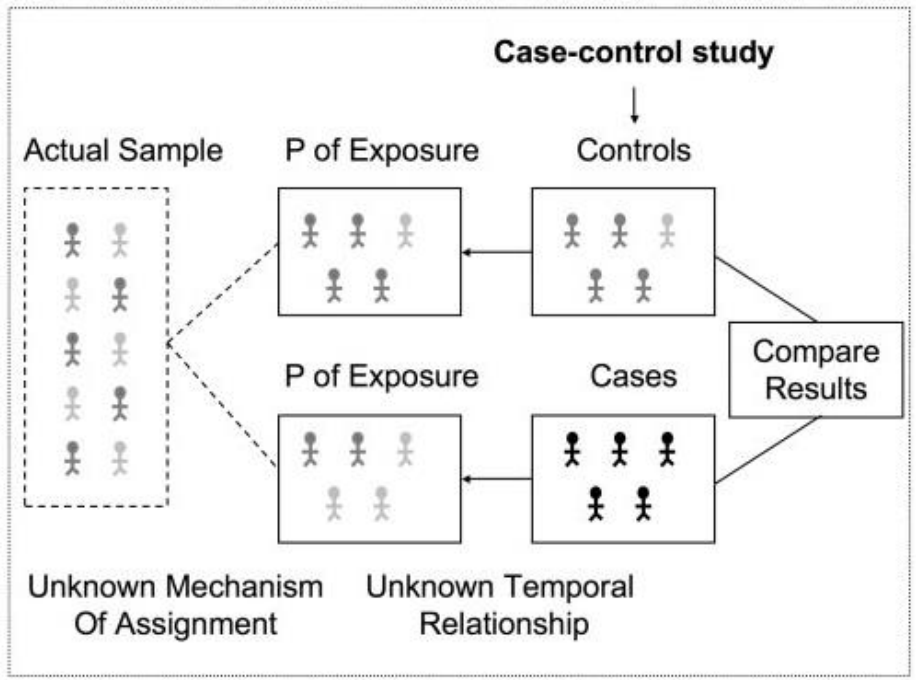
I soggetti sono scelti dalla popolazione di potenziale interesse per un quesito specifico

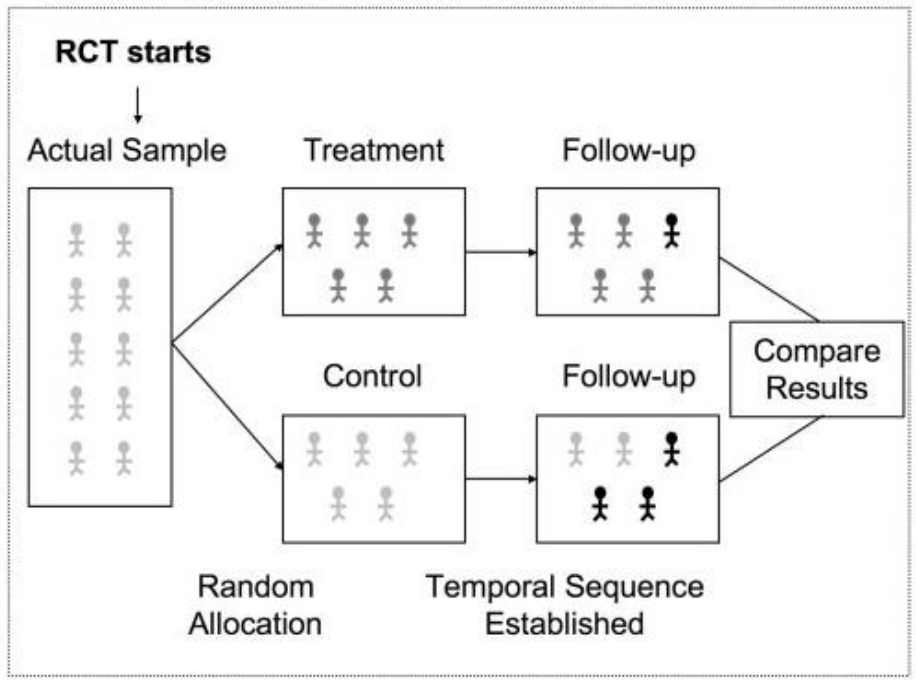
I soggetti sono selezionati sulla base di un **oucome**

I soggetti sono selezionati sull base di un **esposizione**

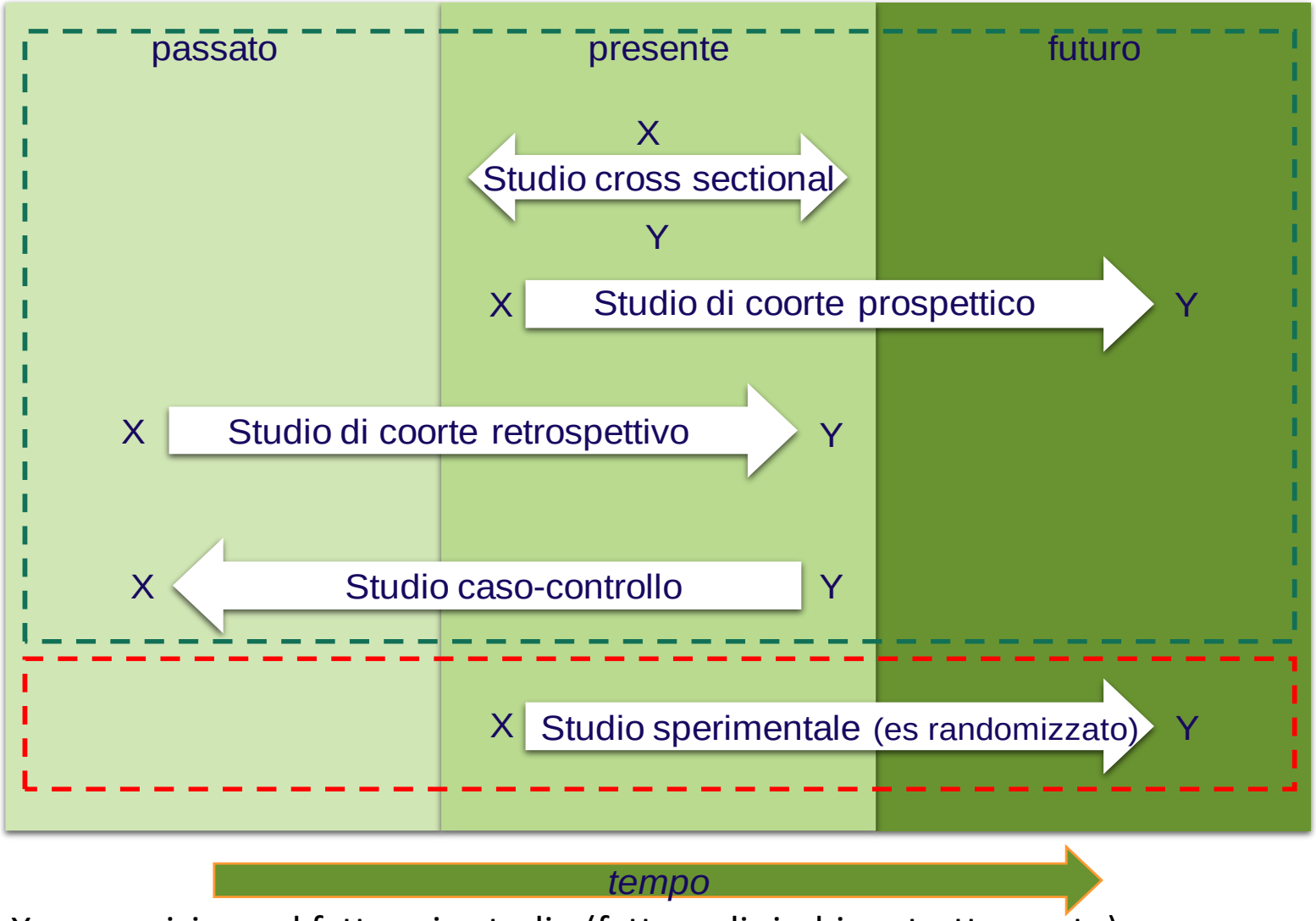

Cross-sectional study
Actual Sample
Unexposed
Snapshot
Exposed
Snapshot
Compare Results
Unknown Mechanism Of Assignment
Unknown Temporal Relationship


Cohort study starts
Actual Sample
Unexposed
P of Disease
Exposed
P of Disease
Compare Results
Unknown Mechanism Of Assignment
Temporal Sequence Established


Case-control study
Actual Sample
P of Exposure
Controls
P of Exposure
Cases
Compare Results
Unknown Mechanism Of Assignment
Unknown Temporal Relationship


RCT starts
Actual Sample
Treatment
Follow-up
Control
Follow-up
Compare Results
Random Allocation
Temporal Sequence Established

...il tempo ...



X = esposizione al fattore in studio (fattore di rischio o trattamento)

Y = effetto dell'esposizione (malattia o guarigione)

Esempio

to estimate the **proportion** of women of childbearing age who are breast-feeding in a given population at the time of the survey

Region and country	Year of survey	Sample size	Percentage of women aged 15–49 years currently breast-feeding
Africa			
Egypt	1980	8788	34.3
Ghana	1979–80	6125	37.7
Kenya	1977–78	8100	43.2
Latin America and the Caribbean			
Colombia	1976	5378	17.1
Mexico	1976	7310	19.8
Venezuela	1977	4361	15.3
Asia and Oceania			
Bangladesh	1975–76	6513	51.1
Indonesia	1976	9155	15.9
Pakistan	1975	4996	40.5

[a] Data from United Nations (1987)

*The World Fertility Surveys (WFS) were national surveys of **human reproductive behaviour** conducted in about 40 developing and 20 developed countries in the late 1970s. Among other aspects of reproductive behaviour, these surveys collected information on **breast-feeding practices** (United Nations, 1987). Table shows the percentages of women aged 15–49 years who were breast-feeding around the time of these surveys in selected countries.*

Esempio

to estimate the proportion of women of childbearing age who are breast-feeding in a given population at the time of the survey

Region and Country (sample size)	Year of survey	Years of schooling			
		Zero	1–3	4–6	7+
Africa					
Egypt (8788)	1980	21.2	19.5	16.3	10.2
Ghana (6125)	1979–80	21.3	n.a.	19.2	15.7
Kenya (8100)	1977–78	19.6	17.4	15.2	12.5
Latin America and the Caribbean					
Colombia (5378)	1976	11.9	11.4	8.3	5.3
Mexico (7310)	1976	12.9	10.9	8.3	3.8
Venezuela (4361)	1977	11.6	10.0	6.7	3.5
Asia and Oceania					
Bangladesh (6513)	1975–76	34.4	30.4	n.a.	n.a.
Indonesia (9155)	1976	28.4	27.0	24.7	13.7
Pakistan (4996)	1975	22.0	n.a.	19.8	n.a.

[a] Data from United Nations (1987)

n.a. = data not available because of small sample sizes.

*The World Fertility Surveys (WFS) were national surveys of **human reproductive behaviour** conducted in about 40 developing and 20 developed countries in the late 1970s. Among other aspects of reproductive behaviour, these surveys collected information on **breast-feeding practices** (United Nations, 1987).*

*In the World Fertility Surveys, breast-feeding practices were examined in relation to **socioeconomic factors** such as **mother's education***

In all countries where the comparison could be made, ***breast-feeding duration decreased consistently with increasing educational level of the mother.***

Scopo

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

studi trasversali descrittivi o di prevalenza

- la **prevalenza** di un fenomeno (malattia o fattore di rischio)

studi trasversali analitici

- le **associazioni** fra fattori di **esposizione** e condizioni di salute o di malattia (**outcome**)
- **l'accuratezza diagnostica** dei test in ambito clinico tramite il confronto fra il test diagnostico in studio e il gold-standard diagnostico di riferimento, su un gruppo di pazienti con sospetto di malattia

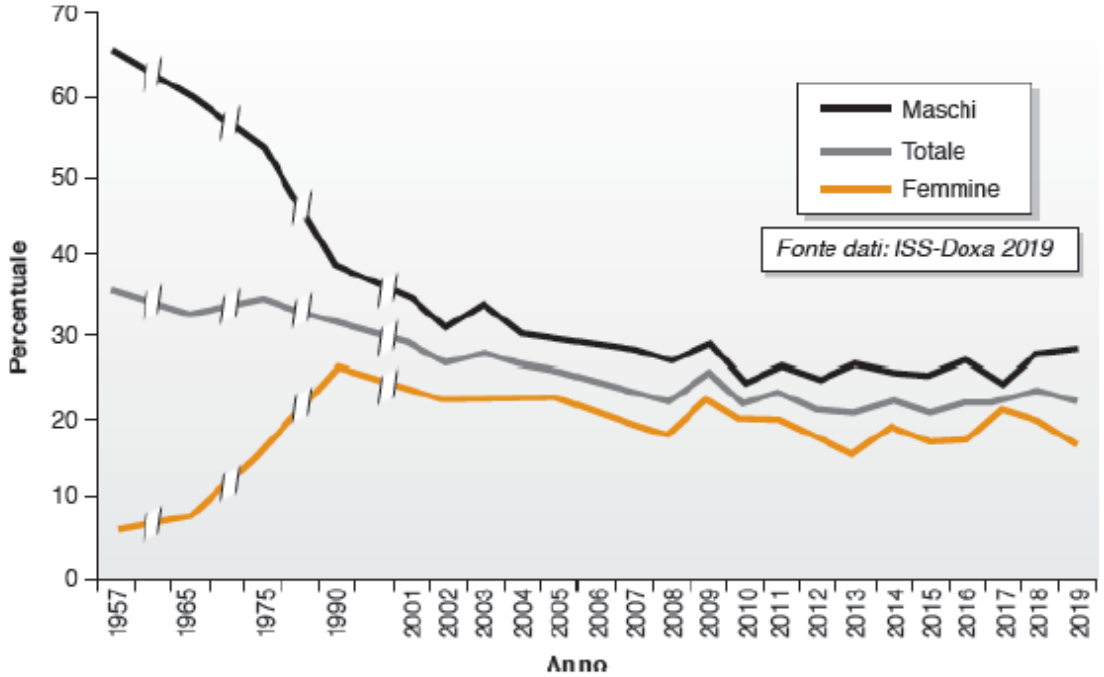
Tipo particolare : studio trasversale ripetuto

In un sottotipo di studio trasversale, noto come **studio trasversale ripetuto** (o seriale), la raccolta dati è condotto sulla stessa popolazione target a **diversi punti temporali**.

Ad ogni tempo, gli investigatori selezionano un campione diverso della popolazione target

E' possibile ripetere lo studio trasversale per analizzare i cambiamenti della popolazione nel tempo (nota come variazione aggregata nel tempo).

Non possono essere utilizzati per esaminare il cambiamento individuale (come in uno studio di coorte).



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

Quando sono utili?

Gli studi 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**

Adatti a **monitorare gli stili di vita**

Non adatti a studiare le **malattie infettive**, soprattutto a ricorrenza epidemica come l'influenza

Metodologie di Pianificazione

Definizione della Popolazione e del Campione:

È fondamentale definire chiaramente la **popolazione di interesse** (ad esempio, una determinata fascia di età, un gruppo geografico specifico) e selezionare un **campione rappresentativo**.

Scelta delle Variabili:

Bisogna decidere quali variabili (demografiche, cliniche, comportamentali) raccogliere e come misurarle. La selezione accurata delle variabili è cruciale per la validità dei risultati.

Metodi di Raccolta Dati:

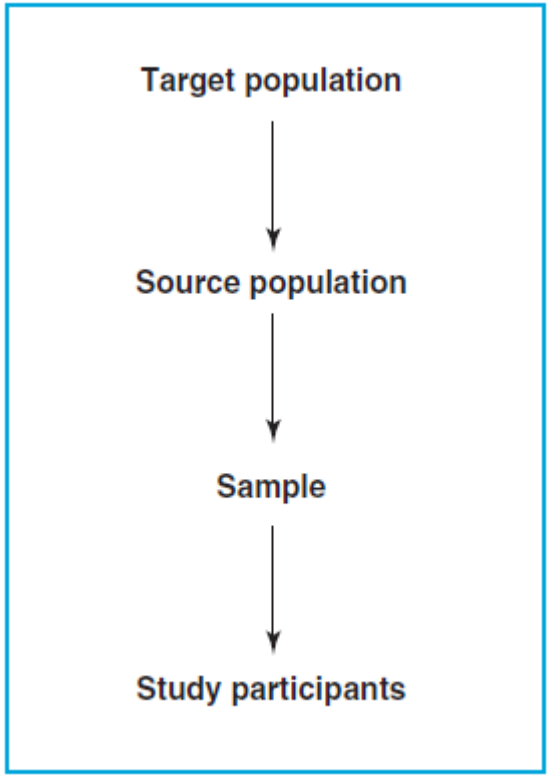
I dati possono essere raccolti tramite ***questionari***, ***interviste***, esami fisici, analisi di laboratorio o altre tecniche appropriate per il contesto dello studio.

Controllo dei Fattori di Confondimento:

È essenziale pianificare l'analisi dei possibili **fattori di confondimento** e considerare strategie per minimizzarli, come l'uso di stratificazione o l'inclusione di variabili di controllo.

Definizione della popolazione

- Idealmente acquisire e registrare sistematicamente le informazioni di **tutti i membri** di popolazione interesse (censimento).
- Selezionare un campione piccolo per rappresentare la popolazione.

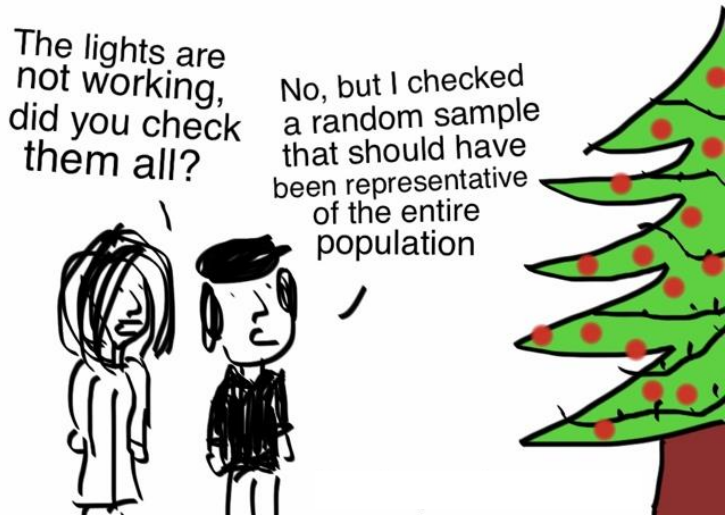


The population to which the main results of the study will be extrapolated

For practical and logistic reasons, the source population is generally more limited than the target population.

If the source population is still too large, a *representative sample* has to be selected

RAPPRESENTATIVITÀ



Metodo di campionamento

Diverse metodologie di campionamento:

- **Campionamento probabilistico**: ogni unità della popolazione ha una determinata possibilità di essere selezionata nel campione (Campionamento casuale semplice, campionamento stratificato e campionamento a più stadi)

- **Campione non probabilistico**: alcuni elementi della popolazione non hanno alcuna possibilità di essere selezionati, o in cui la probabilità di selezione non può essere determinata con precisione (campionamento per convenienza e il campionamento per quote)

Esempio – campionamento probabilistico semplice

*A cross-sectional survey was performed on **random sample**s of women in a high-risk area for cancer of the cervix uteri (Nuuk, Greenland) and in a low-risk area (Nykøbing Falster) of Denmark to assess the **prevalence** of **infections** by specific types of human papillomavirus (**HPV**) and herpes simplex virus (**HSV**) infection. The Danish Central Population Registry is a computerized record of everyone who was alive in 1968 or who was born in or immigrated into Denmark thereafter and includes information on vital status and emigration. A sample of 800 women aged 20–39 years, born in Greenland and residing in the municipality of Nuuk/Godthåb, was drawn **at random** from this population registry. Similarly, a random sample of 800 women aged 20–39 years, born in Denmark and resident in Nykøbing Falster municipality, was also drawn from the same registry*

proper sampling frame (i.e., the computerized list)

relatively small geographical areas

↓

feasible

(Kjaer et al.*, 1988).*

Esempio – campionamento probabilistico stratificato

The seroprevalences of immunoglobulin G (IgG), M (IgM), and A (IgA) ***antibodies*** *to* **Helicobacter pylori** *were assessed by enzyme-linked immunosorbent assay techniques in a survey conducted in the western part of Copenhagen County (Denmark). In 1982, an* ***age- and sex-stratified sample*** *consisting of 4807 men and women born in the years 1922, 1932, 1942, and 1952 (i.e., aged 30, 40, 50 or 60 years) and residing in the western part of Copenhagen County was randomly drawn from the Danish Population Registry, in which all persons living in Denmark are registered*

A stratified random sample involves dividing the population into distinct subgroups according to some important characteristics, such as sex, age or socioeconomic status, and selecting a random sample out of each subgroup. Each subgroup is known as a *stratum* and a separate random sample is selected in each one.

eight sex and age strata were formed and a random sample selected within each stratum

1, males born 1922;
2, males born 1932;
3, males born 1942;
4, males born 1952;
5, females born 1922;
6, females born 1932;
7, females born 1942;
8, females born 1952),

(Andersen et al.*, 1996)*

Metodo di camp

studio trasversale per indagare la **prevalenza della stress lavorativo** tra i dipendenti di un'azienda.

Decido di selezionare i partecipanti semplicemente chiedendo a chiunque sia disponibile durante una riunione di lavoro o durante le pause pranzo di completare un questionario.

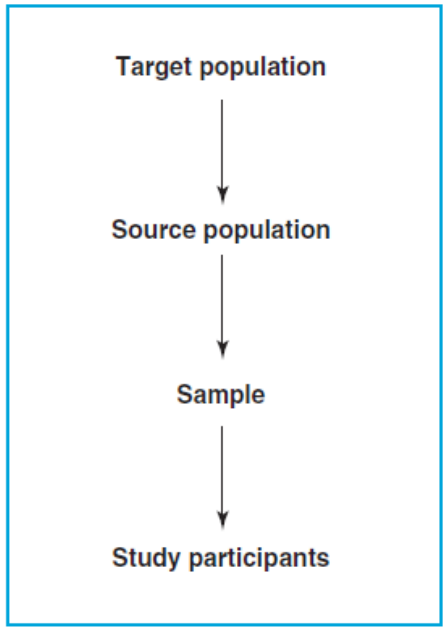
Problemi di generalizzabilità
Bias di selezione

Diverse metodologie d

- **Campionamento prob**
  possibilità di essere sel
  campionamento strati

- **Campione non probabilistico**: alcuni elementi dell on hanno alcuna possibilità di essere selezionati, o in cui la probabilità di selezione non può essere determinata con precisione (campionamento per convenienza e il campionamento per quote)

Non-response



Participants are usually a subset of the initial random sample

Minimize the non-responder rate

Compare respondents and non-respondents with respect to basic characteristics such as age, sex, residence and socioeconomic status

What response level should be considered as acceptable in a survey?

A response rate of 85% might be unacceptable

Metodi di raccolta dati:
Strumento di indagine: validità e riproducibilità

Il **disegno del questionario** è **estremamente importante**.

Molti strumenti diversi anche per lo stesso scopo.

Strumenti devono essere **validati**. Nuovi questionari dovrebbero essere testati per la **riproducibilità e la validità**.

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.

Misure

Grazie allo studio trasversale possiamo stimare:

- la **prevalenza** di un fenomeno (malattia o fattore di rischio)

- le **associazioni** fra fattori di **esposizione** e condizioni di salute o di malattia (**outcome**)

Prevalenza

Casi di malattia presenti in un determinato istante in una popolazione, indipendentemente da quando sono insorti, sono definiti **casi prevalenti**

La **prevalenza è una proporzione** che esprime **il numero di eventi rilevati** (malattie, fattori di rischio, ecc.) **sul totale dei soggetti** esaminati

Essendo una proporzione, **varia tra 0 e 1** (ovvero tra 0% e 100%)

Prevalenza

- Prevalenza puntuale
- Prevalenza periodale

$$\text{Prevalenza puntuale} = \frac{\text{numero casi prevalenti al tempo t}}{\text{popolazione a rischio al tempo t}}$$

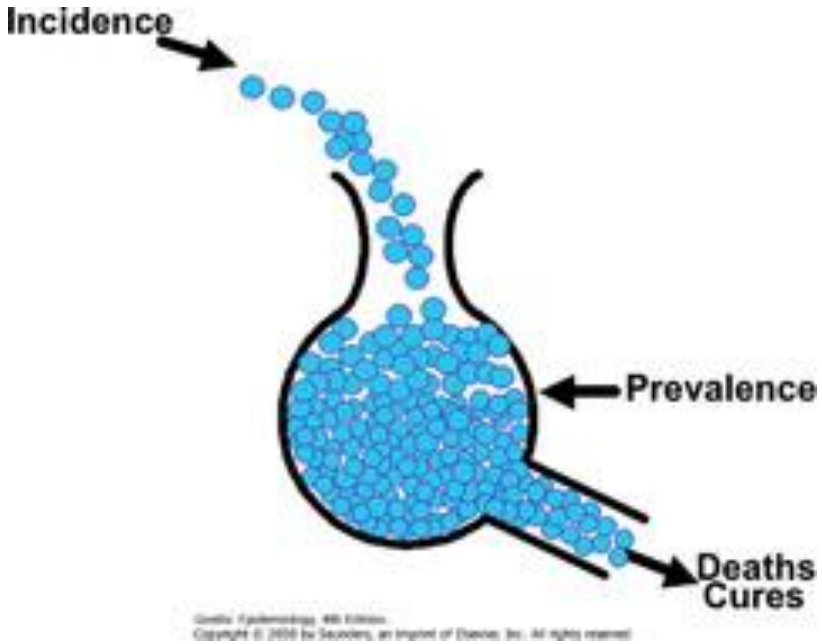
Malati+sani al tempo t

$$\text{Prevalenza periodale} = \frac{\text{numero casi prevalenti nel periodo}}{\text{popolazione a rischio}}$$

Media della dimensione della popolazione durante il periodo

Incidenza vs prevalenza

- **Prevalenza** : misura della **presenza** di una malattia/o di un alto evento di interesse
- **Incidenza**: misura dell'**occorrenza** di un nuovo caso di malattia /o di altro evento



Incidence: the rain arriving

Prevalence: the water in the puddle, new and old

Period Prevalence: during a period

Point Prevalence: at one point in time

The water draining away into the soil or into drains reduces the puddle (i.e. the prevalence), just as recovery or death reduces the number of patients with a problem.

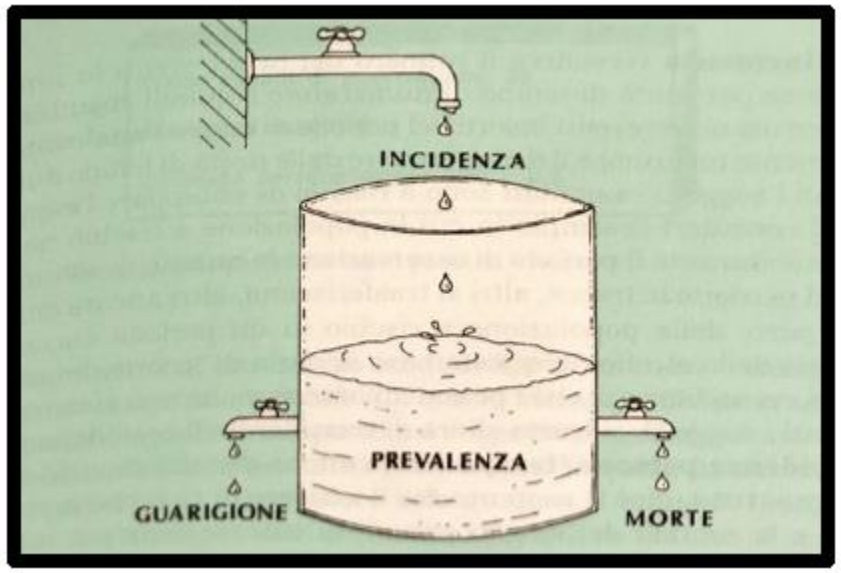
Prevalenza

La prevalenza rende l'idea di **quanto è presente** una malattia nella popolazione

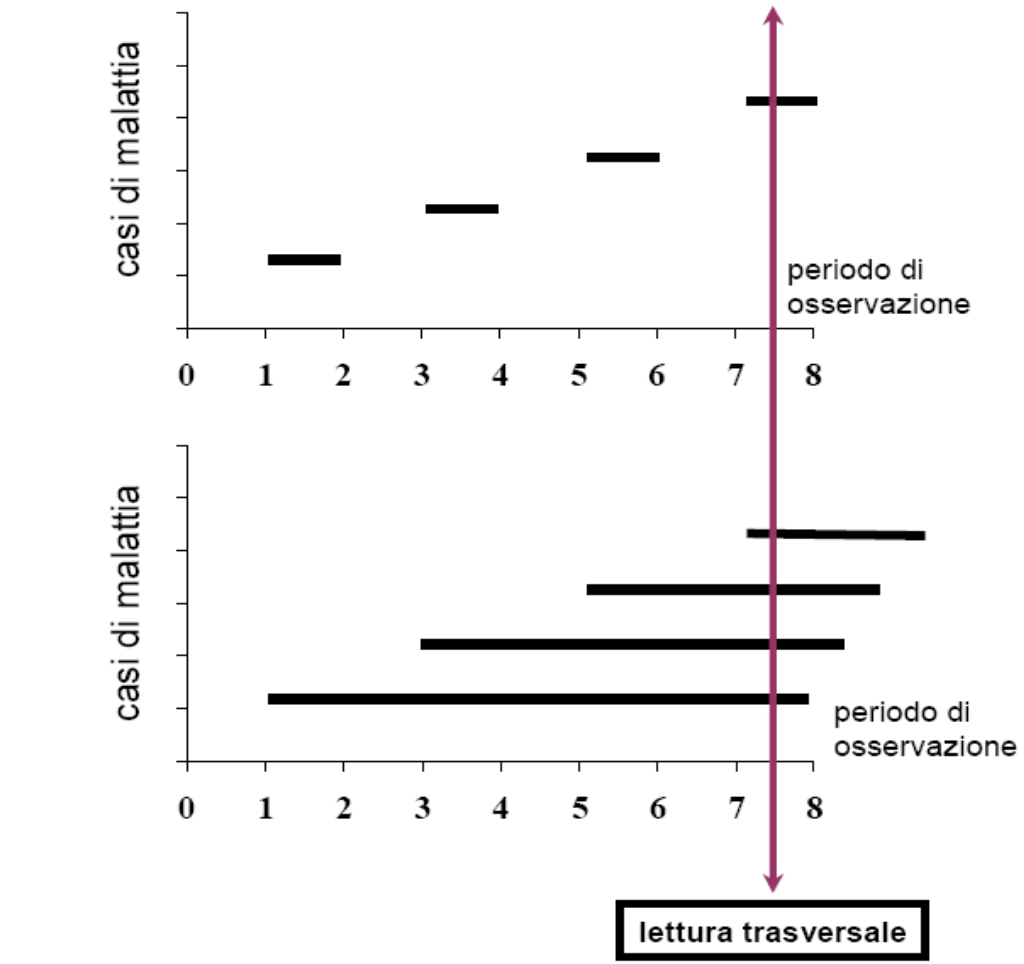
Il numero di casi prevalenti dipende:

- 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

E' solitamente elevata per malattie croniche con bassa letalità
E' solitamente bassa per malattie acute, da cui si può guarire o con alta letalità

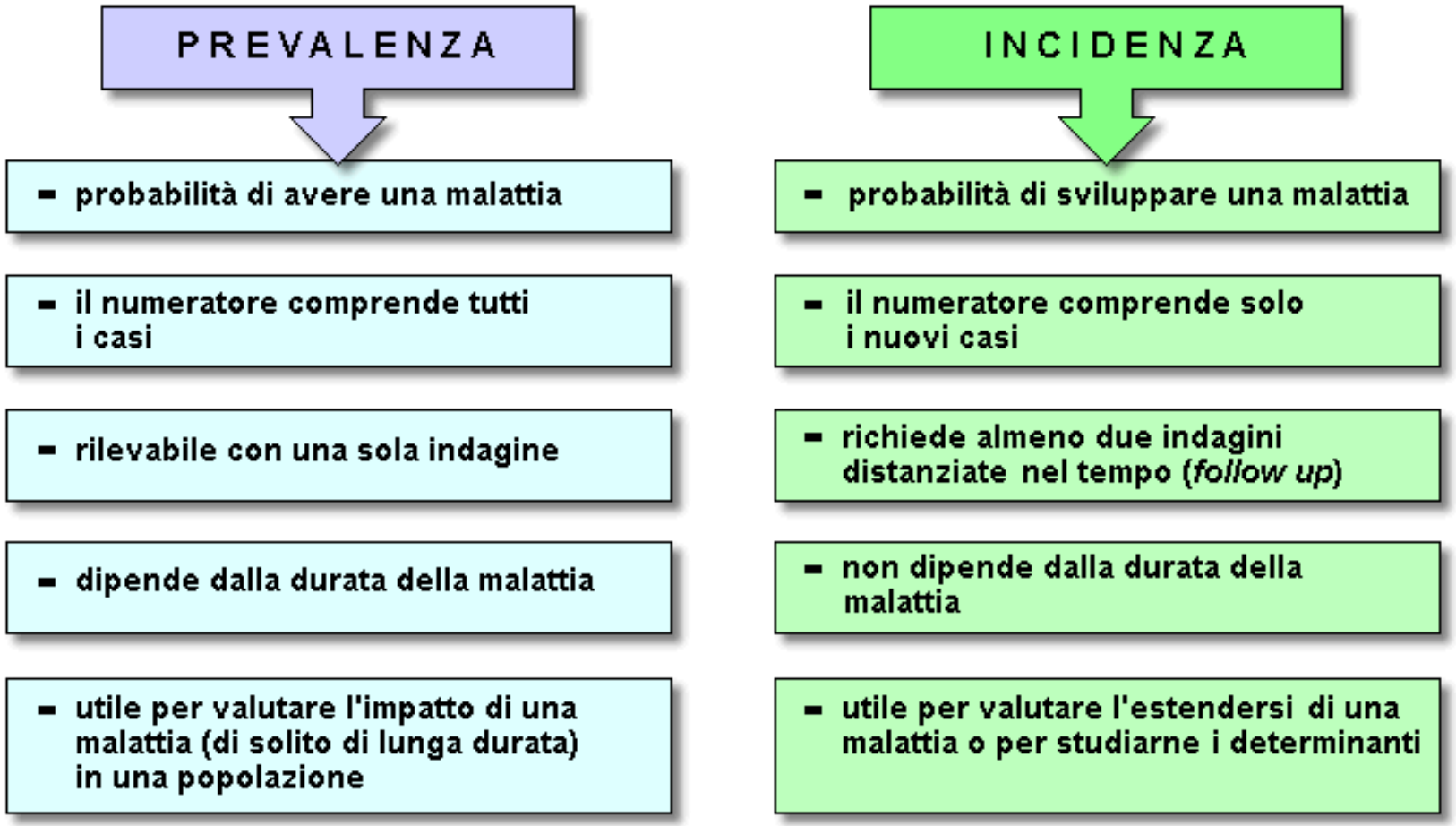


Prevalenza vs Incidenza



durata dello stato di malattia

Prevalenza vs Incidenza



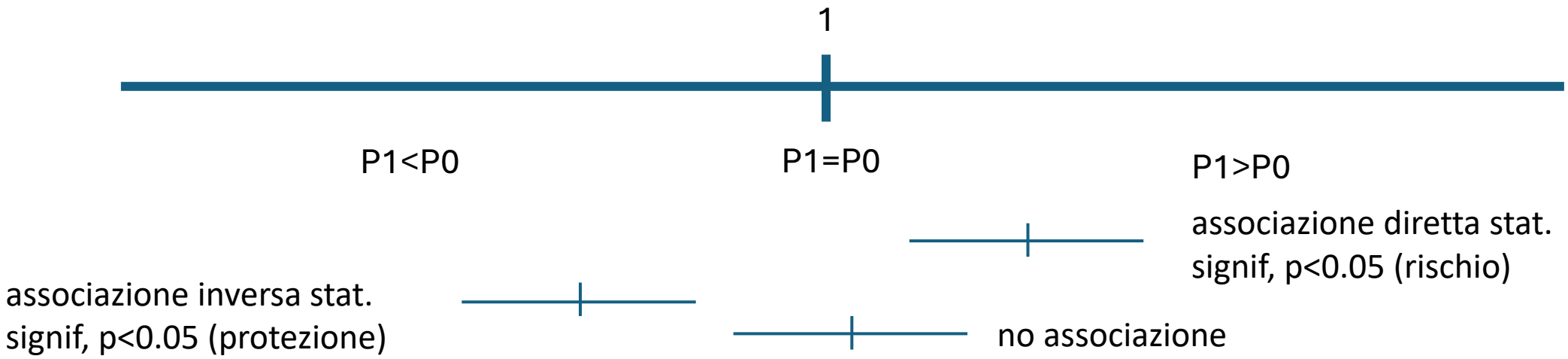
Esempio

*Suppose that a cross-sectional survey was carried out to assess the **prevalence of breast cysts** in a particular female population. A sample of **5891 women** randomly selected from that population were examined and a total of **201** were found to **have breast cysts**.*

The prevalence of breast cysts in this population at the time of the survey was thus: 201 / 5891 = 3.4%.

