Milan, July 30, 2015

The principles of frequentist and Bayesian medical statistics

Paolo Bruzzi
Clinical Epidemiology Unit
National Cancer Research Institute
Genova - Italy

Statistical Mantra

- A study **must** have an adequate size

  – To warrant an adequate "power" to the study

  (i.e. to reduce the risk of a false negative result
  false negative: an effective treatment is not recognised)

  – To obtain precise estimates of the effects of the experimental therapy

Conventional Statistical Rules

- *A study **must** have an adequate size*
- Required Size, based on:
  - Significance level (usually 5%)
  - Power (usually 80-90%)
  - **Minimal clinically worthwhile difference**

Sample Size in cancer clinical trials

In trials in early disease, cumulative mortality from 10% to 70%: **500-5000** pts

In trials in advanced disease, cumulative mortality from 50% to 90%: 3**00-1000** pts

Conventional Statistical Rules

- *A study **must** have an adequate size*
- *Required Size: Usually Hundreds/Thousands of patients*
- In many rare cancer conditions: **NOT POSSIBLE**
  - Incidence
  - Age
  - Molecular variant
  - Stage

Statistical Mantra

- A study **must** have an adequate size

Unjustified Implication

- If an adequate size cannot be attained, (RARE CANCERS) no methodological ties

Small size → **Poor quality**

Poor Quality?

- (Study protocol)
- (Classified as Phase II trials)
- No Randomised controls
- Opaque selection of cases
- Primary endpoint: Objective response
- No statistical plan

First point to stress

The organization of a trial of **small** size requires **more** care in

- Protocol preparation
- **Study design/methodology**
- **Statistical design**
- Addressing Clinical Organizational issues

…than a standard size trial

Methodological issues

- *Statistical Power*

- Study Design
- Bias in evaluating outcome (double blind)
- Endpoint

VALIDITY!

Study Design

- Uncontrolled trial/Historical Controls
  - Well Kown Biases
  - Sufficient if outstanding benefit
  - Necessary if control group unethical

Careful and transparent methodology

Need of guidelines/research

Study Design

- *Uncontrolled trial/Historical controls*
- Randomised Controls

WHY NOT?

RCT's in rare cancers

Pro's

- **VALIDITY**
- **CREDIBILITY**

Con's

- Moderate loss in power
- Often no standard (untreated control group?)
  - Ethics?
  - Acceptance?

Trials in Rare Cancers

If, despite International cooperation/Prolonged accrual

- it is possible to assemble (in a reasonable time) only **a limited number of patients**,
- and the efficacy of a new treatment is not outstanding …

What can be done?

Recent developments

- Bayesian Statistics
- Surrogate endpoints
- New types of systematic reviews
- Adaptive trials

Common beliefs

Frequentist probability

- Objective
- «Hard»
- Useful to analyse experiments
- Scientific

Bayesian Probability

- Subjective
- «Soft»
- Inappropriate to analyse experiments
- Not scientific

WRONG

Differences between Conventional (Frequentist) and Bayesian Statistics

- Meaning of probability
- Use of prior evidence

Frequentist Probability

Probability **of an observation**

(given a hypothesis)

Bayesian Probability

Probability **that a hypothesis is true**

(given observation and prior knowledge)

Frequentist Probability

Probability **of the observed difference** (if the experimental therapy does not work)

Bayesian Probability

Probability **that the experimental therapy works/doesn't work** (given observed difference and prior knowledge)

Frequentist Probability

Definitions and implications

Probability

Definitions (Wikipedia, from Merriam Webster)

…the measure of the likeliness that an **event will** occur

Probability = *measure of the likeliness that an event will occur?*

- Probability that next number from a roulette will be red =18/37 ≈ 50%
- Probability that a man has a cancer in his prostate ≈ 25/100 =25%
- Probability that my home team (Genoa) had won last game = <1/million ≈ 0

Probability = *measure of the likeliness that an event will occur?*

- Probability that next number will be red
- Next Number: red = event
- Likeliness =18/37 ≈ 50%

Theory:

Red Numbers / Total Numbers = Proportion

Experiment

Red Numbers/Plays = **Frequency**

Probability = *measure of the likeliness that an event will occur?*

Probability of an event = proportion

Estimation = frequency = events/plays

Probability = *measure of the likeliness that an event will occur?*

- *Probability that next number from a roulette will be red =18/37 ≈ 50%*
- Probability that a man has a cancer in his prostate ≈ 25/100 =25%
  - Proportion: 25 prost.c./100 adult men
  - Estimation: Frequency of examined men in whom I do find a prostate c.:

  1 every 4 men (25%)

Probability = *measure of the likeliness that an event will occur?*

- *Probability that next number from a roulette will be red =18/37 ≈ 50%*
- *Probability that a man has a cancer in his prostate ≈ 25/100 =25%*
- Probability that my home football team (Genoa) had won last game= <1/million ≈ 0

Proportion? Frequency?

