Trieste, 24 settembre 2019 GIM – Riunione Annuale

#### La tecnica

# (intervento sulla metodologia degli studi)

Paolo Bruzzi - Genova

#### DISCLOSURES

Fees for

- Courses on Clinical Reasearch Methodology/ Seminars/Lectures: Bristol Meier Squibb, Roche, Novartis, Merck Serono, Lilly, Glaxo, Ipsen, Sanofi, EISAI, Tesaro

- Advisory Boards & Consultations: Amgen, Cellgene, Lilly

No Stock

## Contenuti

- 1. Breve panoramica degli sviluppi metodologici della ricerca clinica in Oncologia
- 2. Opportunita' per i nuovi studi GIM
- Proposte di studi metodologici nel gruppo GIM

## **Evolution of medical thinking**

**Pre-Scientific Medicine** 

'700-'900 Empirical Medicine

50's Randomized Clinical Trial

1992 Evidence Based Medicine

>2000 PRECISION MEDICINE?



# Pragmatic (Complexity addressed through statistics)

• Proof by falsification (HO)

Preclinical work + Clinical observations

## Clinical rationale

Preclinical work + Clinical observations





#### **INTERPRETATION**

Preclinical work + Clinical observations



#### <u>CLINICAL TRIALS</u> = PHASES

#### **INTERPRETATION**

## Rigid separation between phases

 Phase I -> MTD -> Dose increases in subsequent groups of patients

• Phase II -> Activity -> Uncontrolled Trial

• Phase III -> Efficacy -> RCT



#### Research Evidence - 1950-mid 80's

- Multiple studies with contrasting results
- Heterogeneous (and often poor) quality
- Small size
- Pubblication bias

## **Evolution of medical thinking**

Pre-Scientific Medicine



>2000 PRECISION MEDICINE?

## EBM greatest achievement

To focus the scientific discussions in Medicine on

a) Internal Validity

b) External Validity

of the available evidence

## Internal Validity

- = Absence of BIAS (=Accuracy)
- Results of trial = Truth +/- Chance

Tools

- Randomization
- Masking/Hard Endpoints
- Intention to treat
- Statistical Plan (data torturing)

## Preclinical work + Clinical observations **Clinical rationale** CAL TRI LS **INTERPRETATION**

## **External Validity**

### Generalizability?

#### How can we use the study results?

#### a) Demonstration of a GENERAL principle (Explanatory trials)



## NOTE

A peculiarity of the Randomized Clinical Trial:

- In the experiments in ALL the other empirical sciences, the researchers try to create the ideal conditions for demonstrating or rejecting the study hypothesis
- In the RCT, this is considered a limitation (and should be avoided ?!?)

# Application of the results of a trial to clinical practice

Requirements

- Unselected patients (Generic incl. criteria)
- Unselected centers (future users)
- Large Sample Size to detect small effects

Large RCT

Pooled analysis of multiple trials RCT -> EBM in Oncology (& Cardiology) – GOLDEN AGE

- <u>Rigid Treatment Protocols</u>
  - Drugs Doses Cycles
  - Changes only for toxicity or failure (progression)

#### Generic Selection Criteria

- Site (e.g. Stomach)
- Histology (ADK vs Lymphoma)
- Stage (early vs late)

RCT -> EBM in Oncology (& Cardiology) – GOLDEN AGE

Huge Trials on heterogeneous populations looking for moderate/minimal effects

#### Key words

- Protection from Bias (Randomization + ITT)
- Representativeness (patients, centers)

## Guiding Principle: Equity

- Public Health Perspective (es. vaccinations)
- Small effects on large populations = large benefits
- <u>Simple therapies</u>, that can and must be used in all patients by all doctors (thrombolisis, Tamoxifen)
- Unselected patients
- Generalizality = Applicability on a large scale

## **GIM** Trials

- Planned and conducted according to the EBM standards
  - Large Size
  - Unselected Study populations
  - Hard Endpoints
  - Adequate follow-up
  - Skilled Centers

Overall, good-to-optimal quality

How to use the results of trials in clinical decisions?