Misura di associazione: prevalence ratio o odds ratio

Negli studi trasversali come misure di associazione si usano i **Prevalence Ratio** o **Odds Ratio** con i corrispondenti **intervalli di confidenza (IC) al 95%** per gli **esposti rispetto ai non esposti** di un certo evento (outcome)



P1=prevalenza negli esposti
P0=prevalenza nei non esposti

Esempio

*Suppose that a cross-sectional survey was carried out to assess the **prevalence of breast cysts** in a particular female population. A sample of 5891 women randomly selected from that population were examined and a total of 201 were found to have breast cysts.*

*Suppose that the investigators wished to assess whether the **prevalence of breast cysts** was associated with **having ever used oral contraceptives**.*

	Lifetime use of oral contraceptives		Total
Breast cysts	**Ever used**	**Never used**	
Yes	124	77	201
No	3123	2567	5690
Total	**3247**	**2644**	**5891**

Prevalence of breast cysts among ever-users = 124 / 3247 = 3.8%

Prevalence of breast cysts among never-users = 77 / 2644 = 2.9%

Prevalence ratio = 1.3

prevalence ratio is a good estimate of the incidence rate ratio only if

the prevalence of the outcome of interest among those unexposed is **low** (less than 10%)

the duration of the disease is the same among those who were exposed and those who were unexposed to the factor of interest.

Prevalence Ratio= 1.3

the prevalence of breast cysts was 30% higher in everusers of oral contraceptives compared to never-users.

Odds ratio

	Exposed	Unexposed	Total
Cases	a	b	n_1
Controls	c	d	n_0
Total	m_1	m_0	N

Odds of exposure in the cases = a/b
Odds of exposure in the controls = c/d

The *odds (of exposure) ratio* can then be calculated as

$$\text{Odds (of exposure) ratio} = \frac{\text{Odds of exposure in the cases}}{\text{Odds of exposure in the controls}} = \frac{a/b}{c/d}$$

Esempio

Note: the odds ratio will yield a good estimate of the prevalence ratio only if the <u>baseline prevalence of the condition is low</u>

Suppose that a cross-sectional survey was carried out to assess the prevalence of breast cysts in a particular female population. A sample of 5891 women randomly selected from that population were examined and a total of 201 were found to have breast cysts.

*Suppose that the investigators wished to assess whether the **prevalence of breast cysts** was associated with **having ever used oral contraceptives**.*

we can calculate
the odds of having ever used oral contraceptives among women with ('cases') and without ('controls') breast cysts.

	Lifetime use of oral contraceptives		Total
Breast cysts	Ever used	Never used	
Yes	124	77	201
No	3123	2567	5690
Total	3247	2644	5891

Odds of exposure to oral contraceptives among 'cases'
124/77 = 1.61
Odds of exposure to oral contraceptives among 'controls'
3123/2567 = 1.22

Odds ratio = 1.61 / 1.22 = 1.3

Esempio

A cross-sectional survey was carried out among women attending a university health service to investigate the determinants of ***cervical human papillomavirus (HPV) infection****. A sample of 467 women were asked to complete a self-administered questionnaire on* ***socio-demographic variables*** *and* ***sexual behaviour*** *at the time of their visit to the clinic. A polymerase chain reaction DNA amplification method was used to detect HPV infection. The prevalence of HPV infection was then examined in relation to marital status and lifetime number of male sexual partners*

	Number of male sexual partners		Total
	10+	1	
HPV-positive	70	19	89
HPV-negative	32	71	103
Total	**102**	**90**	**192**

Prevalence among women with 10+ partners = 70/102 = 68.6%

Prevalence among women with one partner = 19/90 = 21.1%

Prevalence ratio = 68.6% / 21.1% = 3.3

we can calculate the odds of having had 10 or more partners ('exposure') among HPV-positive ('cases') and HPV-negative ('controls') women as:

Odds of exposure among the cases = 70/19 = 3.68
Odds of exposure among the controls = 32/71 = 0.45

Odds ratio = 3.68/0.45 = 8.2

Prevalence ratio=3.3

≠

Odds ratio = 8.2

inappropriate to take the odds ratio as a measure of relative prevalence, because in this example the baseline prevalence of HPV infection is relatively high (21.1% in women who reported only one partner).

(Ley et al.*, 1991)*

Calcolo del campione

I calcoli sono diversi per uno studio trasversale descrittivo o uno studio analitico trasversale.

Obiettivo stimare la prevalenza di un particolare risultato.

Ipotesi su:
valore presunto del tasso di prevalenza
margine di errore desiderato
livello di significatività

Precisione della stima

Obiettivo confrontare due tassi di prevalenza

Ipotesi su
I tassi di prevalenza i due gruppi di studio (esposti e non esposti),
livello di significatività
potenza statistica

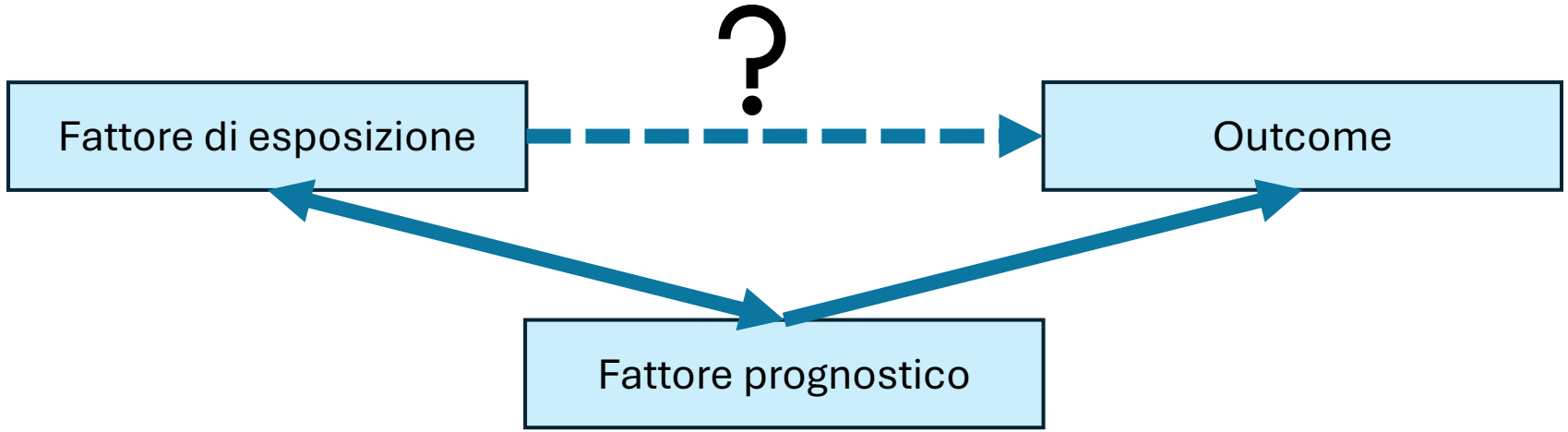
Differenza dei tassi di prevalenza

Confondimento

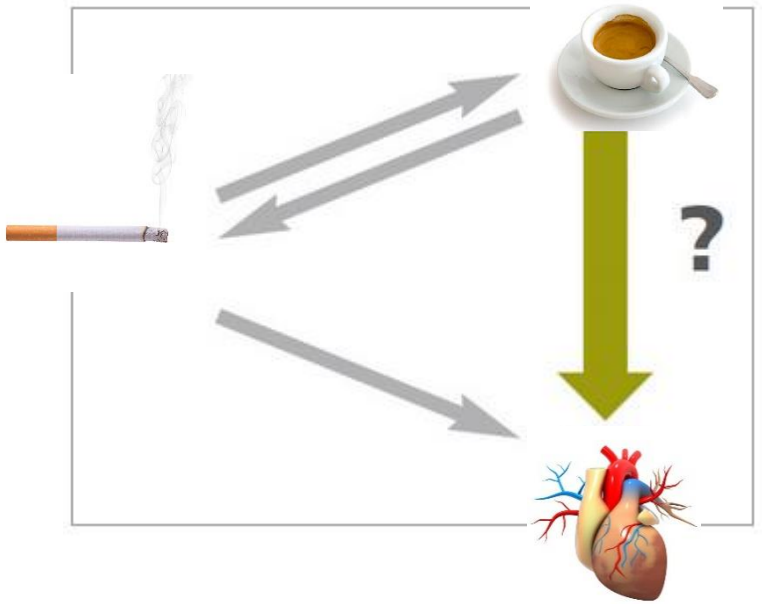
Fattore confondente: variabile associata sia all'outcome che all'esposizione.

Il confondente porterà ad avere associazioni statistiche distorte tra esposizione e outcome.

Il confondente è solitamente una variabile prognostica che allo stesso tempo risulta essere associata al fattore di esposizione



Confondimento



Fattore confondente: variabile associata sia al fattore di interesse che alla variabile di risposta studiata

Il confondente è solitamente una variabile prognostica che allo stesso tempo risulta essere associata al trattamento

[dati fittizi]

		infarto si	infarto no	totale
caffè	si	215	1085	1300
	no	80	1220	1300
	totale	295	2305	2600

fra i 1300 bevitori di caffè, se ne sono ammalati 215 (16.5%)

fra i 1300 non bevitori di caffè, se ne sono ammalati 80 (6.1%)

RR= 2.7

[dati fittizi]

		CAFFE' SI infarto si	CAFFE' SI infarto no	CAFFE' NO infarto si	CAFFE' NO infarto no	totale
FUMO	si	200	800	20	80	1100
	no	15	285	60	1140	1500
	totale	215	1085	80	1220	2600

RR= 1

RR= 1

RRadj= 1

Punti di Forza

1. **Semplicità e Velocità**:
   - Gli studi trasversali sono relativamente facili e rapidi da condurre, poiché vengono raccolti i dati in un unico momento.
   - Non richiedono follow-up a lungo termine come altri tipi di studi longitudinali.
2. **Costo Contenuto**:
   - Essendo uno studio che raccoglie informazioni in un solo punto temporale, i costi sono generalmente inferiori rispetto ad altri studi più complessi.
3. **Rappresentazione della Popolazione**:
   - Forniscono una visione istantanea della distribuzione di caratteristiche in una popolazione.
   - Possono essere utilizzati per ottenere informazioni sulla prevalenza di determinate condizioni o fattori di rischio.

Punti di Debolezza

1. **Causalità Non Determinabile**:

   Poiché i dati sono raccolti in un singolo punto nel tempo, non è possibile stabilire relazioni causali tra le variabili. Si può solo osservare una correlazione.

2. **Bias di Selezione**:

   Se la popolazione campione non è rappresentativa dell'intera popolazione, i risultati potrebbero essere distorti.

3. **Rischio di Confondimento**:

   I fattori confondenti possono influenzare i risultati. Ad esempio, variabili non misurate o sconosciute potrebbero alterare le conclusioni.

4. **Impossibilità di studiare eventi rari**

5. **Impossibilità di studiare l’incidenza**

Riassunto punti di forza e debolezza

TABLE 1] Strengths and Weaknesses of Cross-Sectional Studies

Strengths	Relatively quick and inexpensive to conduct
	No ethical difficulties
	Data on all variables are only collected at one time point
	Multiple outcomes and exposures can be studied
	Easy for generating hypotheses
	Many findings can be used to create an in-depth research study
Weaknesses	Unable to measure the incidence
	Difficult to make a causal inference
	Associations identified might be difficult to interpret
	Unable to investigate the temporal relation between outcomes and risk factors
	Not good for studying rare diseases
	Susceptible to biases such as nonresponse bias and recall bias

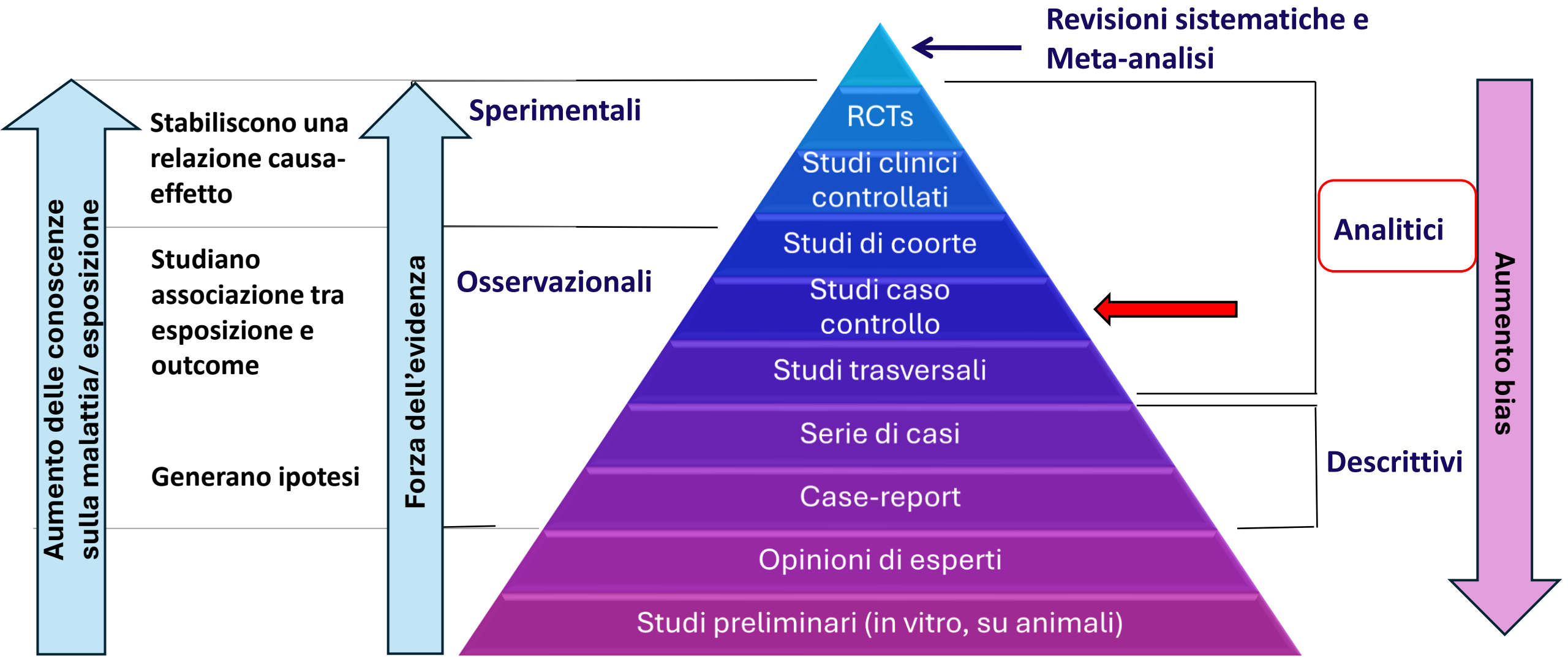
Conclusioni

- Gli studi trasversali sono un'importante metodologia di ricerca, particolarmente utili per ottenere una fotografia istantanea di una popolazione in relazione a variabili specifiche.
- Sebbene presentino alcuni limiti, come la difficoltà nel determinare causalità, la loro capacità di fornire informazioni rapide e poco costose è un vantaggio notevole in molte aree della ricerca.
- Una pianificazione accurata, una gestione appropriata dei dati e un'analisi statistica robusta sono fondamentali per ottenere risultati validi e significativi.

Studi caso-controllo: punti di forza/debolezza, metodologie di pianificazione, qualità metodologica, conduzione e analisi

(E. Rulli)

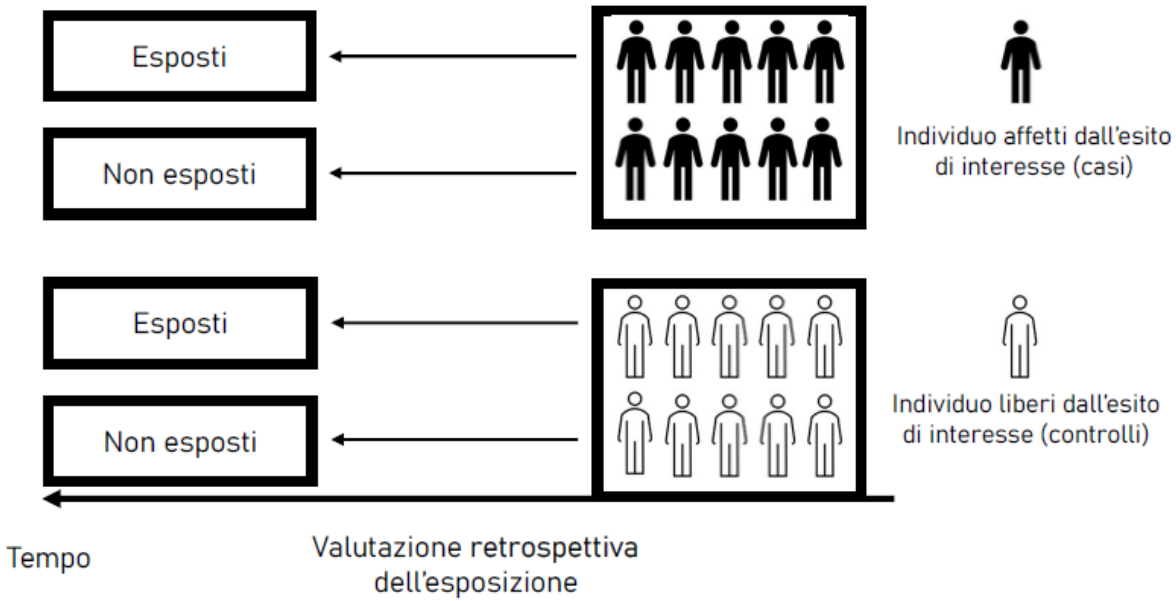
Gerarchia dei disegni di studio



Caso controllo

Lo studio **caso-controllo** è un tipo di studio osservazionale retrospettivo nel quale i soggetti sono selezionati sulla base del fatto che abbiano (casi) o non abbiano (controlli) sperimentato l'evento in studio.

I due gruppi sono quindi confrontati rispetto alla storia i esposizione al fattore di interesse.



	Esposti	Non esposti
Casi	a	c
Controlli	b	d
Totale	?	?

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}}$$

Assunzione

In uno studio caso-controllo i casi e i controlli vengono immaginati come generati da un'unica coorte durante un certo periodo.

Finalità

Finalità eziologiche: i casi e i controlli originati da una coorte di individui **esenti dalla malattia** in studio **e a rischio di contrarla** e rappresentati da quella porzione della coorte che **sperimenta** (casi) **o no** (controlli) la **malattia** in studio; per entrambi viene analizzata l'esposizione pregressa al fattore che si sospetta ne sia causa.

Finalità prognostiche e valutative: i casi e i controlli originati da una coorte di pazienti alla prima **manifestazione della malattia in studio** e rappresentati da quella porzione della coorte che **sperimenta** (casi) **o no** (controlli) uno dei possibili **esiti della malattia stessa**; per entrambi vengono considerate le caratteristiche pregresse (fattori prognostici) o il trattamento (fattori terapeutici) che si sospetta abbiano condizionato l'insorgenza dell'esito.

Tipologie

Su base ospedaliera: i casi e i controlli sono selezionati tra coloro diagnosticati e/o trattati in ambito ospedaliero.

⇧ semplici da realizzare e poco costosi.

⇩ l'impossibilità di identificare la popolazione che ha generato i casi inclusi →approccio particolarmente vulnerabile per la selezione dei casi.

Su base di popolazione: i casi e i controlli sono selezionati tra coloro che insorgono da una ben definita popolazione bersaglio durante un certo periodo di osservazione.

⇧ vantaggio di evitare distorsioni nella selezione dei casi

⇩ più difficili da realizzare se la popolazione bersaglio non è coperta da un valido sistema di registrazione dei casi della malattia in studio.

Su base ospedaliera - esempio

*The relationship between use of **conjugated estrogens** and the **risk of endometrial cancer** was examined among 188 white women aged 40–80 years with newly diagnosed endometrial cancer and 428 controls of similar age hospitalized for non-malignant conditions requiring surgery at the Boston Hospital for Women Parkway Division, Massachusetts, between January 1970 and June 1975. The data on drug use and reproductive variables were extracted from hospital charts and from the medical records of each woman's private physician. Thirty-nine per cent of the cases and 20% of the controls had used conjugated estrogens in the past.*

the cases were white women, aged 40–80 years, who were admitted to a certain hospital in Boston from January 1970 to June 1975 with a first diagnosis of endometrial cancer.

(Buring et al.*,1986)*

Su base popolazione - esempio

A population-based case–control study was carried out in Spain and Colombia to assess the ***relationship*** *between* ***cervical cancer*** *and* ***exposure to human papillomavirus*** *(HPV), selected* ***aspects of sexual and reproductive behaviour***, ***use of oral contraceptives***, ***screening practices***, ***smoking***, *and possible interactions between them.*
The study included 436 incident cases of histologically confirmed invasive squamous-cell carcinoma of the cervix and 387 controls of similar age randomly selected from the general population that generated the cases.

An ***active case-finding network was organized with periodic visits to all hospitals***, *clinics and pathology departments in the public and private sector in each study area to identify and interview the cases before any treatment was applied.* ***All*** *cervical intraepithelial neoplasia (CIN) III cases diagnosed during the study period were also identified and the histological slides were reviewed by a panel of pathologists to ensure completeness of recruitment of the invasive cancer cases.*

Cases taken from a defined population over a fixed period of time.

In population-based case–control studies, it is essential to ensure completeness of case-finding.

(Muñoz et al.*, 1992a)*

Selezione dei casi

Casi incidenti/prevalenti

- Casi incidenti (soggetti con nuova diagnosi)
- Casi prevalenti (diagnosi avvenuta in passato)

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 reclutamento 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

Appaiamento o matching

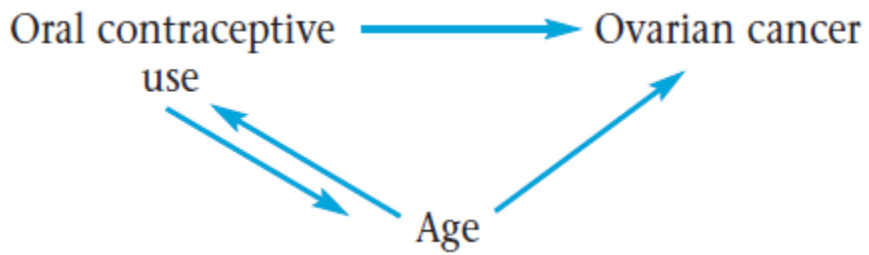
Uno o più controlli sono selezionati per ciascun caso in base alla somiglianza rispetto a caratteristiche diverse dall'esposizione oggetto dell'indagine.

L’appaiamento rende **simili casi e controllo** rispetto alla **caratteristica** in questione in modo tale che questa non possa influenzare l’effetto di interesse

Poiché casi e controlli sono simili sulle variabili corrispondenti, la loro **differenza** per quanto riguarda lo **stato della malattia** può essere **attribuibile** a differenze in alcuni **altri fattori**.

Esempio

We are interested in examining the relationship between current **use of oral contraceptives** and **ovarian cancer**



It is appropriate to **match on age**, since age is associated with the exposure of interest (current oral contraceptive use) and is an independent risk factor for ovarian cancer.

age is a confounding factor.

Failure to match, or otherwise control, for age would result in a **biased assessment** of the effect of oral contraceptive use.

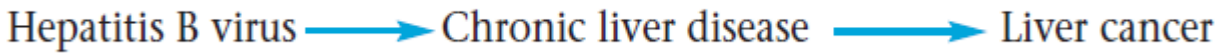
Overmatching

Quando i controlli sono scelti in modo da essere simili ai casi per una caratteristica e questo appaiamento tende a **mascherare l'associazione della malattia** con l'esposizione dell'interesse, si dice che i casi e i controlli sono **overmatched**.

Questo può accadere quando i controlli vengono abbinati ai casi per una caratteristica che fa parte del percorso attraverso il quale la possibile causa di interesse porta alla malattia.

Esempio

Imagine a case–control study conducted in West Africa to investigate the role of **hepatitis B virus** in the etiology of **liver cancer** in which controls were matched to cases on the basis of **previous history of liver disease**.



If chronic liver disease is on the pathway between hepatitis B infection and liver cancer, matching on that condition would result in an **underestimation of the effect of the virus** on the occurrence of liver cancer, since controls would have been made similar to the cases in relation to this variable.

Selezione 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*

Esempio Bias da Selezione: caso-controllo con base ospedaliera

	TUMORE polmone	
Abitudine al fumo	*Casi*	*Controlli*
Fumatori	80	80
Non Fumatori	20	20
Totale	100	100

→ Nessuna associazione

Il fumo non è un fattore di rischio del tumore al polmone?

Nei reparti di medicina sono ricoverati molti individui con **bronchite cronica ostruttiva**, **enfisema e altre malattie associate al fumo**

Selezione 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*

Non partecipazione

Come per i casi, anche per i controlli è importante raccogliere informazioni sui motivi della mancata partecipazione e, ove possibile, di ottenere ulteriori informazioni sulle loro caratteristiche sociodemografiche (ad esempio, sesso, età, status socioeconomico)

*A large multi-centre case–control study was conducted in **high- and low-risk areas** of Italy to evaluate the role of **dietary factors** in the **etiology of gastric cancer** and their contribution to the marked geographic variation in mortality from this cancer within the country. All patients with new histologically confirmed gastric cancer diagnosed between June 1985 and December 1987, resident in the study areas, and aged 75 years or less were eligible as cases. A total of 1129 eligible cases were identified in surgery and gastroenterology departments and outpatient gastroscopic services of private and public hospitals.*

*Controls were randomly selected from population lists within five-year age and sex strata. A total of 1423 population-based controls were sampled, of whom 1159 (**81%**) were successfully interviewed using the same structured questionnaire as for the cases*

Recruitment centre	Sampled No. (%)	Recruited No. (%)	Excluded due to	
			Refusal No. (%)	Poor health No. (%)
Cagliari	118 (100)	108 (91.5)	8 (6.8)	2 (1.7)
Cremona	61 (100)	51 (83.6)	5 (8.2)	5 (8.2)
Florence	547 (100)	440 (80.4)	74 (13.6)	33 (6.0)
Forli	291 (100)	259 (89.0)	20 (6.9)	12 (4.1)
Genoa	205 (100)	137 (66.8)	17 (8.3)	51 (24.9)
Imola	74 (100)	61 (82.4)	10 (13.5)	3 (4.1)
Siena	127 (100)	103 (81.1)	6 (4.7)	18 (14.2)
Total	**1423 (100)**	**1159 (81.4)**	**140 (9.9)**	**124 (8.7)**

[a] Data from Buiatti *et al.* (1989a).

Quanti controlli per ogni caso?

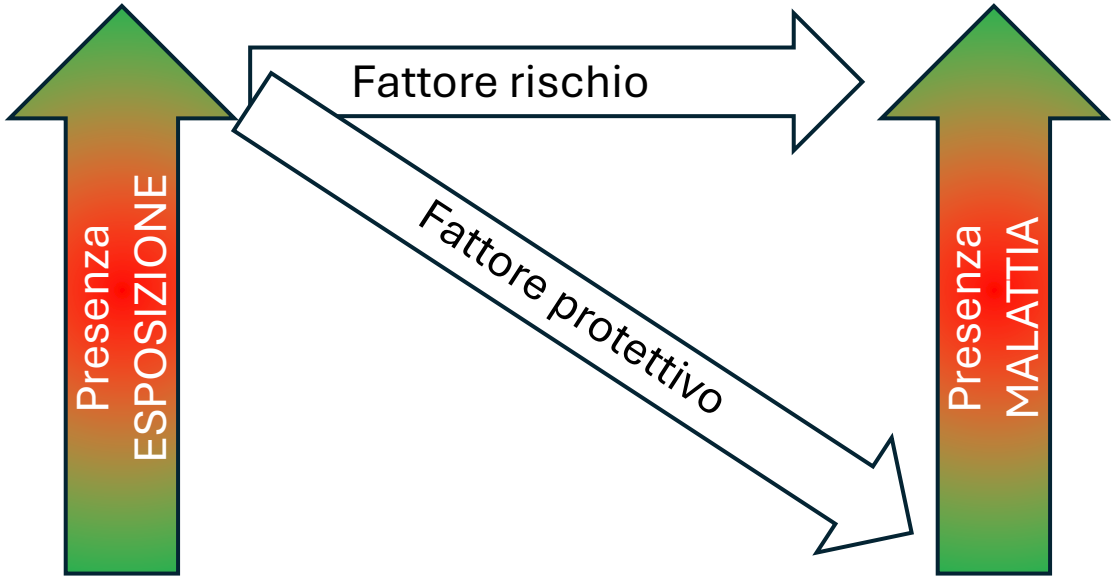
Il rapporto caso-controllo ottimale è 1:1.
Il test del chi quadrato per l'indipendenza più potente se il numero di casi è uguale al numero dei controlli.