Probability = *measure of the likeliness that an event will occur?*

- If my home football team (Genoa) could play again the last game a million times it would not win once

- Theoretical Proportion
- Theoretical Frequency

= Hypothesis

Probability = *measure of the likeliness that an event will occur?*

- Probability that next number from a roulette will be red =18/37 ≈ 50%

IF (Hypothesis)...

…the roulette works fine

(TRUE Frequency = 50%)

Probability = *measure of the likeliness that an event will occur?*

- *Probability that next number from a roulette will be red =18/37 ≈ 50%*
- Probability that a man has a cancer in his prostate ≈ 25/100 =25%

IF (Hypothesis)…the true prevalence of prostate cancer in Western adult men is 25%

Probability = *measure of the likeliness that an event occurs?*

- *Probability that next number from a roulette will be red =18/37 ≈ 50%*
- *Probability that man has a cancer in his prostate ≈ 25/100 =25%*
- Probability that my home football team (Genoa) had won last game= <1/million ≈ 0

IF (Hypothesis)… Genoa does not change his players

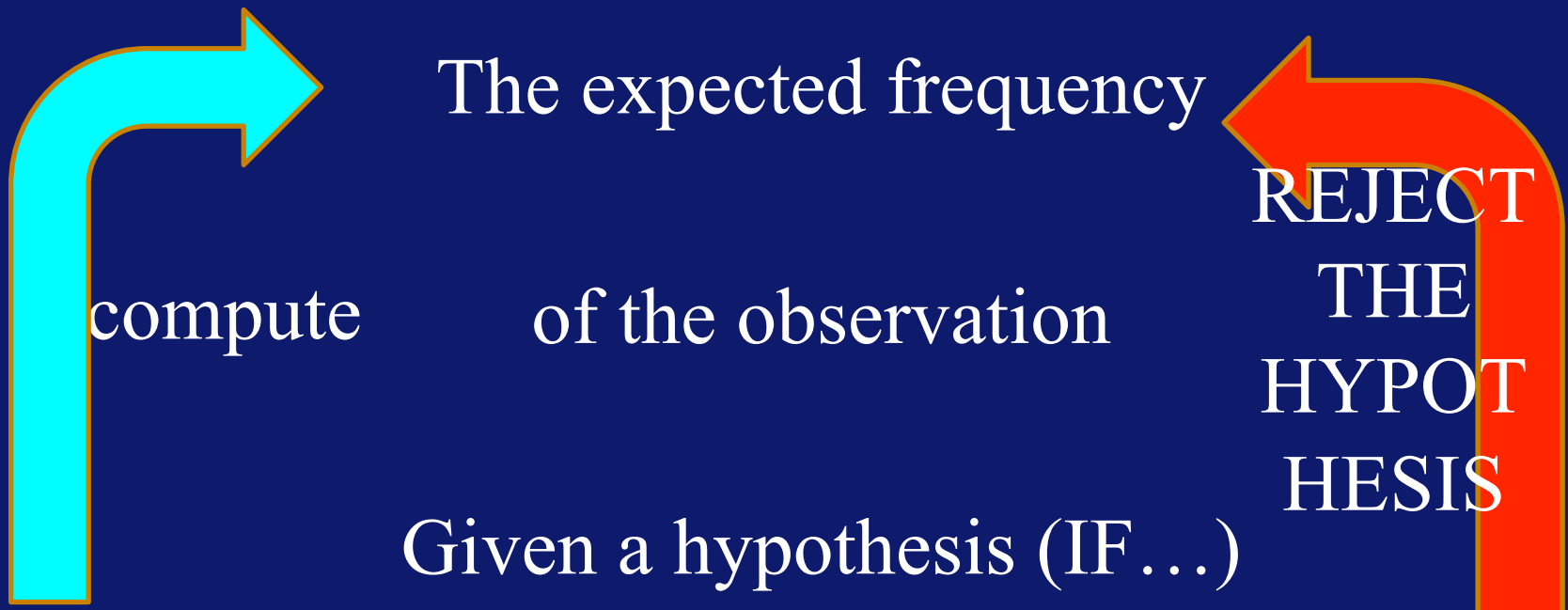
FREQUENTIST PROBABILITY

The expected frequency

of the observation

given a hypothesis (IF…)

FREQUENTIST TEST OF HYPOTHESIS



Other examples of frequentist probabilities

- If the roulette works fine (reds = 50%):
  - Probability that next 3 numbers are red =12%
  - Probability that 1/5 numbers are red=19%
  - Probability that 1/10 numbers are red = 1%

TESTS of HYPOTHESIS

- Does the roulette work fine (reds = 50%)?
  - Next 3 numbers are all red P=12% ?
  - 1/5 numbers are red P=19% ?
  - 1/10 numbers are red P = 1% (2% if 2-sided)

Too rare: I reject the hypothesis that the roulette works fine

This is what medical statistics is (was) all about!

