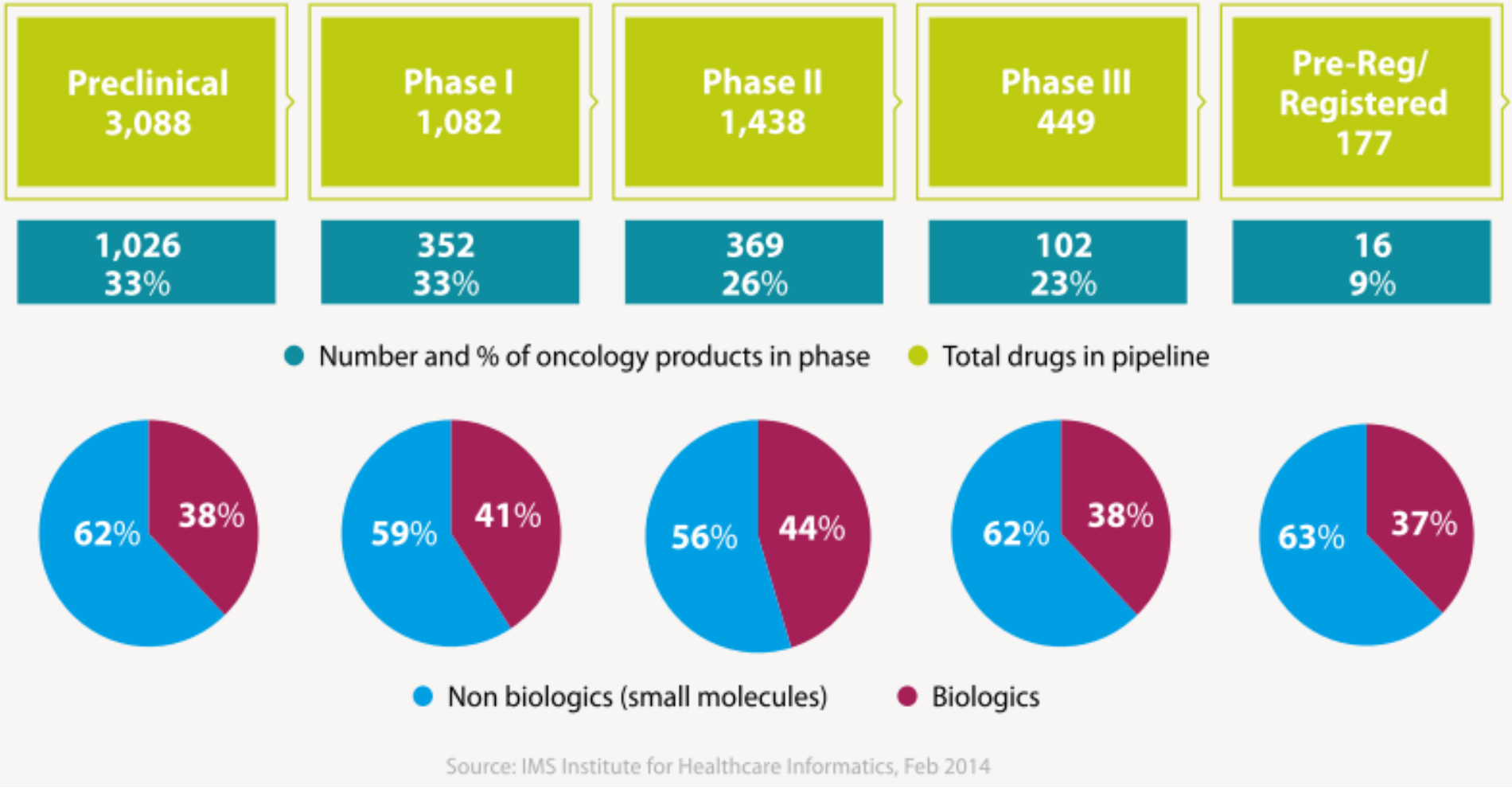
RCT: caratteristiche e criticità

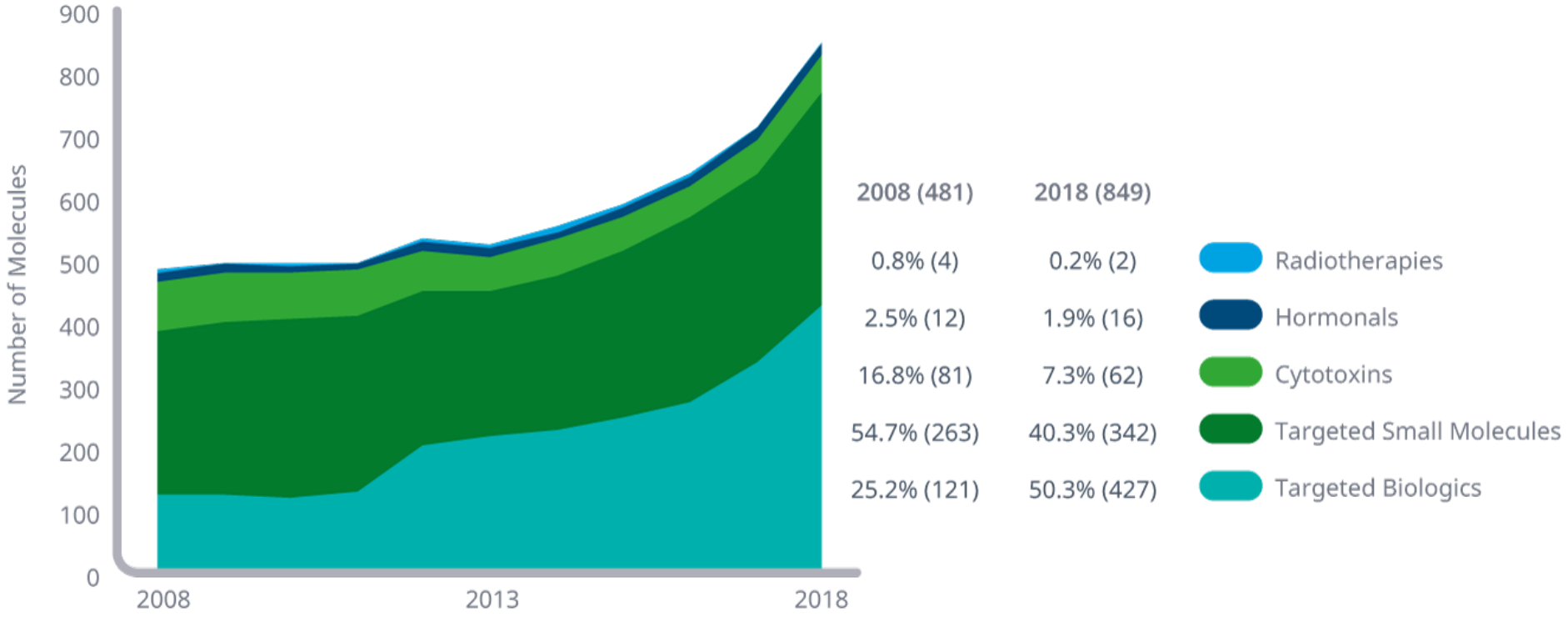
Maria Cecilia Giron
Dipartimento di Scienze del Farmaco
Edificio di Farmacologia
Email: cecilia.giron@unipd.it
Tel. 049-8275091

30% of all drugs in the research pipeline are oncology drugs



Source: IMS Institute for Healthcare Informatics, Feb 2014

The late-stage oncology pipeline included 849 molecules in 2018, up 77% since 2008, due to the increasing number of targeted therapies

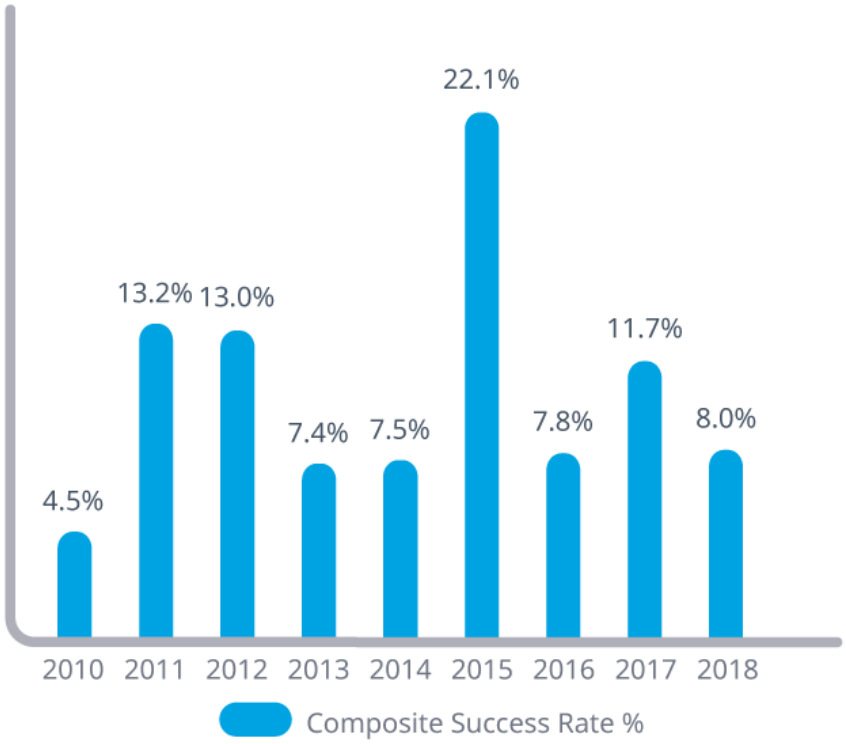


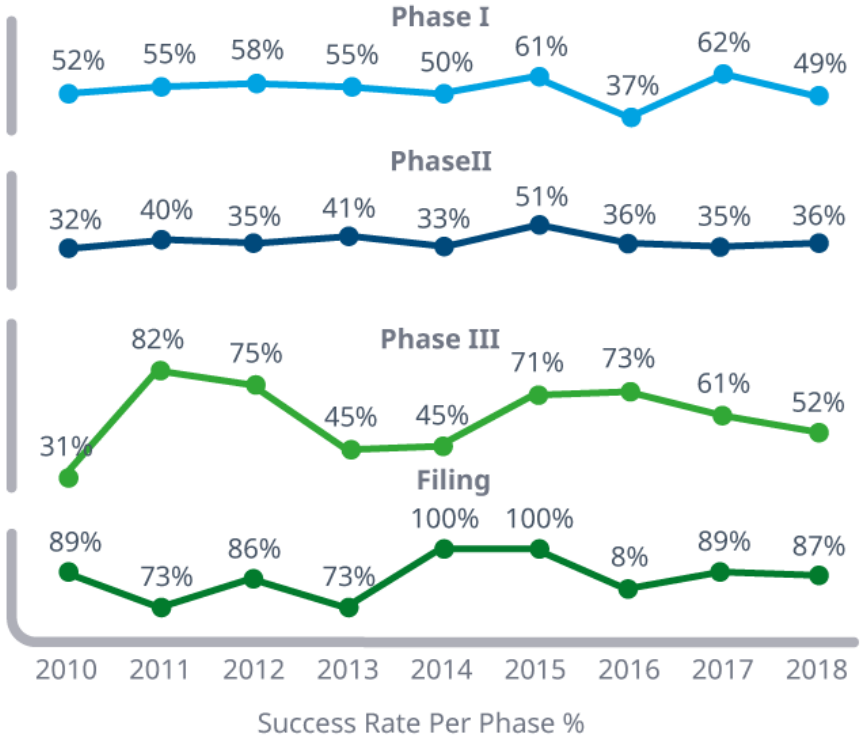
Source: IQVIA Pipeline Intelligence, Dec 2018; IQVIA Institute, May 2019

Chart notes: Late phase pipeline includes trials in Phase II or higher for the most advanced indication. Phase I/II trials are included as Phase II.