Se è problematico reclutare un gran numero di casi e può essere più facile reclutare più controlli per lo studio. Si può aumentare il numero di controlli per aumentare la potenza statistica

Il rapporto ottimale è di **4 controlli: 1 caso.**
con controlli aggiuntivi (superiori a quattro) l'aumento della potenza statistica è limitata

Misura di associazione

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

Misura associazione

Non conoscendo la numerosità del gruppo degli esposti/trattati, non è possibile stimare l'incidenza di evento, ma è possibile solo calcolare la probabilità (odds) per i casi e per i controlli di essere stati esposti/trattati.

La misura di associazione da utilizzare è quindi l'**odds ratio**

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

Esempio

cervical cancer cases were 76% more likely to have never attended school than controls.

*In the case–control study illustrated in the previous example, the **risk of cervical cancer** was examined in relation to **education** (schooling). Data from Spain and Colombia were pooled in this analysis*

	Schooling		Total
	Never ('exposed')	**Ever ('unexposed')**	
Cervical cancer cases	119 (*a*)	317 (*b*)	436 (n_1)
Controls	68 (*c*)	319 (*d*)	387 (n_0)
Total	**187 (m_1)**	**636 (m_0)**	**823 (*N*)**

[a] Data from Bosch *et al.* (1992)

Odds ratio = (119 / 317) / (68 / 319) = 1.76

95% confidence interval = 1.24–2.46

$\chi^2 = 11.04$, 1 d.f.; $P = 0.0009$

(Confidence intervals and test statistics for the odds ratio were calculated as shown in Appendix 6.1.)

Odds ratio in individual - matched

		Controls		Total
		Exposed	Unexposed	
Cases	Exposed	*r*	*s*	*a*
	Unexposed	*t*	*u*	*b*
	Total	***c***	***d***	***N/2***

r = case exposed and control exposed (+ +)
s = case exposed but control not exposed (+ –)
t = case not exposed and control exposed (– +)
u = case not exposed and control not exposed (– –)

The *matched odds ratio* can be calculated as

Odds ratio = *s/t* (provided *t* is not equal to 0)

This odds ratio calculation considers only the **discordant** pairs.
It can be explained intuitively: pairs where both case and control were exposed or where both were unexposed give no information about the relationship of the exposure to disease

Esempio

*A case–control study was carried out in Canada to **assess whether artificial sweeteners**, particularly saccharin, increased the **risk of bladder cancer**. Newly diagnosed cases of bladder cancer that occurred among residents in the provinces of British Columbia, Nova Scotia and Newfoundland between April 1974 and June 1976 were identified through provincial cancer registries and cooperative pathologists and urologists.*
*A total of **821** eligible cases were ascertained, and **632** of these were personally interviewed in their homes using a structured questionnaire. Reasons for failure to interview included death (56), refusal (65), too ill to be interviewed (25), and refusal of permission by the attending physician (34). Most interviews were done within three months of diagnosis, and all within six months. For each case, an individual **matched on sex, age (within 5 years)**, and neighborhood residence was interviewed*

		Controls		Total
		Exposed	Unexposed	
Cases	Exposed	468 (*r*)	87 (*s*)	555 (*a*)
	Unexposed	73 (*t*)	4 (*u*)	77 (*b*)
	Total	**541 (*c*)**	**91 (*d*)**	**632 (*N/2*)**

[a] Data from Howe *et al.* (1977)

Matched odds ratio = 87/73 = 1.19

95% confidence interval for the matched odds ratio = 0.86–1.65

McNemar's $\chi^2 = 1.23$; $P = 0.27$.

(The calculation of confidence intervals and significance tests for matched case–control studies is explained in Breslow & Day (1980)).

(Howe et al.*, 1977)*

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

Esempio

*The relation between use of **oral contraceptives** by young women and their **risk of breast cancer** was investigated in a population-based case–control study conducted in Los Angeles County. The cases were patients with histologically confirmed breast cancer, first diagnosed between July 1972 and May 1982, diagnosed before age 37 years, and without a prior history of malignancy. A total of 510 eligible cases were identified through the local population-based cancer registry, of whom 458 were still alive at the time of the first contact through their doctors. Physicians gave permission to contact 393 (86%) of these patients. Of these, 26 could not be located and 37 refused to be interviewed. Thus, completed questionnaires were obtained from 330 patients. Sixteen of these patients were later excluded because no suitable individually matched control was found.*

Only 62% (314/510) of all eligible patients were included in the final analysis.

Selection bias?

Low participation levels can introduce bias if cases who used oral contraceptives were more or less likely to participate in the study.

If, for instance, users of oral contraceptives were more likely to have a less aggressive form of breast cancer than non-users and, hence, a better survival, this would lead to **over-estimation** of the effect of oral contraceptives since a high proportion of the deaths would have occurred among non-users

(Pike et al*., 1983).*

*The possible association between **oral contraceptive** use and the **risk of breast cancer** at young ages (under 45 years) was investigated in a population-based case–control study conducted in Sweden and Norway. In Norway, where notification of all cancer diagnoses is mandatory, cases were identified from population-based cancer registries. A total of 114 eligible women were identified of whom 105 (92%) participated. For each case who agreed to participate, two controls were chosen from an up-to-date national population register. Potential controls were mailed a request to participate. If an answer was not received within four weeks or if the control refused to participate, a new control was selected. Nine controls were never located; 34 never answered the letter; 38 refused to participate; 4 were either temporarily abroad and could not be reached or had mental disorders. Thus, to obtain two controls for each case, it was eventually **necessary to select 295 controls from the population register**. Only **72%** of the women with whom contact was sought were interviewed.*

non-response because some eligible controls cannot be traced or because they refuse to participate

the control series may not be representative of the population from which the cases arise.

Selection bias?

(Meirik et al*., 1986)*

*In the study described, an introductory **letter** with a brief description of the **aim and scope** of the study was sent initially to cases and controls. If they agreed to participate, they were interviewed personally by specially trained professional female interviewers.*

the aim of the investigation was explained to the women involved. This may have **increased recall bias**, particularly since the study was carried out during a time of great public concern about oral contraceptives and breast cancer.

Recall bias?

accurate measurements of past exposures are usually difficult to obtain, and the degree of accuracy and **completeness of these measurements may be different for cases and controls**

because patients with the disease under study may be inclined to answer questions more carefully than control subjects

independent source of information (e.g., medical records) may help to determine whether there was a systematic difference in recall by cases and controls.

if women using oral contraceptives are more likely than nonusers to examine their breasts, or to have them examined by a physician or nurse, or to undergo mammography, diagnostic bias may be introduced.

oral contraceptive users are more investigated and therefore more likely to be diagnosed with breast cancer than non-users

Diagnostic bias

Strength and weaknesses

- **The main advantages of these studies are:**

1. They are efficient in time and cost (at least compared with prospective cohort studies)
2. They provide the possibility to investigate a wide range of possible risk factors.
3. They are particularly suitable to investigate rare diseases or diseases with a long induction period.

- **The main disadvantages of these studies are:**

1. It may be difficult to select an appropriate control group (*selection bias*).
2. It is difficult to obtain accurate unbiased measures of past exposures (*information bias*).
3. The temporal sequence between exposure and disease may be difficult to establish (*reverse causality*).
4. They are not suitable for investigating rare exposures (unless the exposure is responsible for a large proportion of cases, i.e., the population excess fraction is high).

Confondimento

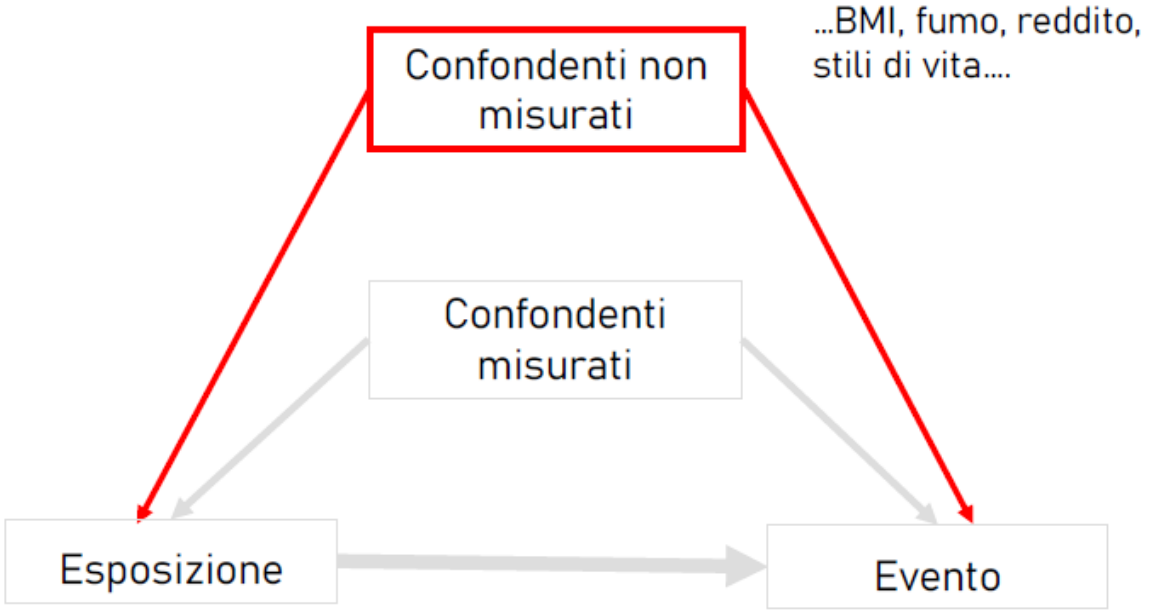
La forza dell'errore introdotto dipende da:

1. Proporzione di soggetti con il confondente nei differenti livelli di esposizione

Tanto più la prevalenza del confondente è sbilanciata tra esposti e non esposti, tanto maggiore sarà l'errore.

2. Forza e direzione dell'associazione tra confondente ed evento

Tanto più forte è l'associazione tra confondente ed evento, tanto maggiore sarà l'errore

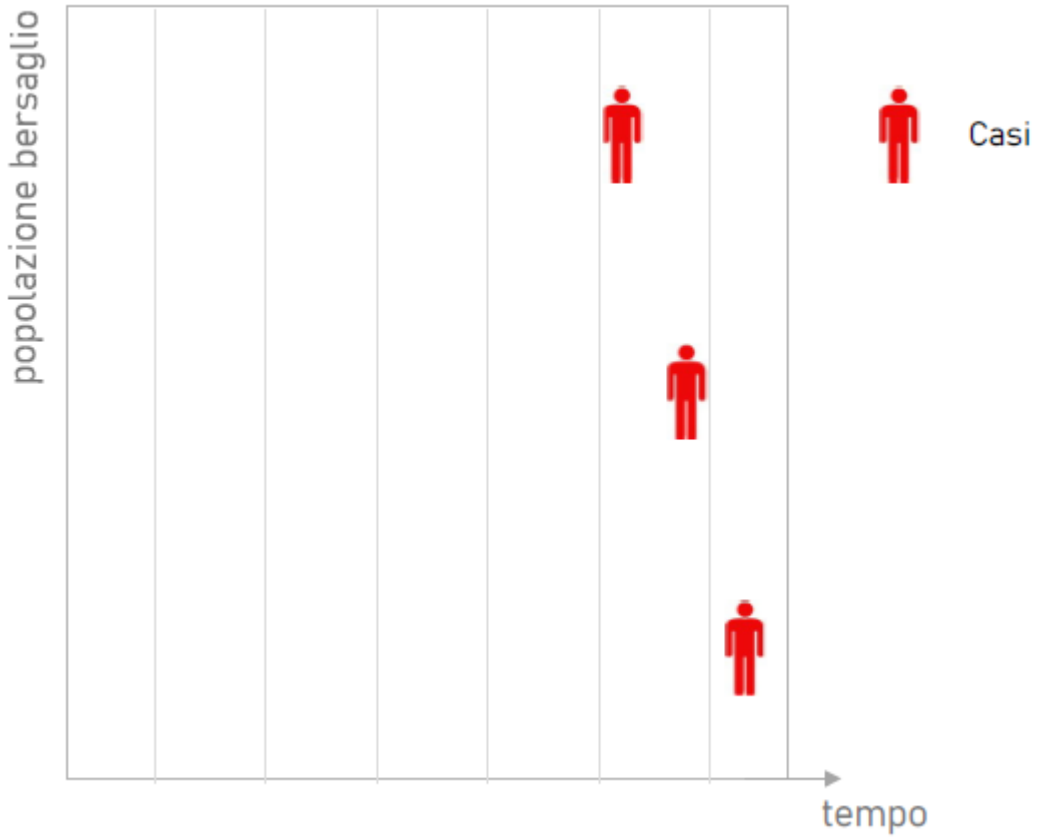


Come eliminare un fattore confondente…

Gli studi case-only (entro paziente)

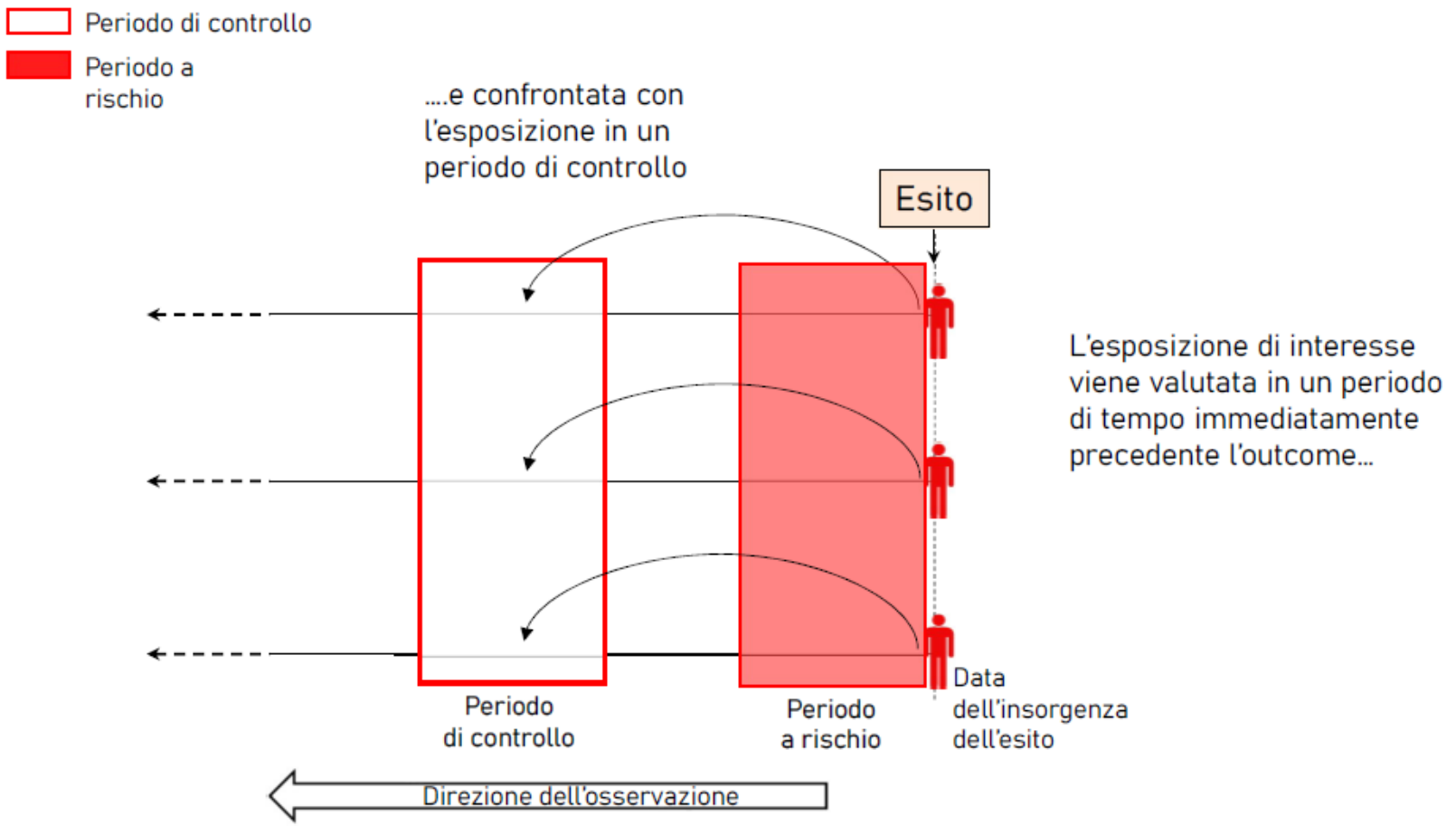
Si utilizzano i casi come controlli di sé stessi

I fattori di rischio costanti nel tempo sono automaticamente tenuti sotto controllo dal disegno stesso.



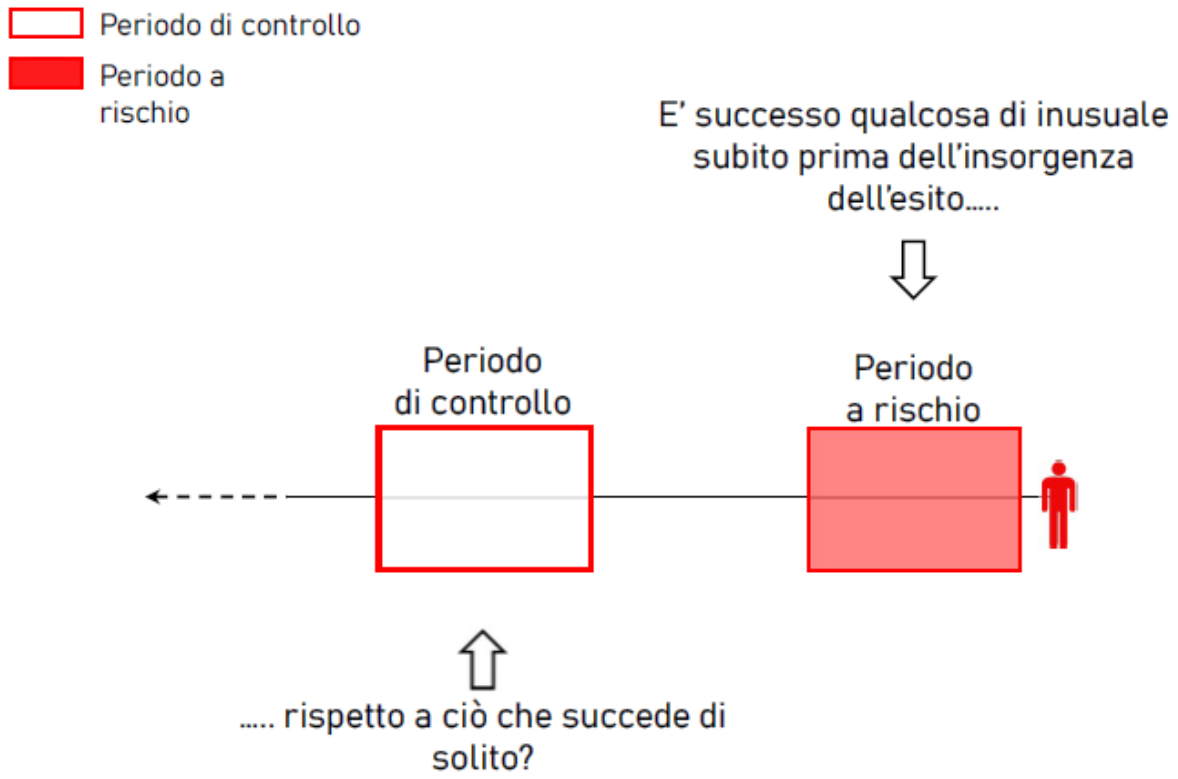
on courtesy of prof.Matteo Franchi

Gli studi case-only – case crossover



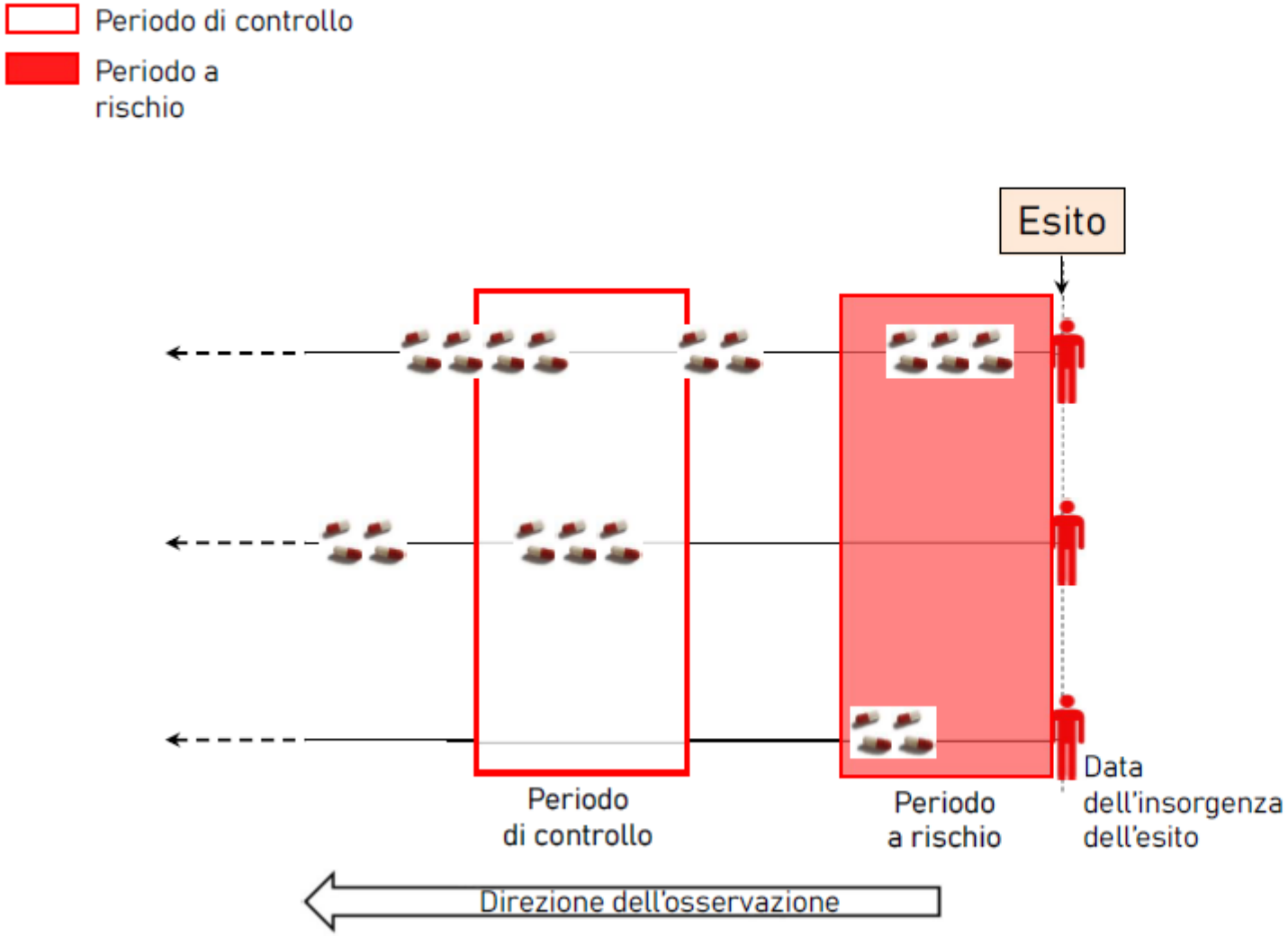
on courtesy of prof.Matteo Franchi

Gli studi case-only – case crossover



on courtesy of prof.Matteo Franchi

Gli studi case-only – case crossover



… un'efficiente applicazione del casecrossover si ottiene quando l'esposizione di interesse è di natura transitoria…

		Periodo di controllo: SI	Periodo di controllo: NO
Periodo a rischio	SI	a	b
	NO	c	d

OR = b/c

Rapporto tra il numero di casi con esposizione discordante

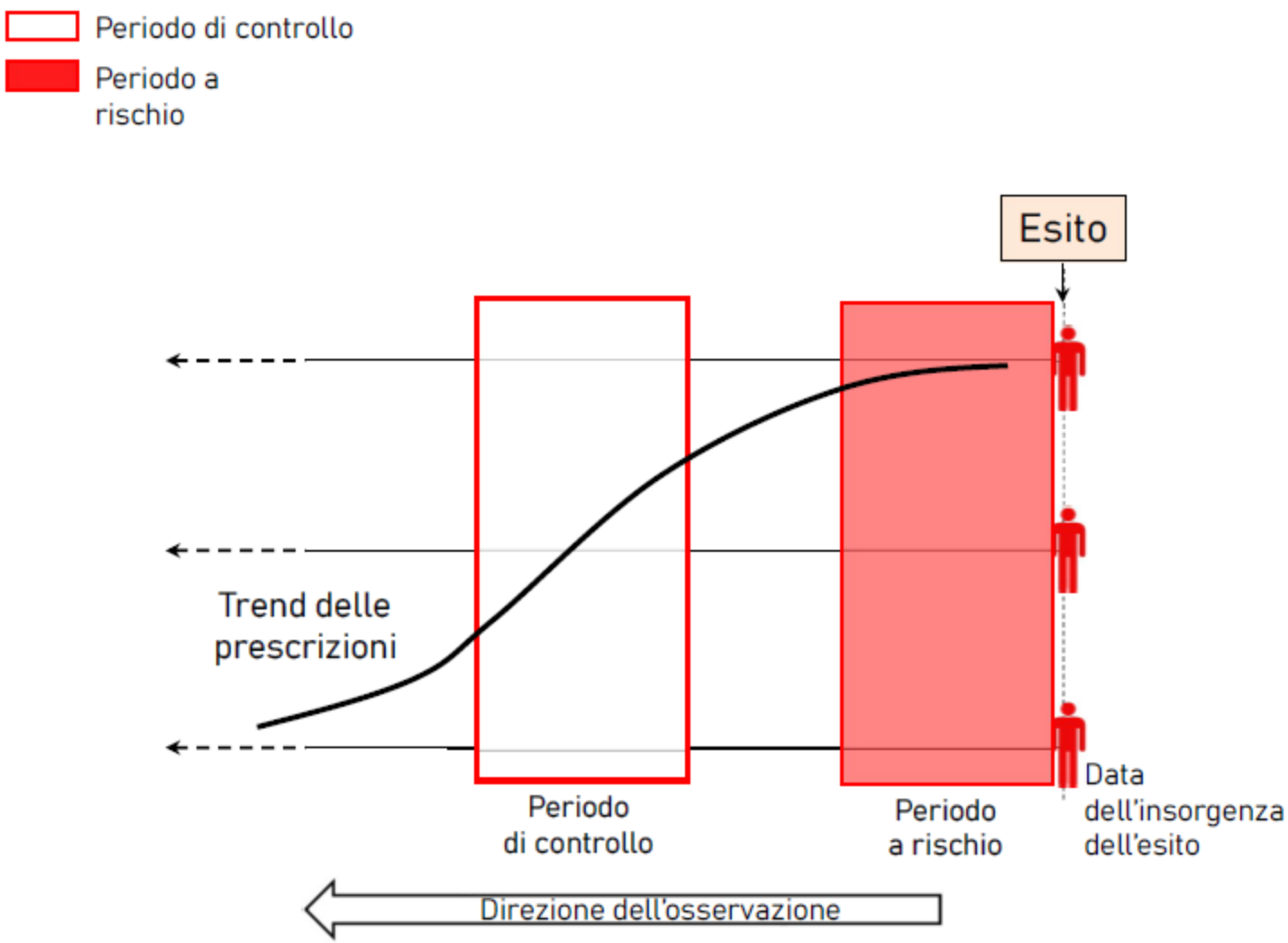
on courtesy of prof.Matteo Franchi

Gli studi case-only – case crossover

Assunti:

1. L’esito in studio deve essere di tipo acuto
2. L’esposizione deve essere di natura transitoria
3. La scelta dell’ampiezza delle finestre deve basarsi su razionali biologici
4. Non devono essere presenti trend nelle prescrizioni di interesse

on courtesy of prof.Matteo Franchi



L'OR calcolato col disegno case-crossover dipende sia dall'effetto dell'esposizione sia dal trend temporale delle prescrizioni

on courtesy of prof.Matteo Franchi

Esempio Warfarin e rischio di sanguinamento gastrointestinale

5.1 Data source and motivating example

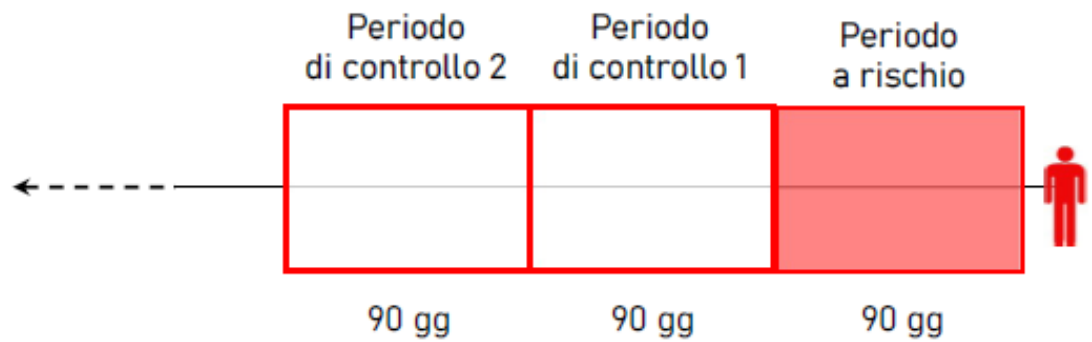
The illustration was based on a previously published case-control study of the effect of the anticoagulant warfarin on the risk of gastrointestinal bleeding, using data from the United Kingdom's General Practice Research Database (GPRD).[41,42] This example was selected specifically because of the presence of randomised controlled trials of this association. Briefly, we identified all first-ever cases of gastrointestinal (GI) bleed recorded in the GPRD from 2000 through 2005. All subjects had at least 3 years of clinical data in the GPRD at the time of this first GI bleed, with this date being the index date (T_0). The characteristics of this GI bleed case series, including the prevalence of potential confounding factors and the presence of drug channelling, have been extensively reported in previous studies.[7,42]

Drug exposure was defined by prescriptions for warfarin issued by general practitioners. Because warfarin doses are often adjusted dynamically during the time period covered by a single prescription,[43,44] duration is often not recorded or can be unreliable when it is recorded. The most frequent recorded durations for prescriptions are 28 or 30 days, although some are listed as being intended to cover 90 days. As the recommended length of warfarin therapy is 3 months for its most common indication (venous thromboembolism) and since prescription duration is difficult to assess for warfarin, we used 90 days as the primary time window of interest.[43] Exposure to warfarin was operationally defined as being at least one prescription of warfarin written to the subject during this time window (the risk period). The two time periods immediately preceding the 90-day risk period were selected as control periods.