1. Set a hypothesis (null hypothesis, H_0)
2. Do the study
3. Compute the probability (frequency, P) of the observed results if H_0 is true
4. If p is large, (usually >5%) do not reject
5. If p is small (usually <5 %) reject H_0

Conventional Statistical Reasoning in Medicine

1. Starting Hypothesis = Null Hypothesis, H_0):

New treatment = Standard (no treat.)

2. To demonstrate that new treatmt is effective

H_0 must be rejected

3. To reject H_0

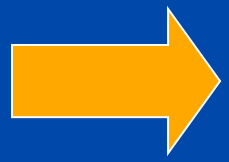
Only the results of the trial can be used

-> Trials of large size

Scientific Method in Medicine

Standard Therapy

Laboratory or trials in other diseases

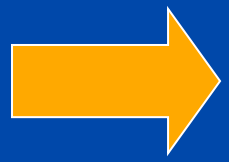


New Standard?

Scientific Method in Medicine

Standard Therapy

Laboratory or trials in other diseases

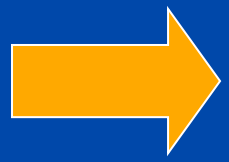


Clinical TRIAL

Scientific Method in Medicine

Standard Therapy

Laboratory
or cl. studies



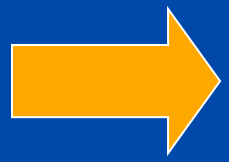
H0: Exp. No better than standard

Clinical TRIAL

Scientific Method in Medicine

Standard Therapy

Laboratory
or cl. studies



H0?

Clinical TRIAL

H1: Exp better

New Standard Th.

Scientific Method in Medicine

Standard Therapy

Laboratory
or cl. studies

H0?

Clinical TRIAL

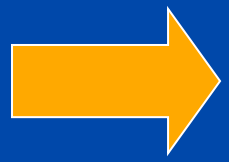
H1?

New Standard Th.

Scientific Method in Medicine

Standard Therapy

Laboratory
or cl. studies



H0?

