BIOSTATISTICAL METHODS

FOR TRANSLATIONAL & CLINICAL RESEARCH

Phase 0 Trials:

EARLY-PHASE CLINICAL TRIALS

Drug development is the process of finding and producing therapeutically useful pharmaceuticals and turning them into **effective and safe medicines**. It is a complex process starting with screening chemicals to identify a lead compound, going through lots of works in toxicology, pharmacodynamics, and pharmacokinetics, and phases of clinical trials.

A successfully completed development and testing program results in lots of information about appropriate doses and dosing intervals, and about likely effects and side effects of the treatment. It is a process carried out by "sponsors" (mostly pharmaceutical companies) and is ultimately judged by "regulators" (e.g. FDA of the United States).

There is no aspect of drug development and testing without participation and contributions from biostatisticians. Statisticians and biostatisticians are also becoming more active in the shaping of the pharmaceutical projects.

There are statisticians even on “the other side of the table”; for many years FDA has employed statisticians and biostatisticians to assist in its review process. At medical centers, biostatisticians participate in protocol designs as well as protocol reviews.

Steps to New Drug Discovery

Get idea for drug target

↓

Develop a bioassay

↓

Screen chemical compounds in assay

↓

Establish effective and toxic amounts, In Vitro & In Vivo

↓

File for approval as an Investigational New Drug (IND)

After an IND is applied, it's the starting of the clinical phase (research with human subjects).

PHASES OF CLINICAL TRIALS

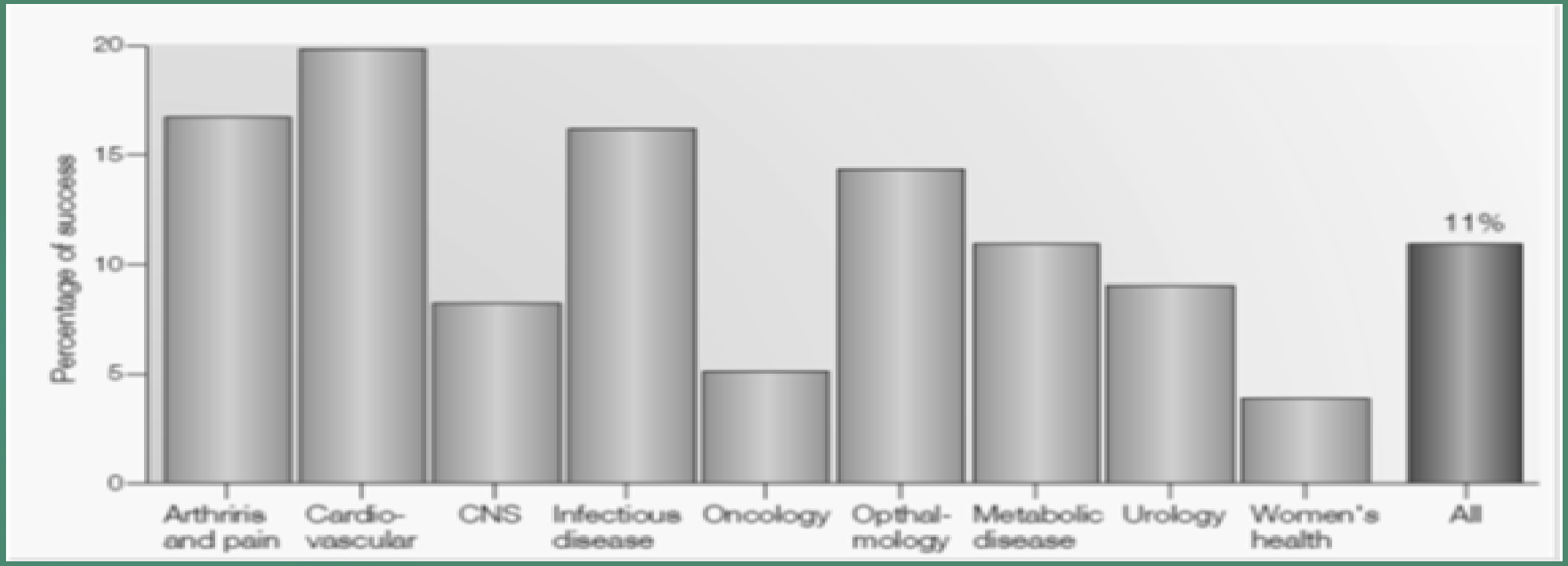
- **Phase I:** First human trial to focus on safety
- **Phase II:** Small trial to evaluate efficacy
- **Phase III:** Large controlled trial to demonstrate efficacy prior to FDA approval
- **Phase IV:** Optional, post-regulatory approval, to provide the medicine's more comprehensive safety and efficacy profile