Esempio Warfarin e rischio di sanguinamento gastrointestinale

Periodo di controllo

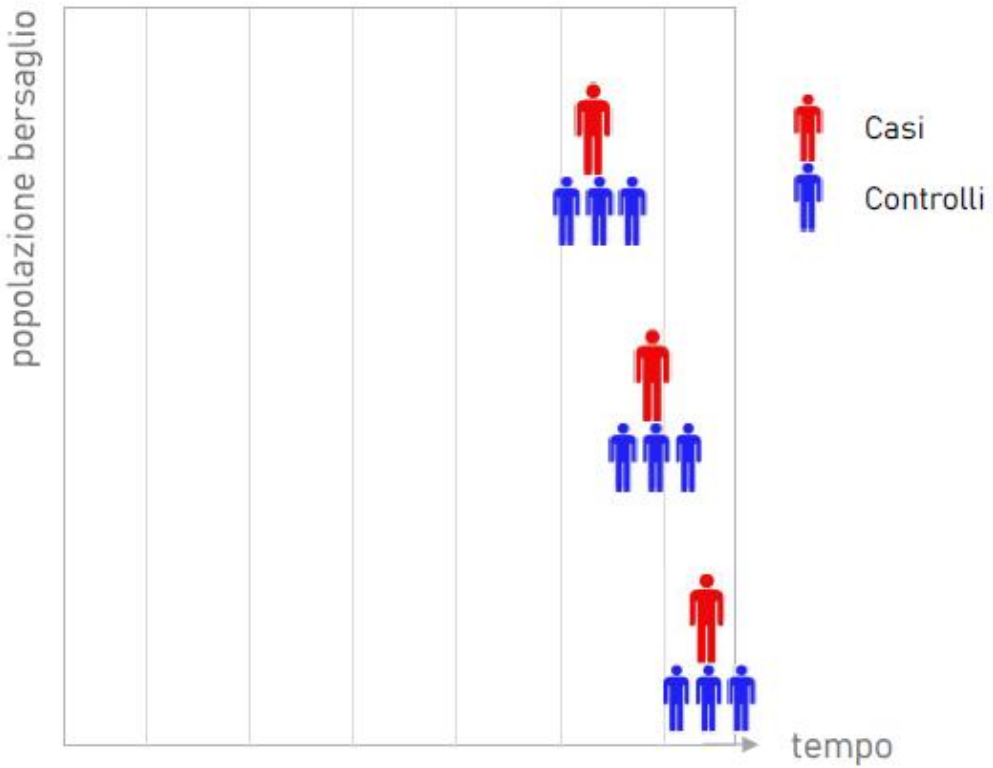
Periodo a rischio



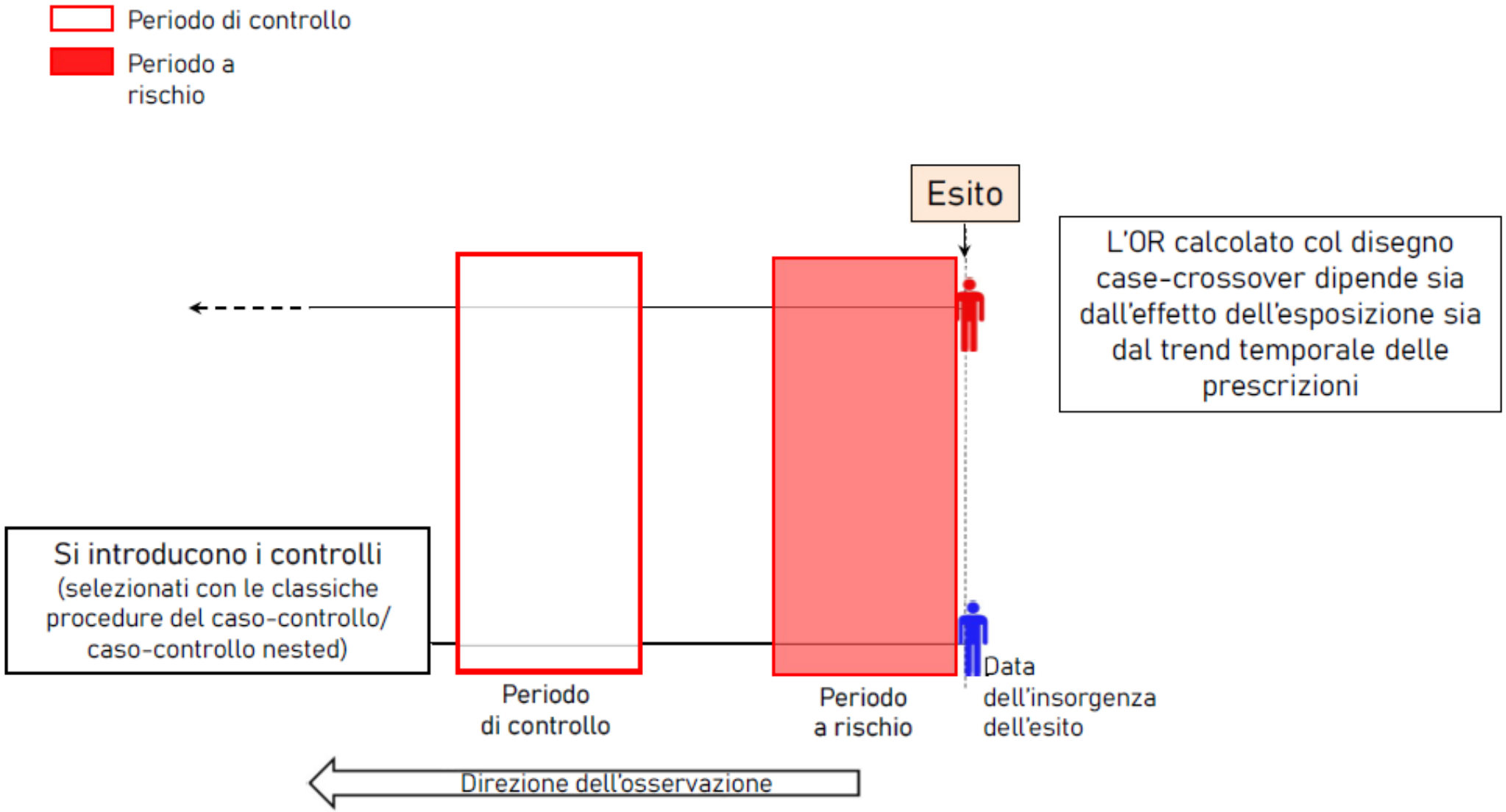
OR=1.60 (1.14-2.23)

on courtesy of prof.Matteo Franchi

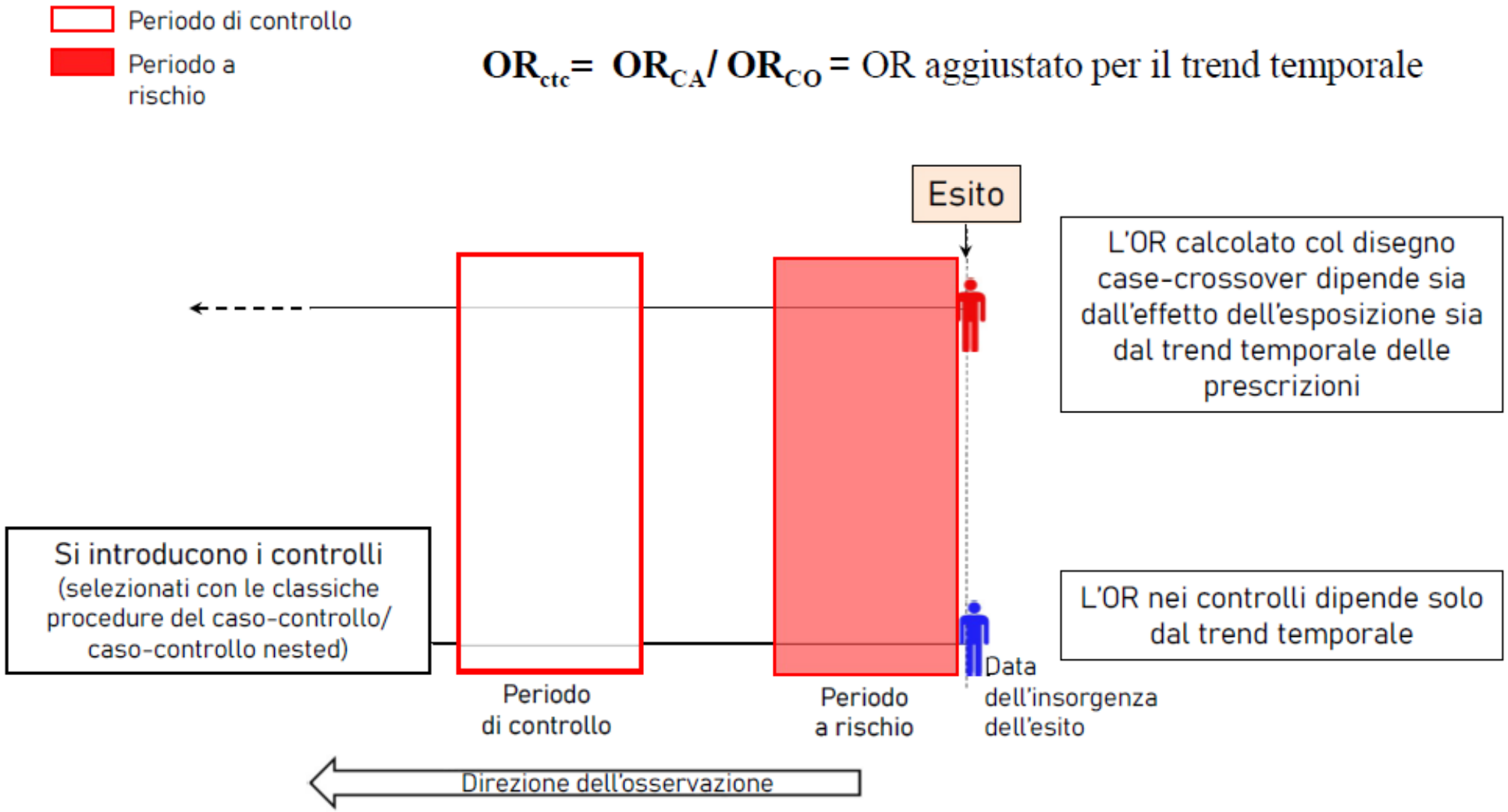
Studi case-time-control



on courtesy of prof.Matteo Franchi



on courtesy of prof.Matteo Franchi



on courtesy of prof.Matteo Franchi

Esempio Warfarin e rischio di sanguinamento gastrointestinale

Periodo di controllo

Periodo a rischio

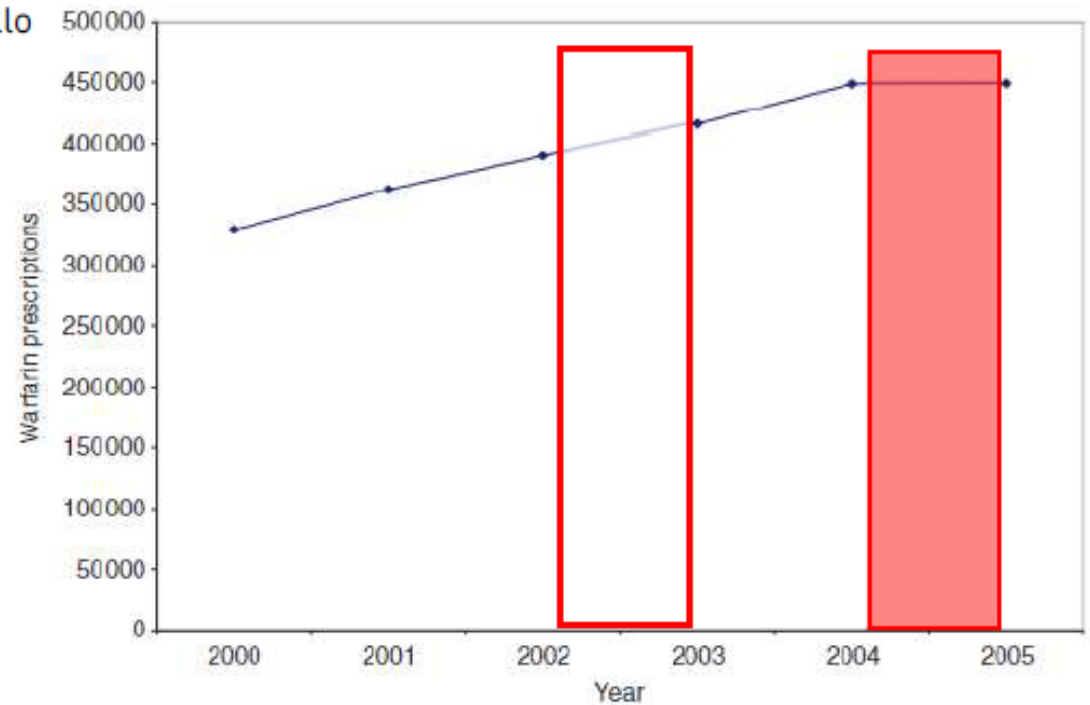


Figure 3 Number of warfarin prescriptions issued by calendar year in the study database. Data from the United Kingdom General Practice Research Database, 2000–2005.

Uso di warfarin nel periodo a rischio vs. uso di warfarin nel periodo di controllo:
OR=2.07 (1.71-2.52)

...L'aumento del rischio osservato in questa finestra potrebbe inglobare l'effetto derivante da un maggior utilizzo di warfarin negli anni più recenti...

Applicando il disegno case-time-control:
OR=1.72 (1.08-2.43)

on courtesy of prof.Matteo Franchi

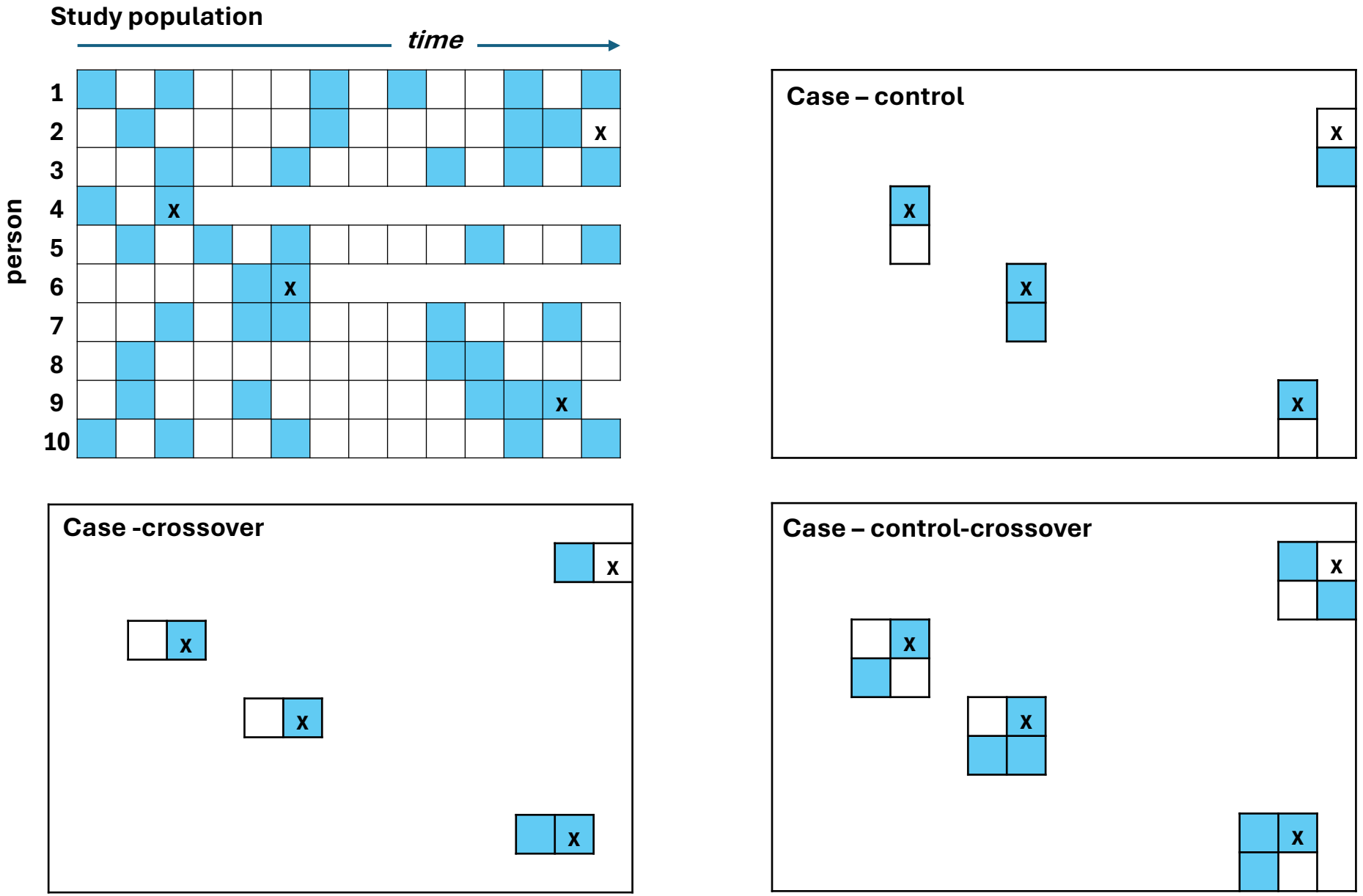


Fig. 3 — Relationships among designs to study *acute effects*. An hypothetical study population consits of 10 persons that can be observed over a period of 15 days. Shaded squares are exposed days, X marks an event. It is assumed that exposure can be classified correctly for each day. For explanation of study designs see text

Studi case-time-control

Assunti:

- Stesse assunzioni del case-crossover
- Il trend temporale nell’esposizione è lo stesso per casi e controlli

on courtesy of prof.Matteo Franchi

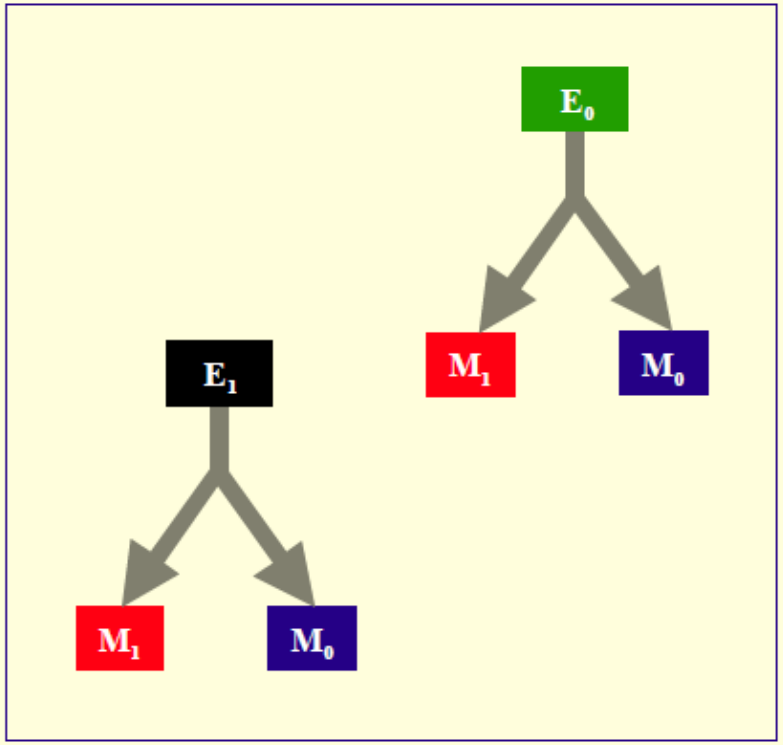
Studi di coorte, con coorte parallela: punti di forza/debolezza, metodologie di pianificazione, qualità metodologica, conduzione e analisi

(M. Cinquini)

STUDI CONTROLLATI CON CONTROLLI STORICI

Se il problema insito degli studi non controllati è l'assenza di un gruppo di controllo, la via più razionale per salire nella scala gerarchica della solidità dell'evidenza è quello di affiancare alla serie di casi un adeguato gruppo di controllo. Se questo è rappresentato dai pazienti precedentemente trattati con altri farmaci, lo studio è definito controllato con controlli storici.

Ad esempio, per valutare l'efficacia del trattamento dell'emicrania con un antidolorifico di nuova generazione, un medico potrebbe confrontare la frequenza con la quale un gruppo di pazienti trattati con il nuovo farmaco riferisce l'attenuazione dei sintomi, con quella riferita dai pazienti che lo stesso medico aveva precedentemente trattato con altri farmaci.



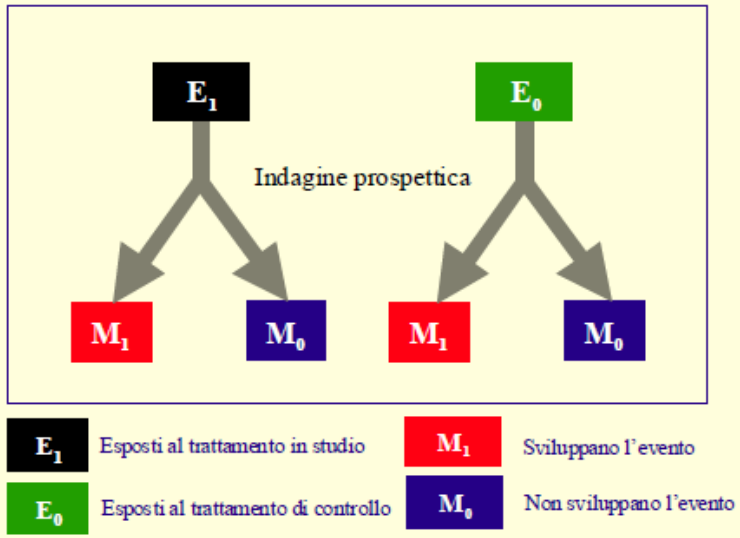
La scelta di controlli storici risponde a criteri di economicità e praticità. Tale considerazione può essere estremamente importante quando lo studio è rivolto a patologie rare e quando il beneficio atteso dal trattamento è elevato. Tuttavia, una simile scelta comporta spesso difficoltà d'interpretazione dei risultati in considerazione del fatto che nel tempo possono verificarsi cambiamenti, anche importanti, della storia naturale delle malattie.

STUDI CONTROLLATI CON CONTROLLI CONCORRENTI: L'EPIDEMIOLOGIA ANALITICA

Se il problema insito negli studi con controlli storici è la presenza di un gruppo di controllo reclutato in condizioni non paragonabili a quelle dei casi, la via più razionale per salire nella scala gerarchica della solidità dell'evidenza è quello di reclutare i casi e i controlli nello stesso arco temporale e con le stesse modalità.

Riprendendo gli schemi tradizionali di classificazione degli studi di epidemiologia analitica, le due principali categorie sono:

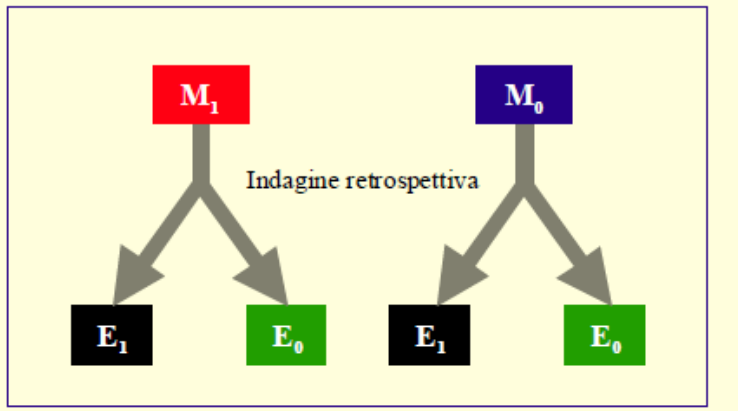
gli studi di coorte, che consistono nella selezione degli individui in base all'intensità di esposizione al fattore in studio (ad esempio i trattati e non trattati con una certa terapia) e nella loro successiva osservazione nel tempo per registrare l'insorgenza degli eventi che si suppone siano causati dall'esposizione.



E_1 Esposti al trattamento in studio — M_1 Sviluppano l'evento

E_0 Esposti al trattamento di controllo — M_0 Non sviluppano l'evento

gli studi caso-controllo, che consistono nella selezione degli individui in funzione della presenza o meno dell'evento di interesse (ad esempio una malattia o un effetto avverso) e nella loro anamnesi per valutare retrospettivamente l'intensità di esposizione ai fattori che si suppone siano causa dell'evento.



E_1 Esposti al trattamento in studio — M_1 Sviluppano l'evento

E_0 Esposti al trattamento di controllo — M_0 Non sviluppano l'evento

STUDI DI COORTE

Definizione

Lo studio di coorte

- Nell’ambito degli studi osservazionali, lo studio di coorte, o studio di follow-up, è considerato lo studio analitico per eccellenza.

- Infatti permette di osservare l’insorgenza della patologia dopo l’avvenuta esposizione, di misurare quindi i tempi di esposizione e di ottenere stime di associazione tra l’esposizione e la probabilità di contrarre una determinata patologia.

- Tale procedura consiste nel confronto tra gruppi, denominati “coorti”, costituiti da soggetti esposti e da soggetti non-esposti.

- **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.**

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 in maniera diversa

STUDI DI COORTE

Metodologia

- Questo disegno di studio viene spesso adottato anche nell'ambito degli studi sperimentali o semi-sperimentali, per esempio per valutare la diversa comparsa di ricadute in gruppi di pazienti sottoposti a trattamenti diversi e quindi per confrontare l'efficacia di tali trattamenti.

- Sulla base della selezione dei soggetti si distinguono diverse tipologie di coorti, tra cui le due principali sono:

a) la coorte chiusa

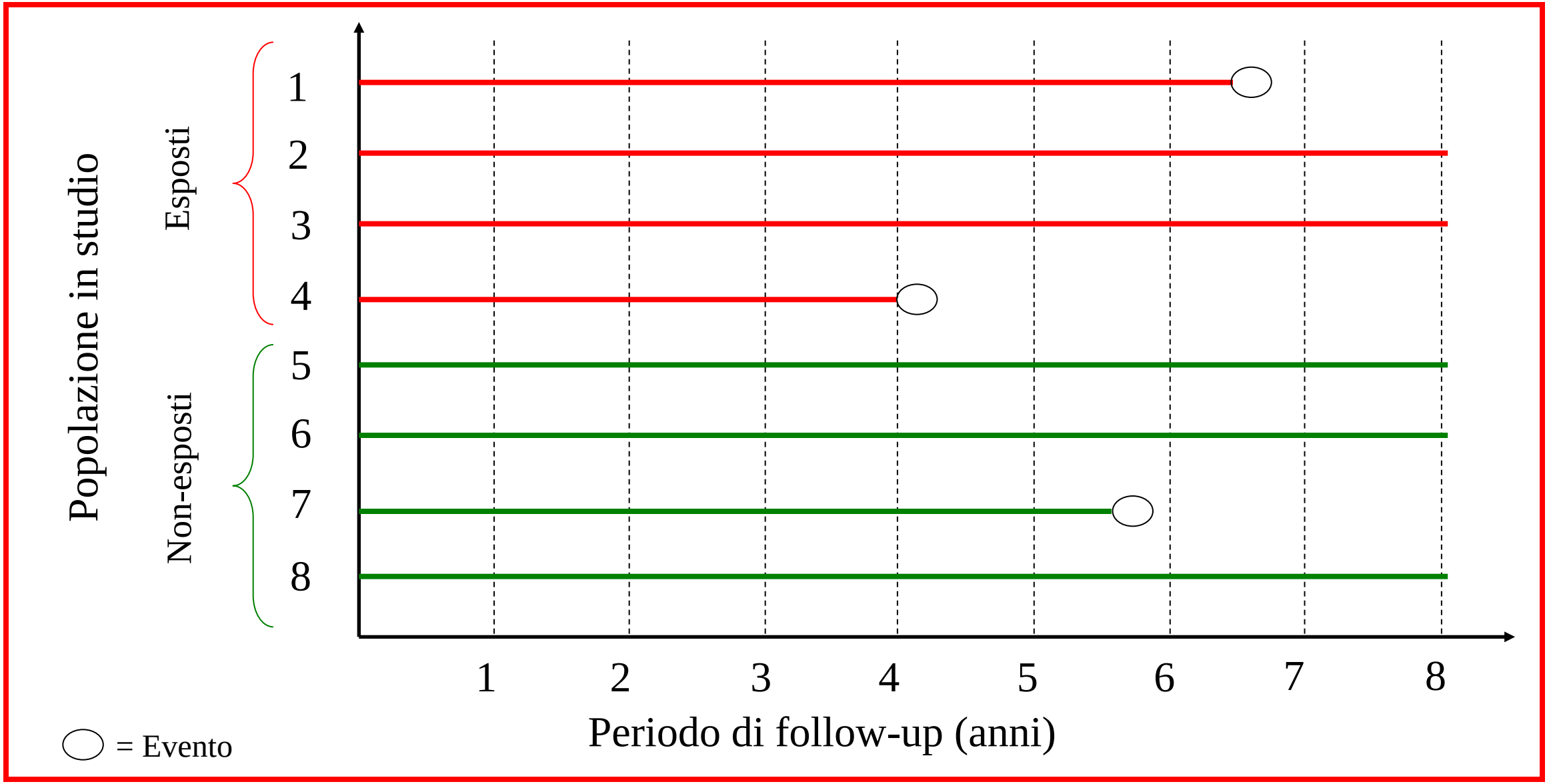
b) la coorte aperta.

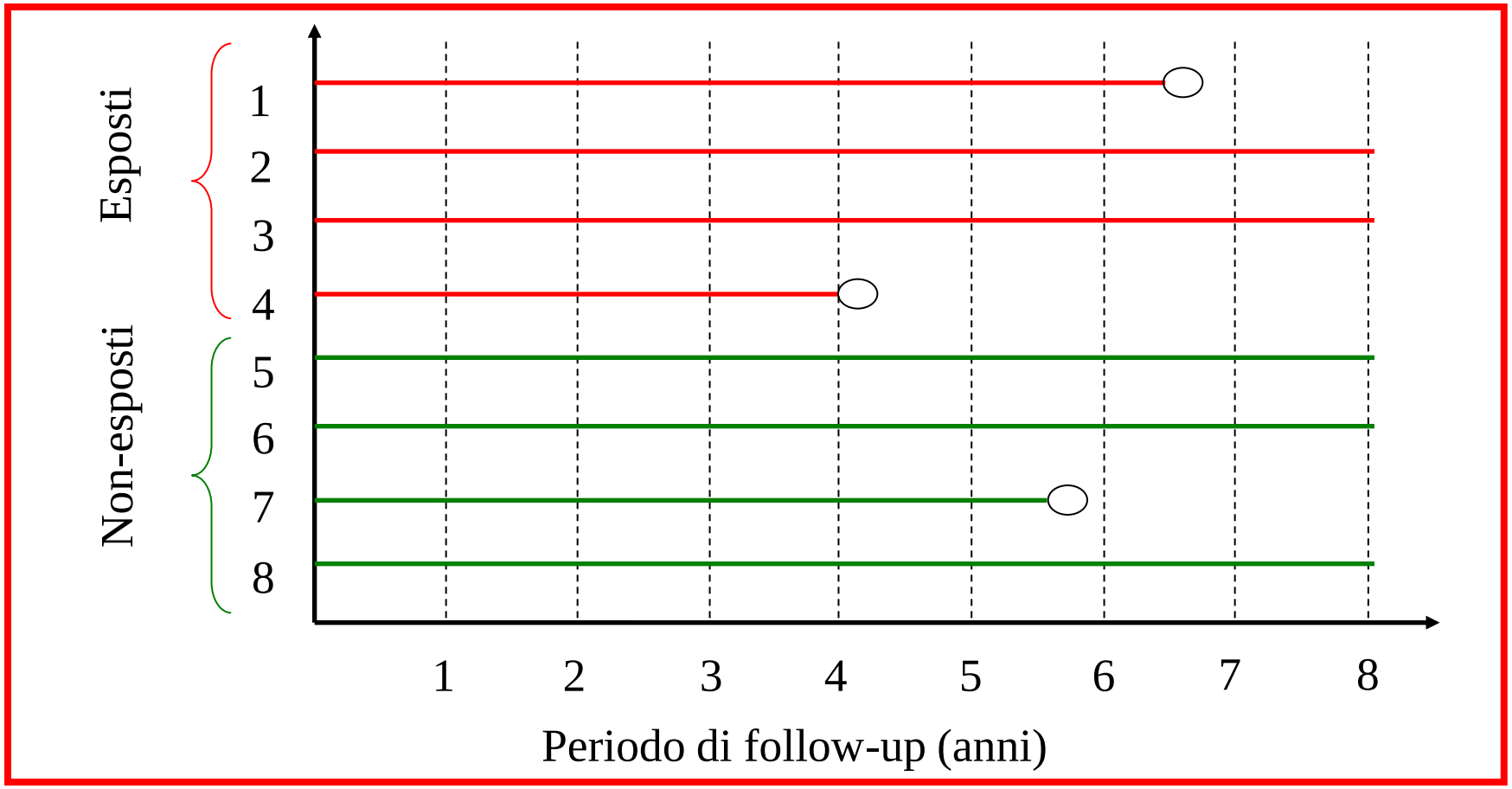
- Nella coorte chiusa il campione in studio viene identificato interamente in un preciso istante temporale.

- Solitamente, vengono identificati simultaneamente i fattori di interesse per lo studio, ovvero la presenza di esposizioni in soggetti sani o il tipo di trattamento per quanto riguarda gli studi clinici.