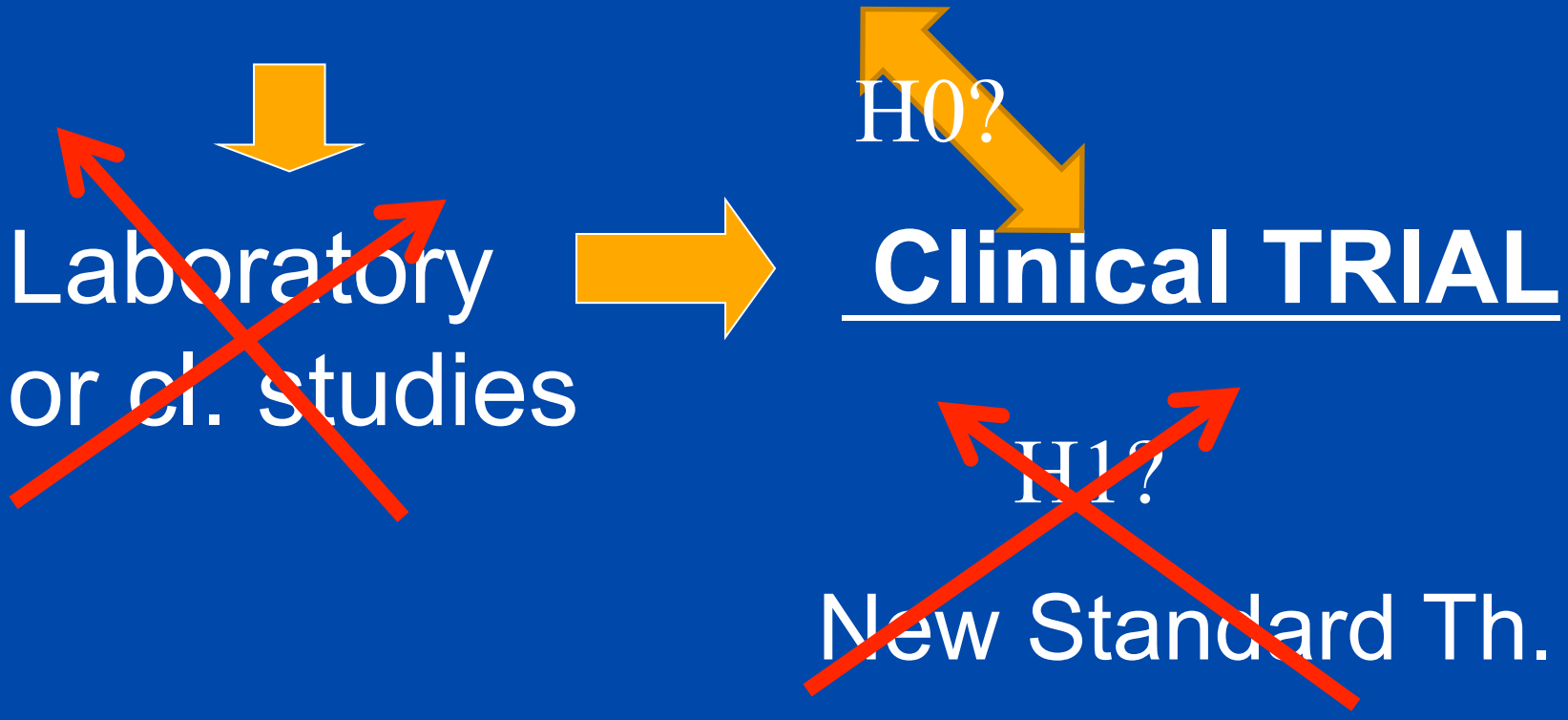
Clinical TRIAL

H1?

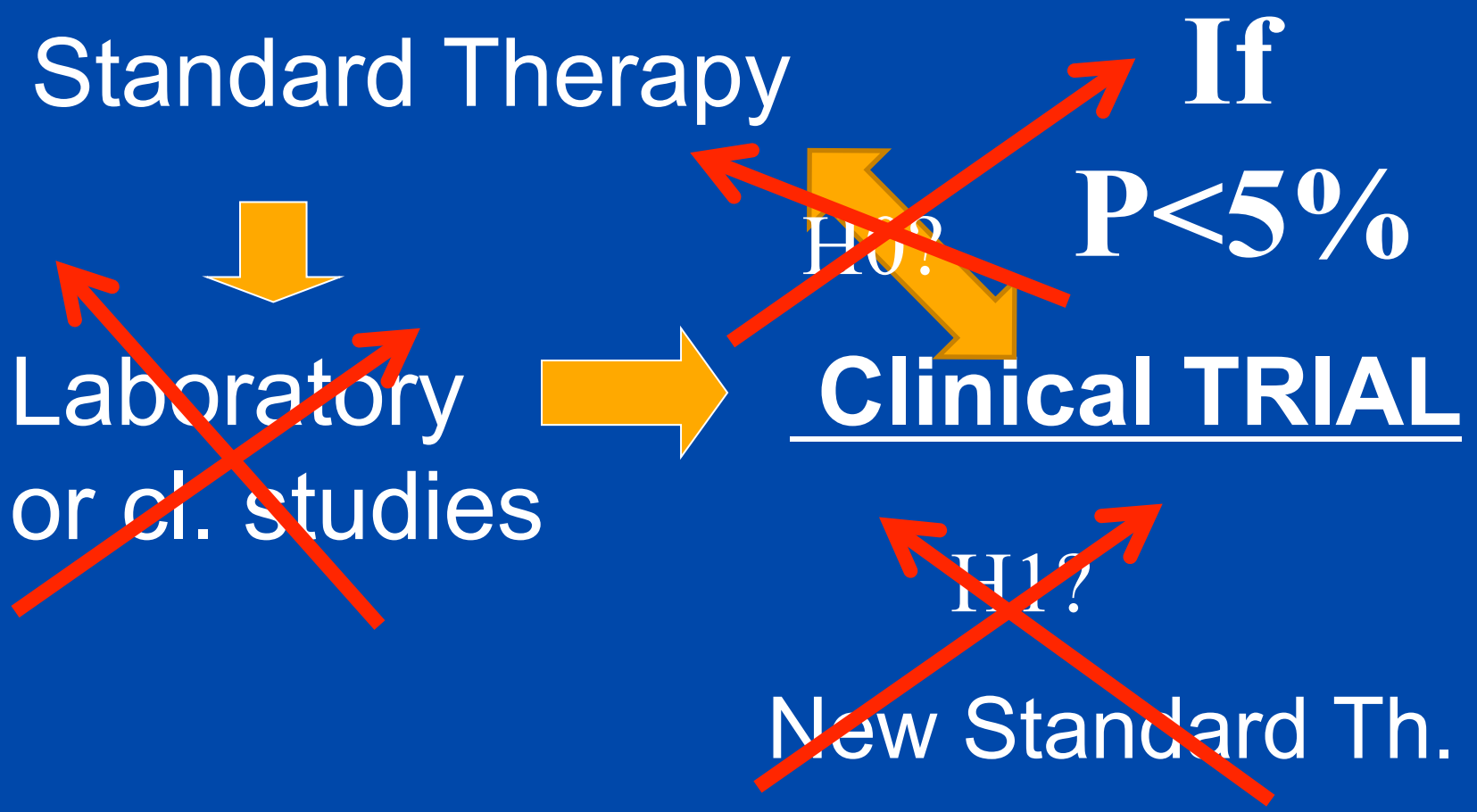
New Standard Th.