Phase I and II clinical trials present special difficulties because they involve the use of agents whose spectrum of toxicity and likelihood of benefits are poorly understood/defined. There were "pre-clinical" studies – e.g. In Vivo & In Vitro experiments and bioassays – but the subjects were animals (In Vivo) or human tissues (In Vitro). And inferences across species are never easy, nor precise.

In recent years, with participation from statisticians, lots of efforts have been focused on the design of Phase I clinical trials; e.g. Bayesian Designs. Progresses have been made in the analysis of Phase II trials – including the use of Meta Analysis.

The process, from Phase I to Phase III, has been in practice for decades; then it was discovered – more recently - that we seemed to get into some kind of troubles, lots of resources have been wasted.

Success Rates from First-in-Man Drug Developments

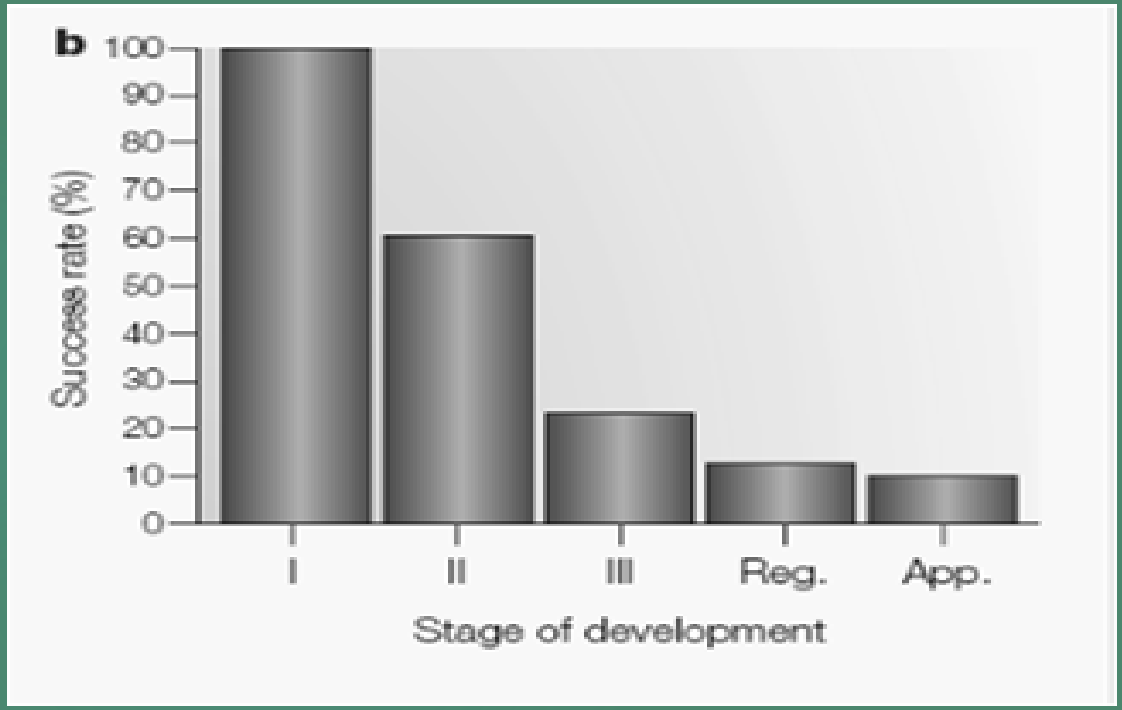


Data from 10 biggest drug companies from 1991-2000
(*Kola And Landis; Nature Reviews Drug Discovery 2004*)

Currently, only 10%-15% of investigational new drug (IND) applications to the Food and Drug Administration (FDA) result in clinically approved agents, and in Oncology it is only 5%. This is a very serious problem, since the development of a new agent is a lengthy and expensive process and many of these agents fail relatively late in that process after lots of money have been invested in Phase I and Phase II trials.

Most drugs fail in late stages of development…particularly in Oncology

Rates of success for compounds entering first in man that progress to subsequent phase



- 70% of oncology drugs that enter Phase 2 fail to enter Phase 3
- 59% of oncology drugs that enter Phase 3 fail
- Risk of failure may be higher for novel targeted agents

Kola & Landis;
Nature Reviews Drug Discovery 2004

Then a new type of Clinical Trials was started, Clinical Trials Phase 0 (zero), which saved a large number of those failures – especially those with unfavorable PD and PK characteristics.

PHARMACOLOGY BASICS

The action of drugs on the human body is called pharmacodynamics (PD) and what the body does with the drug is called pharmacokinetics (PK)

Phase 0 Clinical Trials have been evolved and further improved the last several years with active participation and promotion from statisticians of the National Cancer Institute (NCI). But they might need more contributions from biostatisticians to strengthen the method and the design.

What is a Phase 0 trial?

- First-In-Human trial:
  - Limited number of subjects (5-10); the fewer the better.
  - Low, supposedly non-toxic doses
    - Limited duration of dosing (≈ ≤7 days)
    - One course
    - No therapeutic (or diagnostic) intent
  - Conducted prior to traditional Phase 1 dose escalation,
  - Can be initiated with a less extensive pre-clinical data
- Also referred to as:
  - Pre-phase 1 trial
  - Pilot study
  - Exploratory Investigational New Drug (IND) study

Goal of a Phase 0 Trial

Generate data to increase chance of success of subsequent development of the pharmaceutical agent by eliminating an agent very early in clinical development because of poor pharmacodynamics (PD) or pharmacokinetics (PK) properties or poor bioavailability

"***Bioavailability***" is the ease with which a substance or any nutrient can make its way from the drug you take or the food you eat into your body. When a substance or nutrient is highly "*bioavailable*," it can be digested and absorbed a high percentage of the time and in a dependable way.

Phase 0 Trial Outcomes

- Determine whether a mechanism of action defined in pre-clinical models can be observed in humans (binds to or inhibits its alleged target).
- Provide human PK/PD data for an agent prior to definitive Phase 1-2 testing
- Refine biomarker assay using human tumor tissue and/or surrogate tissue

… Phase 0 Trial Outcomes

- Evaluate human PD and/or PK (e.g., bioavailability) of two or more analogs directed at the same target and possessing practically the same properties *in vitro* and in animal models, helping to select the most promising candidate for further development.
- Evaluate in humans an agent's bio-distribution, binding characteristics and target effects using "micro-dosing" and a variety of novel imaging technologies

Microdosing (or micro-dosing) is a technique for studying the behavior of drugs in humans through the administration of doses so low ("sub-therapeutic") they are unlikely to produce whole-body effects, but high enough to allow the cellular response to be studied.

STUDY DESIGNS

Step #1: Measuring Outcome

To identify a primary endpoint/outcome; for example, a PD endpoint. This endpoint will be measure both before and after the treatment by the agent – Aim is to measure "change"; values are often analyzed on the log scale. To better reflect the biological effect of the agent, it is more ideal to use tumor tissue assay (from biopsy); however, blood test could be used as a surrogate. If blood (or peripheral blood mononuclear cells, PBMCs) assay is used, multiple pre-treatment and post-treatment values could be obtained.

Step #2: Estimating Pre-treatment Variability

If possible, intra-patient variability is more preferred but only when we have multiple values before treatment (e.g. Blood assays). First calculating the variance within each patient, then weighted averaging across patients (weight is (n-1)). For tumor tissue assays, the pre-treatment variance is the inter-patient variance because each provides only one value. The out come of this step is a Standard Deviation (SD) representing pre-treatment variability.

Step #3: Defining a Response

To measure the "treatment effect", say for the tumor PD assay, we use the difference between the pre-treatment and the post-treatment values (Pre-Post, both on the log scale). In order to qualify as a "Response", this effect must satisfy a "biologic criterion" and a "statistical criterion". Investigators have to determined if the change is biologically or clinically important; statisticians have to determine if the change is real.

The Biologic Criterion depends upon characteristic of the biologic target of the agent.
To meet the Statistical Criterion, the change (i.e. "treatment effect") must be statistically significant at the 5% or 10% level (generally one-sided because the anticipated effect is in one direction). That is the PD effect (Pre-Post) divided by the pre-treatment SD must exceed the t-percentile at the corresponding degree of freedom (the one used to obtain SD)

For example, for some dose, we may enroll only three patients; the inter-patient variance has 2 degrees of freedom. A response is reached when the PD effect (Change = Pre-Post) divided by the pre-treatment SD exceeds 1.8 for 10% significance or 2.3 for 5% significance (one-sided, one-sample t-test).

Step #4: Choices by Investigator

(1) The number of dose level, usually 2-3

(2) At each dose level, investigator may set a threshold for the number of patients that must demonstrate a PD response, in order for the dose level to be judged as yielding a promising biologic effect. This threshold is often set at "2".

(3) The investigator may set a target "PD Response Rate" (for the agent, at that dose, across population).

Goal of the Design:

The probability to detect the target PD response rate at a dose level is the statistical power of the design. At each level, depending on the target PD response rate, the design could be one-stage or two-stage.

The goal of the design is to reach a pre-determined statistical power, conventionally 90%, using as few as patients possible.

If investigators have firm idea on a target rate and simply looking for a dose, they can focus on that same rate in all three doses. If so, chosen rate is conventionally 50%. Otherwise, one can test 3 doses at 3 different target rates to explore a Dose-Response relationship.

Example 1: High Dose
Target Rate: 80%
Design: 3 patients
Verification:

$$\begin{aligned}\text{Power} &= \Pr(2 \text{ or } 3 \text{ responses}) \\ &= 1 - \Pr(0 \text{ response}) - \Pr(1 \text{ response}) \\ &= 1 - (.2)^3 - (3)(.8)(.2)^2 \\ &= 1 - .008 - .09 = .896, \text{ almost } 90\%\end{aligned}$$

It works

Example 2: Medium Dose
Target Rate: 60%
Design A: try 3 patients
Verification:

$$\begin{aligned}\text{Power} &= \Pr(2 \text{ or } 3 \text{ responses}) \\ &= 1 - \Pr(0 \text{ response}) - \Pr(1 \text{ response}) \\ &= 1 - (.4)^3 - (3)(.6)(.4)^2 \\ &= 1 - .064 - .288 = .648 \text{ or } 64.8\%\end{aligned}$$

It does not work, under powered

Same Medium Dose

Target Rate: 60%

Design B: raise to 4 patients

Verification:

$$\begin{aligned}\text{Power} &= \Pr(2 \text{ responses or more}) \\ &= 1 - \Pr(0 \text{ response}) - \Pr(1 \text{ response}) \\ &= 1 - (.4)^4 - (4)(.6)(.4)^3 \\ &= 1 - .026 - .154 = .818 \text{ or } 81.8\%\end{aligned}$$

It still does not work. The next step would be raising to 5 patients, or trying a two-stage design.

Same medium Dose Medium

Target Rate: 60%

Design C: 2 stages: 3 patients, then enroll two more if there is exactly one response from the first 3.

Verification:

$$\begin{aligned}
\text{Power} &= \Pr(2 \text{ or } 3 \text{ in stagge I}) + \Pr(1 \text{ in stage I}).\Pr(\text{ at least 1 in stage II}) \\
&= [1 - \Pr(0 \text{ in stage I}) - \Pr(1 \text{ in stage I})] + \Pr(1 \text{ in stage I}).[1 - \Pr(0 \text{ in stage II})] \\
&= [1 - (.4)^3 - (3)(.6)(.4)^2] + [(3)(.6)(.4)^2][1 - (.4)^2] \\
&= (1 - .064 - .288) + (.288)(1 - .16) \\
&= .648 + (.288)(.84) \\
&= .890 \text{ or } 89\%
\end{aligned}$$

It works; you need 3 or 5 patients, not necessarily 4 or 5 patients as in one-stage design

Why the 3+2 two-stage design is more preferred than one-stage designs?

Its "Expected Size" is 3 + (.288)(2) = 3.57which is less than 5 patients need in the one-stage design – even less than 4 patients in the one-stage design which is still under powered. In this calculation, (.288) is the probability that we are required to enroll and test the 2 patients in stage 2.

Example 3: Low Dose
Target Rate: 40%
Design: 2 stages: 5 patients, then enroll three more if there is exactly one response from the first 5.
Verification:

$$
\begin{aligned}
\text{Power} &= \Pr(\geq 2 \text{ in stagge I}) + \Pr(1 \text{ in stage I}).\Pr(\text{ at least 1 in stage II}) \\
&= [1 - \Pr(0 \text{ in stage I}) - \Pr(1 \text{ in stage I})] + \Pr(1 \text{ in stage I}).[1 - \Pr(0 \text{ in stage II})] \\
&= [1 - (.6)^5 - (5)(.4)(.6)^4] + [(5)(.4)(.6)^4][1 - (.6)^3] \\
&= (1 - .078 - .259) + (.259)(1 - .216) \\
&= .663 + (.259)(.784) \\
&= .867 \text{ or } 86.7\%
\end{aligned}
$$

It works; almost there, we could raise the power a little more but it does not seem necessary (because we are almost there).

Summary:

Just find a design to get 90% or almost 90% power to detect success (2 responses) at each predetermined target rate. This design needs as few patients as possible.

Ethical Concerns:

- No therapeutic intent or chance of benefit
- Possibility of pre-treatment and post-treatment tissue biopsies

(These can impede accrual)

Good Features:

Informed consent

Low risk (micro-dosing)

Not excluding from other trials, such as the next Phase I.

In one-stage designs, we enroll and test the whole group simultaneously because
(1) We need to determine pre-treatment variability, especially using tumor tissue assays.
(2) It is not necessary to wait for the result from one subject before starting the next one.
(3) In addition, it would take too long, unnecessarily, to complete the trial if we test one subject at a time and waiting for the result.

That maybe different for the second stage of two-stage designs.

A SMALL CHANGE TO TWO-STAGE DESIGNS

If the determination of “response status” in the second stage of a (conventional) two-stage design does not take a long time, we would not need the second subject if the first subject is a response. We would not need the third subject unless the first two were both non-responses. In other words, instead of, for example, a 5+3 design (as in Example 3), we could design it as “5 + up to 3”.

Of course, it would take more time to complete the trial but not much longer because the size of the second stage is rather small, no more than 3, and we could stop after the first subject if it was a success. In addition, pre-treatment variability has been determined from subjects in the first stage.

Conventional Two-Stage 5+3:

Target Rate: 40%

Design: 2 stages: 5 patients, then enroll three more if there is exactly one response from the first 5.

Verification:

$$\begin{aligned}
\text{Power} &= \Pr(\geq 2 \text{ in stagge I}) + \Pr(1 \text{ in stage I}).\Pr(\text{ at least 1 in stage II}) \\
&= [1 - \Pr(0 \text{ in stage I}) - \Pr(1 \text{ in stage I})] + \Pr(1 \text{ in stage I}).[1 - \Pr(0 \text{ in stage II})] \\
&= [1 - (.6)^5 - (5)(.4)(.6)^4] + [(5)(.4)(.6)^4][1 - (.6)^3] \\
&= (1 - .078 - .259) + (.259)(1 - .216) \\
&= .663 + (.259)(.784) \\
&= .867 \text{ or } 86.7\%
\end{aligned}$$

It works; almost there, the expected size is:

5+(.259)(3) = 5.777

If we employ the newly proposed design, “5 + up to 3”:
The power is unchanged; expected sample size is a smaller.

Power = (.663) + (.259)(.4) + (.259)(.6)(.4) + (.259)(.6)(.6)(.4) = .868
Expected Size = 5 (.259)(.6) + (.259)(.6)(.6) = 5.508 vs. 5.577

OPTIMAL TWO-STAGE DESIGNS

The solution much more simple than Simon's two-stage design for Phase II clinical trials because of the narrower range stage sizes. One can, for example, consider 3-5 subjects in stage I and 2-3 subjects in stage II – for a total of no more than 15 combinations. Then computing and ruling out under-powered combinations; among the combinations with adequate power, the optimal design is the design with smallest expected sample size.

Suggested Exercises:

#1. Consider a one-stage design with a target rate of 60%. Calculate the power for the design with 5 subjects.

#2. Consider a two-stage design with a target rate of 40%; (a) Calculate the power for the design with 4 subjects enrolled in stage 1 and 3 more subjects enrolled in stage 2; (b) Calculate the expected size.