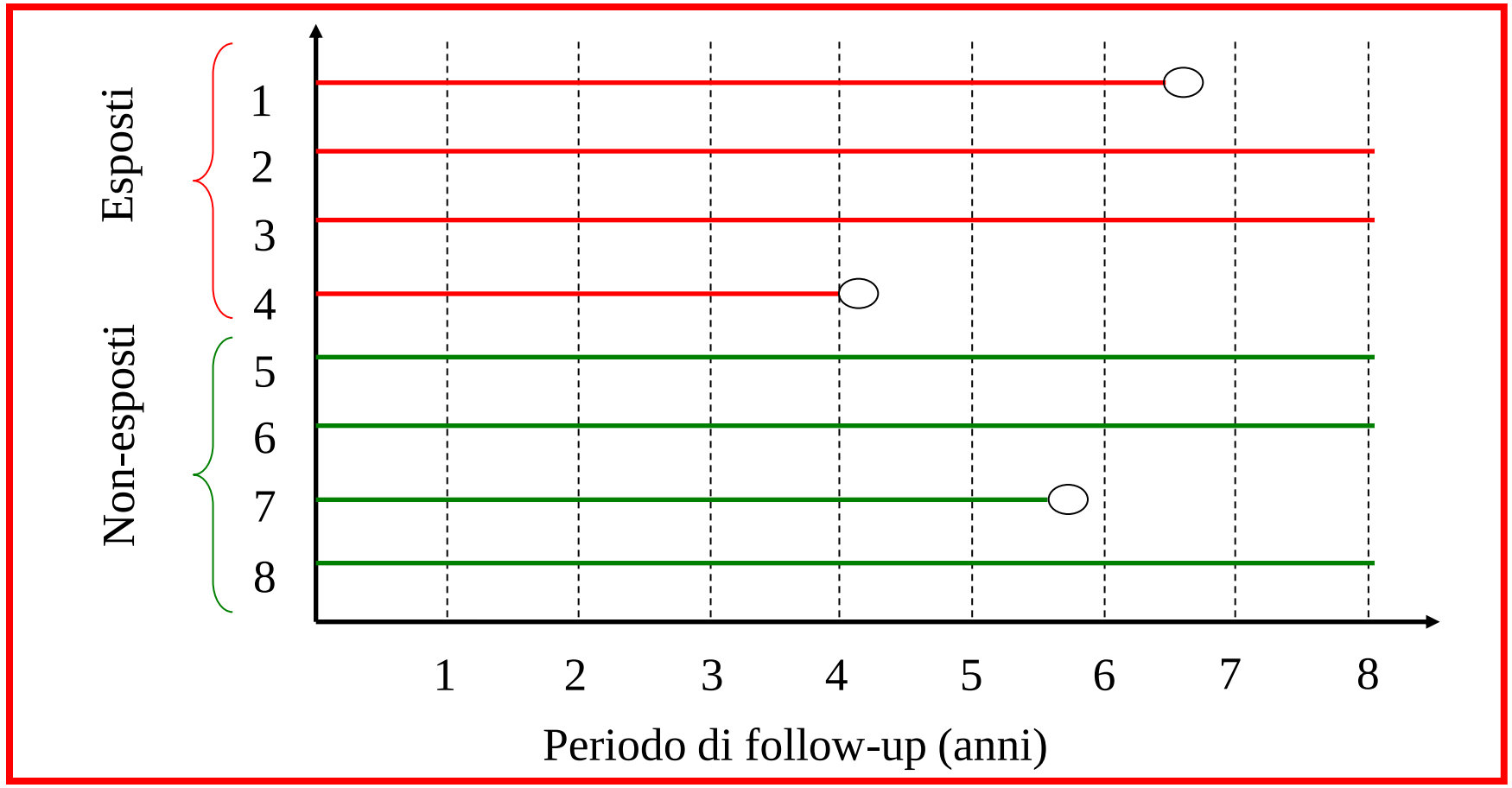
- I soggetti vengono quindi seguiti nel tempo per rilevare la comparsa dell'evento di interesse (patologie, decesso, ricadute o recidive negli studi clinici).

ESEMPIO SCHEMATICO DI UNO STUDIO DI COORTE CHIUSA

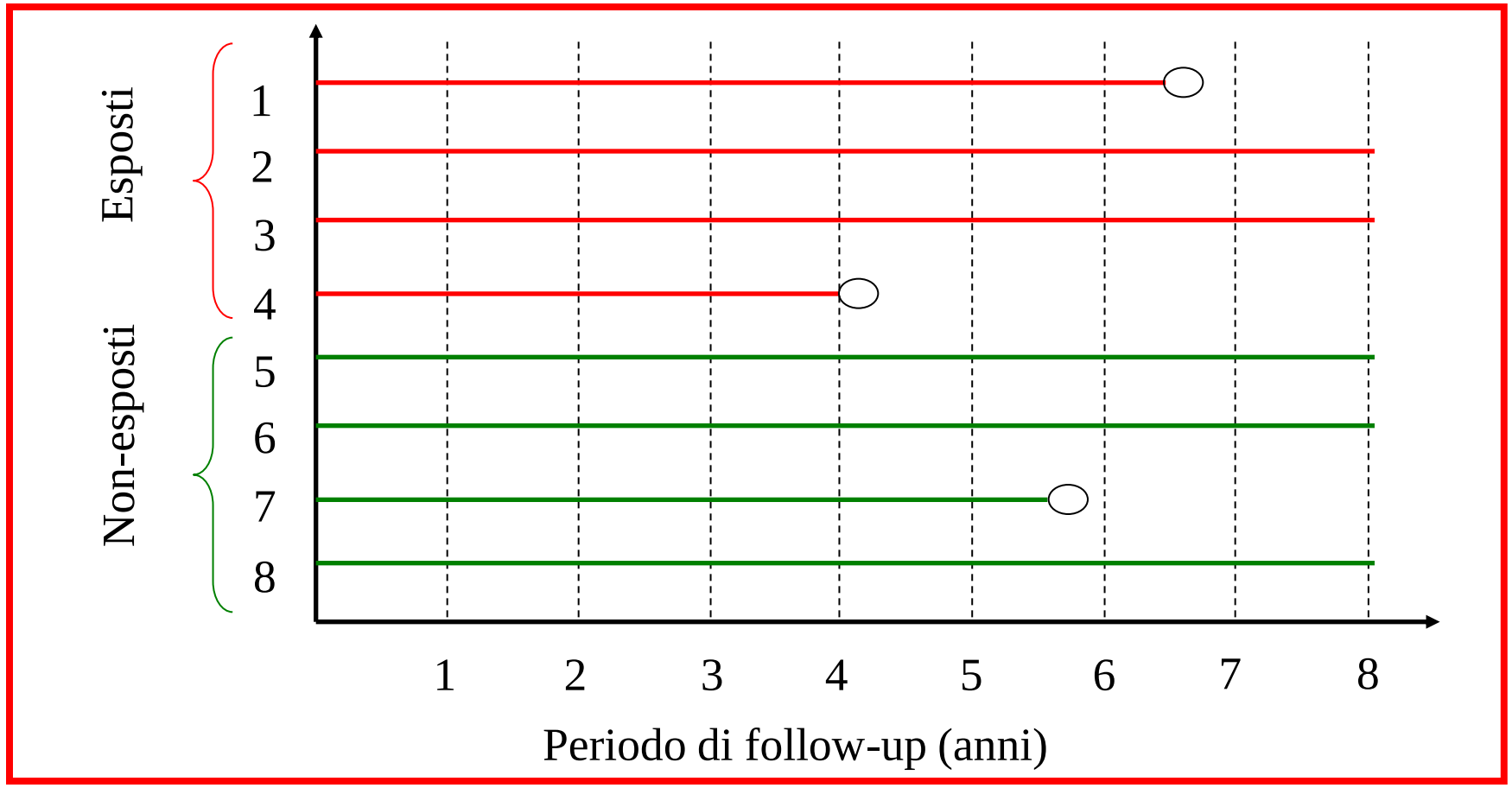




- Nell'esempio sopra illustrato il rischio di ammalarsi negli esposti durante il periodo di osservazione (8 anni) sarà quindi pari al 50% (2 eventi osservati su 4 soggetti in studio), mentre il corrispondente rischio nei soggetti non esposti sarà del 25% (1 evento osservato su 4 soggetti).



- Si noti come le stime di rischio siano condizionate alla durata del tempo di osservazione (periodo di follow-up). Infatti se tale periodo fosse stato di soli tre anni si sarebbero osservati 0 eventi in entrambi i gruppi di esposizione e quindi le corrispondenti stime di rischio sarebbero state entrambe pari a zero.



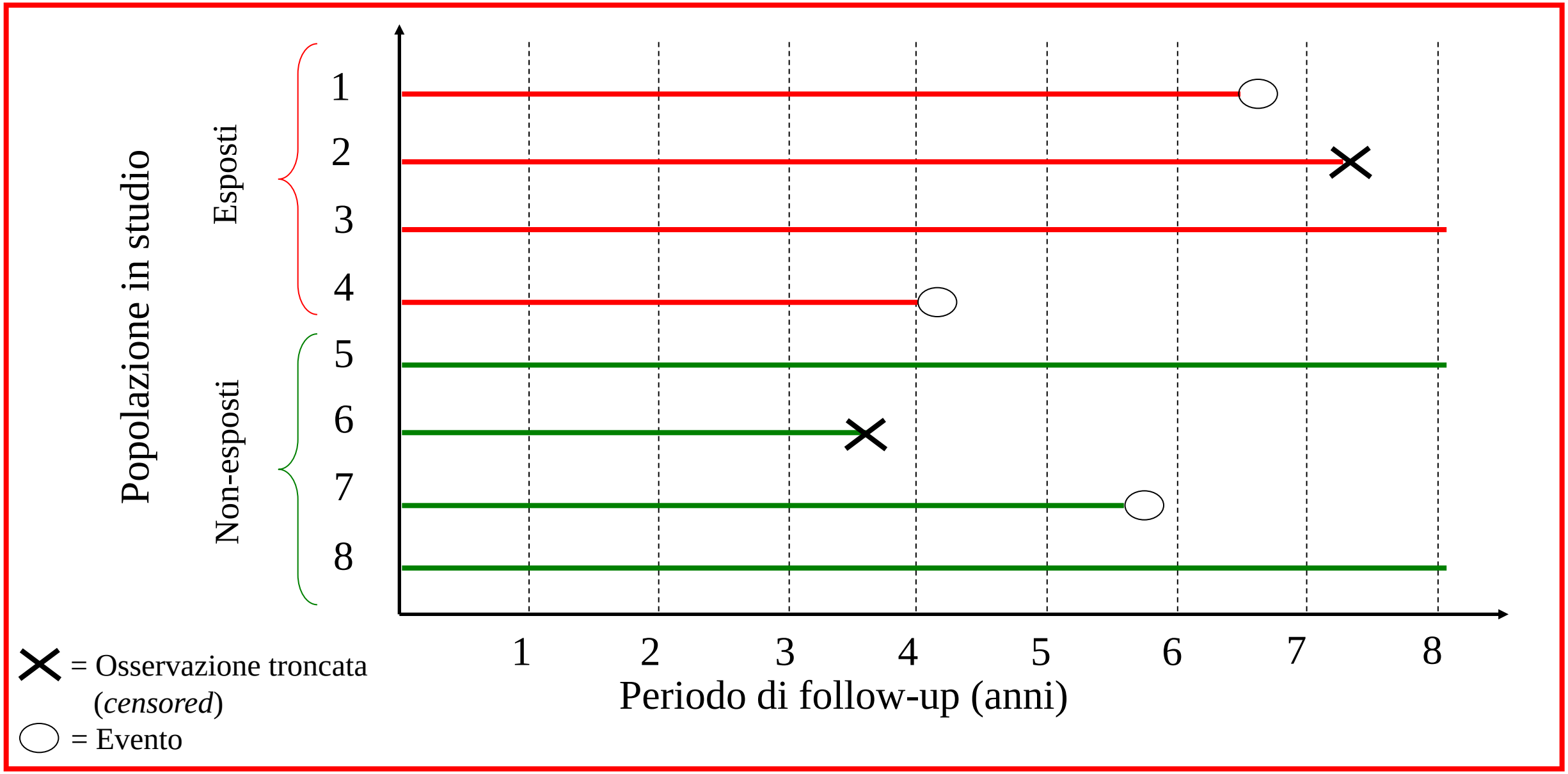
- Se invece il tempo di osservazione fosse stato di 6 anni, si sarebbe osservato un rischio del 25% in entrambi i gruppi, e di conseguenza la stima di RR sarebbe stata pari a 1.

- In una coorte aperta, al contrario del caso precedente, la perdita del soggetto durante il follow-up può avvenire anche per motivi diversi dalla fine del periodo di osservazione o dal manifestarsi dell'evento di interesse.

- Il soggetto può risultare "perso di vista" (ad esempio per fenomeni di migrazione), oppure può decedere per cause diverse da quella in studio.

- In tal caso il tempo di osservazione si definisce troncato (*censored)*.

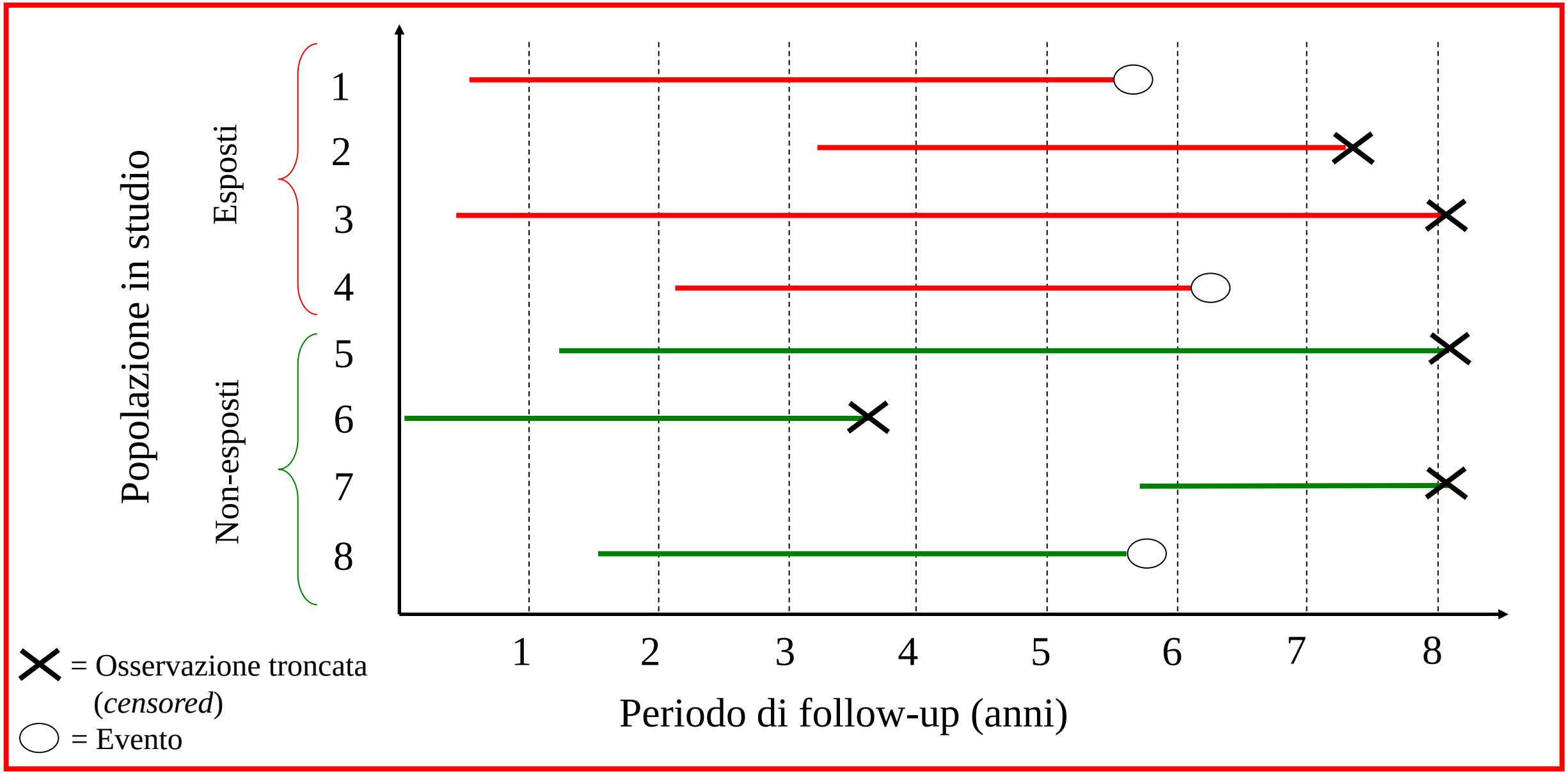
ESEMPIO SCHEMATICO DI UNO STUDIO DI COORTE APERTA

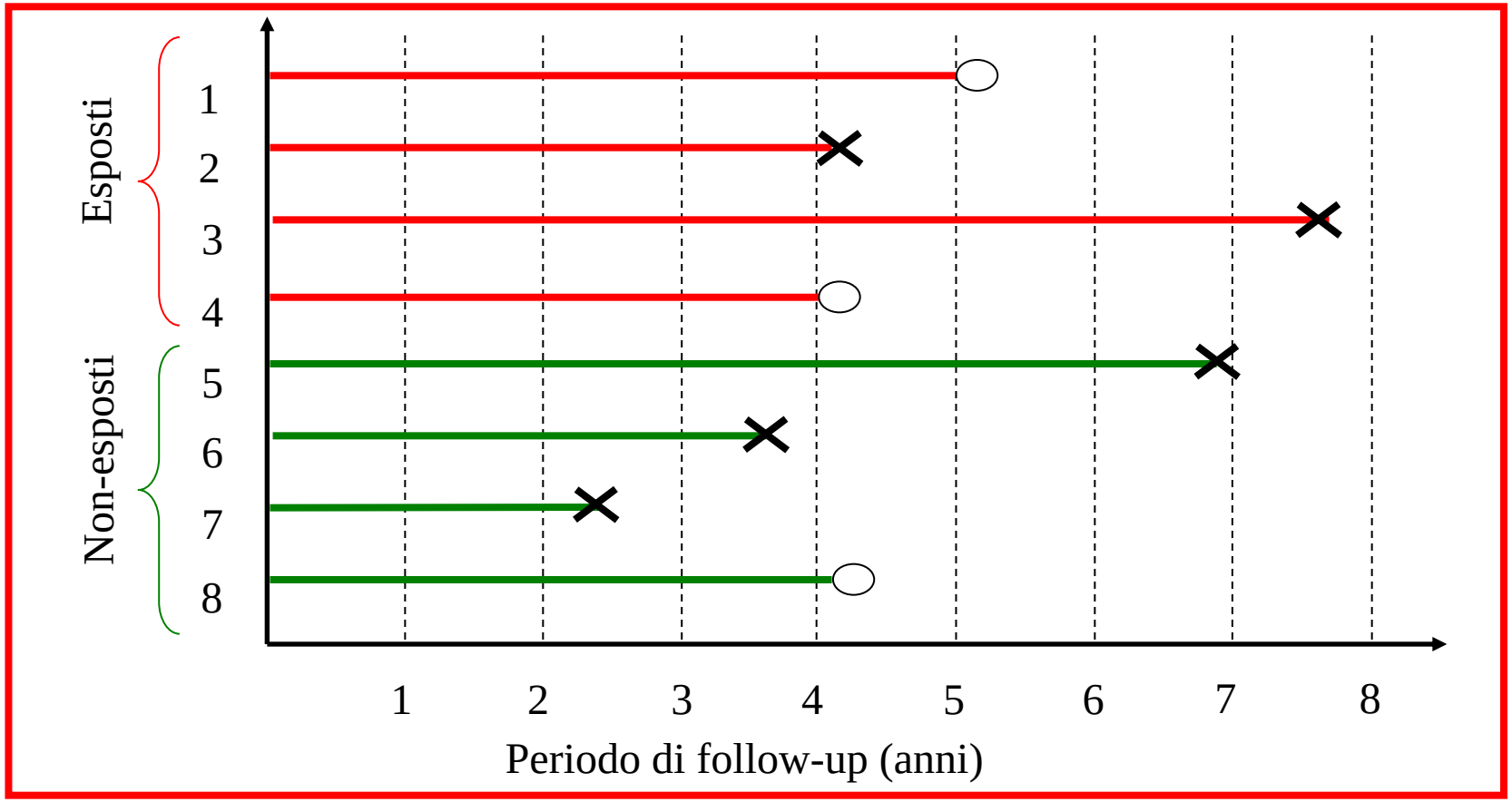


- Un caso particolare, ma molto comune di coorte aperta, è la coorte dinamica, che è costituita da individui che possono cambiare nel tempo, per movimenti naturali, anagrafici o amministrativi.

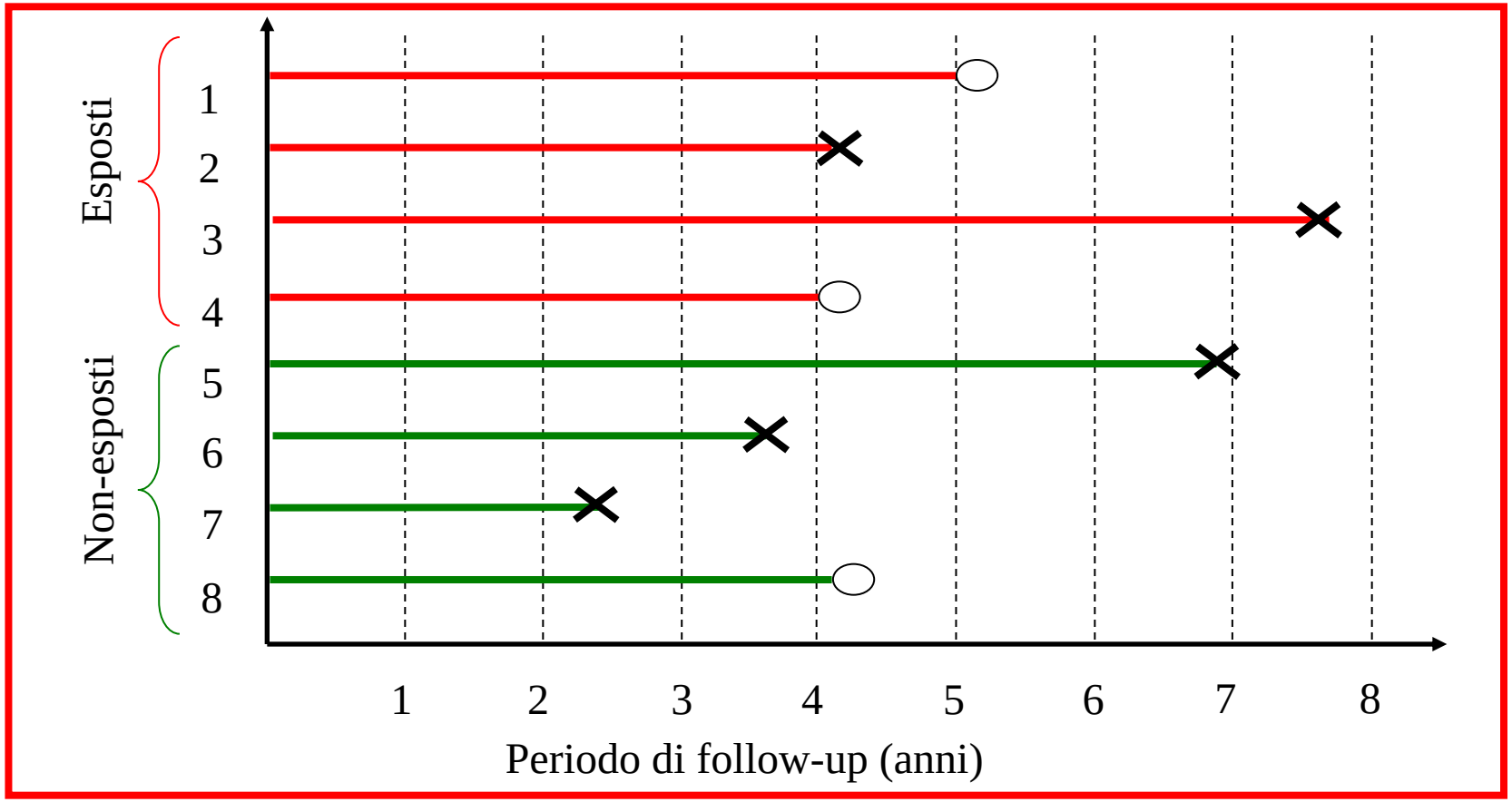
- Un esempio può essere rappresentato dalla maggior parte delle coorti che prendono parte a uno studio clinico.

ESEMPIO SCHEMATICO DI UNO STUDIO DI COORTE DINAMICA





- In genere la presenza del troncamento impedisce di produrre stime dirette del rischio, dato che i tempi di osservazione per i diversi soggetti sono diversi.



- Una possibilità per ottenere stimatori di rischio relativo è quella di stimare un'altra fondamentale grandezza utilizzata in Epidemiologia, ovvero il **Tasso**.

STUDI DI COORTE

Punti di forza/debolezza

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**

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.

IL CONFONDIMENTO
IN UNO STUDIO DI COORTE

- Nell'analisi di dati epidemiologici occorre valutare la presenza di variabili che, se associate sia alla variabile risposta (incidenza, mortalità, ecc...) che al fattore in studio (esposizione, trattamento, fattori genetici, ecc...) possono produrre distorsioni nelle stime di associazione.

- Tali fattori prendono il nome di **confondenti**, il fenomeno viene denominato "**confondimento**" e l'eventuale distorsione indotta nelle stime viene denominata "**bias da confondimento**".

- Esempio di confondimento generato da una variabile categorica (sesso) nell'ambito di uno studio di coorte in cui anche l'esposizione è riportata su scala dicotomica (presenza o assenza).

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Morti	**Pers/anno**	**Morti**	**Pers/anno**	**Morti**	**Pers/anno**
Esposti	**108**	**44870**	**30**	**3218**	**78**	**41652**
Non-Esposti	**51**	**21063**	**44**	**11699**	**7**	**9364**

$RR_T = 0.99$ $(0.71, 1.4)$

$RR_1 = 2.5$ $(1.6, 3.9)$

$RR_2 = 2.5$ $(1.2, 5.4)$

$RR_T = 0.99$ $(0.71, 1.4)$ $RR_1 = 2.5$ $(1.6, 3.9)$ $RR_2 = 2.5$ $(1.2, 5.4)$

- Perché vi sia confondimento occorre che la stima di RR ottenuta nell'analisi dei dati raggruppati (analisi *pooled*) sia diversa da quella derivante dall'analisi stratificata per i livelli del confondente.
- Se però le stime di RR risultassero diverse tra i vari strati del fattore in esame vi sarebbe indicazione che tale variabile modifichi l'effetto dell'esposizione (o del trattamento).
- In tal caso vi sarebbe **interazione** tra le variabili e non confondimento.
- Nell'esempio sopra riportato le stime di RR nei due strati del confondente sono identiche (RR=2.5) e molto diverse dalla stima *pooled* (RR=0.99), per cui non vi è dubbio che ci si trovi in presenza di confondimento.

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Morti	Pers/anno	Morti	Pers/anno	Morti	Pers/anno
Esposti	108	44870	30	3218	78	41652
Non-Esposti	51	21063	44	11699	7	9364

- Perché una variabile si comporti da confondente è necessario che sia associata sia alla variabile risposta che al fattore in studio, (nell'esempio riportato, sia all'esposizione che al tasso).

- Nell'esempio sopra riportato ciò può essere evidenziato stimando i tassi nelle due categorie di esposizione, separatamente per i maschi e per le femmine.

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Morti	Pers/anno	Morti	Pers/anno	Morti	Pers/anno
Esposti	108	44870	30	3218	78	41652
Non-Esposti	51	21063	44	11699	7	9364

Maschi **Femmine**

Esposti

$$\lambda_{E,M} = \frac{30}{3218} = 932.3 \cdot 10^{-5} \qquad \lambda_{E,F} = \frac{78}{41652} = 187.3 \cdot 10^{-5}$$

Non-Esposti

$$\lambda_{NE,M} = \frac{44}{11699} = 376.1 \cdot 10^{-5} \qquad \lambda_{NE,F} = \frac{7}{9364} = 74.8 \cdot 10^{-5}$$

- I tassi nei maschi sono più elevati dei corrispondenti tassi nelle femmine entro ogni categoria di esposizione (associazione tra stimatore di rischio e variabile di stratificazione).

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Morti	Pers/anno	Morti	Pers/anno	Morti	Pers/anno
Esposti	108	44870	30	3218	78	41652
Non-Esposti	51	21063	44	11699	7	9364

- Inoltre le femmine contribuiscono al gruppo degli esposti con molte più persone-anno rispetto ai maschi, mentre il loro contributo al gruppo dei non-esposti è simile a quello dei maschi (associazione tra esposizione e variabile di stratificazione).

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Morti	Pers/anno	Morti	Pers/anno	Morti	Pers/anno
Esposti	108	44870	30	3218	78	41652
Non-Esposti	51	21063	44	11699	7	9364

- In maniera più intuitiva, si può affermare che il confondimento è dovuto alla presenza di molte femmine tra gli esposti, che, presentando tassi inferiori a quelli dei maschi, hanno mascherato l'effetto dell'esposizione.

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

- **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

MODIFICAZIONE D'EFFETTO IN UNO STUDIO DI COORTE

- La modificazione d'effetto si produce quando una variabile interagisce con l'esposizione, per cui tale fenomeno, viene anche denominato interazione.

- La presenza di interazione può essere evidenziata dalla presenza di un diverso andamento dello stimatore di rischio entro le categorie della variabile di interazione.

- Contrariamente a quanto avviene per il confondimento, il ricercatore non può produrre stime aggiustate per l'effetto di tale variabile (stime comuni d'effetto), in quanto l'associazione tra esposizione e rischio è diversa nei gruppi a confronto.

- Un esempio di tale fenomeno è illustrato nella tabella seguente, in cui gli eventi di interesse sono rappresentati dai nuovi casi di malattia osservati in una coorte ipotetica.

	Tutti i soggetti		Strato 1 Maschi		Strato 2 Femmine	
	Casi	Pers/anno	Casi	Pers/anno	Casi	Pers/anno
Esposti	391	769309	189	478383	202	290926
Non-Esposti	119	358341	78	242043	41	116298

$RR_T = 1.5$ $(1.2, 1.9)$

$RR_1 = 1.2$ $(0.94, 1.6)$

$RR_2 = 2.0$ $(1.4, 2.8)$

$RR_T = 1.5$ $(1.2, 1.9)$ $RR_1 = 1.2$ $(0.94, 1.6)$ $RR_2 = 2.0$ $(1.4, 2.8)$

- Nell'intera coorte in studio il rischio di ammalarsi sembra associato all'esposizione, essendo lo stimatore di RR superiore a 1 in modo statisticamente significativo.

- Stratificando per sesso, si osserva una differenza di rischio relativo tra maschi e femmine, poiché i primi presentano un lieve eccesso di rischio non significativo (RR = 1.2), mentre nelle femmine l'esposizione porta a un raddoppio del rischio (RR = 2.0).

- In sostanza i risultati suggeriscono la presenza di interazione tra sesso ed esposizione, nel senso che le femmine sarebbero più suscettibili all'esposizione rispetto ai maschi .

MODELLO DI REGRESSIONE DI POISSON CON INTERAZIONE (CENNI)

- La presenza di interazione non permette di ottenere stime comuni di RR tra i diversi livelli del modificatore d'effetto, per cui in genere conviene produrre stime separate.

- Tuttavia può essere conveniente verificare se l'evidenza di modificazione d'effetto sia attribuibile o meno alla fluttuazione statistica .

- Per tale scopo esistono alcuni test formali, oppure, in alternativa, si può fittare un modello di Poisson con un termine di interazione e testarne la significatività statistica .

ALTRE 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.

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

STUDI DI COORTE

Qualità metodologica

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
- Uncontrolled case series: varie; nessuna raccomandata da Cochrane

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*

Misure di accuratezza diagnostica e validazione di un test

(E. Rulli)

Test medico

Test medico è una procedura volta a

individuare,

diagnosticare,

caratterizzare,

monitorare

una condizione medica specifica

Validazione clinica di un test

- Scopo primario: valutare le caratteristiche del test nell'eseguire una diagnosi o diagnosticare una malattia
- Le tipologie di studi dipendono dallo scopo del test
    - diagnosi,
    - screening,
    - stadiazione
    - Sostituzione di un test esistente o in aggiunta

Caratteristiche di un test

Lo scopo è di trovare dei test che riescano a predire con **accuratezza** la presenza attuale o lo sviluppo futuro di un determinato outcome.

I test per loro stessa natura non sono infallibili, ovvero possono portare a dei falsi positivi ed a dei falsi negativi.

Ci si affida ai test infatti per produrre delle **stime accurate** ma non delle certezze assolute.

Per valutare la bontà di un test di screening o diagnostico, è possibili utilizzare una serie di indici quali **sensibilità, specificità, accuratezza, valore predittivo positivo e negativo**

Reference

La valutazione della performance di un'indagine diagnostica richiede il confronto tra i risultati dell'indagine e uno standard di riferimento

gold standard **reference standard**

In diagnostica oncologica, l'esempio tipico è quello della verifica, in un campione di n pazienti, del risultato di un'indagine radiologica (**test**) rispetto al referto istopatologico (**reference standard**)

Caratteristiche di un test

Quando si confronta il risultato di un test di screening con il **reference standard** oppure con l'accertamento di una malattia, è possibile calcolare delle misure di accuratezza.

VERO POSITIVO (VP): è risultato positivo al test ed effettivamente ha la malattia

VERO NEGATIVO (VN): è risultato negativo al test ed effettivamente NON ha la malattia

FALSO POSITIVO (FP): è risultato positivo al test ma NON ha la malattia

FALSO NEGATIVO (FN): è risultato negativo al test ma in realtà ha la malattia

		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

Indici test

Per poter valutare l'utilità dei risultati di un test è necessario conoscere:

SENSIBILITA'

probabilità di ottenere una **classificazione positiva**, dato che la **malattia** è davvero **presente**.

"dato un individuo con la malattia, qual è la probabilità che il test risultati positivo?".

SPECIFICITA'

probabilità di ottenere una **classificazione negativa**, dato che la **malattia** è davvero **assente**.