Report: Global Oncology Trends 2019 – Therapeutics, Clinical Development and Health System Implications. IQVIA Institute for Human Data Science, May 2019

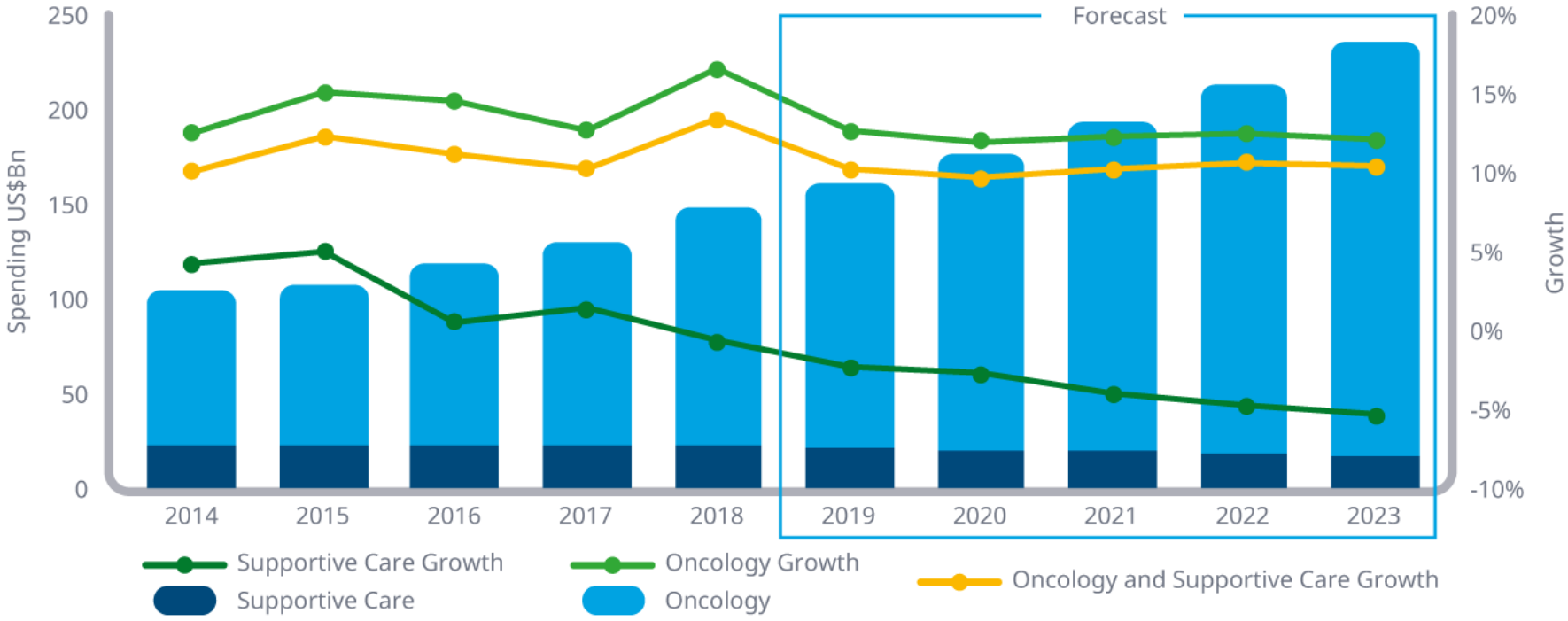
Oncology clinical trials have a high risk of failure with a composite success rate of 8% in 2018, slightly lower than the average since 2010





Source: IQVIA ARK R&D Intelligence; IQVIA Institute, Apr 2019
Chart notes: Composite Success Rate = Phase I x Phase II x Phase III x Regulatory Submission.
Report: Global Oncology Trends 2019 – Therapeutics, Clinical Development and Health System Implications. IQVIA Institute for Human Data Science, May 2019

Oncology spending will reach nearly $240 billion through 2023, growing 9–12%

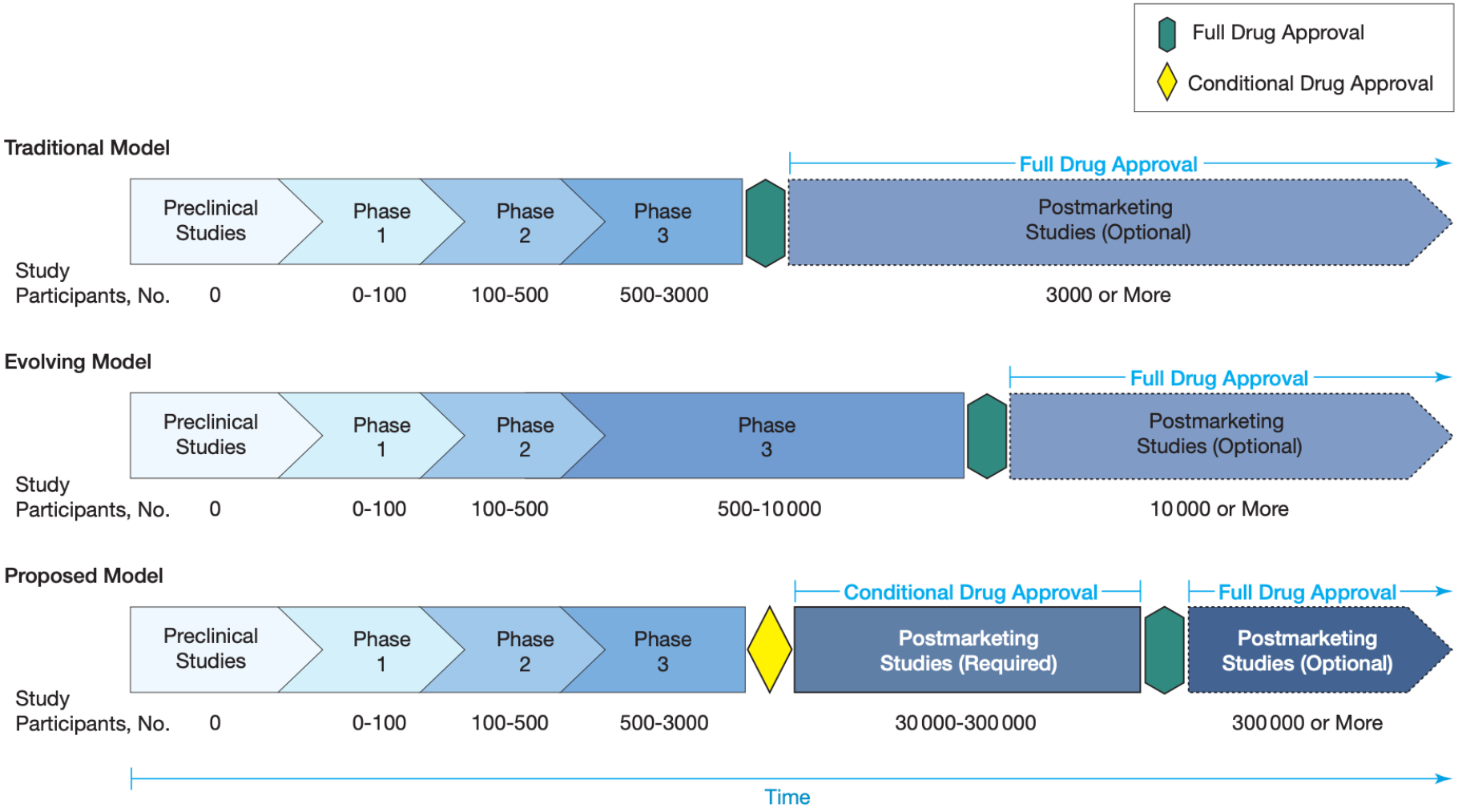


Source: IQVIA Institute, Apr 2019

Chart notes: Therapeutic oncologics include those classified by EphMRA (European Pharmaceutical Market Research Association) as cytotoxics in the L1 or L2 classes, as well as radiotherapeutics (V3C) and specific molecules classified elsewhere but used primarily in cancer (lenalidomide, aldesleukin, pomalidomide). Supportive care includes anti-nauseants and cancer detox agents (A4A and V3D), erythropoietins (B3C), GM-CSF white blood cell boosters (L3A), other interferon therapies used in cancer (L3B excluding multiple sclerosis drugs), and bisphosphonates used to prevent bone metastases (M5B4).

Report: Global Oncology Trends 2019 – Therapeutics, Clinical Development and Health System Implications. IQVIA Institute for Human Data Science, May 2019

Figure. Alternative Models for Studying Drug Safety



Top row, historical approach; middle row, where the current system is evolving toward now; and bottom row, proposed approach. Phase 1 indicates dose escalation, usually in healthy study participants; phase 2, dose ranging (usually first time in patients); phase 3, pivotal trials for registration; phase 4, postmarketing (not always required).

Elementi critici degli studi RCT

I risultati sono validi?

Validità interna

- Forza disegno sperimentale
- Randomizzazione
- Cecità
- Gruppo controllo
- Completezza Follow-up
- Analisi sottogruppi

I risultati sono rilevanti ?

Rilevanza del dato

- Statistica
- Clinica

Applicabilità

I risultati sono estrapolabili?

Validità esterna

- Rappresentatività del campione (Criteri inclusione/esclusione)
- Comparabilità degli ambienti
- Disponibilità dei trattamenti

Studi in oncologia

net clinical benefit

accelerated approval

reimbursement price

Strategia standard per aumentare la probabilità di registrazione di un farmaco

1. Trial(s) sovradimensionato/i
2. Endpoint primario surrogato (EFS, PFS)
3. Analisi ad Interim nel momento piu' favorevole per il trattamento sperimentale
4. Trial interrotto per significativita' statistica
5. Cross-over dei pazienti di controllo al trattamento sperimentale (“*motivi etici”)*
6. L'assenza di effetto sulla sopravvivenza viene attribuito al cross-over
7. Pubblicazione risultati e pressioni su EMA

Event Free Survival, Progression Free Survival

IMPRECISION MEDICINE

For every person they do help (blue), the ten highest-grossing drugs in the United States fail to improve the conditions of between 3 and 24 people (red).

1. ABILIFY (aripiprazole)
Schizophrenia

2. NEXIUM (esomeprazole)
Heartburn

3. HUMIRA (adalimumab)
Arthritis

4. CRESTOR (rosuvastatin)
High cholesterol

5. CYMBALTA (duloxetine)
Depression

6. ADVAIR DISKUS (fluticasone propionate)
Asthma

7. ENBREL (etanercept)
Psoriasis

8. REMICADE (infliximab)
Crohn's disease

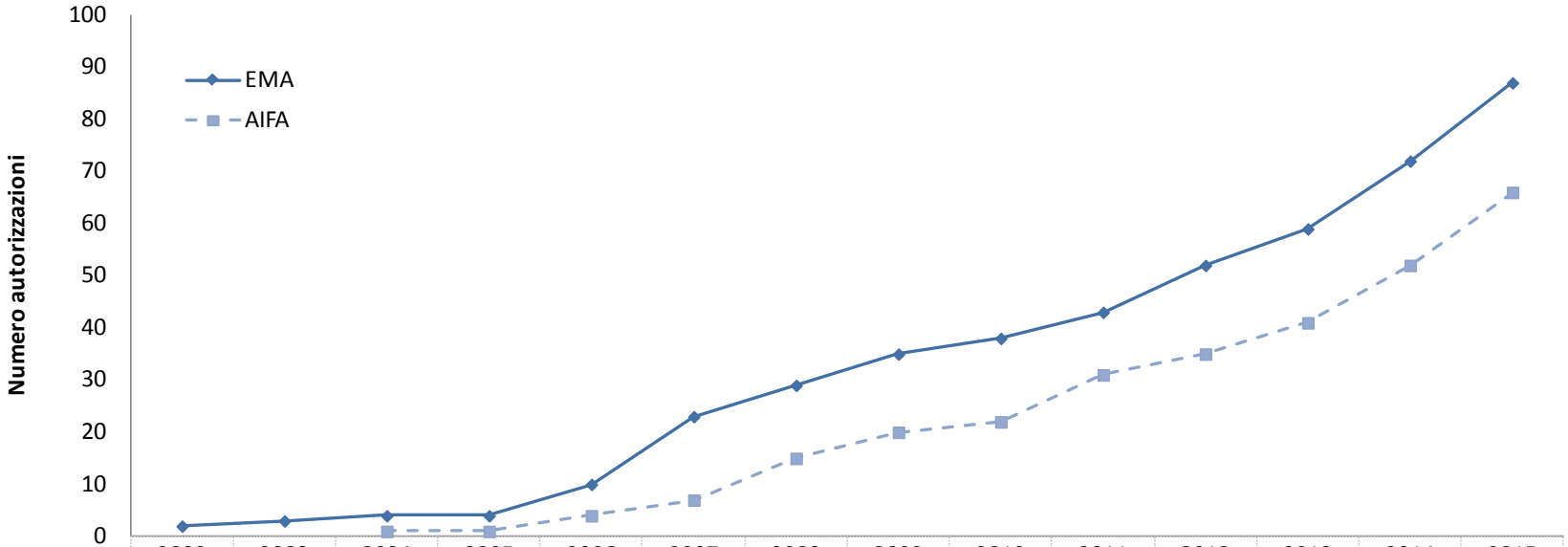
9. COPAXONE (glatiramer acetate)
Multiple sclerosis

10. NEULASTA (pegfilgrastim)
Neutropenia

Based on published number needed to treat (NNT) figures. For a full list of references, see Supplementary Information at go.nature.com/4dr78f.