Scientific Method in Medicine

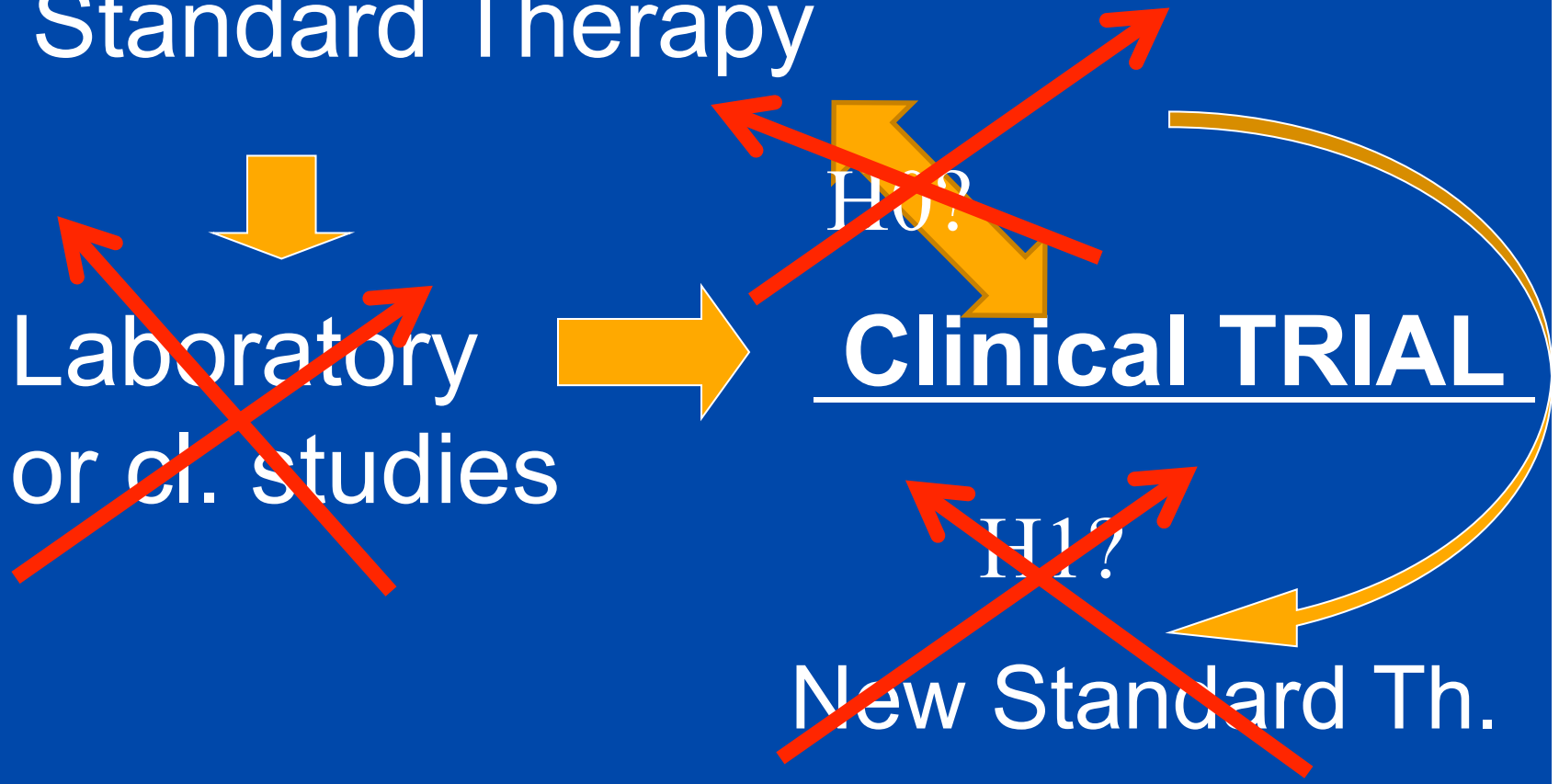
Standard Therapy



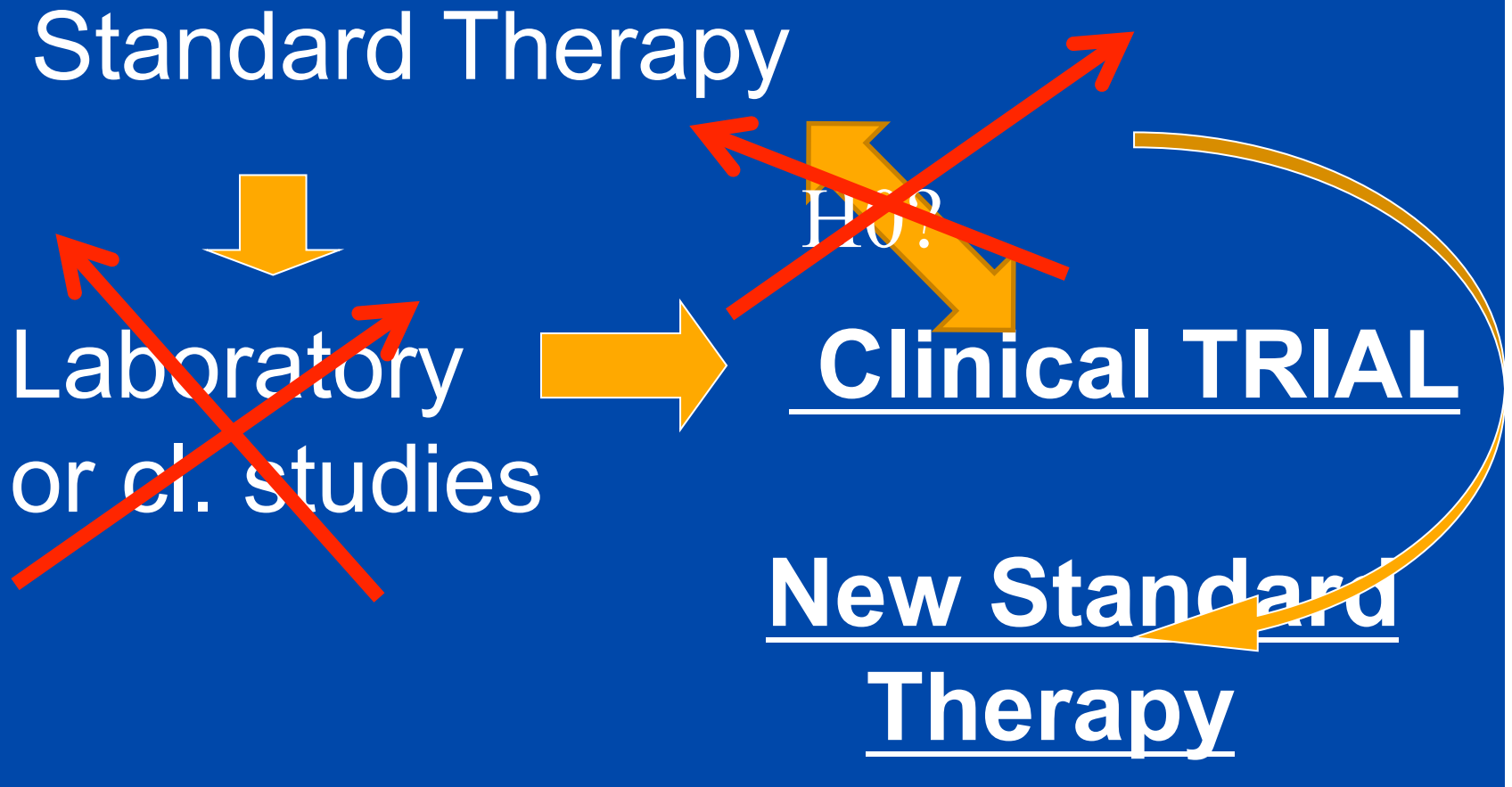
Scientific Method in Medicine



Scientific Method in Medicine



Scientific Method in Medicine



Advancement of knowledge in Medicine (conventional statistics)

- Dominant theory is true (=standard therapy is better) until sufficient evidence becomes available **against** it
- To this purpose, only evidence collected within one or more trials aimed at falsifying it can be used
- **No use of**
  - **External evidence**
  - **Evidence in favor of…**

How to interpret the results of a study

- Internal Validity
- Biological Plausibility
- Internal Coherence
- External Consistency
  - Direct
  - Indirect

Null Hypothesis (H0): the new drug is identical to the standard (if no standard, completely ineffective)

- Biological Rationale
- Preclinical studies (disease models)
- Evidence of activity in Phase II
- Evidence of activity within same trial
- Efficacy in other diseases with similar.
- Efficacy in other stages same disease

The 2 Reasons why large numbers of patients are needed in clinical trials

- **Outstanding efficacy seldom observed**

- **Any knowledge outside the primary analysis of the clinical trial is ignored in the design and analysis of the trial**

Hypothetical Example

- As a statistician, I'm asked to design 2 separate trials in the same rare disease, squamous gastric cancer (no standard treat.)
- Study A:

Experimental therapy: Radiochemotherapy

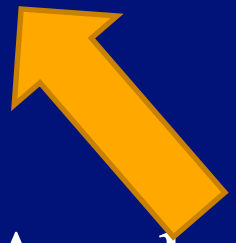
  - Effective in squamous cancers of other sites
  - Phase II trial; Response Rate 60%

- Study B
- Experimental therapy: Intercessory prayer

Squamous gastric cancer

Planning a trial of

RT+CTX



Analysing its results

(p value)

Squamous gastric cancer

Squamous gastric cancer

Planning a trial of

RT+CTX Intercessory prayer

Analysing its results

(p value)

Squamous gastric cancer

Results of the 2 trials

RT+CTX

20% reduction in deaths P=0.15

Intercessory prayer

20% reduction in deaths =0.15_

Treatment x next patient with SGC?