"dato un individuo che NON ha la malattia, qual è la probabilità che il test risultati negativo?"

Indici test

SENSIBILITA' = veri positivi/totale malati

		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

SPECIFICITA' = veri negativi/totale sani

		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

Esempio: sensibilità

Sensibilità della mammografia e della risonanza magnetica (RM) a contrasto dinamico nel riconoscimento di lesioni tumorali mammarie in pazienti candidate a mastectomia.

Sono state considerate **99** mammelle in **90** pazienti sottoposte a mastectomia monolaterale (n = 81) o bilaterale (n = 9).

Lo standard di riferimento, costituto dall'esame istopatologico, ha identificato **188** lesioni tumorali.

mammografia: **124** veri positivi e **64** falsi negativi

RM: **152** veri positivi e **36** falsi negativi.

Sensibilità mammografia: 124/(124 + 64) = 0.660

RM 152/(152 + 36) = 0.809

Diremo quindi che la sensibilità (per lesione) della mammografia è risultata pari a 66.0%, mentre quella della RM è risultata pari a 80.9%.

Il tasso o frazione dei FN mammografia: 0.340 o 34.0%

RM: 0.191 o 19.1%

		Malattia
		Presente
Risultato del test	**Positivo**	Veri Positivi (VP)
	Negativo	Falsi Negativi (FN)
	Totale	Totale malati

Esempio: specificità

Screening del carcinoma polmonare mediante TC a bassa dose.

Su un totale di **1611** soggetti asintomatici sottoposti al primo round di screening, sono risultati positivi alla TC **186** soggetti (ulteriormente studiati con TC ad alta risoluzione), **21** dei quali sono stati sottoposti a biopsia. **13** soggetti sono risultati affetti da carcinoma polmonare.

In assenza di tumori diagnosticati tra il primo e il secondo round di screening, abbiamo:

1425 VN (1611 soggetti totali meno 186 positivi)

173 FP (186 positivi meno 13 VP).

La specificità sarà quindi 1425/(1425 + 173) = 1425/1598 =0.892 = 89.2%

In questo caso è considerata una lesione per soggetto. Lesione e soggetto coincidono nell'unità statistica.

		Malattia
		Assente
Risultato del test	**Positivo**	Falsi Positivi (FP)
	Negativo	Veri Negativi (VN)
	Totale	Totale sani

Sensibilità e **specificità** (come pure tasso dei falsi negativi e tasso dei falsi positivi) dipendono dalle caratteristiche tecniche dell'indagine, dall'abilità del Valutatore e della sua équipe (tecnici, infermieri ecc.) nell'eseguirla e dalla capacità del Valutatore nell'interpretarla.

Non sono influenzate dalla prevalenza di malattia nella popolazione indagata

l'attendibilità del
risultato positivo o negativo

Indici test

Valore predittivo positivo (VPP)

Il valore predittivo positivo rappresenta la probabilità che un individuo sia **malato** dato che è risultato **positivo al test**.

"dato che il test è risultato positivo, qual è la probabilità che l'individuo sia davvero malato?"

Valore predittivo negativo (VPN)

Il valore predittivo negativo a probabilità che un soggetto **non abbia la malattia** dato che è risultato **negativo al test**.

"dato che il test è risultato negativo, qual è la probabilità che l'individuo NON abbia davvero la malattia?"

Indici test

VPP = veri positivi/totale test postivi

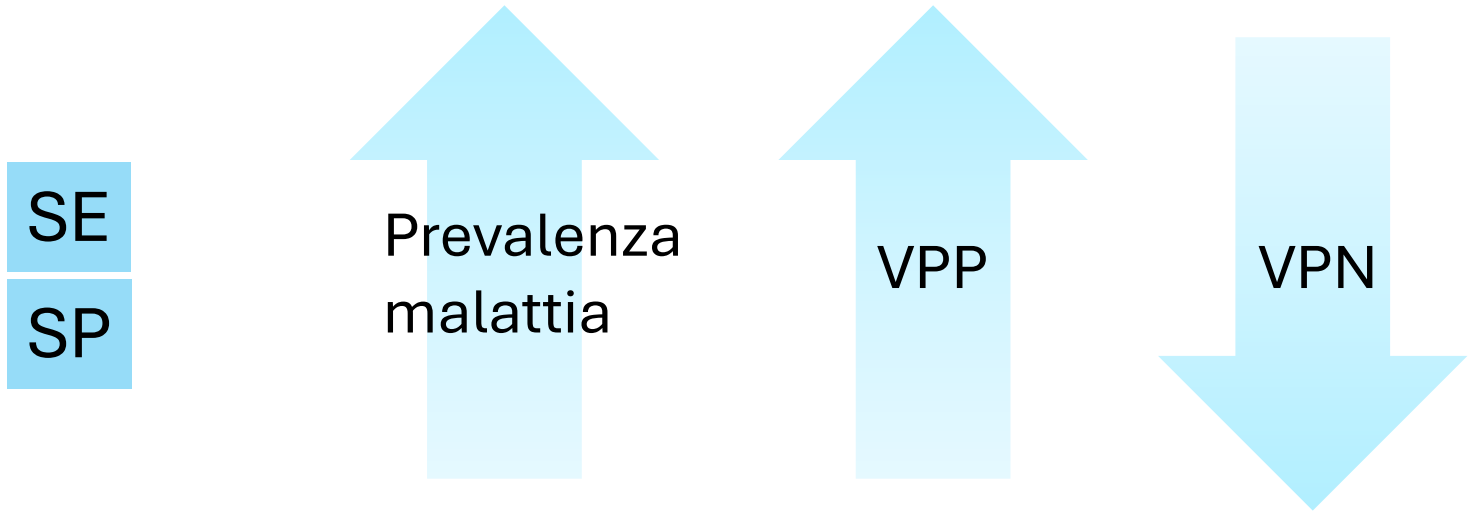
		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

VPN = veri negativi/totale test negativi

		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

I **valori predittivi** non dipendono solo dalle caratteristiche tecniche dell'indagine e dall'abilità dell'équipe nell'eseguirla e interpretarla.

Infatti, rimanendo costanti sensibilità e specificità, essi si modificano in funzione della **prevalenza di malattia**:



Se il campione di soggetti sottoposti all'indagine è **interamente composto da affetti dalla malattia**,
VPP= 1.0 (ovvero 100%)
anche con una sensibilità bassissima (purché diversa da 0)
VPN=0 (ovvero 0%) anche con una specificità altissima (anche pari a 1.0, ovvero al 100%).

Specularmente, se il campione di soggetti sottoposti all'indagine è **interamente composto da non affetti dalla malattia:**
VPN=1.0 (ovvero 100%) anche con una sensibilità bassissima (purché diversa da 0)
VPP=0 (ovvero 0%) anche con una specificità altissima (anche pari a 1.0, ovvero al 100%).

non c'è sensibilità, per quanto alta, che possa diagnosticare una malattia in soggetti sani e non c'è specificità, per quanto alta, che possa diagnosticare l'assenza di malattia in soggetti malati.

Indici test

Accuratezza (ACC)

L’ accuratezza indica la proporzione di osservazioni che sono state **correttamente classificate** come veri positivi o veri negativi.

“qual è la probabilità che un test sia corretto?”

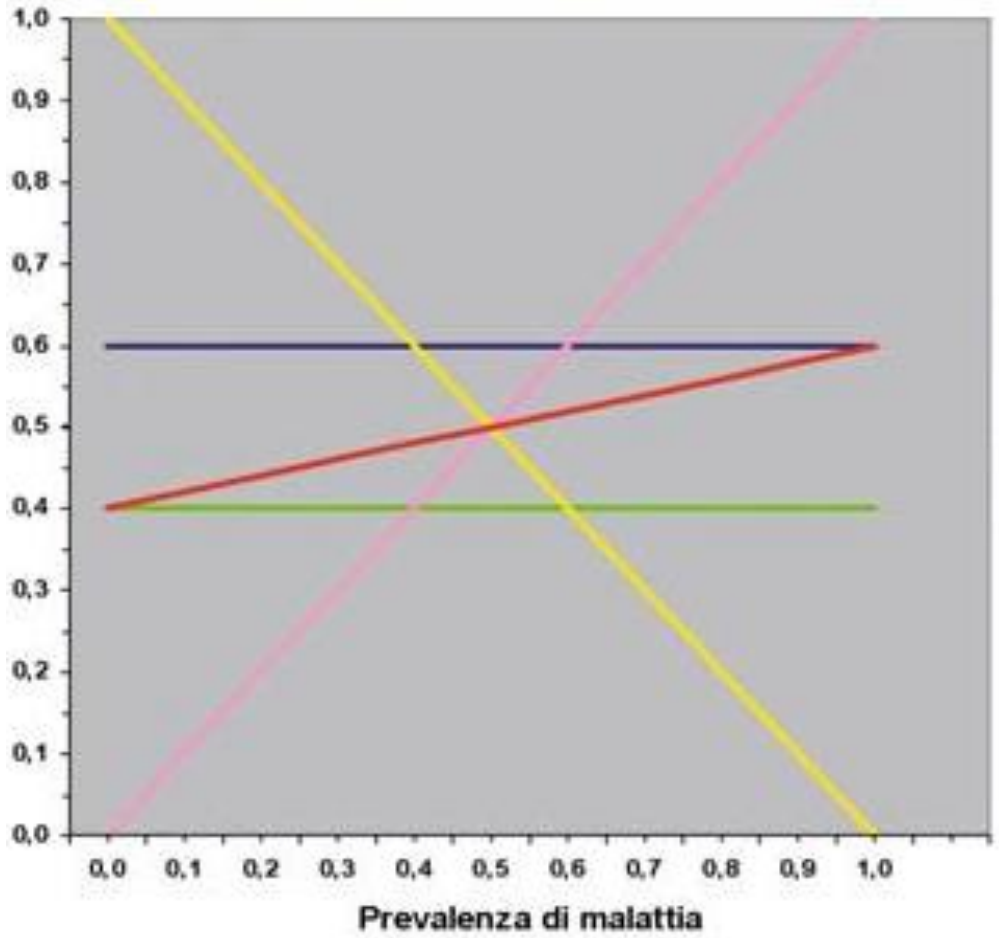
È una sorta di “media” tra sensibilità e specificità, pesata in funzione della prevalenza.

Indici test

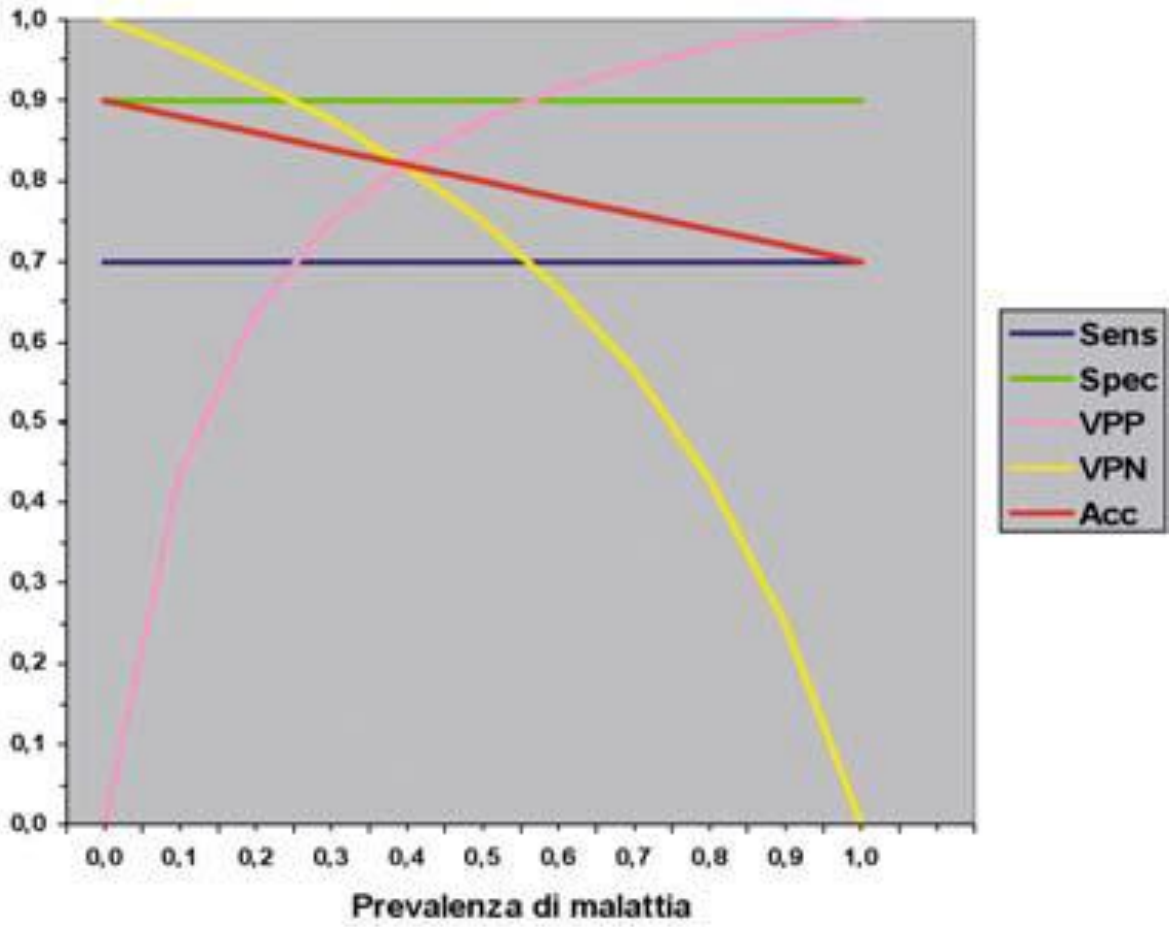
ACC = veri positivi+veri negativi/totale sottoposti a test

		Malattia		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	Veri Positivi (VP)	Falsi Positivi (FP)	Totali test positivi
	Negativo	Falsi Negativi (FN)	Veri Negativi (VN)	Totali test negativi
	Totale	Totale malati	Totale sani	Totale sottoposti al test

Indici test



SE=0.6
SP=0.4



SE=0.7
SP=0.9

Esempio: Valori predittivi in mammografia clinica e di screening.

Con una **sensibilità del 95%** e una **specificità dell'80%** applicate a **1.000** donne **sintomatiche** con nodulo palpabile, supponendo una **prevalenza** di malattia del **50%,** avremo:

		Carcinoma		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	4750	1000	5750
	Negativo	250	4000	4250
	Totale	5000	5000	1000

VPP = 4750/(4750 + 1000) = 0.826 = 82.6%;
VPN 4000/(4000 + 250) = 0.941= 94.1%.

IMPATTO:
Avremo così mediamente solo un approfondimento diagnostico con eventuale **biopsia in una donna sana ogni quasi 5 donne con carcinoma** (4750/1000 =4.75).

La donna con un nodulo palpabile che alla fine si dimostrerà benigno non riterrà inutili o pericolosi approfondimenti diagnostici anche invasivi.

Esempio: Valori predittivi in mammografia clinica e di screening.

Con una **sensibilità del 95%** e una **specificità dell'80%** applicate a **10000** donne **asintomatiche,** supponendo una **prevalenza** di malattia del **0.3%,** avremo:

		Carcinoma		
		Presente	**Assente**	**Totale**
Risultato del test	**Positivo**	285	1940	2225
	Negativo	15	7760	7775
	Totale	300	9700	10000

VPN =7760/(7760 +15) = 0.998 = 99.8%

VPP =285/(285 + 1940) = 0.128= 12.8%.

IMPATTO:
In pratica, dovremo mediamente richiamare per approfondimenti con eventuale biopsia quasi **7 donne sane** (1940/285 = 6.8) **prima di arrivare a diagnosticare un carcinoma**.
Il tasso di richiamo sarà molto elevato, pari al 22.25% (2225/10000).

L'effetto sarà quello del falso allarme (se a ogni round si richiama il 20-25% delle donne, in 4-5 round tutte le donne saranno mediamente richiamate).

Oltre le considerazione sul carico di lavoro e sui costi economici derivati, le donne perderanno fiducia nel programma di screening.

Sensibilità e specificità assumono rilevanza diversa secondo la prevalenza e la gravità di malattia della popolazione indagata.

Nello studio di **soggetti sintomatici** cercheremo di avvalerci di indagini ad **alta sensibilità** e saremo disposti ad accettare una specificità non elevata che potrà essere controbilanciata da ulteriori passi dell'iter diagnostico.

Viceversa, nello studio di **soggetti asintomatici** (screening) cercheremo di avvalerci di indagini ad **alta specificità**, accettando anche il prezzo di una minore sensibilità.

Infatti, mentre nel primo caso l'obiettivo prioritario è diagnosticare la presenza di malattia sintomatica, potenzialmente avanzata, nel secondo caso l'obiettivo della diagnosi di malattia asintomatica deve essere controbilanciato dalla limitazione del numero di accertamenti inutili in quote rilevanti della popolazione sottoposta a screening.

Una visione generale della dipendenza dei valori predittivi dalla prevalenza di malattia può essere ottenuta rappresentando la probabilità **post-test di malattia** (ovvero la probabilità di malattia dopo l'esecuzione dell'indagine con risultato positivo o negativo)
in funzione della probabilità **pre-test di malattia** (ovvero la prevalenza di malattia).

La **probabilità post-test** di malattia nel caso di **risultato positivo** dell'indagine equivale al **VPP**,

mentre la probabilità post-test di malattia nel caso di **risultato negativo** dell'indagine equivale al **complemento a 1 del VPN**.

Probabilità pre-test di malattia*:* probabilità che il paziente sia affetto dalla malattia.

Assenza di altre informazioni → **prevalenza di malattia**

Es. programmi di screening la probabilità pre-test sarà equivalente alla prevalenza di malattia nella popolazione generale

In diagnostica clinica, la **probabilità pre-test** → **prevalenza nella popolazione generale modificata** dai criteri selettivi del medico richiedente l'indagine, sulla base della sequenza anamnestico-clinica

Teorema di Bayes

la probabilità che il risultato di un'indagine corrisponda o meno alla presenza di malattia dipende dalla **probabilità pre-test** e dalla **potenza dell'indagine**

odds di malattia post-test = LR positivo × *odds di malattia pre-test*

ODDS
Valutando la probabilità di malattia in un campione di 10 soggetti che include 3 malati, diremo che la *frequenza di malattia rispetto al campione* è di 3/10, ovvero di 0.3 (pari al 30%), mentre l'odds di malattia è 3/7, ovvero 0.43. L'odds rappresenta quindi il rapporto tra malati e sani, ossia quanti malati ci sono per ogni sano.

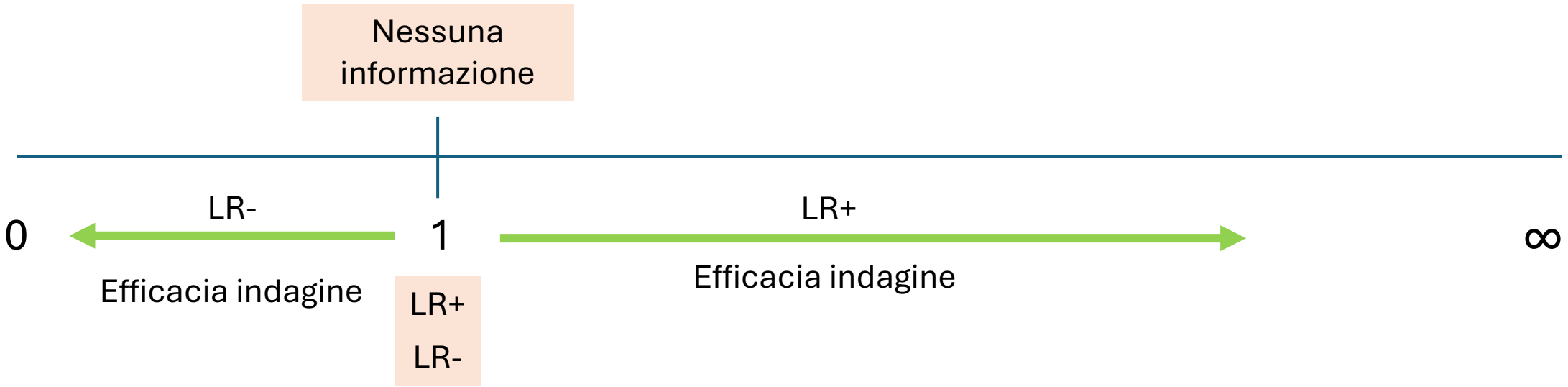
Indici test

Likelihood positivo (LR +)=sensibilità/(1 – specificità)

"di quanto il risultato positivo dell'indagine incrementa la probabilità di malattia ?"

Likelihood negativo (LR -)=(1 – sensibilità)/specificità

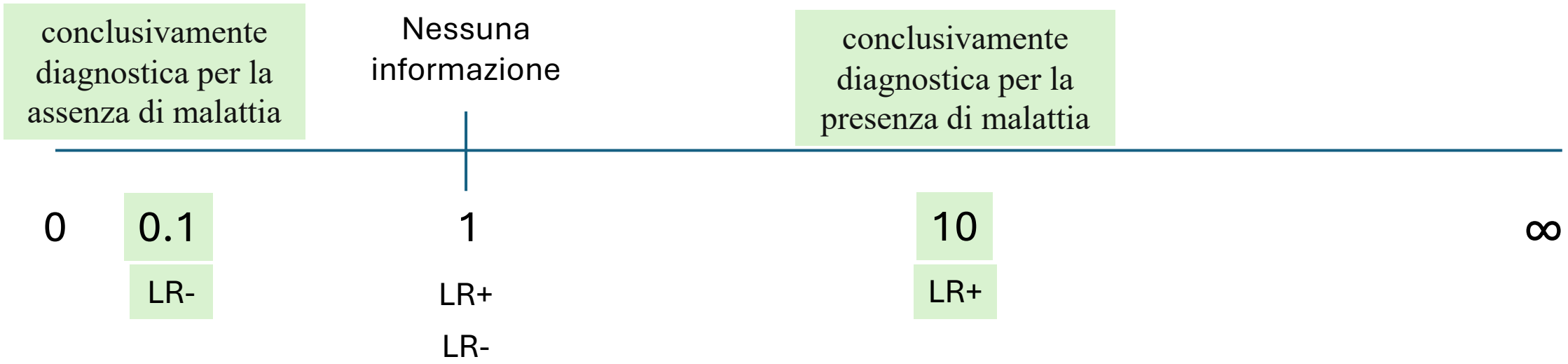
"di quanto il risultato negativo dell'indagine riduce la probabilità di malattia ? "



Indici test

Likelihood positivo (LR +)=sensibilità/(1 – specificità)

Likelihood negativo (LR -)=(1 – sensibilità)/specificità



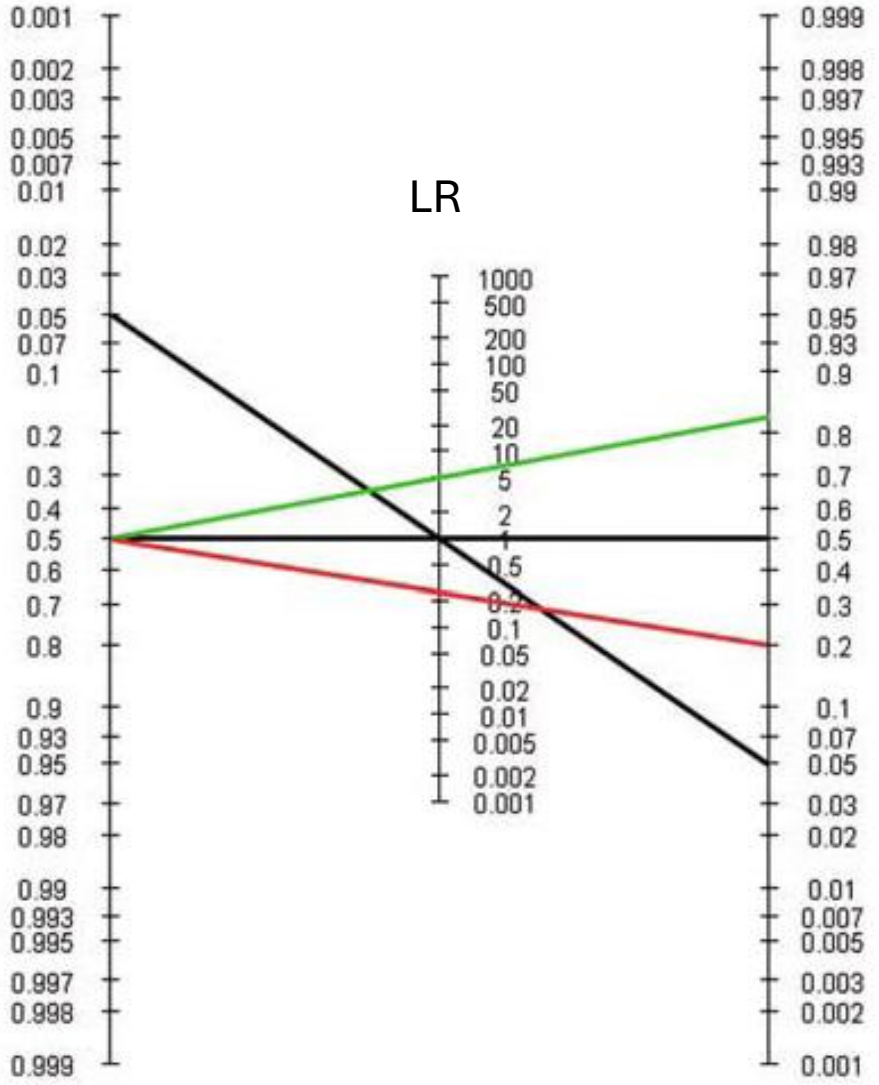
Nomogramma di Fagan

Il *nomogramma bayesiano di Fagan* trasforma la probabilità pre-test in quella post-test mediante semplice proiezione geometrica, senza necessità di calcoli. La pendenza della retta sul nomogramma consente di *vedere* graficamente la potenza dell'indagine.

Con una LR+=5 trasforma pre-test=0.5 (incertezza assoluta) in Post-test=0.8 (alta probabilità di malattia)

LR- =0.35 trasforma P.pre-test = 0.5 in P. post-test=0.2.

LR=1 non cambia la P



[Fagan, 1975]

Riassunto indici

Tabella 1.3. Indici che misurano la performance di un'indagine diagnostica

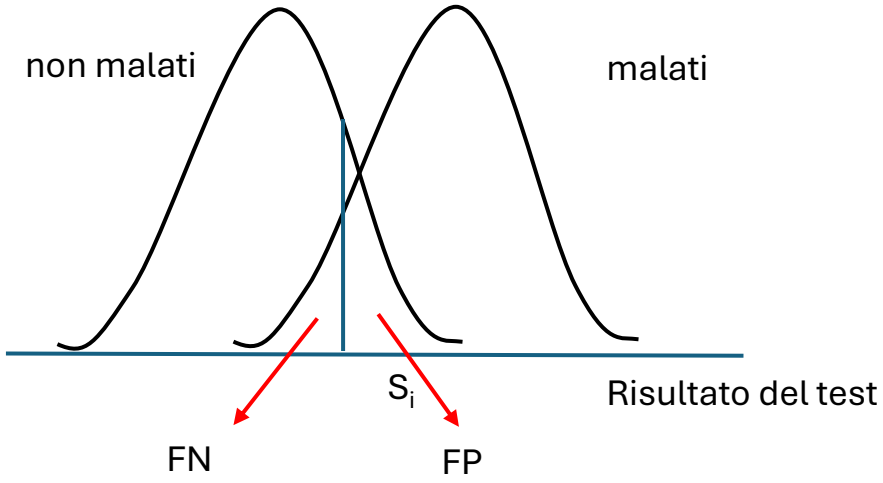
Indice	Definizione	Formula	Dipendenza dalla prevalenza di malattia
1. Sensibilità (o frazione dei VP)	Capacità dell'indagine di individuare la malattia	VP/(VP + FN)	No
2. Specificità (o frazione dei VN)	Capacità dell'indagine di individuare l'assenza della malattia	VN/(VN + FP)	No
3. Valore predittivo positivo	Attendibilità del risultato positivo	VP/(VP + FP)	Sì
4. Valore predittivo negativo	Attendibilità del risultato negativo	VN/(VN + FN)	Sì
5. Accuratezza diagnostica	Attendibilità globale dell'indagine	(VP + VN)/ (VP + VN + FP + FN)	Sì
6. Frazione dei FN	Peso dei FN rispetto agli affetti dalla malattia	FN/(FN + VP) = (1 – Sensibilità)	No
7. Frazione dei FP	Peso dei FP rispetto ai non affetti dalla malattia	FP/(FP + VN) = (1 – Specificità)	No
8. Rapporto di verosimiglianza del risultato positivo (LR positivo)	Incremento della probabilità di malattia in presenza di risultato positivo	Sensibilità/ (1 – Specificità)	No
9. Rapporto di verosimiglianza del risultato negativo (LR negativo)	Riduzione della probabilità di malattia in presenza di risultato negativo	(1 – Sensibilità)/ Specificità	No

LR = *likelihood ratio*.

La *prevalenza di malattia*, essendo il rapporto tra il numero di soggetti affetti e il totale del campione studiato, è pari a (VP + FN)/(VP + VN + FP + FN).

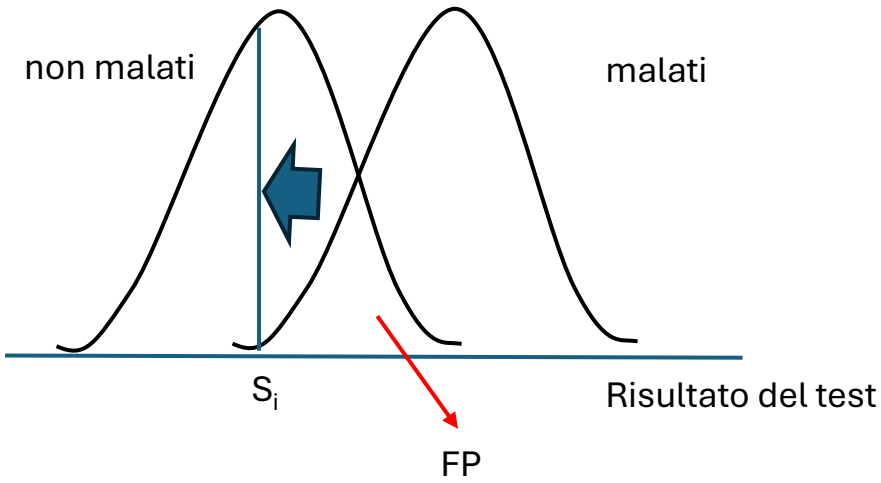
Indici test

Molti test forniscono dati su una scala numerica continua



Qualsiasi soglia scelta (S_i) comporta una quota di diagnosi errate (FP o FN)

Indici test



Per massimizzare SE basterebbe stabilire una soglia molto bassa

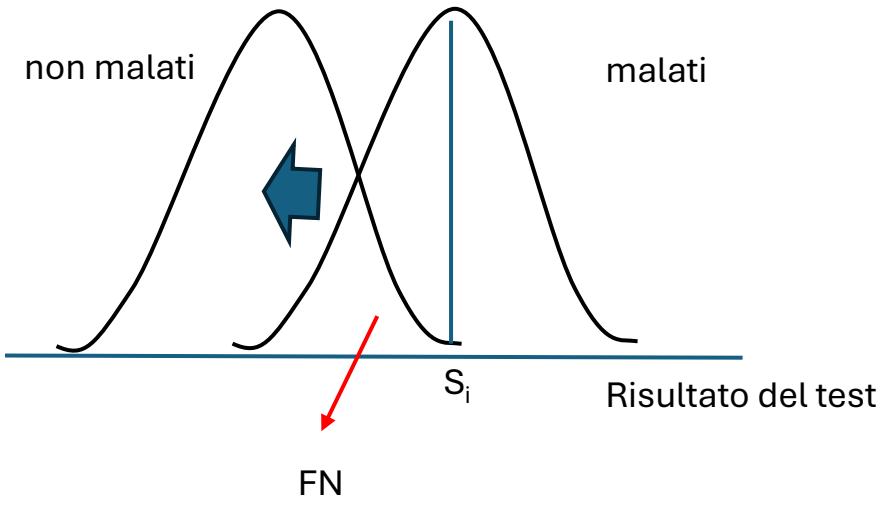
nessun FN → SE=100%

Es. pressione sistolica per definire ipertensione.

Se stabilisco soglia a 100 mm Hg, certezza che tutti gli ipertesi sono diagnosticati

Ma… diagnosi di ipertensione nella maggior parte dei normotesi.

Indici test



Per massimizzare SP basterebbe stabilire una soglia molto alta

nessun FP → SP=100%

Es. pressione sistolica per definire ipertensione.