Drugs approval in EMA & AIFA

Figura 1.9.1. Confronto farmaci autorizzati EMA ed approvati AIFA (dato cumulato 2002-2015)



	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
EMA	2	3	4	4	10	23	29	35	38	43	52	59	72	87
AIFA			1	1	4	7	15	20	22	31	35	41	52	66

ATC IV	Farmaco	Principio attivo	Classe	Innovazione terapeutica	Data decisione CTS	Data G.U. (data efficacia)	Data scadenza requisito
L01XC	YERVOY	Ipilimumab	H	Importante	30/10/2012	09/03/2013	08/03/2016
L02BX	ZYTIGA	Abiraterone	H	Potenziale	15/11/2012	06/04/2013	05/04/2016
M09AB	XIAPEX	Collagenasi di clostridium histolyticum	H	Potenziale	06/03/2013	14/03/2013	13/03/2016
L01XC	ADCETRIS	Brentuximab vedotin	H	Potenziale	02/12/2013	08/07/2014	07/07/2017
L01XC	PERJETA	Pertuzumab	H	Importante	02/12/2013	08/07/2014	07/07/2017
L04AX	REVLIMID	Lenalidomide	H	Potenziale	13/02/2014	30/09/2014	29/09/2017
J05AX	TIVICAY	Dolutegravir	H	Potenziale	10/03/2014	02/11/2014	01/11/2017
J04AK	SIRTURO	Bedaquilina	H	Potenziale	11/03/2014	01/10/2014	30/09/2017
L01XC	KADCYLA	Trastuzumab emtansine	H	Potenziale	07/04/2014	11/10/2014	10/10/2017
L01CD	ABRAXANE	Nab paclitaxel	H	Importante	07/04/2014	21/02/2015	20/02/2018
J05AX	SOVALDI	Sofosbuvir	A	Importante	15/05/2014	20/12/2014	19/12/2017
L01XE	XALKORI	Crizotinib	H	Potenziale	09/06/2014	11/04/2015	10/04/2018
J05AE	OLYSIO	Simeprevir	A	Potenziale	10/11/2014	24/02/2015	23/02/2018
J05AX	DAKLINZA	Daclatasvir	A	SI	16/02/2015	05/05/2015	04/05/2018
R07AX	KALYDECO	Ivacaftor	A	SI	16/02/2015	05/05/2015	04/05/2018
J05AX	HARVONI	Ledipasvir/ Sofosbuvir	A	SI	24/03/2015	14/05/2015	13/05/2018
J05AX	VIEKIRAX	Ombitasvir/ Paritaprevir/ Ritonavir	A	Importante	21/01/2015	24/05/2015	23/05/2018
J05AX	EXVIERA	Dasabuvir	A	Importante	21/01/2015	24/05/2015	23/05/2018
V10XX	XOFIGO	Radio ra 223 dicloruro	H	Potenziale	13/05/2014	11/06/2015	10/06/2018
L04AX	IMNOVID	Pomalidomide	H	SI	18/02/2015	20/08/2015	19/08/2018
L01XX	ZYDELIG	Idelalisib	H	SI	18/02/2015	11/09/2015	10/09/2018
L01XE	IMBRUVICA	Ibrutinib	H	SI	13/07/2015	05/01/2016	04/01/2019

42 innovative drugs approved (12 in oncological area)

EFFICACY negli STUDI CLINICI

Le strategie di disegno, analisi e presentazione dei trials sui nuovi farmaci forniscono un'immagine distorta (troppo ottimistica) dell'efficacia di questi trattamenti

come verificare criticamente questi risultati ?

Clinical Governance

*"La **Clinical Governance** è un sistema attraverso cui le organizzazioni sanitarie sono responsabili del continuo miglioramento della qualità dei loro servizi e della salvaguardia di elevati standard di assistenza attraverso la creazione di un ambiente in cui possa svilupparsi l'eccellenza dell'assistenza sanitaria".*

- *responsabilità*
- *trasparenza*
- *coinvolgimento / partecipazione*
- *etica e valore del lavoro*


Education & Training
Risk Management
Clinical audit
Clinical Governance
Openness
Clinical effectiveness
Research and Development

Linee guida

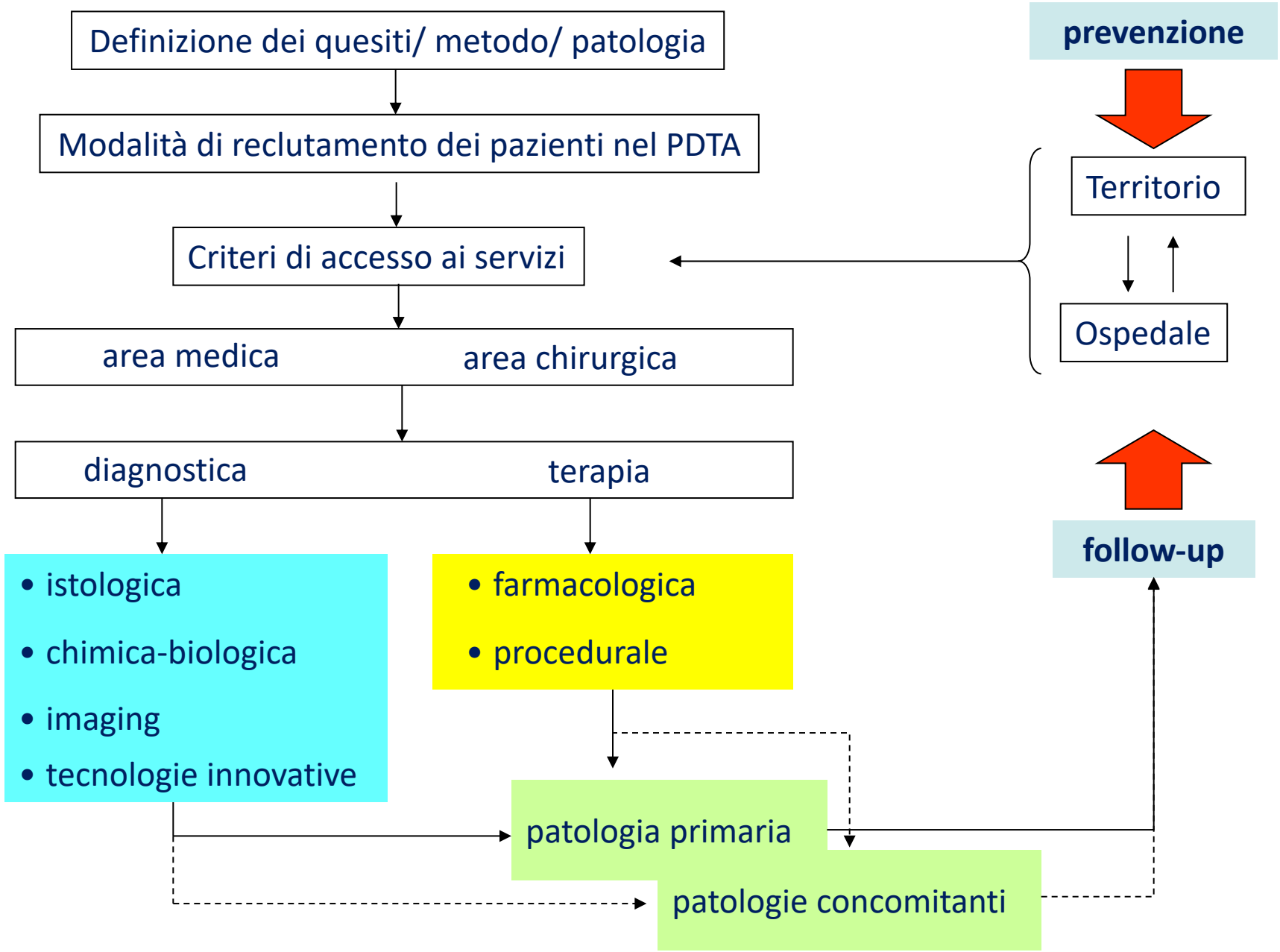
le linee guida nascono come strumento di governo clinico e ben si prestano a trattare quesiti strettamente terapeutici, che a loro volta si avvalgono di studi clinici randomizzati e controllati

	Sanità pubblica	Clinica
Principale ambito	Popolazioni	Individuo
Enfasi	Prevenzione	Diagnosi
	Promozione della salute	Trattamento
	Approccio olistico alla comunità	Approccio olistico al paziente
Paradigma	Interventi rivolti all'ambiente, agli stili di vita, ai sistemi sanitari, ecc.	Cura medica
Criteri di specializzazione	Analitico (epidemiologia)	Organico (cardiologia, ecc.)
	Per tipologia di popolazione e setting (es. salute occupazionale)	Gruppi di pazienti (es. pediatria)
	Per funzioni di valutazione, di sviluppo di politiche, ecc.	Per funzioni tecniche (es. radiologia)

Differenze fra approccio di sanità pubblica e approccio clinico ai problemi di salute (Fineberg 1990)