- Search of the evidence (Trials)
- Synthesis of the evidence
  - = <u>Systematic Revision</u>

## **Evidence Based Medicine** at its best

- Chance -> Large Numbers
- evidence C
- direct, randomised evidence
- Evidence Symesis -> Average effect -> Applicable to all patients and centers
- Methodological Quality >>> clinical quality

## ≈2000

The Evidence Based Medicine philosophy dominates the medical scene

- Large RCT's
- Reporting of trials (CONSORT)
- Systematic Revisions with meta-analyses
- Evidence-Based Guidelines (GRADE)
- Rigid Frequentist Perspective

## **Evidence Based Medicine**

 $\sim$ 

## Cookbook Medicine?

## Evidence Based Medicine >2000 The model creaks

- Do treatments have the same effect
  - In all centers?
  - In all patients?
- Rare Diseases?
- Rare Variants of frequent diseases?
- 1 drug -> 1 trial
- RCT's often ethically questionable (due to plausible extrapolations)

## Evidence Based Medicine >2000 The model creaks

- Do treatments have the same effect in all centers?
- Skills? E.g. Surgery?
- Experience? E.g. Radiology?
- Facilities & Instruments? E.g. Radiotherapy?
- Organizzation? E.g. Senology?
- Efficiency? E.g. Waiting Times

Increasing evidence of heterogeneity in quality of care which affects outcomes

## Evidence Based Medicine >2000 The model creaks

Do treatments have the same effect in all centers?

- Do treatments have the same effect on all patients?

 Average effect in average patient (included in the trials)
 How to use this result? 2 options:

a) Extrapolation to patients not incl. in trial

b) Interpolation to ALL patients incl. in trial

 Average effect in average patient (included in the trial)
 How to use this result? 2 options:

a) Extrapolation to patients not incl. in trial
FORBIDDEN! Include in trials future users!
(Old age, co-morbidities, other treatments, etc)

 Average effect in average patient (included in the trials)
 How to use this result? 2 options:

b) Interpolation to ALL patients incl. in trial

## Is the same treatment appropriate for ALL patients?

- Average effect in average patient Individual patient?
  - Age?
  - Comorbidities?
  - Genetics?
  - Stage?
  - Biology of disease?
  - Previous therapies?
  - etc.

Huge Variability in frequency and prognosis

• Average effect in average patient



### EBM & Subgroup analyses (S.A.)

- 1° phase: S.A. considered with suspicion
  - Many poorly conducted S.A.
  - Not necessary: average effect is fine for estimates of effectiveness (Cardiology, BC, Prevention)
- 2° phase: S.A. may provide important information
  - Properly conducted S.A. are unbiased but
    - Require many patients (Meta-analyses...)
    - From controlled studies (...of randomised trials)

## Subgroup analyses

- Methodology has become standardised
  - Careful planning to prevent selection and assessment biases
  - Test for interaction = H0: the (lack of) effect is the same in all subgroups
  - Multiplicity controlled (Exploratory vs confirmatory analyses, Corrections of p values)

No problems with large datasets (frequent diseases)
**Clinical Question:** Tools for clinical decision 'Opinion' Trial System.Revision+Meta-analysis +Subgroup Analyses **EB** Recommendations Where is the problem?

# **Conventional Phase III Trials**

- Inclusion Criteria -> Fixed for the entire trial
- Therapy -> Same Protocol for the entire trial
- Statistics -> Prevenine addetailed statistical plan
- Sample Size -> Predetermined Huge

### Same rigidity in other research areas?

• NO! Trial and Error!

• Biology, Engineering, Chemistry, Physics

The results of the experiment and new (external) knowledge are used to continuously update study design & methods Why flexibility was (is) not accepted in phase III trials ?