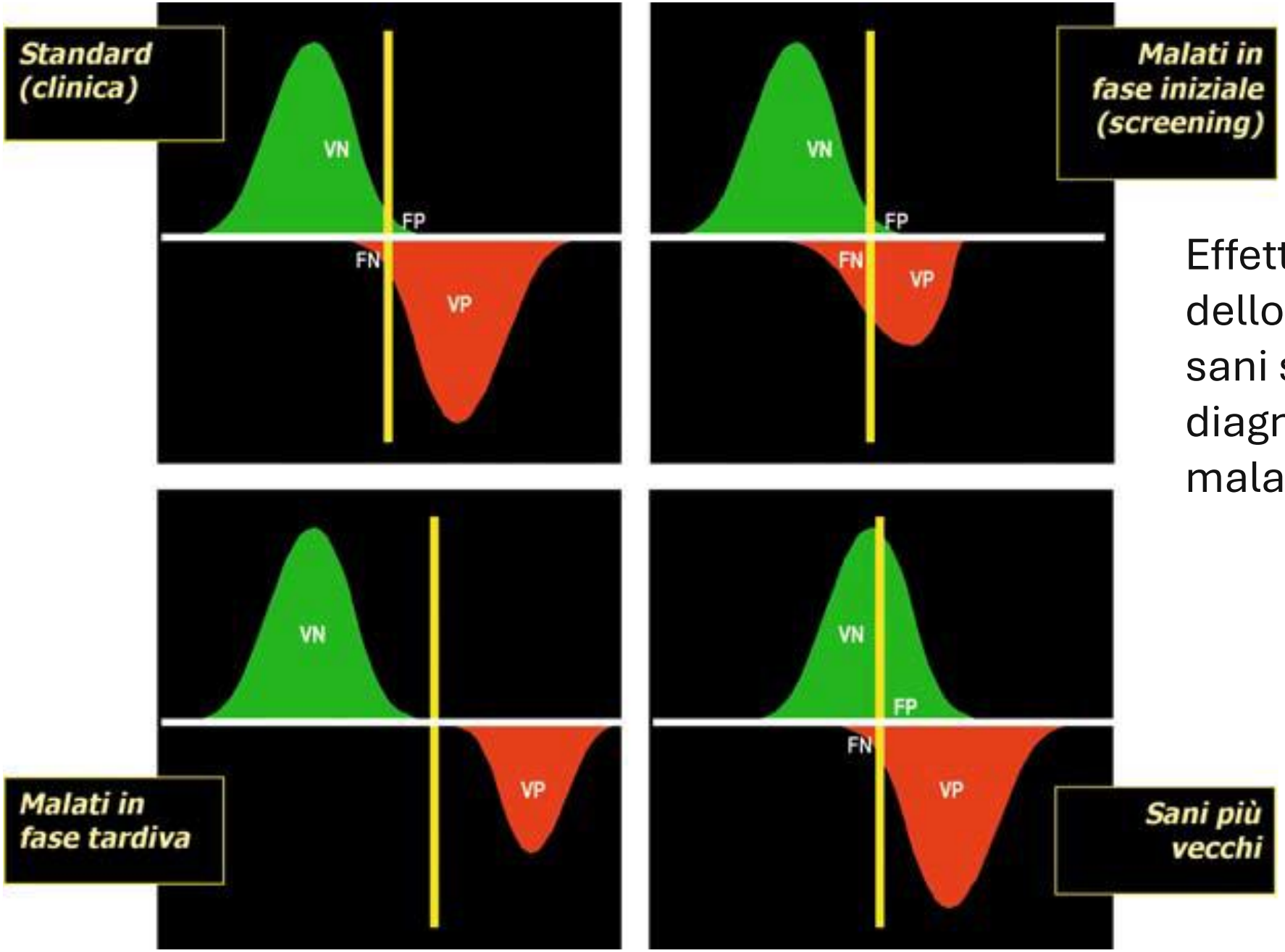
Se stabilisco soglia a 200 mm Hg, certezza che tutti i normotesi siano identificati come tali dal test

Ma…esclusione dalla diagnosi di ipertensione nella maggior parte degli ipertesi.

Ottimizzazione del cutoff

La soglia potrà essere **ottimizzata** →minimizzare gli errori→ **minor numero di falsi positivi e negativi**

Tuttavia, la selezione di casi da sottoporre ai test modifica l'impatto di tale soglia.

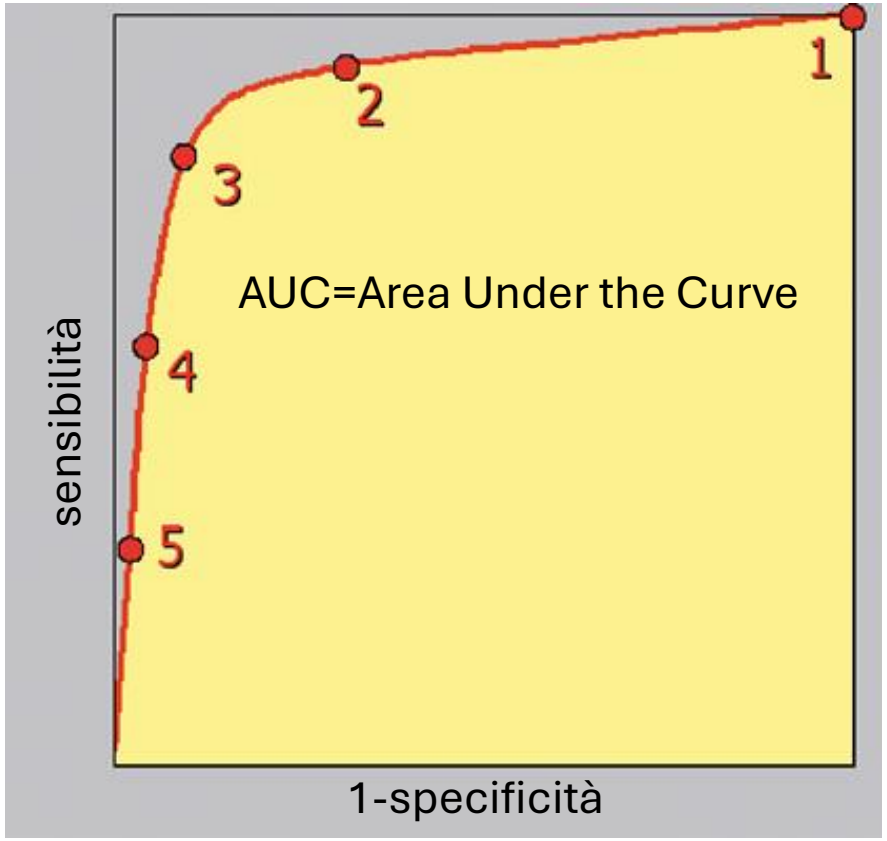


Effetto della modificazione dello spettro di malati e di sani sottoposti a un’indagine diagnostica per una data malattia

CURVA ROC (Receiver Operating Characteristics)

Valutazione grafica e matematica della performance complessiva del test e del cut-off che fornisce il risultato migliore e/o più appropriato alla situazione clinica specifica

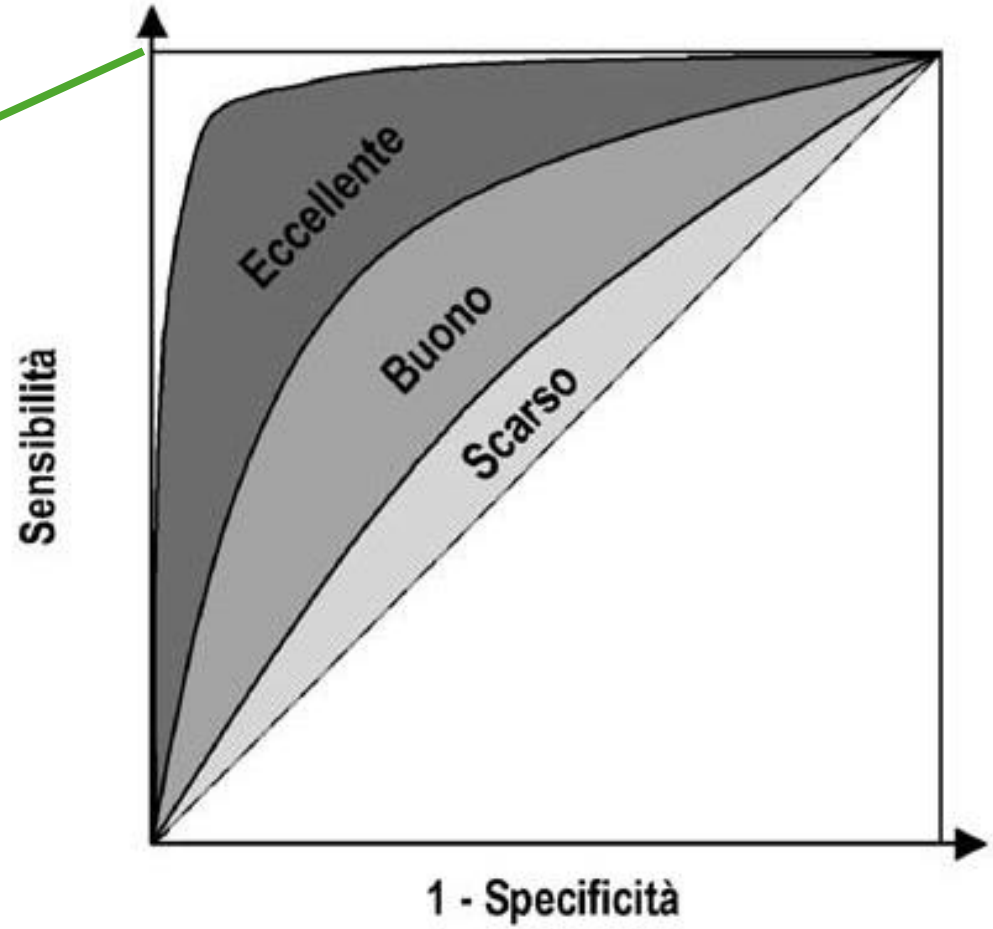
Grafico di sensibilità e 1-specificità ad ogni possibile soglia del test.



CURVA ROC (Receiver Operating Characteristics)

Miglior punto
SE=1
SP=1

Possibile confrontare performance di test differenti

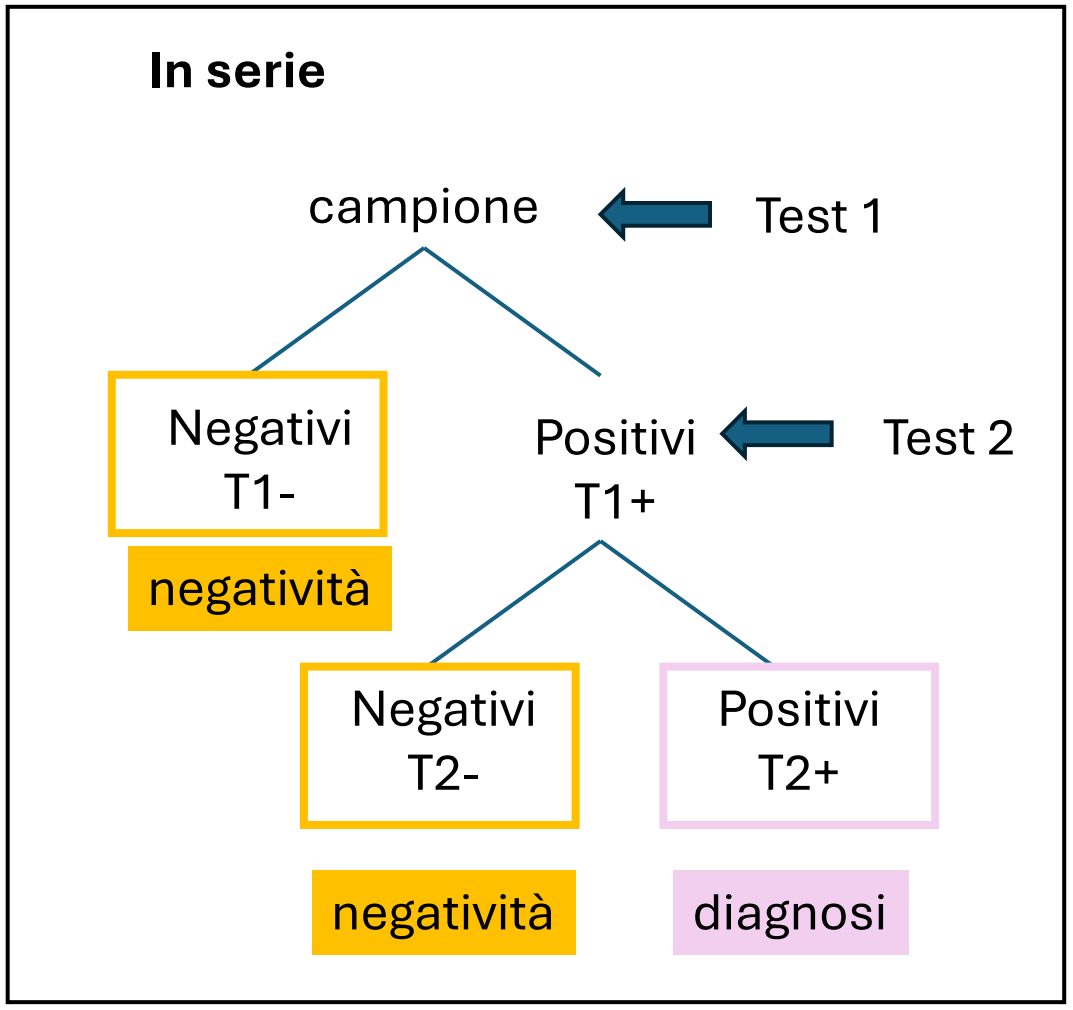


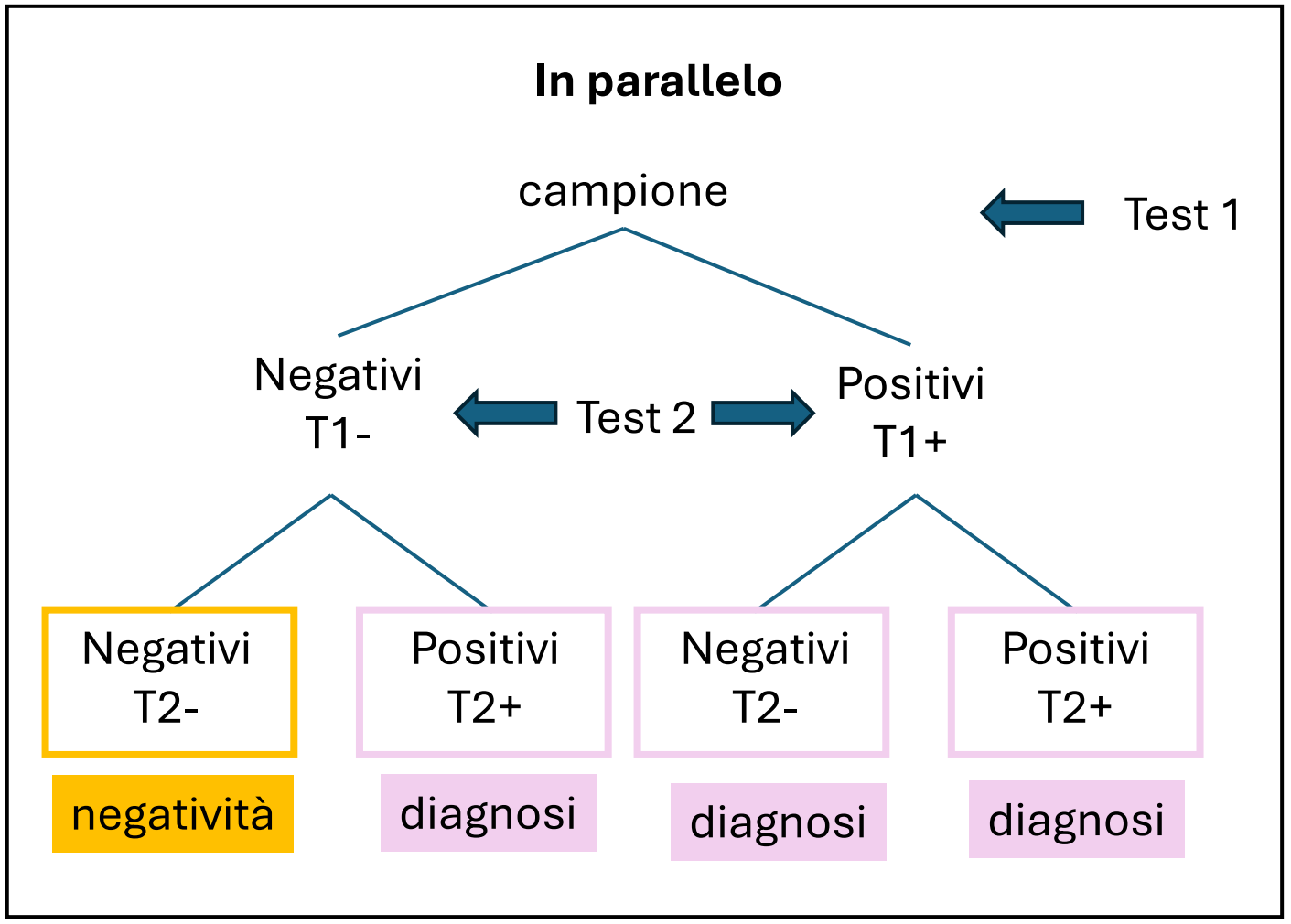
Tipologie di studi

Anche se il focus di questi studi è sulla validazione, i disegni possono variare

- Possibili obiettivi
  - Determinare e performance di un test
    - Ex: to assess the accuracy of V/Q scans for acute pulmonary embolism
  - Comparare le performance di vari test
- I test possono essere comparati testa a testa o valutati in aggiunta
- È possible anche estendere l'obiettivo oltre la validazione clinica, guardando all'ultilità clinica

Impatto su aggiunta di un test





Impatto su aggiunta di un test

In serie

Alcuni VP al test 1 sono classificati come FP dopo il test 2

SE procedura test in serie

<SE test 1

<SE test 2

Alcuni FP al test 1 sono classificati come VN dopo il test 2

SP procedura test in serie

>SP test 1

>SP test 2

Da usare quando previlegiare SP
Quando il test 2 è invasivo e si vuole sottoporre a pz con un sospetto dato dal primo test

In parallelo

Alcuni FN al test 1 sono classificati come VP al test 2, alcuni FN al test 2 sono classificati VP con il test 1

SE procedura test in serie

>SE test 1

>SE test 2

Alcuni VN al test 1 e alcuni VN al test 2 sono classificati come FP alla sequenza

SP procedura test in serie

<SP test 1

<SP test 2

Da usare quando previlegiare SE
Quando è necessario avere la diagnosi in tempi brevi

Tipologie di studi

- Molti test coinvolgono un valutatore che deve dare un’interpretazione

- Multi-reader, multi-case sono lo scenario più comune

- Quando più lettori sono coinvolti in uno studio che confronta due test, ci sono vari approcci.

Appaiato e non appaiato

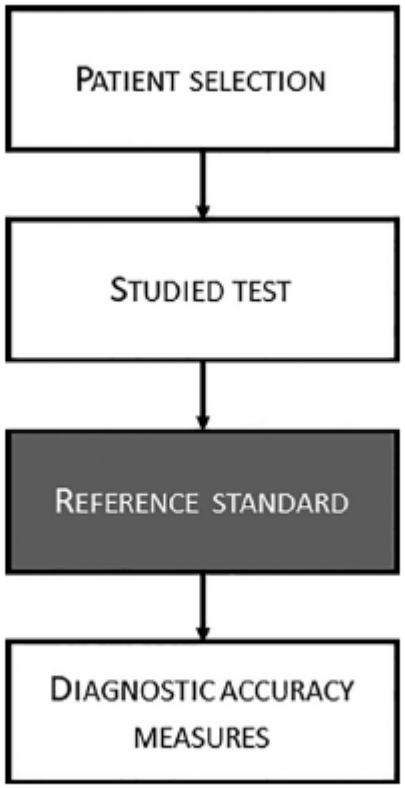
DISEGNO APPAIATO

- I valutatori forniscono interpretazioni per entrambi i test medici.
- I pazienti ricevono ogni test medico.
- vantaggio: maggiore potenza statistica perché i valutatori e i pazienti stanno agendo come propri controlli.

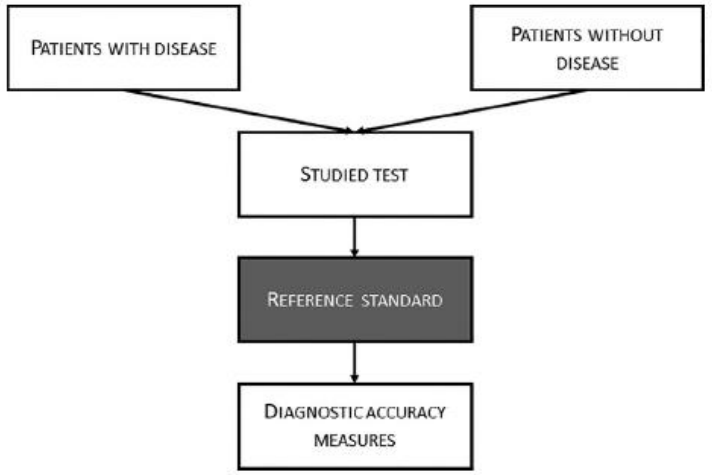
DISEGNO NON APPAIATO

- un diverso insieme di valutatori fornisce interpretazioni per ogni test.
- un insieme differente dei pazienti riceve ogni prova medica.

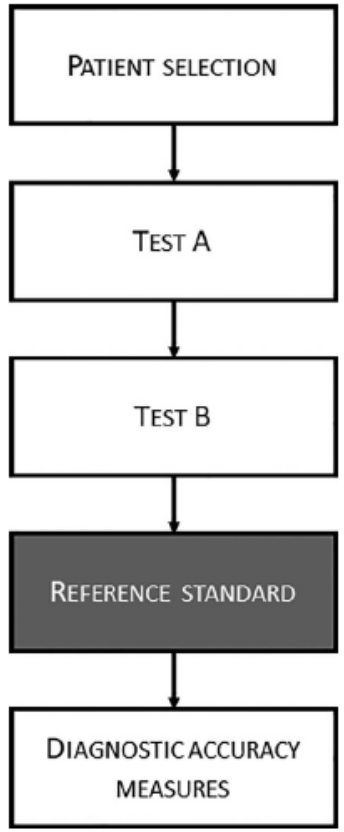
Esempi di disegni di studi



Diagnostic accuracy cross-sectional studies.



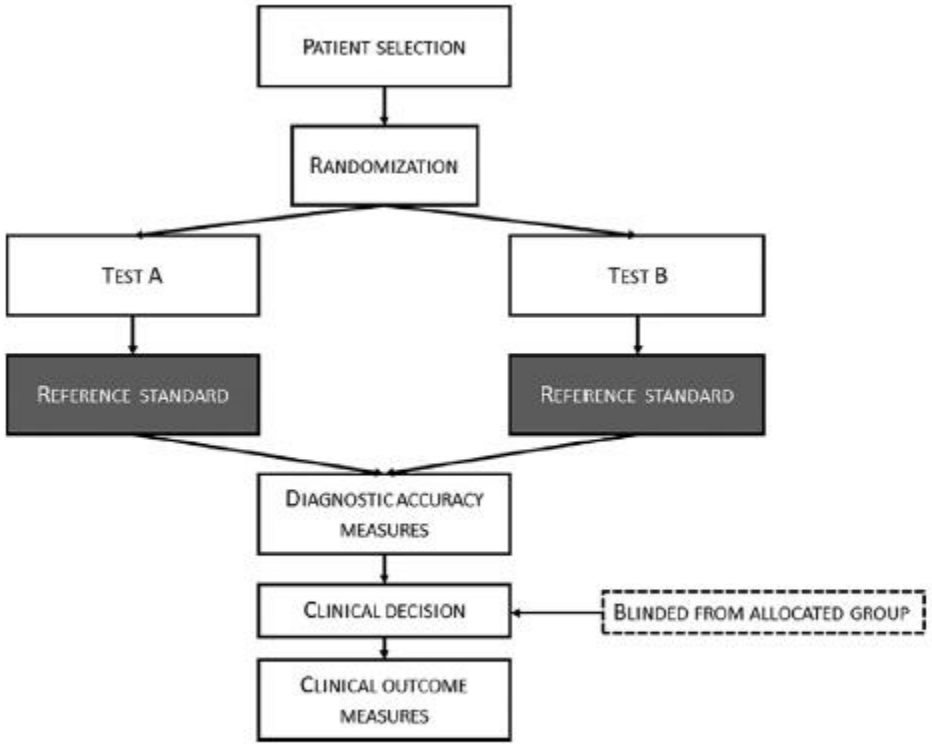
Diagnostic accuracy case-control studies.



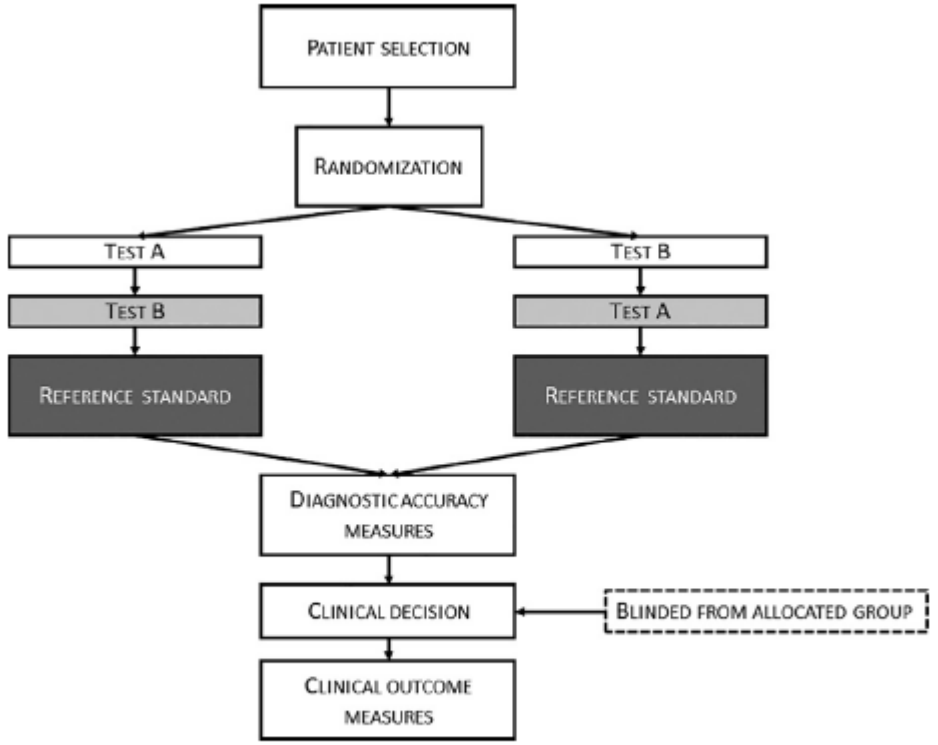
Nonrandomized comparative diagnostic accuracy study.

J A. Bullen, CHEST 2020

Esempi di disegni di studi



Randomized diagnostic accuracy study—option 1.



Randomized diagnostic accuracy study—option 2.

J A. Bullen, CHEST 2020

Impatto dei disegni

Valutare l'accuratezza dei test medici è una parte fondamentale del processo decisionale nel mondo dell'assistenza sanitaria.

Gli studi di accuratezza consentono di valutate le caratteristiche del test come

- sensibilità,
- specificità
- i valori predittivi negativi e positivi.

Gli studi di accuratezza consentono ai ricercatori di descrivere i tipi di errori che un test può essere incline a fare e quanto spesso accadono.

Questo tipo di informazioni è fondamentale per i medici che devono in ultima analisi, integrare i risultati dei test medici nella gestione del paziente

Valutazione dei risultati degli studi

Come applicare il test diagnostico a un paziente specifico

- Il test è disponibile, conveniente e accurato nel setting di applicazione?
- Una stima clinicamente ragionevole delle probabilità pre-test del paziente può essere fatta dall'esperienza personale, dalle statistiche sulla prevalenza, dai database di pratica o dagli studi primari?
- I pazienti dello studio sono simili ai pazienti della pratica clinica?
- Quanto è attuale lo studio che stiamo analizzando
  - Dopo la pubblicazione dello studio, sono state apportate modifiche?

Valutazione dei risultati degli studi

La probabilità post-test influenzerà la gestione del paziente specifico?

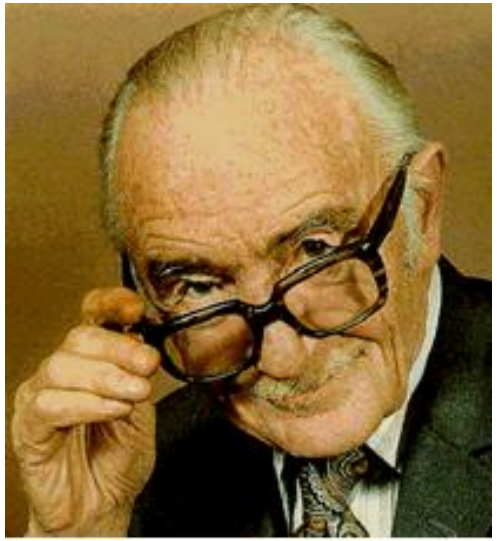
- Il risultato potrebbe spostare il giudizio del clinico : per esempio, i risultati del test potrebbero fermare tutti i test ulteriori? Cioè, escludere la malattia, o al contrario permettere di fare una diagnosi definitiva del disturbo e passare alla scelta di opzioni di trattamento appropriate.
- Il paziente sarà disposto ad effettuare il test?
- I risultati del test aiuteranno il paziente a raggiungere i propri obiettivi?

Before ordering a test,
decide what you will do

if it is positive or negative.