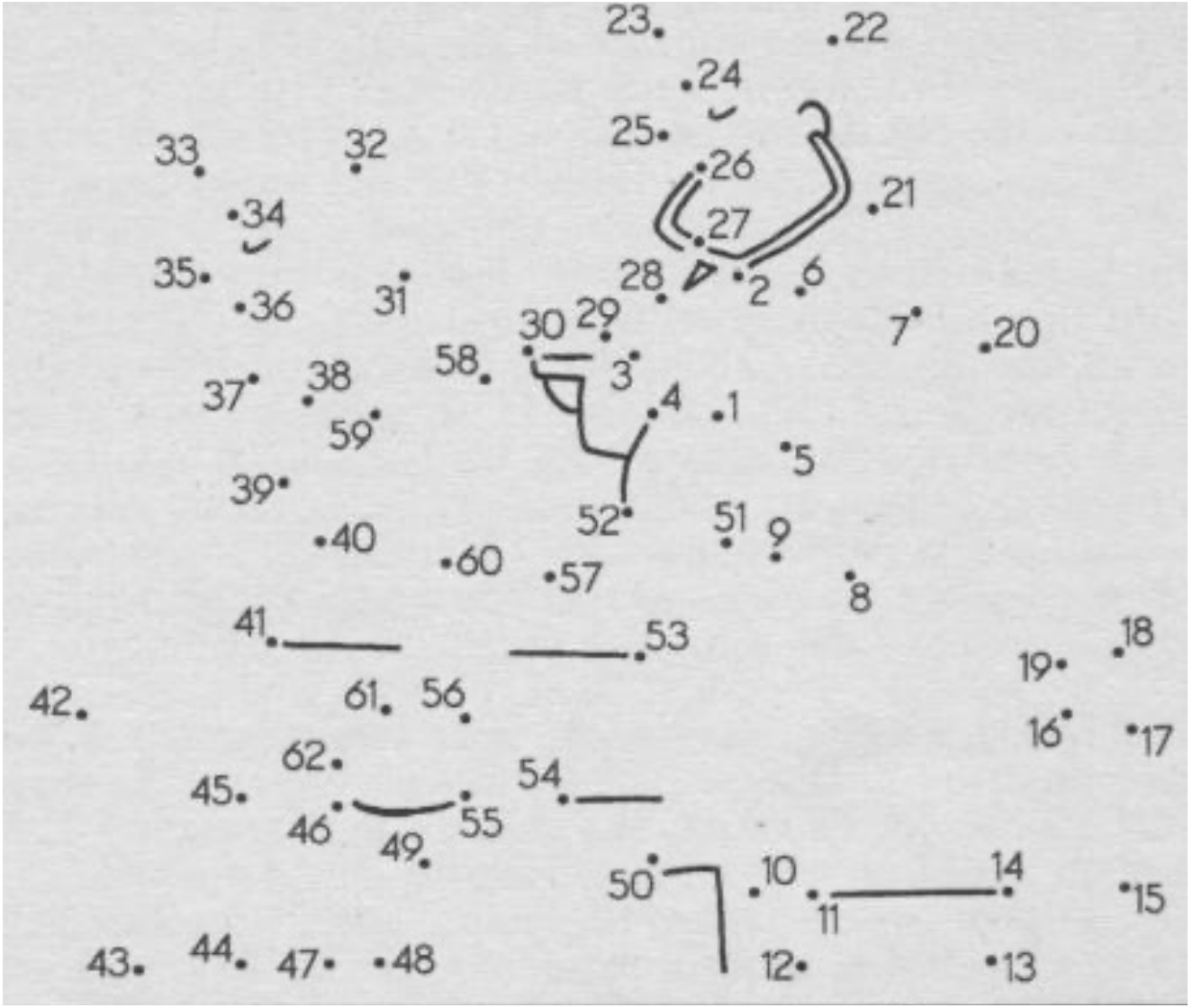
http://www.iss.it/binary/lgmr2/cont/Manuale_PNLG.1234439852.pdf

Il percorso diagnostico terapeutico assistenziale (PDTA)

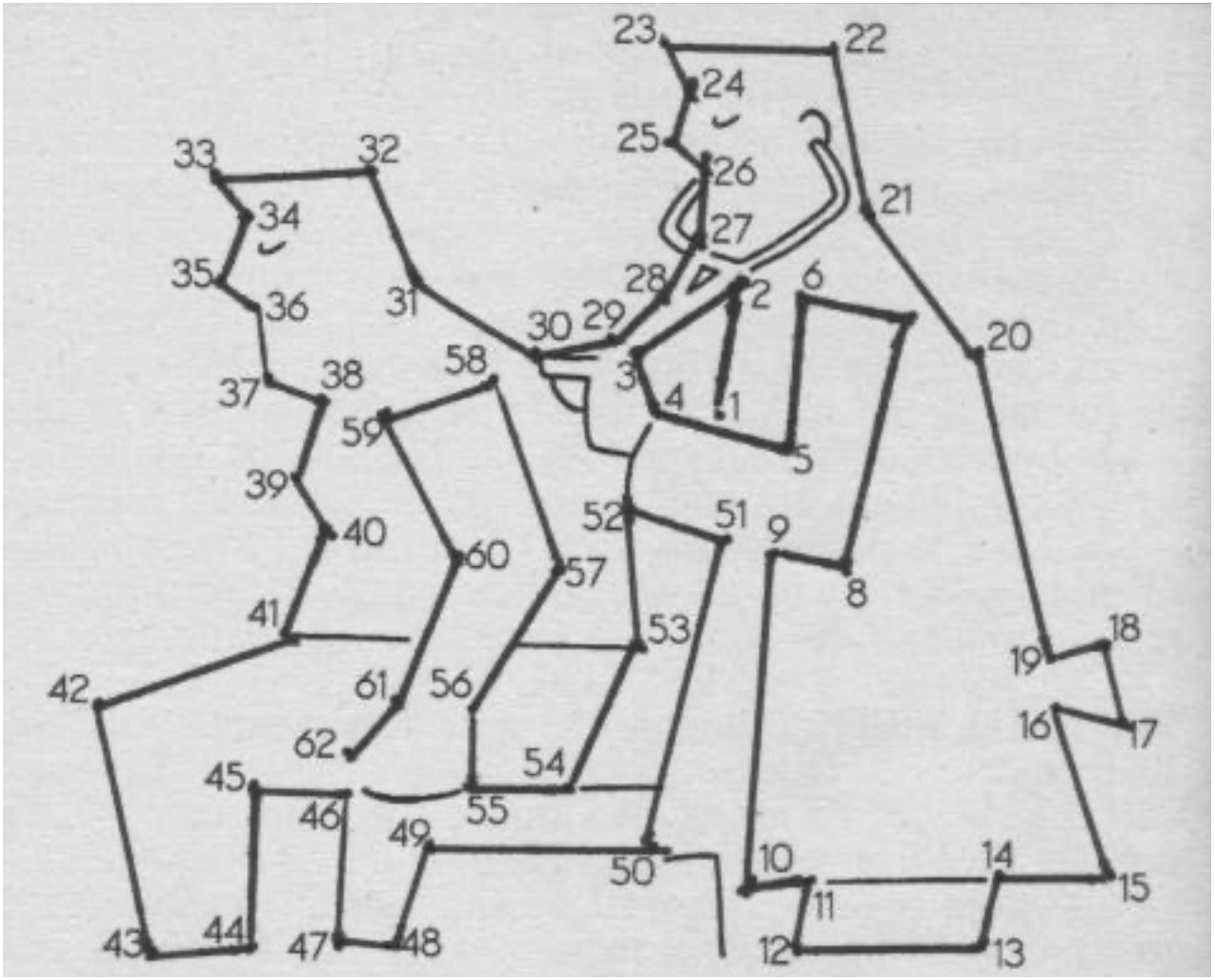


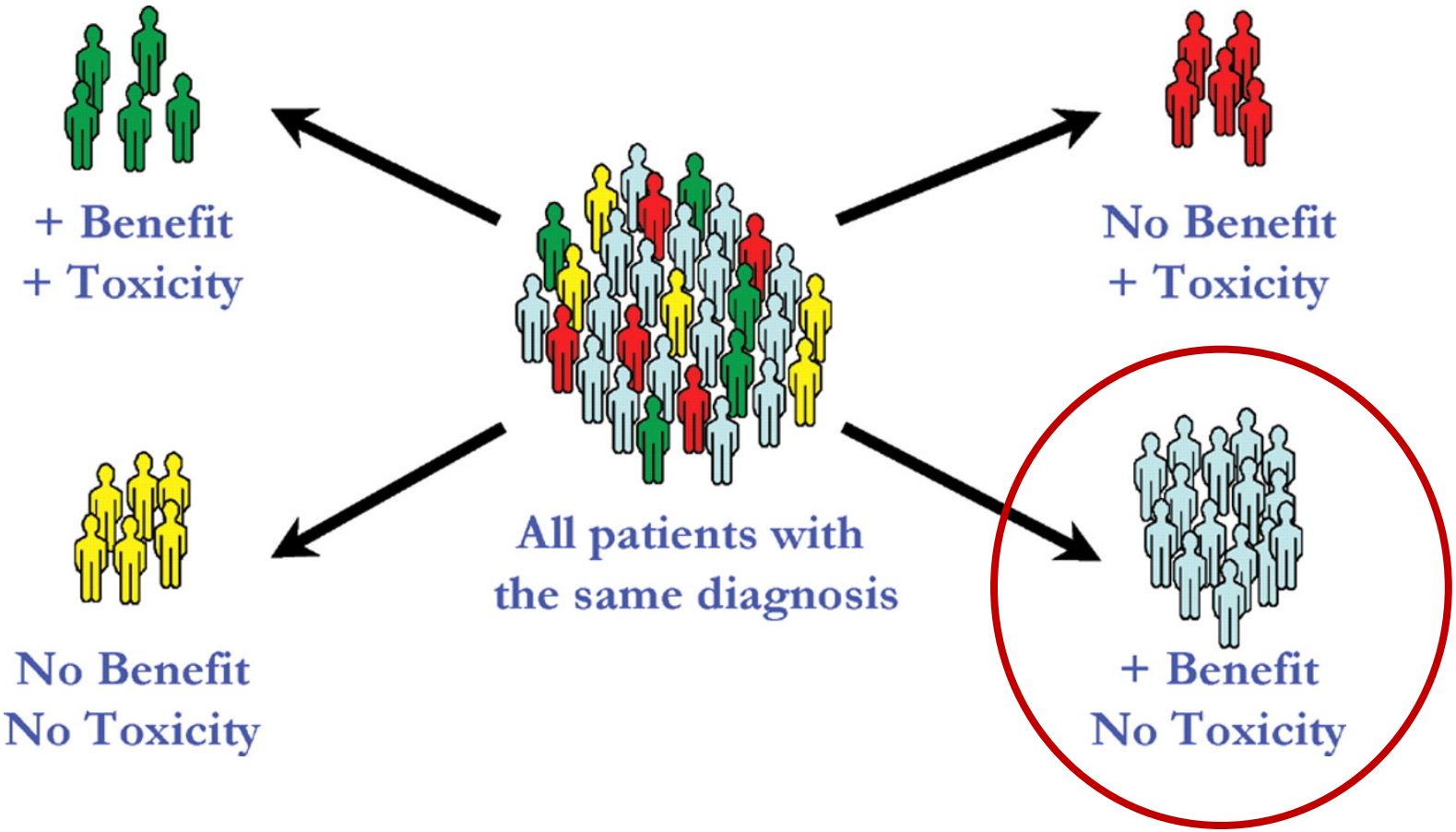
Il percorso diagnostico terapeutico assistenziale (PDTA)

PDTA : analisi complessa ma percorsi tracciati



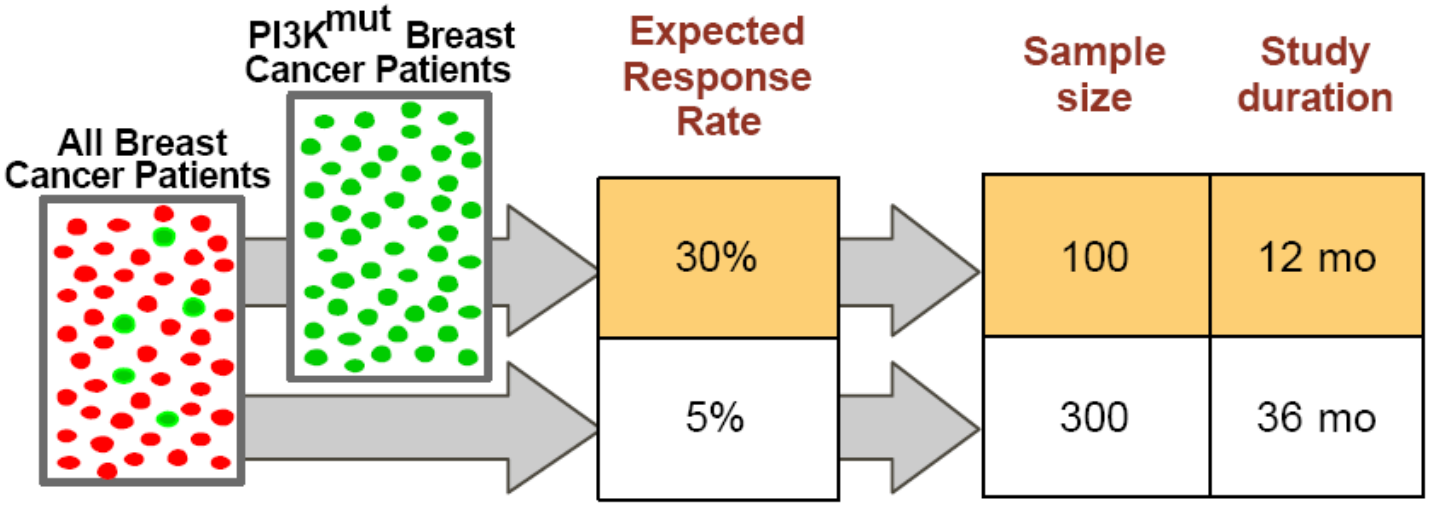
PDTA : il quadro generale diventa chiaro quando si analizzano tutte le condizioni prevedibili