Frequentist P

Probability **of the observed difference** if either therapy does not work = 15%

Bayesian Probability

Probability **that either therapy works a lot/ works a little/does not work ?**

Is it the same for the two treatments?

Differences between Conventional and Bayesian Approaches

- *Meaning of probability*
- **Use of prior evidence**

Conventional P

Probability of the observed difference (**if the experimental therapy does not work**)

Bayesian Probability

Probability that the experimental therapy works/doesn't work (given observed difference and **prior knowledge**)

Conventional (frequentist) statistical reasoning

Experimental evidence

Bayesian statistical reasoning

Experimental evidence + Previous Knowledge

Example

		Mortality
Tumor X	**Nil vs A**	**15% vs 10%**
N=2000		**P = 0.0001**

H0 Rejected: A is effective in X

Example

		Mortality
Tumor X	*Nil vs A*	*15% vs 10%*
N=2000		*P = 0.0001*
Tumor Y	**Nil vs A**	**15% vs 7.5%**
N= 240		**P=0.066**

H0 not rejected: A not shown effective in y

Prior Information:
X and Y are BRAF+

		Mortality
Tumor X	*Nil vs A*	*15% vs 10%*
N=2000		*P = 0.0001*
Tumor Y	Nil vs A	15% vs 7.5%
N= 240		**P=0.066**

Prior Information:

X and Y are BRAF+

A = Anti BRAF

Mortality

Tumor X *Nil vs A* *15% vs 10%*

N=2000 *P = 0.0001*

Tumor Y Nil vs A 15% vs 7.5%

N= 240 **P=0.066**

INTERPRETATION?

Interpretation of the two trials

CONVENTIONAL

Tumor X: P = 0.0001

Tumor Y : P= 0.066

Efficacy of treatment A

- proven in X
- undemonstrated in Y

Interpretation of the two trials

CONVENTIONAL

Efficacy of treatment A is proven in X, undemonstrated in Y

BAYESIAN

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

in tumor X: 90%

in tumor Y: 90%

Disadvantages of Bayesian Statistics

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

Conceptual Advantages of Bayesian Statistics

- Reflects human reasoning ("common sense")
- It is focused on estimates of effect
- Provides a conceptual framework for medical decision making
- **IT IS TRANSPARENT**

Practical Advantages of Bayesian Statistics in rare tumors

1. No need to set the sample size in advance

Adaptive designs: enrol patients until sufficient evidence in favour or against efficacy

2. When strong a priori evidence is available

and trial results are in agreement with it

Smaller sample size is necessary – You can stop any time

Prior evidence in Bayesian statistics

Note: The difference between Bayesian and conventional statistics decreases with increasing strength of the empirical evidence

Rare Tumors!

Prior evidence in Bayesian statistics

- **Needed in order to compute posterior probability**

Prior evidence in Bayesian statistics

- *Needed in order to compute posterior probability*
- **It must be transformed into a probability distribution (mean, median, standard deviation, percentiles, etc)**

Prior evidence in Bayesian statistics

- *Needed in order to compute posterior probability*
- *It must be transformed into a probability distribution*
- **Based on**
  - Objective information
  - Subjective beliefs
  - Both