If both answers are the same

don't do the test



Archie Cochrane

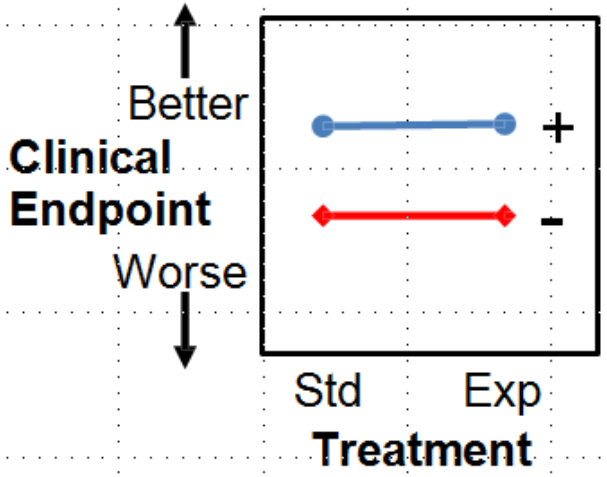
Sviluppo e validazione di un modello prognostico

(V. Torri)

Prognostic vs. Predictive biomarker

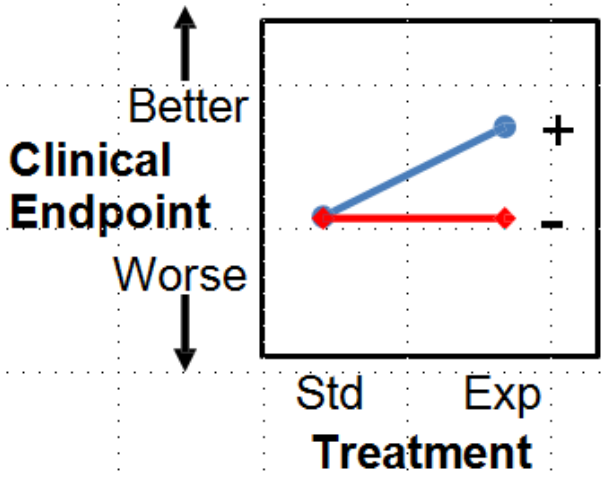
- prognostic biomarkers
  - used to identify likelihood of a clinical event, disease recurrence or progression in patients who have the disease or medical condition of interest,
- predictive biomarkers
  - used to identify individuals who are more likely than similar individuals without the biomarker 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:

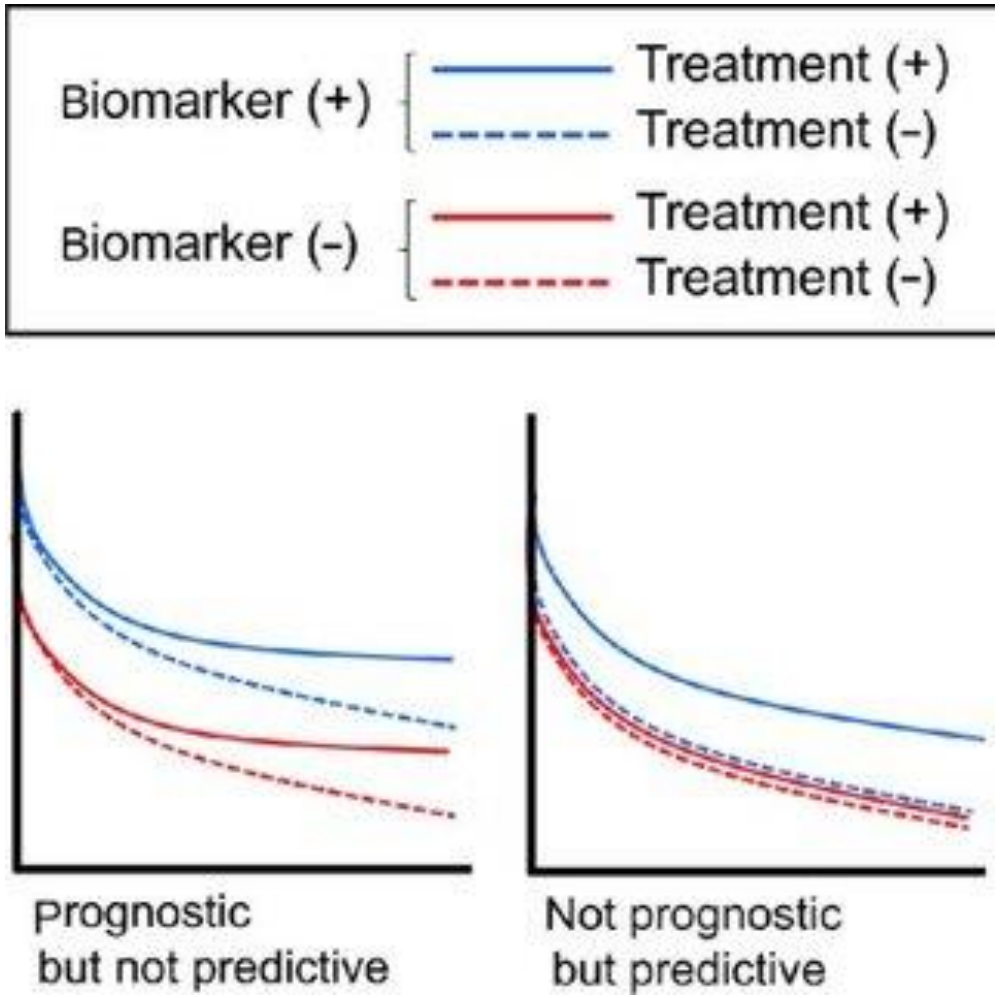
- associated with **disease outcome**
- risk assess (+,-) to stratify for treatment



Predictive:

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

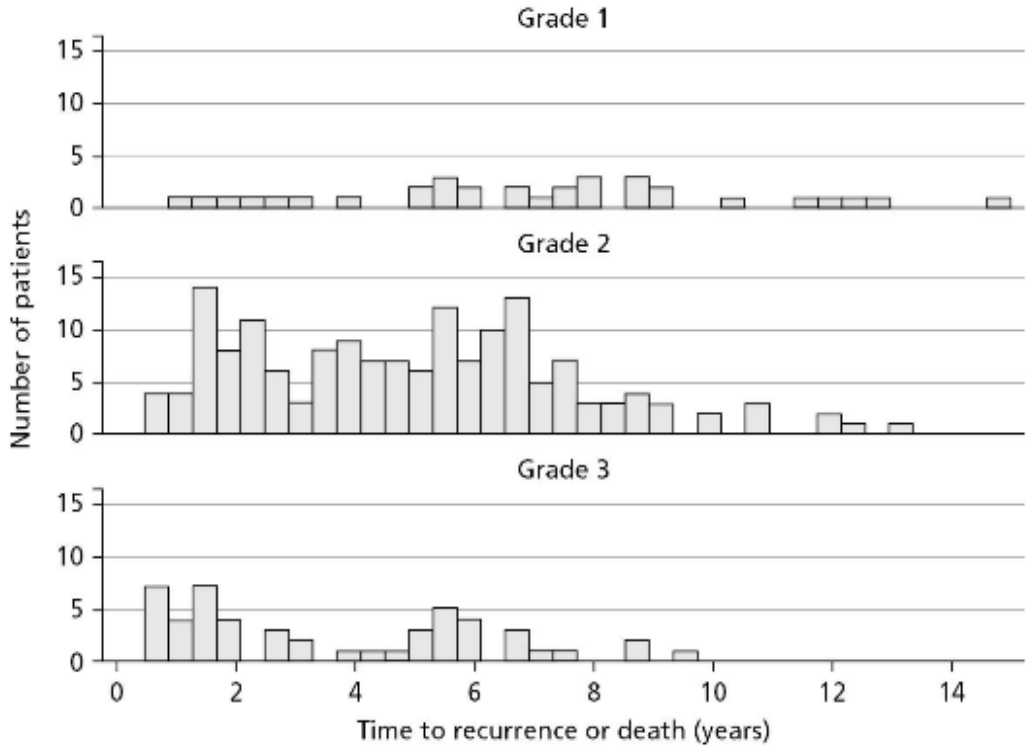
Prognostic vs. Predictive biomarker



Prognostic model: definitions

- Different terms can be used for prognostic model:
  - risk score
  - risk classification tool
  - prognostic score or index
  - clinical prediction model.
    - The term clinical decision (or prediction) rule is best reserved for a simplified or categorised prediction model that explicitly incorporates one or more risk thresholds to guide clinical decisions, e.g.
    - initiate treatment if a patient's predicted outcome risk is above 10%,
    - and withhold the treatment otherwise)

Models vs. Variables



"The median isn't the message" for individual prognosis: evidence of wide and overlapping distributions of survival times across tumour grades in 246 breast cancer patients treated with tamoxifen over a possible 7 years of follow up.

Journal of Clinical Oncology. 12(10), 2086-93. Copyright © 1994

- In daily practice, a more individualized approach to prognosis inevitably requires multiple prognostic factors (referred to as 'predictors') to be utilized in combination, to tailor outcome predictions for each individual conditional on their unique set of observed characteristics (referred to as predictor values).

- A prognostic model is a formal combination of multiple predictors in order to make individualized predictions about a particular outcome (endpoint) of interest from a given state of health (startpoint) over a specified time interval.

  - For continuous outcomes, a prognostic model predicts an individual's expected (mean) outcome value by a particular time point.
  - For binary or time-to-event outcomes, a prognostic model predicts an individual's outcome risk (probability) by a particular time point (or time points).

Definition and purpose of prognostic models

- Prognostic models are regression models relating to patients outcome
- They are typically developed using a multivariable regression framework, which provides an equation to estimate an individual's expected
  - outcome value (for continuous outcomes)
  - or outcome risk (for binary or time-to-event outcomes)
- based on their values of multiple prognostic factors (such as age and smoking, symptoms and signs, imaging and electrophysiology results, biomarkers and genetic information).
- In this context, the included prognostic factors are better known as predictors, independent variables, or covariates, which are more generic terms

Definition and purpose of prognostic models

- Prognostic models 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.

Example of prognostic model equation and individualised survival curve prediction

Context: A prognostic model predicting the risk of recurrent venous thromboembolism after cessation of initial therapy following a first venous thromboembolism[22]

(a) Developed model equation

Probability of recurrence by time t is defined by $F(t) = 1 - S_0(t)^{\exp(\text{risk score})}$ where

$$\text{risk score} = (-0.0105 \times \text{Age}) + (0.545 \times \text{Gender:Male}) + (1.735 \times \text{Site:Proximal DVT}) + (1.756 \times \text{Site:PE}) + (0.701 \times \log(\text{D-dimer})) + (-0.291 \times \log(\text{Lag time}))$$

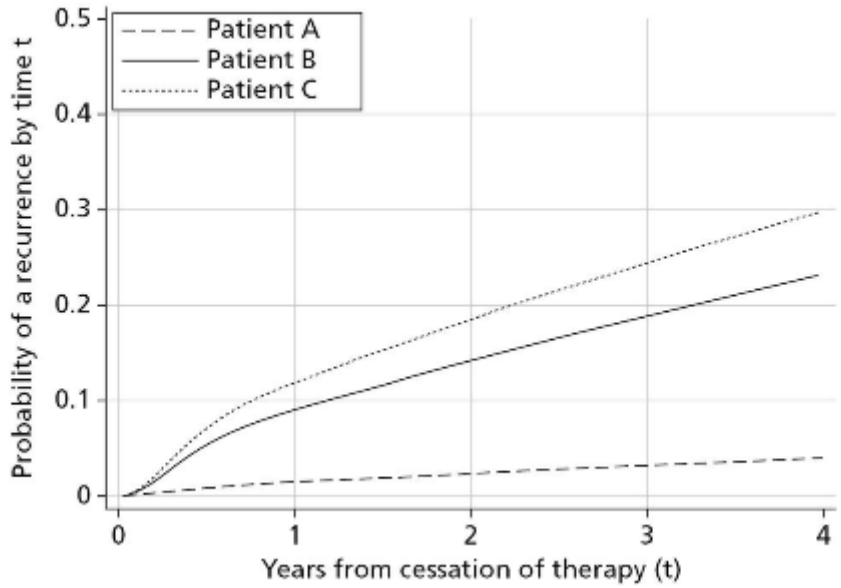
and $S_0(t)$ is the baseline survival available at all time points up to 4 years, e.g. $S_0(6\ \text{months}) = 0.9996$, $S_0(1\ \text{year}) = 0.9993$, and $S_0(2\ \text{years}) = 0.9988$, and the reference site is a distal deep vein thrombosis (DVT).

(b) Applying the model to produce predicted recurrence risk curves for three different patients

Values of predictors for each patient:

Predictor	Patient A	Patient B	Patient C
Age (years)	51	64	74
Gender:Male	1	1	1
Site:Proximal DVT	0	1	0
Site:Pulmonary Embolism (PE)	0	0	1
D-dimer (ng/mL)	275	417.5	747
log(D-dimer)	5.55	6.03	6.62
Lag time (days)	22	29	33
log(Lag time)	3.14	3.4	3.53

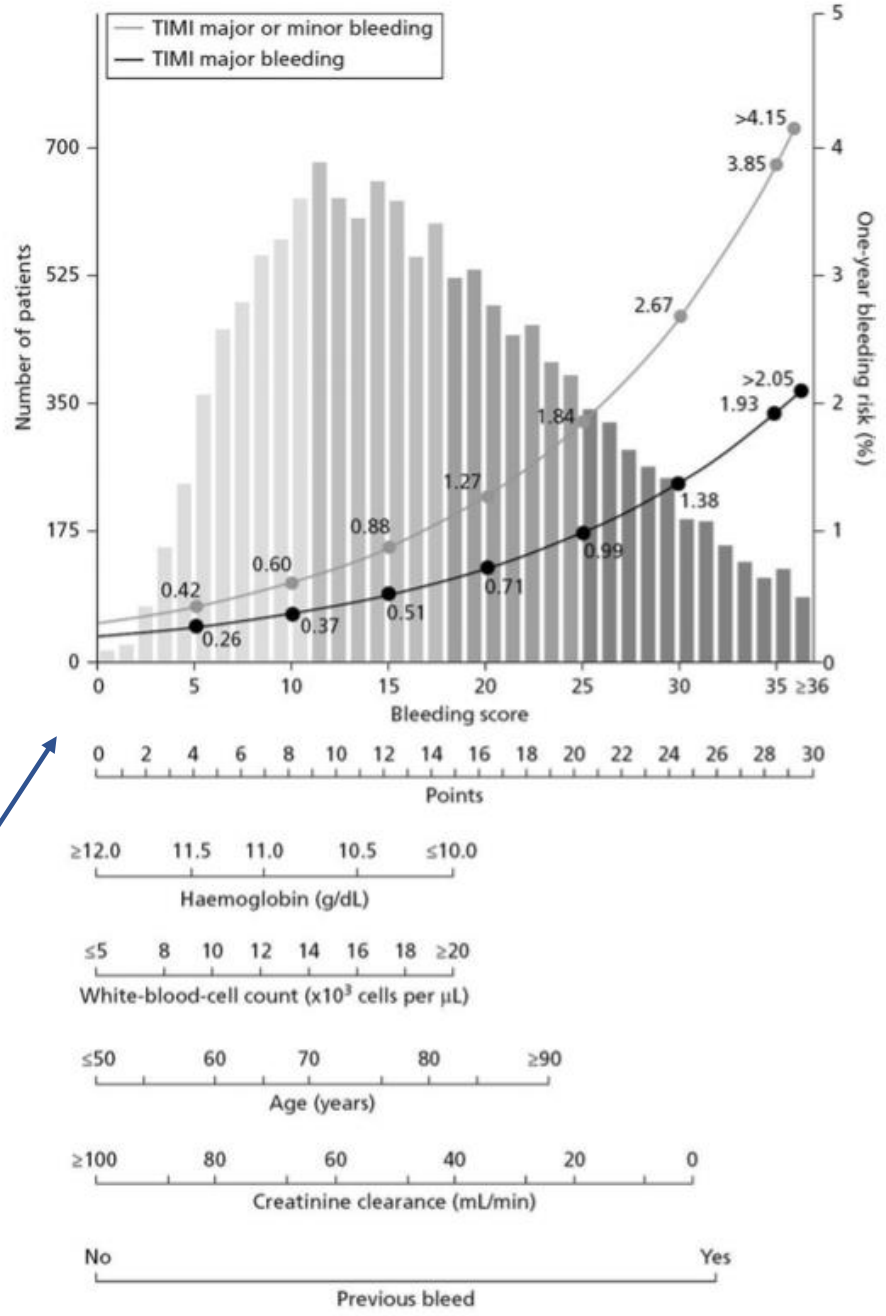
Predicted risk of recurrence over time for each patient:



Example of nomogram

- Prognostic models are sometimes developed using a regression model, including an intercept or baseline hazard term,
  - but then presented in a simplified form or rule (e.g. by rounding predictor effects to whole numbers)
  - or as a nomogram, which provides a graphical representation of the original mathematical equation from which the predicted outcome (risk) can be calculated

Derivation and validation of the predicting bleeding complications in patients undergoing stent implantation and subsequent dual antiplatelet therapy (PRECISE-DAPT) score: a pooled analysis of individual-patient datasets from clinical trials. - The Lancet 2017; 389(10073), 1025-34.



Why are prognostic models important?

- Fundamentally, prognostic models allow health professionals to tailor prognostic information to their patient, with clearer communication.

- Example:
  - “In individuals with similar characteristics to you, the risk of death by one-year is 20%; therefore, for every five patients similar to you, we expect four to be alive by one-year.”

Why are prognostic models important?

- Prognostic models aim to assist (not replace) healthcare professionals with their prediction of a patient's future outcome and to enhance informed decision making together with the patient.
- In particular, they help translate the results of randomized trials examining the effect of interventions on prognosis.
- Considering a binary outcome, then under the common assumption that a particular treatment has a constant relative benefit across all patients (i.e. the risk ratio for treatment versus control is constant across the entire outcome risk range, i.e. 0 to 1),
- the absolute treatment benefit depends on a person's risk of the outcome without treatment;
- those with the highest baseline risk will have the greatest absolute reduction in risk.
    - Expensive therapies or those with harmful potential side-effects may thus be reserved for those considered at higher risk, as estimated by a prognostic model.

Why are prognostic models important?

- In women with osteoporosis the effectiveness of alendronate for reducing the rate of non-vertebral fracture was, when expressed as a rate ratio, constant across subgroups defined by increasing fracture risk according to the FRAX prognostic model.
- However, the reduction in absolute rate due to alendronate increased as the baseline fracture risk increased, therefore those at highest outcome risk defined by FRAX had larger potential reduction in their absolute rate

IRR (95% CI) and RD (95% CI) Comparing Alendronate to Placebo for Nonvertebral Fractures in Tertiles of FRAX Score With FN BMD

		ALN				PLB					
Tertile	Range of FRAX with FN BMD	*n*	Person-years	Events	Rate[a]	*n*	Person-years	Events	Rate[a]	IRR (95% CI)	RD (95% CI)
1	4.75–22.06	1083	4118.06	91	2.21	1071	4080.69	104	2.55	0.87 (0.65–1.15)	−0.34 (−1.0 to 0.33)
2	22.07–34.19	1087	3883.49	129	3.32	1066	3750.81	140	3.73	0.89 (0.70–1.13)	−0.41 (−1.3 to 0.43)
3	34.2–85.36	1066	3453.07	164	4.75	1086	3482.90	197	5.66	0.84 (0.68–1.03)	−0.91 (−2.0 to 0.17)

IRR = incidence rate ratio; CI = confidence interval; RD = risk difference; FRAX = WHO Fracture Risk Assessment Tool; FN = femoral neck; BMD = bone mineral density; ALN = alendronate; PLB = placebo. [a]Rate = fractures per 100 person-years.

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.

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

- 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.
- 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.
    a. what or should be meant by validating a model
    b. why we need to validate models
    c. how we should attempt to validate a model

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

Validating a prognostic model

- two types of validated model, not necessarily associated:
  a. 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.
  b. 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, 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
- 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 (AIC)
  - 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.
    a. absence of clear inclusion and exclusion criteria,
    b. many patients excluded through missing data (which may not be missing at random),
    c. unclear rationale for the choice of treatments
    d. 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 ≥ 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.

R Altman and P Royston, Statistics in Medicine - 2000

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 centers, perhaps by different investigators.

R Altman and P Royston, Statistics in Medicine - 2000

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 modeling 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

Internal validation

- 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.

R Altman and P Royston, Statistics in Medicine - 2000

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 coefficients compatible?
  - How well does the model fit the new data?
  - Is the correct ordering of the prognostic groups preserved?
  - Are the event rates between the prognostic groups significantly different?

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

- Fischl et al.: prognostic index to predict the chance of a patient with acute bronchial asthma requiring hospitalization following initial treatment in the ER.
  - Assessment of patients with acute asthma by this multi-factorial approach should allow the physician to identify those patients who are in need of hospitalisation.
    - patients with predictor index scores of 4 or higher (calculated before therapy) should be considered for prompt hospitalisation.

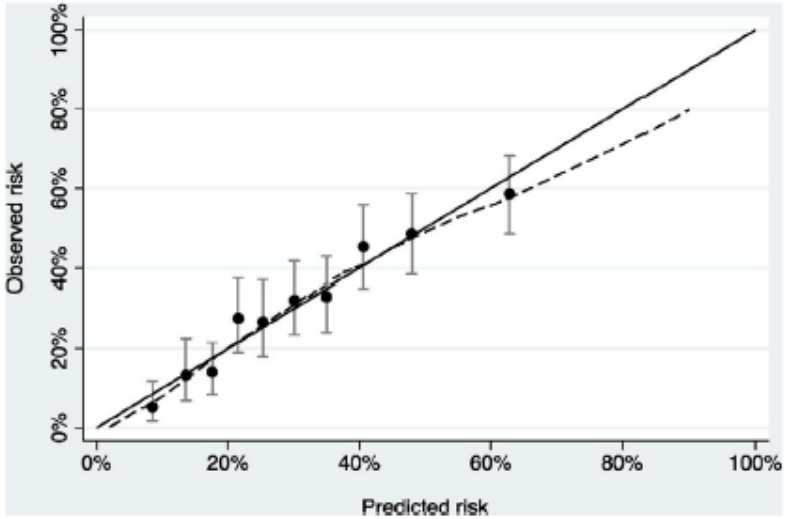
Comparison with validation study

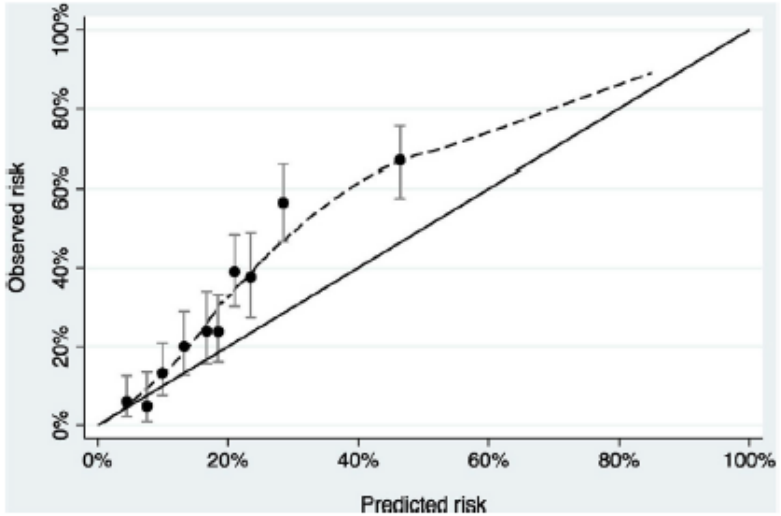
Prognostic classification of relapse or hospitalization in acute asthma. Values in table are proportions (and numbers) of patients who relapsed or were admitted to hospital.

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)	

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?

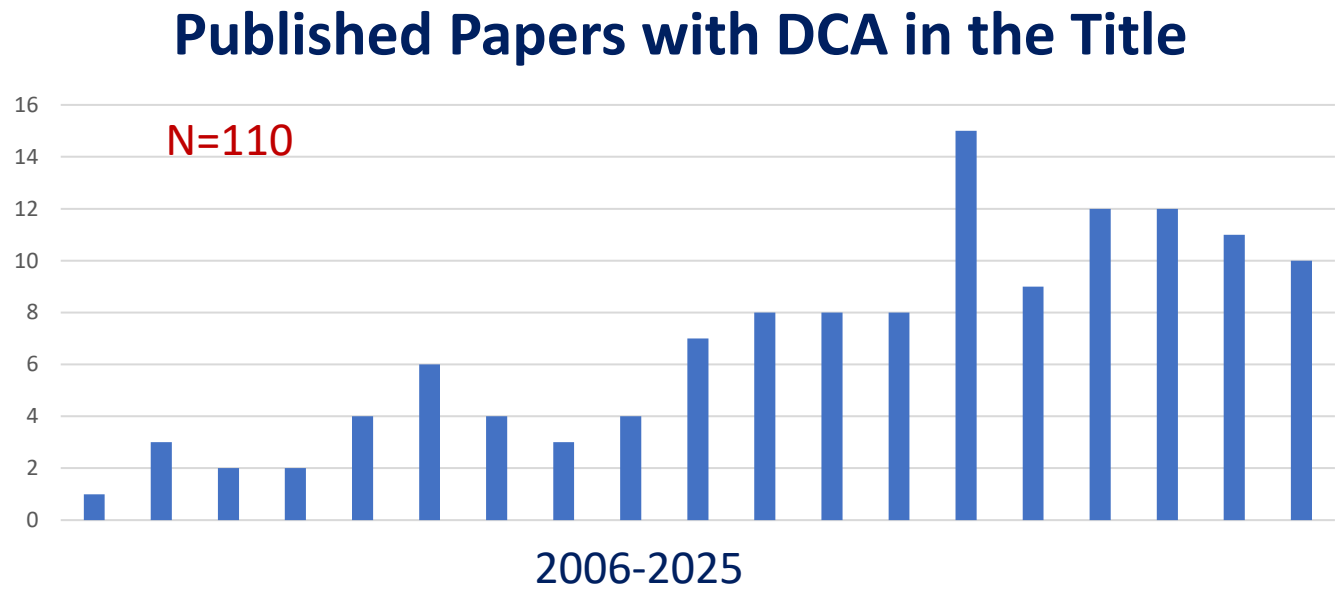
- Discrimination and calibration provide valuable information for researchers.
  - Poor discrimination, demonstrates the predictors in a model are not strongly associated with outcome, such that researchers should consider additional predictors.
  - Poor calibration of a model suggests genuine differences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

Adding Utility ⇨ 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



AJ Vickers et al. Diagnostic and Prognostic Research 2019

What is DCA?

- DCA is a method for evaluating the benefits of a diagnostic test across a range of patient preferences for accepting risk of undertreatment and overtreatment to facilitate decisions about test selection and use.
- .

Why DCA is important

- The area under the curve (AUC) can be interpreted as the probability that in a pair of individuals, one who did and one who did not experience the event, the individual who experienced the event had the higher predicted probability.
- The AUC metric focuses solely on the predictive accuracy of a model;
  - therefore it cannot tell us whether the model is worth using at all or which of two more models is preferable.
  - Take the case where a false-negative result is much more harmful than a false-positive result.
  - A model that had a much greater specificity but slightly lower sensitivity than another would have a higher AUC, but would be a poorer choice for clinical use.
- Decision-analytic methods incorporate information on consequences.

Key concepts

- Probability threshold, namely, a level of diagnostic certainty above which the patient would choose to be treated.
  - captures the relative value the patient places on receiving treatment for the disease, if present, to the value of avoiding treatment if the disease is not present;
    - if the treatment has high efficacy and minimal cost, inconvenience, and adverse effects, then the probability threshold will be low;
    - conversely, if the treatment is minimally effective or associated with substantial morbidity, then the probability threshold will be high.

Key concepts

- The net benefit, or “benefit score” is determined by calculating the difference between the expected benefit and the expected harm associated with each proposed testing and treatment strategy.
  - The expected benefit is represented by the number of patients who have the disease and who will receive treatment (true positives) using the proposed strategy.
  - The expected harm is represented by number of patients without the disease who would be treated in error (false positives) multiplied by a weighting factor based on the patient’s threshold probability.

Key concepts

- The weighting factor captures the patient's values regarding the risks of undertreatment and overtreatment.
- Specifically, the false-positive rate is multiplied by the ratio of the threshold probability divided by 1 – the threshold probability.
  - If the treatment threshold is 10% (0.1) for a patient with possible pneumonia, then the weighting factor applied to the number of patients without pneumonia treated in error would be 0.1/0.9, or one-ninth, minimizing the effect of false-positive results because the burden of unnecessary treatment is low.
  - Conversely, for a patient with a brain mass that is possibly malignant, the probability threshold might be 90% (0.9), leading to a weighting factor of 0.9/0.1, or 9, and greatly increasing the effect of the risk of false-positive

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 Spinal Instability Neoplastic Score (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.

Steps:

1. Chose a value for p_t.
2. Calculate the number of true- and false-positive results using p_t as the cut-point for determining a positive or negative result.
3. Calculate the net benefit of the prediction model.

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

4. Vary p_t over an appropriate range and repeat steps 2 – 3.
5. Plot net benefit on the y axis against p_t on the x axis.
6. Repeat steps 1 – 5 for each model under consideration.
7. Repeat steps 1 – 5 for the strategy of assuming all patients are positive.
8. Draw a straight line parallel to the x-axis at y=0 representing the net benefit associated with the strategy of assuming that all patients are negative.

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.

AJ Vickers, The Spine Jurnal 2021

DCA: working example

- The reduction in the number of unnecessary surgeries per100 patients is calculated as

  (Net Benefit of the model - Net Benefit of reference)/(p_t/(1 - p_t)) × 100.

- This value is net of false negatives, and is therefore the equivalent to the reduction in unnecessary surgeries without a decrease in the number of patients who duly have surgery

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	$(0 - 0 \times (0.25 \div 0.75)) \div 1000 = 0$

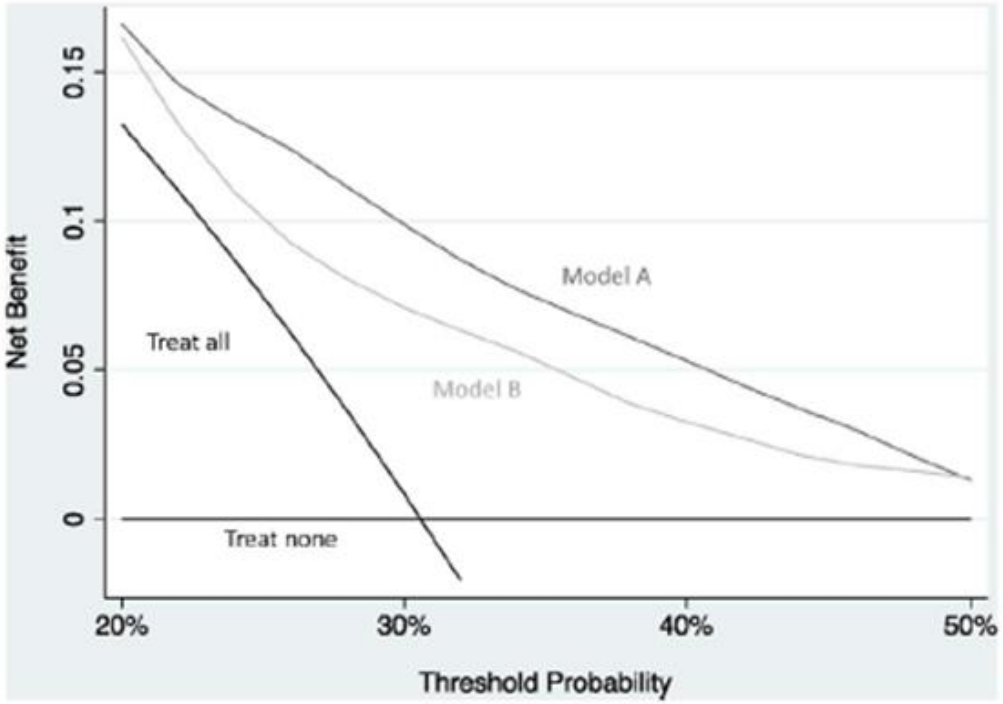
Net benefit is given at a threshold probability of 25% along with that for the clinical alternatives of recommending surgery for all or no patients.

P_t=0.25	Net Benefit Comparison	Reduction of unecessary saurgeries
Model A vs Treat ALL	(0.128-0.075) = 0.053	100*(0.053/0.333) = 16
Model A vs Treat none	(0.128-0.000) = 0.128	100*(0.128/0.333) = 39
Model A vs Model B	(0.128-0.108) = 0.020	100*(0.020/0.333) = 6

AJ Vickers, The Spine Jurnal 2021

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
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.



Backup Slides

Predictive models

Predictive biomarker

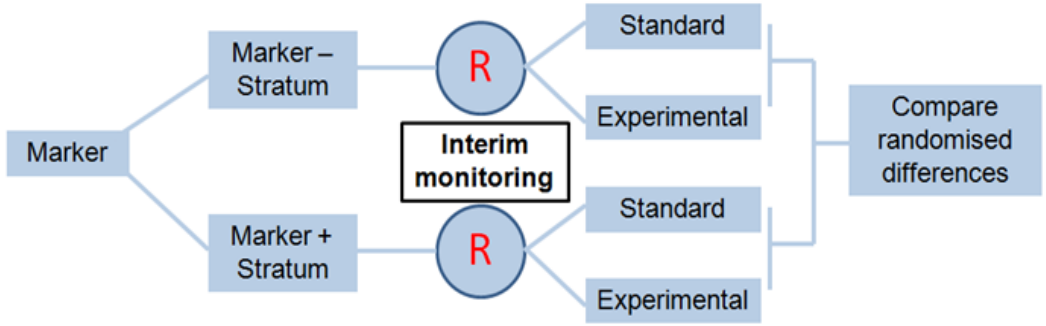
- To identify a predictive biomarker, there generally should be a comparison of a treatment to a control in patients with and without the biomarker.

- 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

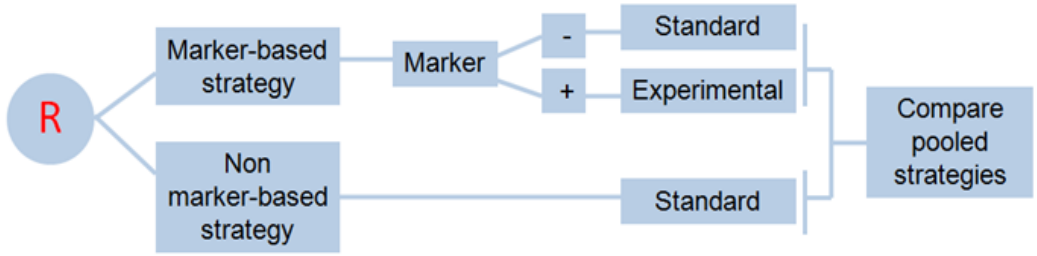
Biomarker-Stratified Design (Full specification)

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

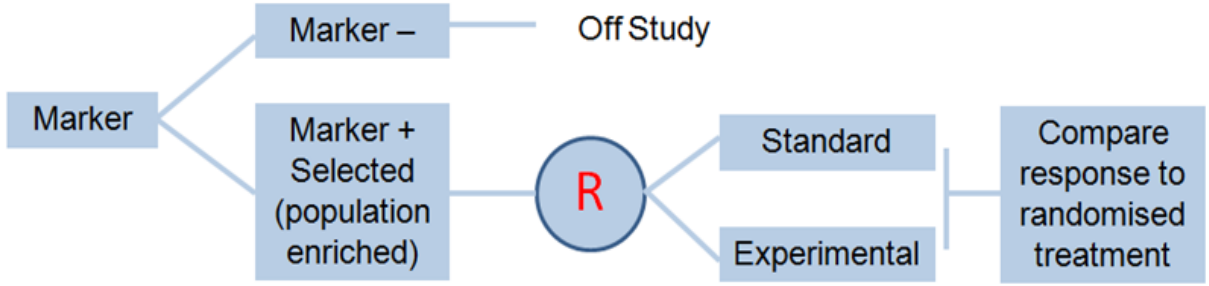


Biomarker-Strategy Design ("Use" vs "Ignore" biomarker)

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.