Marsh, S. et al. Hum. Mol. Genet. 2006 15:R89-93R

The economics of patient selection: PI3K inhibitor development in breast cancer



Biomarker-driven selection will:

- Reduce trial size and associated costs by 67%
- Provide an unequivocal signal of efficacy by eliminating dilution effect of non-responders

Fonte: Novartis

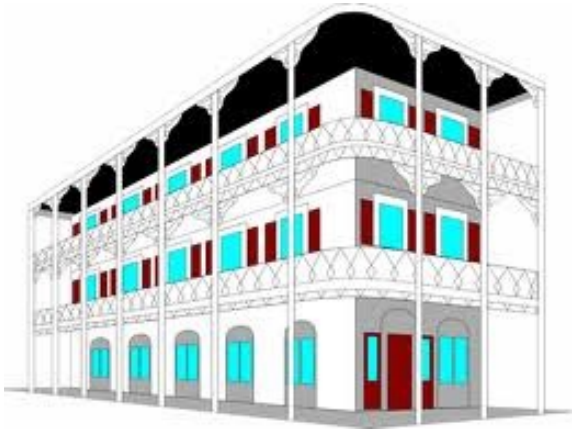
but....

Acquired PIK3CA amplification causes resistance to selective phosphoinositide 3-kinase inhibitors in breast cancer

"Our results suggest a novel mechanism of resistance involving amplification of an activating mutant PIK3CA allele in breast cancer cells ...."

RCTs (efficacy)

Real World (effectiveness)

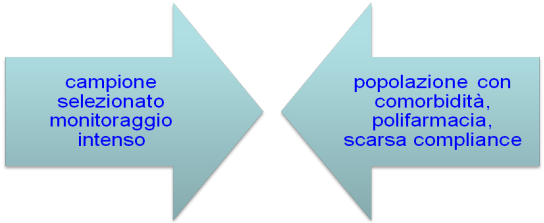


Quale è la relazione fra l'efficacy e l'effectiveness ?

campione selezionato monitoraggio intenso

popolazione con comorbidità, polifarmacia, scarsa compliance

RCTs (efficacy)



Real World (effectiveness)

Study characteristics	Efficacy trial	Effectiveness trial
Research question	Will the intervention work under ideal conditions?	Will the intervention result in more good than harm under usual practice conditions?
Setting	Restricted to specialized centers	Open to all institutions
Patient selection	Selected, well-defined patients	A wide range of patients selected using broad eligibility criteria
Study design	Smaller RCT using parallel group or factorial or other approaches (crossover design)	Large multicenter RCTs using parallel groups or factorial cluster
Baseline assessment	Elaborate and detailed	Simple and clinician friendly
Study intervention	Tightly protocolized using optimal therapy under optimal conditions	Implemented in usual clinical Practice; limited study protocol if any
Co-interventions	Tightly controlled protocol for many aspects of care	All therapy based on local clinical practice/ experience/minimal control
Compliance	Compliance essential	Non compliance expected and considered in sample size / analysis
Analysis	May be done by treatment received where non compliant patients may be removed	Always intention to treat where all patients are included
Data management		
i) Data collection	Elaborate	Minimal and simple
ii) Data monitoring	Detailed and rigorous	Minimal
Study management	Significant interventions and support from research staff	Minimal support and interventions from research team

*Data monitoring refers to the review of source documents and adjudication/verification of outcomes[11]

Quali dati raccogliere NEL SINGOLO PAZIENTE:

1) **la dose "reale"** giornaliera (conoscere ciascuna delle dosi pro/die per stimare la dose media pro/die di "quel " paziente nonchè le eventuali oscillazioni in tale paziente tra dose massima giornaliera e dose minima giornaliera)

2) SOPRATTUTTO la **durata del trattamento**: Intervallo in giorni che intercorre tra inizio trattamento e fine trattamento

---

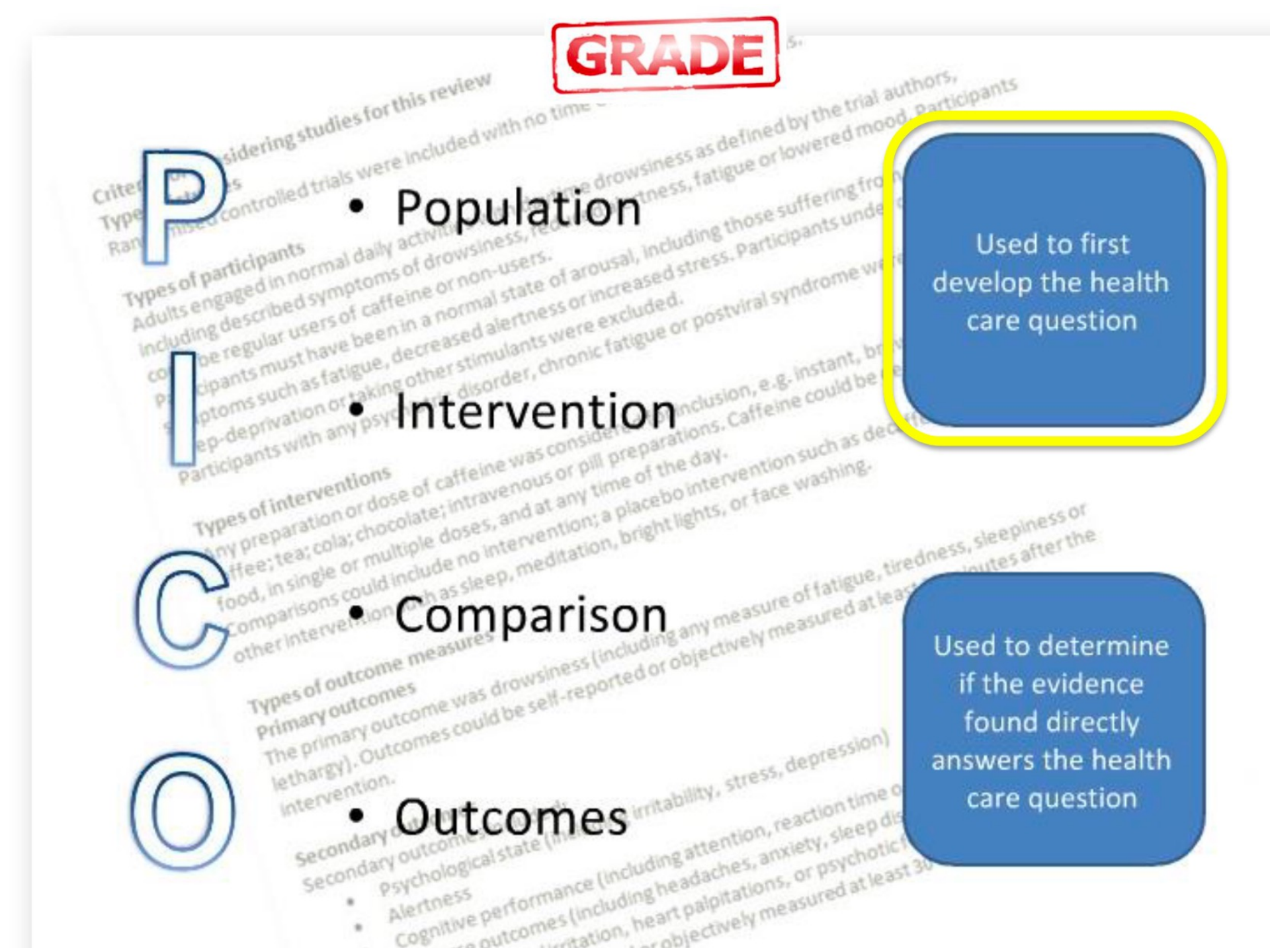
Quali dati ottenere dalla CASISTICA GLOBALE :

1) Analizzare la **distribuzione statistica delle dosi e delle durate**

2) Per quanto riguarda le durate: stimare la durata media, il range (max, min), la distribuzione per fasce di durata e soprattutto costruire la curva di Kaplan Meier riferita alle durate di trattamento

Il percorso verso la decisione terapeutica...

Una volta definito con chiarezza il quesito clinico...


GRADE
P
• Population
I
• Intervention
C
• Comparison
O
• Outcomes
Used to first develop the health care question
Used to determine if the evidence found directly answers the health care question

Ne risulta un quesito posto come: «L'evento I di una popolazione P come è relazionato all'evento O rispetto all'elemento C?»

Un esempio di applicazione di tale metodo può essere: «Il fumo di sigaretta negli uomini di età superiore a 50 anni è legato ad una maggior incidenza di infarto cardiaco rispetto al fumo di sigaro?»

Strutturazione del Quesito Clinico sec. modello P.I.C.O.

P	Nei **P**azienti con…	Specifiche caratteristiche di malattia (stadio, classe di rischio, ecc.)
I	l'**I**ntervento…	Intervento terapeutico oggetto del quesito clinico
C	(è suscettibile di impiego) in **C**onfronto con…	Trattamento altrimenti consi-derabile in alternativa all'inter-vento in esame
O	riguardo agli **O**utcome di beneficio/danno…	Parametri clinico-laboratoristici ritenuti essenziali per la decisio-ne terapeutica

I fattore **P** (problem/patient/population) indica il soggetto del quesito, ossia il gruppo di persone (popolazione) accomunati da almeno un elemento statistico comune (età, sesso, malattia, ecc.). Risponde al quesito: "**Come si può descrivere un gruppo di pazienti simile a quello da trattare?".** I termini di ricerca devono essere bilanciati fra la specificità e la sintesi, in modo da avere un campione il più adeguato possibile. Inoltre è bene considerare la plausibilità della relazione con l'esito *O* in questione.

Il fattore **I** (intervention) indica la caratteristica principale, ossia la condizione, patologia o evento che agisce sulla popolazione *P* in questione (fattori di rischio, condizioni patologiche pregresse, test clinico ecc.). Risponde alla domanda: "**Quale intervento principale deve essere preso in considerazione?**". Il termine deve essere specifico, in modo da escludere fattori minori o confondenti.

Il fattore **C** (comparison/control) indica il termine di paragone con cui va confrontato il fattore *I* in grado di relazionarsi con l'esito *O*. Risponde al quesito: "Qual è l'alternativa principale da confrontare con l'intervento?" o " **Quale sarebbe l'intervento alternativo che si potrebbe applicare?**". la risposta può essere anche nulla.

Il fattore **O** (outcome) è l'esito o il fine della ricerca. Risponde alla domanda: "Cosa si può sperare di ottenere?", oppure "**Su cosa incide realmente questo intervento**?". Deve essere speculare all'eventuale problema e, quindi, essere specifico ma sintetico.

Elementi critici degli studi RCT

I risultati sono validi?

Validità interna

- Forza disegno sperimentale
- Randomizzazione
- Cecità
- Gruppo controllo
- Completezza Follow-up
- Analisi sottogruppi

I risultati sono rilevanti ?

Rilevanza del dato

- Statistica
- Clinica

Applicabilità

I risultati sono estrapolabili?