Prior evidence in Bayesian statistics

No special way to elicit/obtain prior information

No special way to summarize information

- Meta-analytic techniques

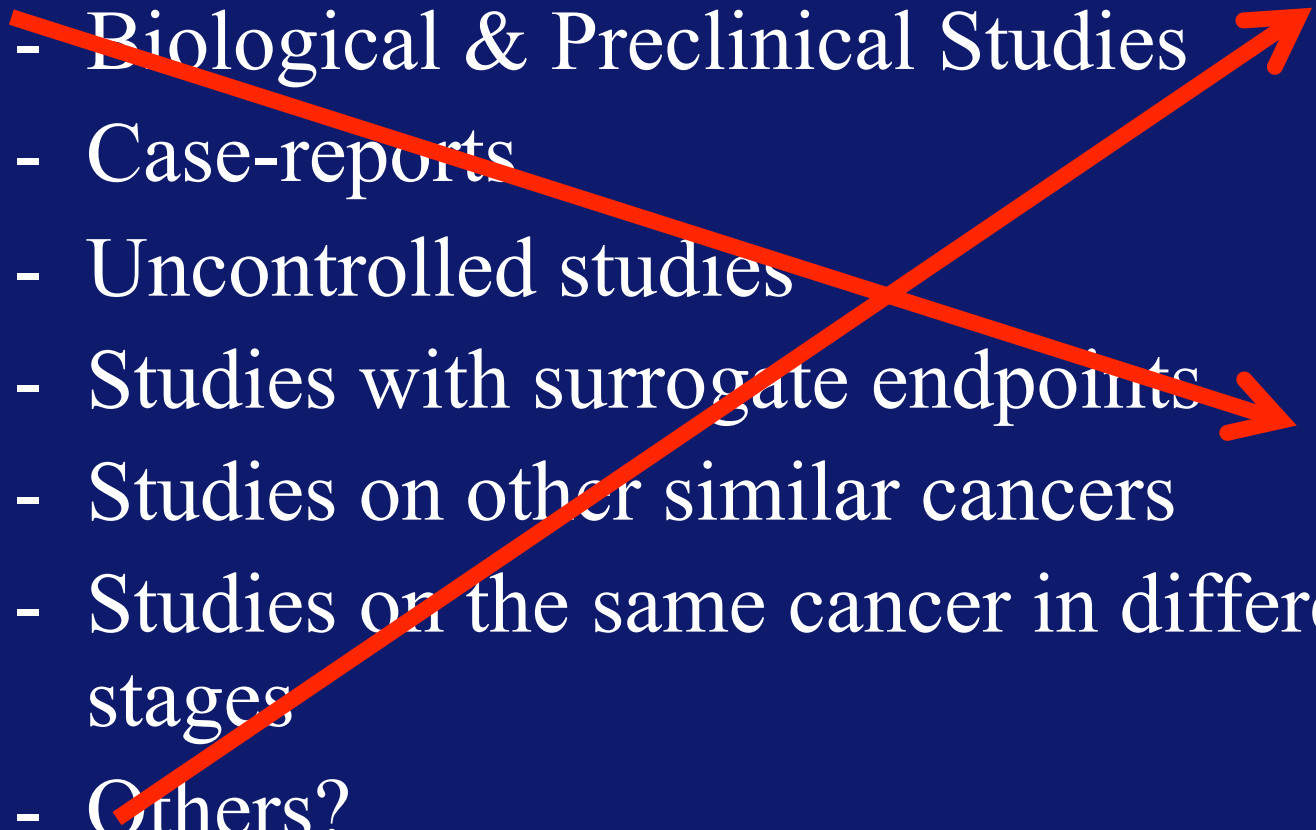
Frequentist - Bayesian

Sources of prior evidence

- *Randomised Trials*
- Biological & Preclinical Studies
- Case-reports
- Uncontrolled studies
- Studies with surrogate endpoints
- Studies on other similar cancers
- Studies on the same cancer in different stages
- Others?

Meta-analyses in frequent tumors

- ***Randomised Trials***
- Biological & Preclinical Studies
- Case-reports
- Uncontrolled studies
- Studies with surrogate endpoints
- Studies on other similar cancers
- Studies on the same cancer in different stages
- Others?



Meta-analyses in frequent tumors

- ***Randomised Trials***

Weighted exclusively based on their size (and quality)

Rare Tumors

- ~~*Randomised Trials*~~
- Biological & Preclinical Studies
- Case-reports
- Uncontrolled studies
- Studies with surrogate endpoints
- Studies on other similar cancers
- Studies on the same cancer in different stages
- Others?

Meta-analyses in rare tumors

Need to use information from studies

- <100% valid
- <100% pertinent to the question of interest
  - Different cancers
  - Different treatments
  - Different endpoints

Prior evidence and clinical trials

Need to develop and validate new (meta-analytic) approaches to summarize prior information in rare tumors

Requirements

- Explicit
- Quantitative
- Reproducible

Efficacy trials in rare tumors

- Uncontrolled efficacy (phase III) trials of high quality
- Randomized activity (Phase II) trials followed by uncontrolled efficacy trials (with historical controls)
- RCT's with surrogate endpoints
- **Adaptive, Bayesian, activity/efficacy RCT's based on unconventional Systematic Reviews**

RCT's in rare cancers

- Loss of power (50% less patients in exp treatment)

Available patients : 100

Cure Rate in controls: 40%

RCT (50 x2):80% power for delta:30% (to 70%)

Uncontr.trial 80% power for delta:21% (to 61%)

RCT's in rare cancers

- Loss of power /Precision

(50% less patients in exp treatment)

Available patients : 100

RCT (50 x2):	Difference +/- 15%
Uncontrolled tr.	Difference +/- 11%
(Histor. Controls)	

Differences between the present and the proposed approach

- Present :
  - Rational but informal integration of the available knowledge
- Proposed (Bayesian)
  - Formal, explicit and quantitative integration of the available knowledge
    - Verifiable quantitative methods
    - Sensitivity analyses
    - Focus on summary effect estimates