- Frequentist Statistics Control of *(obsession with)* false positive rate (alfa error)
- Difficult to replicate (randomised) trials
  - If positive, another RCT unethical
  - If negative, costs (and ethics)
- Inadequate knowledge of biology

### **INADEQUATE KNOWLEDGE OF BIOLOGY**

Conseguences

- Aspecific Drugs (Cytotoxics !?!)
- Unselected patients (Non-small cell lung c., Ascitis, lctus)
- Small expected effects -> Large trials
- No chance to learn from early results (Markers?)

Technical progresses – Scientific Discoveries

# New diagnostic and treatment tools



**New Effects!** 

### New patients

### <u>New patients</u>

- Prognostic subgroups
- -Rare diseases
- Patients with actionable molecular changes (targets)
- "Responders"

### <u>Rarity</u>

#### New Effects

- MTD? Dose-Response? (Phase I ?)
- Activity -> Objective Response? (Phase II?)
- Molecular Therapies (Cancer, Genetic & Inf. Diseases)
  - Stronger effects
  - Target-driven vs site+histology-driven effects
- Immunotherapy (Cancer, Reumathology)
  - Size/Type of benefit? (median OS?, HR, %alive)
  - Large for few vs Small for many
  - Treatment duration/rechallenge/associations
- Surgery
- Devices

- ...

Technical progresses – Scientific Discoveries

# New diagnostic and treatment tools



# New patients New New Effects! Methods?

# EBM in a brave new world!

#### Problem

#### **Methodological challenges**

- 1. More knowledge -> Ethical problems for randomization
- 2. Rarity

- 3. Need to improve trial efficiency and to handle the complexity
- 4. Medical Decision/Patient enpowerment/Costs

# EBM in a brave new world

#### Problem

- 1. More knowledge -> Ethical problems for randomization
- 2. Rarity

- Need to improve trial efficiency and to handle the complexity
- 4. Medical Decision/Patient enpowerment/Costs

#### Methodological challenges

- RCT or uncontrolled trials?
- Trial Design & size? **Surrogate Endpoints?** Bayesian Statistics?
- Flexibility -> Adaptive designs Surrogate Endpoints
- Shift from hypothesis testing to estimation of the effect **Surrogate endpoints**

### Modern methodological challenges

1. RCT's or uncontrolled <u>trials</u>?

### Available evidence and RCT

If a new drug ...

- With a well-identified molecular target
- present in different tumors
- shows, in a RCT, strong clinical effects in one of these tumors

...is it ALWAYS necessary and <u>ethically</u> <u>acceptable</u> to conduct a <u>Standard</u> <u>Randomised TRIAL</u> in each one of the other cancers?

# Imatinib

CML -> Large RCT GIST -> Large uncontrolled trial

#### Other rare indications -> Case Series

(dermatofibrosarcoma protuberans, plexiform neurofibromas, chordomas)

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

### Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation

Paul B. Chapman, M.D., Axel Hauschild, M.D., Caroline Robert, M.D., Ph.D., John B. Haanen, M.D., Paolo Ascierto, M.D., James Larkin, M.D., Reinhard Dummer, M.D., Claus Garbe, M.D., Alessandro Testori, M.D., Michele Maio, M.D., David Hogg, M.D., Paul Lorigan, M.D.,



#### **Overall Survival:** Anti BRAF +Anti-MEK vs Anti BRAF



COMBO450=en corafen ib 450 mg QD + bin imetin ib 45 mg BID; HR=h azard ratio; OS=overall survival; VEM=vemurafen ib 960 mg BID.



#ASCO18 Slides are the property of the author, permission required for reuse.

PRESENTED BY: Reinhard Dummer

Presented By Reinhard Dummer at 2018 ASCO Annual Meeting

#### Anti-B-RAH + Anti- MEK in BRAF- Positive tumors

Melanoma -> Large RCT's



#### JUNE 22, 2017

- Combination of dabrafenib and trametinib gained approval from the Food and Drug Administration (FDA) for the treatment of pts BRAF V600+ metastatic <u>NSCLC</u>.
- That approval was based on results from a three-cohort, multicenter, nonrandomized, open-label study of patients with stage IV NSCLC.
- In this phase 2 study, 36 untreated pts and 57 pretreated pts were assigned to ..... Investigator-assessed objective response rate was the primary endpoint
- At a median follow-up of nine months, the ORR was 61.1 percent in the treatment-naïve group, and 68 percent of patients did not show progression

#### Anti-B-RAF + Anti- MEK in BRAF- Positive tumors

Melanoma -> Large RCT's



# NSCLC BRAF+ -> FDA approved (1-arm trial?)

#### BRAF+ Pt. With c. in another site?

### New paths to drug use

Large RCT in a frequent cancer with the target - Proof of **principle – Toxicity** 

Uncontrolled (but formal) trial(s) in other (rare) cancers with the target

Off label use in individual cases with the target

# New paths to drug use Large RCT in a frequent cancer with the rget - Proof of principle – Toxicity Uncontrolled (but iormal) trial(s) in other Canert with the target Off label use in individuation the target

# Single Arm Trials

a) Without a control group

- b) With a (or more) control group(s)
  - Historical
  - Concurrent

ORIGINAL ARTICLE

#### PD-1 Blockade with Pembrolizumab in Advanced Merkel-Cell Carcinoma





...phase 2, single group, Simon's two-stage..." Response Rate: 14/25 (56%) (Historical RR: 50-60%)

Median PFS: 9 months (*Historical: 3 months*)

Kaufmann HL Lancet Oncology 2016



# Single Arm Trials

- a) Without a control group
  Absolute benefit = "success" rate
  (e.g. % Responders, long-term survivors, etc.)
- Breakthrough drugs
- Otherwise unreliable (large variations)

# Single Arm Trials

a) Without a control group

- b) With a control group
  - Historical
  - Concurrent

# Study Design

- Uncontrolled trial with non-randomized Controls
  - Well Kown Biases
  - Sufficient if outstanding benefit
  - Necessary if control group unethical

#### Careful and transparent methodology Need of methodological guidelines/research

### Modern methodological challenges

1. RCT's or uncontrolled <u>trials</u>?

### 2. <u>Rarity</u>

(Not yet a problem in Breast cancer, but in the future?)

# Rarity?

1. Increasing number of potential predictive factors (rare subgroups)

2. Rare diseases

3. Rare variants of frequent diseases amenable to targeted treatment

# Rare Diseases/Subgroups

- Dominant problem in modern oncology (but also in many other diseases)
- Scotomised by methodologists, statisticians, and EBM-adepts
- Guidelines do provide recommendations, but with uniformly low levels of evidence
- Health agencies are very inconsistent in their policies
- <u>Clinical researchers do not know what to do</u>

## Basic question: How to design a trial in a rare disease?

"Statistics is the same whether the disease is frequent or rare" J. Bogaerts, EORTC



# "Small" trials

#### Solutions?

#### <u>Surrogate endpoints</u>

- Scanty empirical evidence and ill-focused statistical research
- Few SE validated, none for rare diseases
- <u>Trial results + Indirect evidence -> Bayesian</u> <u>approaches</u>

#### Why Large phase III trials?

• Inadequate knowlege of cancer biology

• <u>Frequentist statistical philosophy</u> <u>Control of false positive rate (alfa error)</u>

# **Empirical Approach**


### **Conventional Statistical Reasoning**

To reject H0, only evidence collected within one or more trials aimed at falsifying it can be used -> LARGE SAMPLE SIZE

### <u>No use of</u>

- External evidence

- Evidence in favor of...

# Prior Information: X and Y are BRAF+ -DrugA = Anti BRAF



### Interpretation of the two trials

CONVENTIONAL Tumor X: P = 0.0001Tumor Y : P = 0.066

Efficacy of treatment A Proven in X NOT PROVEN in Y

### Interpretation of the two trials

### BAYESIAN

(Posterior) Probability that treatment A lowers mortality "*significantly*" (HR<0.8)

in tumor X: >90% in tumor Y: >90%

Less direct evidence is needed to convince us if it agrees whith the available indirect evidence

### **Disadvantages of Bayesian Statistics**

- It is (felt as)
  - Subjective
  - Arbitrary
  - Amenable to manipulations (pharma companies?)

# **Current Situation**

- Bayesian Statistics is largely used in early (Phase I &II) trials
- It NEVER formally used in the design and interpretation of efficacy (phase III) trials
- <u>A subconscius, OPAQUE, unconfessed,</u> <u>Bayesian reasoning is behind many</u>
  - Decisions by regulatory agencies
  - <u>Clinical Recommendations</u>
  - <u>Clinical decisions</u>

### JUNE 22, 2017

- Combination of dabrafenib and trametinib gained approval from the Food and Drug Administration (FDA) for the treatment of pts BRAF V600+ metastatic <u>NSCLC</u>.
- That approval was based on results from a three-cohort, multicenter, nonrandomized, open-label study of patients with stage IV NSCLC.
- In this phase 2 study, 36 untreated pts and 57 pretreated pts were assigned to ..... Investigator assessed objective response rate was the primary endpoint<sub>26</sub>
- At a median follow-up of nine months, the ORR was 61.1 percent in the treatment-naïve group, and 68 percent of patients did not show progression

# Advantages of Bayesian Statistics

- Reflects human reasoning ("common sense")
- It is focused on estimates of effect
- Provides a conceptual framework for medical decision making
- <u>It is transparent</u>
- It is flexible and promotes flexibility

### Challenges

1. Increasing number of potential predictive factors (rare subgroups)

2. Rare diseases

3. Flexibility -> Adaptive designs

## **Conventional Phase III Trials**

- Inclusion Criteria -> Seldom modified
- Treatment -> Same Protocol throughout the trial
  Statistic - Statistical Plan predefined in detail
- Sample Size -> Predefined HUGE

# **Conventional Methodology**

Rigid separation between each phase of development of a drug and the next one

- Fase I -> MTD
- Fase II -> Activity -> Instrumental Endpoints
- Fase III -> Efficacy -> Clinical Endpoints

# **Conventional Methodology**

Within each phase/trial

- One primary aim
- One primary endpoint, fixed in advance
- Fixed selection criteria
- Fixed treatment protocol
- Statistical plan pre-specifies sample size, number, time and type (e.g. subgroup) analyses

### Convertionative/Teithodology

Within each phase/trial

- One primary aim
- One primary endpoint, fixed in advance
- Fixed selection criteria
- Fixed treatment protocol
- Statistical plan pre-specifies sample size, number, time and type (e.g. subgroup) analyses

# Adaptive design clinical trial

FDA's Definition:

- prospectively planned
- opportunity for modification
- of one or more specified aspects of the study design and hypotheses
- <u>based on analysis of data (usually interim</u> <u>data) from subjects in the study</u>"



### Seamless Phase II-III trials

- Minimize overall trial time (no stop between phases)
- Flexibility to study crucial aspects
  - -dose finding
  - subroup selection
- All enrolled patients are considered in the final analyses

## Adaptive trials Seamless Phase II-III trials

Statistical Nightmare (accommodate multiple analyses)

• Organizational nightmare (timely flow of samples, tests and data)

Regulatoy nightmare (which trial is being approved by EC's?)

## Challenges

1. Increasing number of potential predictive factors (rare subgroups)

2. Rare diseases

3. Flexibility -> Adaptive designs

4. Shift from hypothesis testing to estimation of the effect

### Aim of a phase III trial

To demonstrate

-> The presence of a benefit (any): P<0.05

#### the efficacy of the experimental drug

To evaluate

-> The size of the effect (if any) Summary indicator Why the shift from hypothesis testing to estimation of the effect?

- Clinical decisions
  - Efficacy vs Toxicity
  - Type of effect (large for few vs small for many)
  - Comparison with other therapies
- Public Health decisions
  - Costs, priorities
- Quite often, at the start of the trial the null hypothesis is not very plausible

## New Concept: Estimand

• Definition:

The estimand is the quantity of interest whose true value you want to know

Example:

- Estimand: Treatment Effect on Overall Survival

### 1 Estimand, Several Estimators

- Estimand: Treatment effect on OS
- •Estimators:
  - -Difference in median OS
  - -Hazard Ratio
  - -Probability of 5yrs Os
  - Probability of 5 yrs OS in <u>Compliers</u>
  - Probability of 5 yrs OS in <u>Responders</u>
- BIASED BUT CLINICALLY RELEVANT

– etc

#### STATISTICS IN MEDICINE

#### **Per-Protocol Analyses of Pragmatic Trials**

Miguel A. Hernán, M.D., Dr.P.H., and James M. Robins, M.D.

The result of an intention-to-treat analysis ... may not be directly relevant for guiding decisions in clinical settings.

Health care professionals and patients would like to have an effect measure that, unlike the intention-totreat effect, is not influenced by the degree of adherence.

### Commentary

"...the protocols of pragmatic trials would benefit from explicit definition of the <u>per</u> <u>protocol effect,</u> ....

... to prioritize the **patient-centeredness** of the research."

# B) Proposte di ricerca - aree

- 1. Personalizzazione delle terapie
- 2. Studi non-controllati
- 3. Endpoint surrogati
- 4. Estimandi

## 1. Personalizzazione delle terapie

- Studio di letteratura e sui dati GIM su fattori che predicono l'efficacia di chemio e ormonoterapia adiuvante
- Costruzione di un algoritmo di supporto decisionale che utilizzi le moderne conoscenze sui fattori prognostico/predittivi

### Stage III colon cancer . Probability of being RF at 5 yrs



■ Observed 3-year DFS

### Stage III colon cancer Probability of being RF at 5 yrs



## 2. Studi non controllati

Quesito 1: Cosa sarebbe successo se in alcuni studi GIM randomizzati si fossero utilizzati controlli storici? – Studio ricerca di gruppi di controllo storici per confronto con gruppi randomizzati

Quesito 2: E' possible costruire una "reference cohort" per gli studi Gim nei quali non fosse eticamente giustificata la randomizzazione?

# 3. Endpoint surrogati

 Utilizzo di un nuovo approccio per la validazione di RFS e PFS nel c. Mammario sfruttando i dati GIM

# 4. Estimando

- Valutazione del guadagno prognostico associato alla risposta obiettiva nel MBC
- Valutazione della distribuzione dei benefici dei trattamenti nel EBC e nel MBC

# C) Disegno di nuovi studi

- Studi adattativi
- Studi di fase I-II-III
- Studi non-controllati
- Studi Bayesiani
- Studi di associazioni/sequenze

#### A TRIAL OF MITHRAMYCIN IN THE TREATMENT OF ADVANCED MALIGNANT DISEASE

#### I. A. SEWELL\* AND H. ELLIS

From the Professorial Surgical Unit, Westminster Hospital, London, S.W.1

Received for publication February 3, 1966

With this information in mind, we have conducted a clinical trial of Mithramycin on a small group (26) of patients with advanced malignant disease. Our dosage regime was based on the most recent reports on the use of the drug (Brown at the time of writing (January, 1966), 4 of the 26 patients undergoing trial have shown a definite quantitative remission.

Objective Response? Uncontrolled trials? Small trials? Selected Patients? No statistical plan!

# Molto Moderno!