Validità esterna

- Rappresentatività del campione (Criteri inclusione/esclusione)
- Comparabilità degli ambienti
- Disponibilità dei trattamenti

Studi in oncologia

net clinical benefit

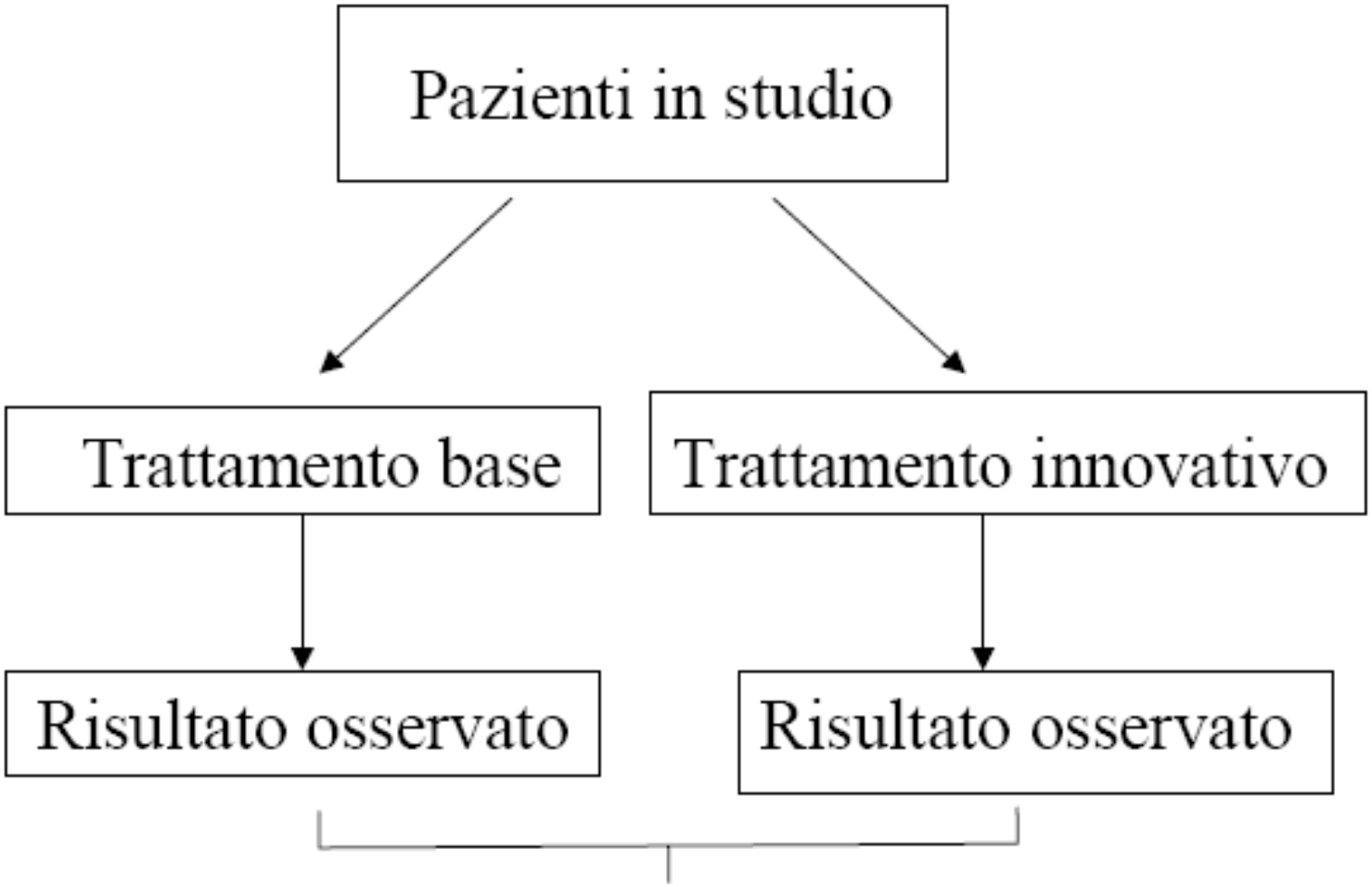
accelerated approval

reimbursement price

Criteri fondamentali per una corretta sperimentazione clinica sui farmaci

Il confronto dei risultati ottenuti nel gruppo sperimentale rispetto a quelli del gruppo di controllo

- ✓ Studi di superiorità
- ✓ Studi di non inferiorità


Pazienti in studio
Trattamento base
Trattamento innovativo
Risultato osservato
Risultato osservato
Test di H0 sulla differenza tra i risultati osservati

Superiorità/Non inferiorità

Esiste un trattamento efficace? → **NO** → **Trial di superiorità verso placebo**

↓

SI

↓

Trial verso miglior trattamento → **Trial di superiorità: il nuovo trattamento più efficace del trattamento di controllo**

↓

Trial di non inferiorità: il nuovo trattamento non meno efficace del trattamento di controllo

↓

Il nuovo trattamento dovrebbe avere altri vantaggi: es. più sicuro, meno invasivo, meno costoso, ecc.

Esempi di trial di non inferiorità ACCETTABILI

TRIAL	VANTAGGI DEL NUOVO TRATTAMENTO
Angioplastica+ stent carotide *vs* chirurgia (Lancet 2006;368:1239)	**Minore invasività**
Amoxicillina 3 gg *vs* 8 gg nella polmonite comunitaria (BMJ 2006;332:1355)	**Minore resistenza, minori costi**
Ibandronato mensile *vs* giornaliero (Clin Ther 2007;29:1116)	**Migliore compliance**
Glargine 1/die *vs* Lispro 3/die (Lancet 2008;371:1073)	**Migliore tolleranza, minor fastidio**

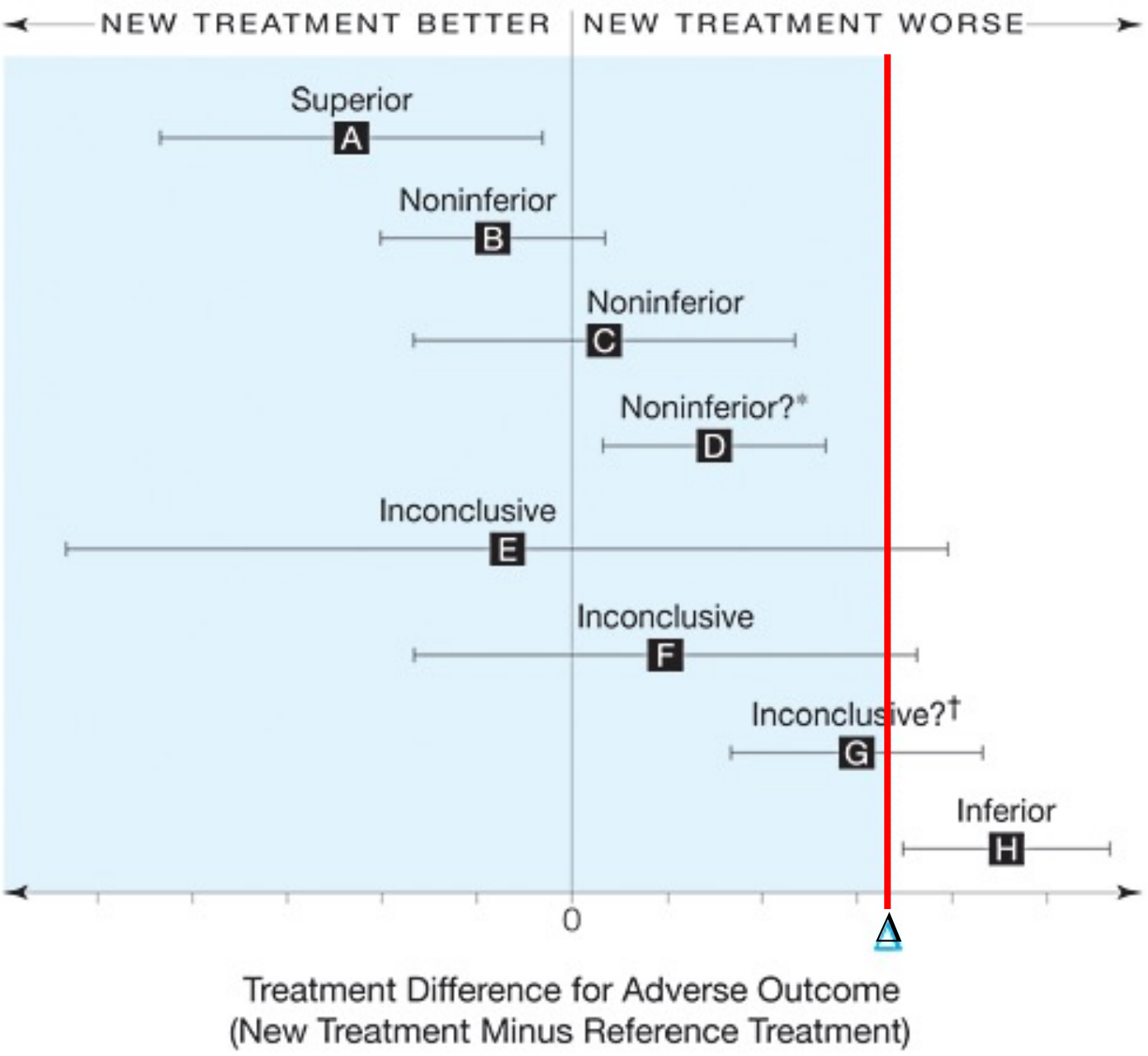


Figure. Possible Scenarios of Observed Treatment Differences for Adverse Outcomes (Harms) in Noninferiority Trials [Piaggio G et al. JAMA 2006; 295:1152-60]

Un dibattito acceso

Non-inferiority trials are unethical because they disregard patients' interests

Silvio Garattini, Vittorio Bertele'

Equivalence trials[1] have been widely used to assess new drugs, but have recently lost ground to a non-inferiority design. This type of trial is usually accepted by regulatory authorities for approval of new drugs or new indications,

but not to the extent that it is recognised as such. For example, if the non-inferiority limit is set at 7·5%, an increase in the incidence of serious events or deaths—say 7% instead of the 5% currently established

Lancet 2007; 370: 1875–77
Published Online October 23, 2007
DOI:10.1016/S0140-6736(07)61604-3

ELSEVIER

Journal of Hepatology 46 (2007) 947–954

Journal of Hepatology

www.elsevier.com/locate/jhep

Review

Methodology of superiority vs. equivalence trials and non-inferiority trials

Erik Christensen*

Clinic of Internal Medicine I, Bispebjerg University Hospital, Bispebjerg Bakke 23, DK-2400 Copenhagen NV, Copenhagen, Denmark

Non-inferiority trials are unethical because they disregard patients' interests

Silvio Garattini, Vittorio Bertele'

Equivalence trials[1] have been widely used to assess new drugs, but have recently lost ground to a non-inferiority design. This type of trial is usually accepted by regulatory authorities for approval of new drugs or new indications, although the US Food and Drugs Administration has raised some concerns.[2] In this paper, we argue that the scientific community should ban non-inferiority and equivalence trials because they are unethical, whatever measures are taken to prevent their methodological pitfalls and inappropriate interpretation of results.[3–8] Exceptions might exist, but we could not identify a situation in which patients can justifiably be entered into a trial that will not provide them with any advantage.

Pretext for looking for non-inferiority

Use of equivalence or non-inferiority rather than superiority designs implies the intention of not trying to prove any additional value of new drugs. However, the declared aim is to expand treatment options for patients with poor tolerance of, or no response to, available products. Drug producers argue that there is no reason to define the benefit-risk profile of new agents as better than those of existing drugs: it is enough to show that they are similar. One does not even need to know whether a new drug with some innovative peculiarities—for example, longer activity—is more effective. The added value rests on the probability of better compliance with, for instance, once-a-day administration compared with a more complex regimen. Similarly, the added value of a more convenient formulation arguably lies in its ease of use. Generally, when a new drug is claimed to have only minor advantages or no advantage over available products, a superiority test is not believed to be necessary; non-inferiority allows new products to compete with older ones on the basis of small differences made to seem but not to the extent that it is recognised as such. For example, if the non-inferiority limit is set at 7·5%, an increase in the incidence of serious events or deaths—say 7% instead of the 5% currently established for the comparator—is not seen as large enough to mark a difference between the new and the control drug. The new drug will therefore be considered non-inferior to the old drug, even if in 1000 patients treated with the former, there could be 20 more deaths than with the latter.

These arguments also apply to equivalence trials, which aim to prove similarity of a new drug to the comparator, since true equivalence is theoretical and is difficult to demonstrate. Equivalence means that a new drug is not much worse than the comparator (as in non-inferiority trials), but also is not much better. Similarity is defined by limits that include a superiority margin as well as a non-inferiority margin. Since equivalence trials explore the differences between control and study treatments in both directions, they provide a more reliable estimate of the relative efficacy of two treatments than do non-inferiority trials. However, use of a non-inferiority limit exposes equivalence trials to the same ethics issues.

No limits to the non-inferiority limit

The wider the non-inferiority interval, the smaller the sample needed. The smaller the sample, the smaller the investment needed to do the trial, as well as the greater the chance of overlooking a difference and concluding non-inferiority. This situation has led to the adoption of extreme hypotheses, which are just arbitrary, yet approved by ethics committees and allowed in the scientific literature; for example, in the COMPASS study,[9] the thrombolytic saruplase was judged equivalent to

Lancet 2007; 370: 1875–77

Published Online
October 23, 2007
DOI:10.1016/S0140-6736(07)61604-3

Mario Negri Institute for Pharmacological Research, Milan, Italy (Prof S Garattini MD, V Bertele' MD)

Correspondence to:
Dr Vittorio Bertele', Mario Negri Institute for Pharmacological Research, Milan, 20156, Italy
bertele@marionegri.it

situation in which patients can justifiably be entered into a trial that will not provide them with any advantage.

Pretext for looking for non-inferiority

Use of equivalence or non-inferiority rather than superiority designs implies the intention of not trying to prove any additional value of new drugs. However, the declared aim is to expand treatment options for patients with poor tolerance of, or no response to, available products. Drug producers argue that there is no reason to define the benefit-risk profile of new agents as better than those of existing drugs: it is enough to show that they are similar. One does not even need to know whether a new drug with some innovative peculiarities—for example, longer activity—is more effective. The added value rests on the probability of better compliance with, for instance, once-a-day administration compared with a more complex regimen. Similarly, the added value of a more convenient formulation arguably lies in its ease of use. Generally, when a new drug is claimed to have only minor advantages or no advantage over available products, a superiority test is not believed to be necessary; non-inferiority allows new products to compete with older ones on the basis of small differences made to seem to benefit patients.

Looking for non-inferiority or overlooking differences?

What is wrong with this approach? Problems arise from the definition of non-inferiority and the statistical criteria for its basis.[3–8] Non-inferiority is a kind of similarity within a limit. The limit is the degree of tolerable inferiority of the new drug compared with the standard treatment. This arbitrary difference in efficacy, the non-inferiority margin or delta, is decided before doing the study. Non-inferiority is judged to have been established when the point estimate and 95% CI of the effect of the new drug do not fall outside the preset non-inferiority margin. A non-inferior test drug could actually be less effective or less safe than the comparator, aim to prove similarity of a new drug to the comparator, since true equivalence is theoretical and is difficult to demonstrate. Equivalence means that a new drug is not much worse than the comparator (as in non-inferiority trials), but also is not much better. Similarity is defined by limits that include a superiority margin as well as a non-inferiority margin. Since equivalence trials explore the differences between control and study treatments in both directions, they provide a more reliable estimate of the relative efficacy of two treatments than do non-inferiority trials. However, use of a non-inferiority limit exposes equivalence trials to the same ethics issues.

No limits to the non-inferiority limit

The wider the non-inferiority interval, the smaller the sample needed. The smaller the sample, the smaller the investment needed to do the trial, as well as the greater the chance of overlooking a difference and concluding non-inferiority. This situation has led to the adoption of extreme hypotheses, which are just arbitrary, yet approved by ethics committees and allowed in the scientific literature; for example, in the COMPASS study,[9] the thrombolytic saruplase was judged equivalent to streptokinase for post-myocardial infarction, even though the saruplase group had 50% more deaths than the control group. Therefore, in absolute numbers, saruplase would be regarded as effective and safe as streptokinase even if there were 35 deaths per 1000 treated in addition to the 70 with streptokinase alone. The test of this questionable hypothesis only required about 3000 patients, in an era in which testing the superiority of tissue-type plasminogen activator over streptokinase involved about 90 000 patients overall in three large clinical trials.[10–12]

The results of trials like COMPASS also arouse concern about the breadth of their confidence intervals. Sometimes these are so wide that what is judged non-inferior from a statistical point of view might actually be questionable from a clinical point of view.[13,14]

Research, Milan, 20130, Italy
bertele@marionegri.it

Unreliable messages from questionable methods

As with a superiority design in a placebo-controlled trial, evidence of non-inferiority to active comparators might allow drugs onto the market that are in fact less acceptable than those in current clinical use. Worse, if the difference between the standard treatment and placebo is small, dependent on the non-inferiority limit, the effect of the supposedly non-inferior drug might actually be close to that of placebo. In any case, the loss in efficacy might be greater than it appears, since the effect of the standard treatment includes that of placebo. For example, if the standard treatment prevents 30% of expected events and the non-inferiority margin allows the new drug to prevent only 20%, the allowable loss in efficacy appears to be a third—but if the placebo effect accounts for 10% of the overall action, half the efficacy could actually be lost. Thus, non-inferiority trials expose patients to clinical experiments without any assurance that the experimental drug is not worse than the standard treatment, and without really exploring whether it is better.

Commercial aims, not patients' interests

Are there specific reasons for allowing a non-inferiority approach? One reason cited is that for patients who do not respond to existing treatments, products with similar activity could offer a useful alternative. The aim is reasonable, but the approach is not. If the target is non-responders to current treatments, why not test the new agents' superiority in this subset, rather than its non-inferiority in the overall population? This approach would meet patients' needs best, but restricts the market that can be targeted by the drug companies.

Another suggested reason is that non-inferior drugs might be better tolerated or easier to use than existing treatments. However, these features are unlikely to be confirmed in non-inferiority trials, since any advantage should translate into better compliance and result in a superior rather than a non-inferior outcome.

Superiority trials are also said to generally take much longer and require many more patients than do non-inferiority trials, delaying the availability of

rhagic strokes after thrombolysis in acute myocardial infarction. In these circumstances, however, a superiority trial would be a preferable way to compare the effectiveness of two treatments in terms of survival without stroke by cumulatively measuring efficacy and safety events.

These examples are intended to show that any question of practical relevance to patients requires a test of superiority. The superiority approach, whether the hypothesis is verified or not, provides information about new drugs in the context of available treatments, whereas the non-inferiority trial does not. From a commercial point of view, to prove non-inferiority of new products is less risky than aiming to establish their superiority. Failure to prove superiority can tarnish the product's commercial image, although it could provide more information for doctors and patients. The non-inferiority approach is likely to overlook differences that might stop the product getting onto the market. A demonstration of non-inferiority leaves the product in a kind of limbo: its place in therapy is not established, although its place on the market is assured.

Enrolling patients in non-inferiority trials betrays their trust

We believe that non-inferiority studies have no ethical justification, since they do not offer any possible advantage to present and future patients, and they disregard patients' interests in favour of commercial ones. This situation betrays the agreement between patients and researchers set out in any fair informed consent form that presents randomised trials as the only ethical way to address clinical uncertainty. Non-inferiority trials claim minor advantages for the test drugs, but do not prove their efficacy compared with older products. Few patients would agree to participate if this message were clear in the informed consent form: as we said before, why should patients accept a treatment that, at best, is not worse, but could actually be less effective or less safe than available treatments?[15]

In conclusion, we believe that non-inferiority trials fail to meet the commitments of good clinical research: "Ask

drug is not worse than the standard treatment, and without really exploring whether it is better.

Commercial aims, not patients' interests

Are there specific reasons for allowing a non-inferiority approach? One reason cited is that for patients who do not respond to existing treatments, products with similar activity could offer a useful alternative. The aim is reasonable, but the approach is not. If the target is non-responders to current treatments, why not test the new agents' superiority in this subset, rather than its non-inferiority in the overall population? This approach would meet patients' needs best, but restricts the market that can be targeted by the drug companies.

Another suggested reason is that non-inferior drugs might be better tolerated or easier to use than existing treatments. However, these features are unlikely to be confirmed in non-inferiority trials, since any advantage should translate into better compliance and result in a superior rather than a non-inferior outcome.

Superiority trials are also said to generally take much longer and require many more patients than do non-inferiority trials, delaying the availability of potentially useful drugs. However, non-inferiority trials do not necessarily need a smaller sample size, which can be the result of selecting a large inferiority margin or of other questionable methodological choices.[1] Moreover, it is our view that a delay in the availability of proven effective drugs is preferable to early availability of potentially advantageous drugs whose real efficacy has not been formally established. Actual efficacy testing might never be done, particularly if patients no longer agree to be randomly assigned to older drugs.

A more convincing approach might be to test non-inferior efficacy for the sake of improved safety. This strategy is reasonable if the outcome events used to measure efficacy have clinical importance similar to those for safety as, for instance, deaths and haemor-

non-inferiority leaves the product in a kind of limbo: its place in therapy is not established, although its place on the market is assured.

Enrolling patients in non-inferiority trials betrays their trust

We believe that non-inferiority studies have no ethical justification, since they do not offer any possible advantage to present and future patients, and they disregard patients' interests in favour of commercial ones. This situation betrays the agreement between patients and researchers set out in any fair informed consent form that presents randomised trials as the only ethical way to address clinical uncertainty. Non-inferiority trials claim minor advantages for the test drugs, but do not prove their efficacy compared with older products. Few patients would agree to participate if this message were clear in the informed consent form: as we said before, why should patients accept a treatment that, at best, is not worse, but could actually be less effective or less safe than available treatments?[15]

In conclusion, we believe that non-inferiority trials fail to meet the commitments of good clinical research: "Ask an important question, and answer it reliably".[16] Although a non-inferiority study reduces research and development costs and commercial risks thereafter, it asks no relevant clinical questions. Randomisation should not even be allowed in such trials, since it is unethical to leave to chance whether patients receive a treatment that is anticipated to provide no extra benefit, but could be less safe and less effective than existing treatment options.

With regard to the reliability of the methods and consequently of the results, the uncertainty surrounding alleged non-inferiority is hard to accept; however small an increase in relative risk, the increase in risk unavoidably implies an absolute excess of adverse events in the population. Sometimes the risk turns out to be significantly greater in the test treatment group, without

this difference necessarily disproving non-inferiority.[17] To expose patients to such risks in the trial and in real life, with no benefit in exchange, is clearly unethical. We hope that these arguments will foster a debate on the issue.

Conflict of interest statement

We declare that we have no conflict of interest.

Acknowledgments

We thank Iain Chalmers, Brian Godman, and Valter Torri for reviewing the manuscript and providing useful suggestions, and Judith Baggott for editorial assistance.

References

1 Bertele' V, Torri, W, Garattini S. Equivocal equivalence. http://www.marionegri.it/page.asp?idp=891441724 (accessed Oct 17, 2006).

2 US FDA's non-inferiority stance could slow sinusitis approvals. SCRIP—World Pharmaceutical News, Sept 18, 2006: 27.

3 Siegel JP. Equivalence and noninferiority trials. *Am Heart J* 2000; **139:** S166–70.

4 James Hung HM, Wang SJ, Tsong Y, Lawrence J, O'Neil RT. Some fundamental issues with non-inferiority testing in active controlled trials. *Stat Med* 2003; **22:** 213–25.

5 Hung HM, Wang SJ, O'Neill R. A regulatory perspective on choice of margin and statistical inference issue in non-inferiority trials. *Biom J* 2005; **47:** 28–36.

6 D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues—the encounters of academic consultants in statistics. *Stat Med* 2003; **22:** 169–86.

7 Snapinn SM. Alternatives for discounting in the analysis of noninferiority trials. *J Biopharm Stat* 2004; **14:** 263–73.

8 Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006; **145:** 62–69.

9 Tebbe U, Michels R, Adgey J, et al. Randomized, double-blind study comparing saruplase with streptokinase therapy in acute myocardial infarction: the COMPASS equivalence trial. *J Am Coll Cardiol* 1998; **31:** 487–93

10 Gruppo Italiano per lo Studio della Sopravvivenza nell'Infarto miocardico. GISSI-2: a factorial randomised trial of alteplase versus streptokinase and heparin versus no heparin among 12 490 patients with acute myocardial infarction. *Lancet* 1990; **336:** 65–71.

11 ISIS-3 (Third International Study of Infarct Survival) Collaborative Group. ISIS-3: a randomised comparison of streptokinase *vs* tissue plasminogen activator *vs* anistreplase and of aspirin plus heparin *vs* aspirin alone among 41 299 cases of suspected acute myocardial infarction. *Lancet* 1992; **339:** 753–70.

12 The GUSTO investigators. An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* 1993; **329:** 673–82.

13 Bertele' V, Torri V, Garattini S. Inconclusive messages from equivalence trials in thrombolysis. *Heart* 1999; **81:** 675–76.

14 Barbui C, Violante A, Garattini S. Does placebo help establish equivalence in trials of new antidepressants? *Eur Psychiatry* 2000; **15:** 1–6.

15 Garattini S, Bertele' V, Li Bassi L. How can research ethics committees protect patients better? *BMJ* 2003; **326:** 1199–201.

16 Yusuf S, Collins R, Peto R. Why do we need some large, simple randomized trials? *Stat Med* 1984; **3:** 409–22

17 Splawinski J, Kuzniar J. Clinical trials: active control *vs* placebo—what is ethical? *Sci Engineer Ethics* 2004; **10:** 73–79.

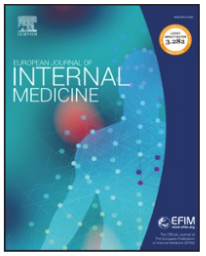
Un dibattito acceso

European Journal of Internal Medicine 67 (2019) e9–e10

Contents lists available at ScienceDirect

European Journal of Internal Medicine

journal homepage: www.elsevier.com/locate/ejim



Letter to the Editor

Looking for safety but overlooking efficacy: Non-inferiority trials of anti-diabetics

ARTICLE INFO

Keywords:
Diabetes
Anti-diabetic medicines
Hypoglycemic drugs
Glycemic control
Randomized clinical trials
Non-inferiority trials
Placebo

https://doi.org/10.1016/j.ejim.2019.07.006
Received 16 April 2019; Accepted 10 July 2019
Available online 16 July 2019

To the Editor-in-Chief,

Testing new interventions against placebo when effective treatments are already available is considered unethical [1] and dangerous for patients [2], so testing non-inferiority which aims at demonstrating that a new intervention is not unacceptably worse than standard clinical practice should be equally unethical [3,4]. The understanding is measure. In essence, the scant outcome of the trial was that dapagliflozin was not effective but at least did not harm patients by causing more heart failures. In any case, according to the non-inferiority hypothesis and the 30% margin selected, the investigators would have taken the results as good even if major cardiovascular events had occurred – say - in 120 every 1000 patients treated with dapagliflozin instead of the 94 given placebo. The question is: in exchange for what?

To the Editor-in-Chief,

Testing new interventions against placebo when effective treatments are already available is considered unethical [1] and dangerous for patients [2], so testing non-inferiority which aims at demonstrating that a new intervention is not unacceptably worse than standard clinical practice should be equally unethical [3,4]. The understanding is that part of the benefit can be given up in exchange for less important advantages – such as better tolerability, convenience, etc. – which are claimed though often not proved. Going one step further, i.e. exploring even whether a treatment is non-inferior to placebo – meaning not unacceptably worse than doing nothing - should be inconceivable from both the ethical and the clinical points of view. Nevertheless, the approach has gained ground, for example in the area of diabetes.

From an anti-diabetic treatment one would expect a reduction of events involving the heart, peripheral vessels and nerves, kidney or eyes. However, cardiovascular events may be associated with both diabetes and some of its treatments [5]. Nonetheless, it is still surprising that clinical trials are often aimed at showing that anti-diabetic medicines do not cause any increase in major cardiovascular events and even more, "that the therapy [does] not result in an unacceptable increase in cardiovascular risk". This is what a 2008 FDA guidance [6] stated and may have triggered a burst of so-called non-inferiority safety trials intended to show that a new anti-diabetic medicine is not (much) more harmful than placebo. We describe a couple of recent sample cases [7–9] which follow several others in the last decade.

The DECLARE–TIMI 58 trial [7] followed more than 17,000 patients with type 2 diabetes and high cardiovascular risk for about four years to test the hypothesis that dapagliflozin (a selective inhibitor of sodium-glucose co-transporter 2) does not raise the relative risk of major adverse cardiovascular events seen with placebo (cardiovascular death, myocardial infarction, or ischemic stroke) by more than 30%. Dapagliflozin did not affect the rate of cardiovascular events (8.8% in the dapagliflozin group and 9.4% in the placebo group; hazard ratio, 0.93; 95% CI 0.84-0.03); however, it did result in a lower rate of cardiovascular deaths or hospitalization for heart failure (4.9% vs. 5.8%; hazard ratio, 0.83; 95% CI 0.73-0.95), which was a co-primary outcome measure. In essence, the scant outcome of the trial was that dapagliflozin was not effective but at least did not harm patients by causing more heart failures. In any case, according to the non-inferiority hypothesis and the 30% margin selected, the investigators would have taken the results as good even if major cardiovascular events had occurred – say - in 120 every 1000 patients treated with dapagliflozin instead of the 94 given placebo. The question is: in exchange for what? Adequate glycemic control?

The CARMELINA trial [8] tested the same hypothesis in almost 7000 patients treated with linagliptin (a dipeptidyl peptidase-4 inhibitor) and followed for about two years. Luckily, here too the investigators did not see the 30% excess of events they were ready to consider acceptable: major cardiovascular events occurred in 12.4% and 12.1% patients respectively in the linagliptin and placebo groups (hazard ratio, 1.02; 95% CI 0.89-1.17). A secondary analysis [9] reported that linagliptin did not even affect the incidence of cardiovascular deaths or hospitalization for heart failure (11.6% compared to 12.1% in the placebo arm; hazard ratio, 0.94; 95% CI 0.82-1.08). In short, like dapaglifozin, linagliptin proved ineffective but harmless.

One first question relates to the trial hypothesis itself. Why should we accept any increase of the cardiovascular events we expect to reduce? This is even more surprising since that expectation is evidence-based: older anti-diabetics (metformin, sulfonylureas, insulin) and a few of the newer ones (empagliflozin, canaglifozin and liraglutide) have already been proved to prevent cardiovascular events, besides ensuring glycemic control. [10] The purpose of trials of new anti-diabetics should rather be to explore whether potentially innovative treatments further reduce cardiovascular risk besides controlling glycemia. However, if we assume that glycemic control is a valid surrogate of cardiovascular benefit for any other anti-hyperglycemic treatments – which is questioned [11] - we should expect a reduction of events, not an increase. Instead, regulators suggest – and clinical investigators adopt - an asymmetric evaluation of the efficacy and safety of anti-diabetics, the former relying on glycemic control, the latter on an acceptable increase in cardiovascular risk. This implies that a reduction of blood glucose is supposed by itself to counterbalance a possible higher incidence of cardiovascular events. Not to mention that the permitted

higher incidence means up to 30% more major cardiovascular events than might be expected with placebo.

Not even the concern about cardiovascular events such as heart failure [5,9] should make clinical investigators satisfied with any excess of major cardiovascular events, however limited, particularly in the absence of evidence of clinical efficacy. Patients would not be happy to have more heart failure episodes – however few - if, despite their perfect glycemia, their risk of myocardial infarction or cerebral stroke or even of dying was no lower. Patients want to be sure they will have fewer major cardiovascular events overall, preferably without or with only a few heart failure episodes. To address this expectation clinical trials should assess how many major cardiovascular events a given anti-diabetic avoids, without increasing the heart failure rate. This implies a superiority test addressing the efficacy of an anti-diabetic intervention in reducing a composite outcome which encompasses major cardiovascular events including heart failure, not just a questionable excess over placebo.

An explicit conclusion of current trials of anti-diabetics should recognize that, at best, those trials proved the experimental drugs did nothing, i.e. were no more effective than placebo, but hopefully with no worse toxicity - which is what non-inferiority versus placebo actually implies. It is disappointing that clinical investigators and ethics committees from so many clinical sites in different countries agreed on this scant outcome after involving so many patients for so long. Potential participants would certainly deny their consent if adequately informed about the scope of trials intended to show that new anti-diabetics are possibly not (very) harmful, though not necessarily beneficial.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

None

editing.

References

[1] World Medical Association. The Declaration of Helsinki. Ethical Principles for Medical Research Involving Human Subjects. 2013https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/, Accessed date: 15 April 2019.

[2] Garattini S, Bertele' V, Banzi R. Placebo? No thanks, it might be bad for me!. Eur J Clin Pharmacol 2013;69:711–4. https://doi.org/10.1007/s00228-012-1383-6.

[3] Garattini S, Bertele' V. Non-inferiority trials are unethical because they disregard patients' interests. Lancet 2007;370:1875–7. https://doi.org/10.1016/S0140-6736(07)61604-3.

[4] Bertele' V, Banzi R, Gluud C, Garattini S. EMA's reflection on placebo does not reflect patients' interests. Eur J Clin Pharmacol 2012;68:877–9. https://doi.org/10.1007/s00228-011-1175-4.

[5] Zannad F, Rossignol P. Dipeptidyl Peptidase-4 inhibitors and the risk of heart failure regression to the truth? Circulation 2019;139:362–5. https://doi.org/10.1161/CIRCULATIONAHA.118.038399.

[6] Food and Drug Administration. Guidance for Industry—Diabetes Mellitus Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/UCM071627.pdf, Accessed date: 15 April 2019.

[7] Wiviott SD, Raz I, Bonaca MP, et al. DECLARE–TIMI 58 investigators. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2019;380:347–57. https://doi.org/10.1056/NEJMoa1812389.

[8] Rosenstock J, Perkovic V, Johansen OE, et al. Effect of Linagliptin vs placebo on major cardiovascular events in adults with type 2 diabetes and high cardiovascular and renal risk. The CARMELINA randomized clinical trial. JAMA 2019;321:69–79. https://doi.org/10.1001/jama.2018.18269.

[9] McGuire DK, Alexander JH, Johansen OE, et al. CARMELINA Investigators. Linagliptin effects on heart failure and related outcomes in individuals with type 2 diabetes mellitus at high cardiovascular and renal risk in CARMELINA. Circulation 2019;139:351–61. https://doi.org/10.1161/CIRCULATIONAHA.118.038352.

[10] National Institute for Health and Care Excellence. Advice: Type 2 Diabetes Mellitus: Medicines Optimisation Priorities. https://www.nice.org.uk/advice/ktt12; 15 January, 2015, Accessed date: 15 April 2019.

[11] Rodriguez-Gutierrez R, McCoy RG. Measuring What Matters in Diabetes. JAMA 2019;321:1865–6. https://doi.org/10.1001/jama.2019.4310.

Vittorio Bertele'*, Rita Banzi, Chiara Gerardi, Silvio Garattini
Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy