

Clinical Trial Design

Steven Hirschfeld., MD PhD
 Captain, United States Public Health Service
 Associate Director for Clinical Research
Eunice Kennedy Shriver

National Institute for Child Health and Human Development



National Institutes of Health
 Building Discovery for Health



Outline

- Why clinical studies
- Rationale
- Types
- Design considerations in general
- Design considerations for type of study
- Adaptive designs
- Factorial designs
- Integrating an approach

2

Definitions

- **Clinical research** is research that directly involves a particular person or group of people
- A **clinical trial** is one type of clinical research that follows a pre-defined plan or protocol. By taking part in clinical trials, participants can not only play a more active role in their own health care, but they can also access new treatments and help others by contributing to medical research.

3

NIH Definition

- **Patient-oriented research:** This type of research involves a particular person or group of people or uses materials from humans. This research can include:
 - Studies of mechanisms of human disease
 - Studies of therapies or interventions for disease
 - *Clinical trials* (see [About clinical trials](#) for more details)
 - Studies to develop new technology related to disease
- **Epidemiological and behavioral studies:** These types of studies examine the distribution of disease, the factors that affect health, and how people make health-related decisions.
- **Outcomes and health services research:** These studies seek to identify the most effective and most efficient interventions, treatments, and services

4

HHS Definition

- **Clinical Trial:** A controlled study involving human subjects that is designed to prospectively evaluate the safety and effectiveness of new drugs or devices or of behavioral interventions (IRB Handbook)
- **OHRP:** Research is a systematic investigation, including research development, testing, and evaluation, that is designed to develop or contribute to generalizable knowledge.

5

Food and Drug Administration

- Applicable law is Food Drug and Cosmetic Act as Amended Title 21 - Chapter 9 - Subchapter V - Part A - Section 355 Subsection 1(A)
 - full reports of investigations which have been made to show whether or not such drug is safe for use and whether such drug is effective in use;
- Applicable regulation is Code of Federal Regulations Title 21 Part 314 Section 314.126
 - The purpose of conducting clinical investigations of a drug is to distinguish the effect of a drug from other influences, such as spontaneous change in the course of the disease, placebo effect, or biased observation.
 - Reports of adequate and well-controlled investigations provide the primary basis for determining whether there is "substantial evidence" to support the claims of effectiveness for new drugs.

6

Rationale for Clinical Trials

- No other way to obtain data
 - Non-clinical data may not apply or cannot be extrapolated
 - Other study results are in different populations, under different conditions or lack informative outcomes
 - Existing patient records may be unstructured, incomplete, inconsistent or inaccessible
- Need to minimize bias and uncertainty

7

Product Development

- For marketing authorization, sufficient data required to support a claim
- Nature of proposed claim drives data requirements
- Products without prior experience require
 - Initial safety data →
 - Proof of concept →
 - Efficacy data
- Trial design varies by stage of product development and therefore intent

8

Initial Safety Data

- What is a biologically and clinically effective exposure?
- What are the likely risks?
- Are the risks acceptable?

9

Proof of concept

- What population will have their lives improved?
- Is the new product as acceptable as the currently available options?
- What is the intended outcome?
- Does the new product produce the intended outcome predictably in the target population?
- Are the risks associated with exposure to the product acceptable in the target population?

10

Efficacy

- Does the product predictably produce intended outcomes with acceptable risk in a target population?
- How does the product compare to alternatives?
- Are the findings sufficiently robust to be generalizable?

11

Outcome Measures

- Outcome measures can be categorized as to how they are expressed
 - Discrete- only particular values are acceptable including yes or no
 - Continuous- any value within a specified range is acceptable
- Outcome measure category will vary according to study goals and available assessment tools

12

Ethical Considerations

- Ethics- minimization or avoidance of harm to others
- To be ethical, a clinical trial must be informative
- For an intervention trial to be informative, a specific question must be asked
- To answer a question, the trial must measure an outcome in an unbiased, valid and interpretable manner

13

Ethical Basis of Clinical Research

Criteria based on Emmanuel, Wendler and Grady (2000)

- Value
- Scientific validity
- Fair subject selection
- Favorable risk-benefit ratio
- Independent review
- Informed consent
- Respect

14

Considerations in trial design

- Generalizability
- Define population
- Define outcome(s)
 - Standard hypothesis- treatment has no effect
- Arms for comparison purposes
- Biological mechanisms
- Define analysis- function of study goals
- Interpretation of outcomes
- Linking data
- Clinicaltrials.gov and other registries
- Obligation to report results

15

Trial Design-Categories

Type	Comment
Exploratory	A study to examine biological or clinical activity but not designed to establish efficacy.
Efficacy Superiority	The test treatment is better than a comparator. The confidence intervals around the measurement for the test treatment and for the comparator should not overlap. For example if the standard treatment shows that median survival for a population is 22 months and the confidence intervals are plus or minus 3 months, then the test treatment must have a lower confidence interval that is greater than 25 months (22 plus 3) to be considered superior. Results of 29 months plus or minus 3 months or 28 months plus or minus 2 months would qualify.
Efficacy Non-inferiority	The test treatment is not worse than a comparator. Exact equivalence is difficult to prove requiring large study populations and precise measures. The usual approach is to consider that a treatment is not worse than an accepted treatment by direct comparison with the understanding that: <ol style="list-style-type: none"> 1. The effect of the accepted treatment is measurable, reproducible, and meaningful. 2. An acceptable difference between the accepted treatment and the new treatment is defined prior to beginning the study and is smaller than the total effect of the accepted treatment. For example if the accepted treatment increases median survival by 6 months and the acceptable difference is 1 month, then the new treatment in direct comparison to the accepted treatment must not differ by more than 1 month and the accepted treatment must have a median survival that is consistent with previous results.

16

Trial Design Elements-Time Frame

Trial Design Elements	Comment
Prospective	A protocol is written and the study is performed following the approval of the protocol. Opportunity to collect all data that is required to test the hypothesis.
Retrospective	A protocol is written to systematically analyze historical data. Often unable to locate all the data required to test the hypothesis.

17

Trial Design Elements- Arms

Arms	
Single	A single series of patients
Multiple	Treatment arms are compared concurrently to one another
Crossover	Patients can change from one treatment arm to another based on predetermined criteria. If the critical evaluation occurs after the crossover it may be difficult to interpret due to factors such as the sequence of therapy having an effect or one therapy having a delayed effect.

18

Trial Design Elements- Controls

Controls	
Historical	<p>Comparison is made with either a specific study that is considered to be an appropriate match for the current therapy or with a valid meta-analysis of a series of previous studies. Major problems are:</p> <ol style="list-style-type: none"> 1.Changes in medical practice (secular effect) that affect results over time 2. Differences in eligibility criteria in different protocols 3. Differences in assessment in different studies 4. Differences in analysis in different studies

19

Trial Design Elements-Placebo Controls

Placebo	
Supportive Care	If no active therapy exists for the patient population then it may be ethical to provide supportive care plus a placebo versus supportive care plus the test therapy. Examples may include testing a symptom benefit therapy.
Add on	All arms receive the same standard therapy with the control arm receiving in addition a placebo and the experimental arm receiving in addition the test therapy.

20

Trial Design Elements- Active Controls

Active	
Standard therapy	Direct comparison between a standard therapy and the test therapy
Add on	All arms receive the same standard therapy with the experimental arm receiving in addition the test therapy.
Withdrawal	All arms receive the same therapy until a predetermined time when one arm has the test therapy withdrawn. The endpoint of interest is usually the appearance of an event that would be prevented if the test therapy were still present. This type of design is often used in measuring the effect of lowering blood pressure.

21

Trial Design Elements- Dose Comparison

Dose Comparison	
Dose Comparison	Different doses of the same drug are compared. Generally a trend in benefit or response that follows the exposure to the drug is considered evidence of activity- the greater the exposure the greater the response

22

Trial Analyses

Analyses	
Single	Study analysis occurs when either a particular time or predefined landmark is reached
Multiple	Study analysis will incorporate one or more interim analyses triggered by events, landmarks or time schedule
Adaptive	The study design may alter in a prospectively defined manner based on the outcome of an interim analysis; for example, a study arm may be closed to accrual, the overall sample size may be increased, or a new study arm may open. All adaptive designs are based on rules described in detail in the original protocol and are not based on protocol amendments developed subsequent to any data analysis.

23

Adaptive Design Examples

For binary responses

- Play-the-winner (PW) rule (Zelen, 1969),
- Randomized play-the-winner (RPW) rule (Wei and Durham, 1978),
- Generalized Pylar urn (GPU) design (WeiF, 1979), and the
- Drop-the-loser (DL) rule (Ivanova, 2003)

For continuous responses

- Linear rank test statistic based design (Rosenberger, 1996),
- Link function based design (Bandyopadhyay and Biswas, 2001),
- Wilcoxon score based design (Bandyopadhyay and Biswas, 2004),
- Utility based design (Atkinson and Biswas, 2005), among others.
- Real life applications of response-adaptive designs are due to Bartlett et al. (1985), Rout et al. (1993), Tamura et al. (1994), Biswas and Dewanji (2004), among others.

24

Factorial Designs

- Multiple outcomes are offered and analyzed
- Sequential steps
- Simplest is a 2 x 2 factorial
 - Treatment A vs Treatment B
 - Subsequently
 - Treatment C vs Treatment D
 - Outcomes analyzed are A + C, A + D, B + C, B + D
 - Need enough patients and enough events to analyze and interpret

25

Relationship between design and analysis

- There are multiple statistical approaches to clinical trial analysis, with the distinctions dependent on whether a normal, or parametric, distribution of results is expected and whether prior information is incorporated
- Most clinical data do not follow a parametric distribution but many analytic plans incorporate a parametric distribution in the assumptions. Consequently real world observations may well vary from inferences from clinical studies

26

Integrated approach to trial design

- The sequence of steps can vary, but an effective approach to designing a clinical trial is to first decide what the question is, then
 - determine whether there is an endpoint or small collection of endpoints that can address the question
 - assess the reproducibility and applicability to the disease condition and patient of measuring the endpoint
 - decide what type of study design would be most resource effective, minimize bias, and maximize certainty
 - select the appropriate elements to write a study protocol

27

Access to new therapies

- Participation in clinical trials is the preferred method to be exposed to new products or new uses for existing products because the data are collected in a systematic and consistent manner and can contribute to the general experience
- Additional mechanisms to access new products exist and are available to individual patients

28

Clinical Trial Outcome Measures

Steven Hirschfeld., MD PhD
 Captain, United States Public Health Service
 Associate Director for Clinical Research
Eunice Kennedy Shriver
 National Institute for Child Health and Human Development



National Institute of Health
 Building Healthy Kids



Outcome Measures= Endpoints

- High level overview
 - Live longer, live better or both
 - In general, observations or reports
 - Reports can be direct from the patient or indirect from someone else
 - For oncology, observations or reports can pertain to the patient, the malignancy or both
 - Caution: Changes in the malignancy may not reflect changes in the patient and vice versa

30

Definitions

The National Institutes of Health Definition Working Group defined the terms 'clinical endpoint', 'biomarker' and 'surrogate endpoint' in 2001 as:

- (1) A **clinical endpoint** is a characteristic or variable that reflects how a patient feels, functions or survives.
- (2) A **biomarker** is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention.
- (3) A **surrogate endpoint** is a biomarker intended to substitute for a clinical endpoint that should predict clinical benefit or harm or lack of both.

31

Biomarker Definition

The Biomarkers Consortium, a public private partnership dedicated to developing biomarkers for general use, defines biomarkers as

“characteristics that are objectively measured and evaluated as indicators of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention. “

32

Biomarker Qualification

- A biomarker needs to be qualified, meaning that the assessment results are reproducible and consistent and independent of whom is performing the assessment or where the assessment is done.
- Qualification usually involves the establishment of standard operating procedures, calibration of the outcome measures, a training procedure and, if applicable, specifications for reagents and equipment.

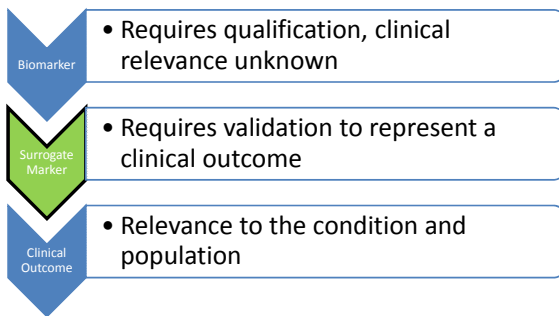
33

Surrogate Marker Validation

- Validation requires clinical studies where the direct measure of the clinical outcome is statistically compared to values of the candidate biomarker.
- Changes in both a positive and negative direction are correlated between the candidate biomarker and the clinical outcome measure and interpreted in the context of plausible biological mechanisms and what is known about the causal pathway of the intended clinical outcome.
- The validation process may not apply to all populations, so should be accepted only for the population in which the surrogate was studied and validated.

34

Hierarchy of Outcomes



35

General Endpoint Characteristics

- Endpoints must be prospectively defined to minimize bias
- The measuring technique must be reproducible and uniform in a patient population
- Serial measurements must be performed consistently with regard to technique and time for all patients at all sites and in all study arms
- Statistical analysis of the endpoints must be prospectively defined
- To be interpretable, must measure a parameter that is clinically relevant or be a validated surrogate

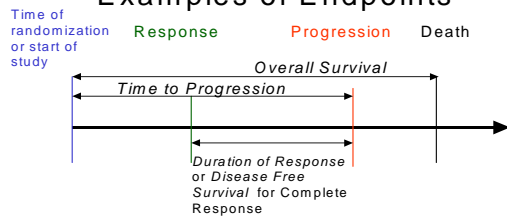
36

Time is a key variable

- Outcomes can be expressed as either
 - Time interval = duration between start of study and a significant event = time dependent
 - End of study is unknown until study complete
 - Usually determined by count of significant events
 - Rate = significant events in a predefined time interval = time independent
 - End of study is known by calendar date and predictable

37

Examples of Endpoints



Note that Overall Survival is for all patients while Time to Progression is only for patients that have progressive disease and Duration of Response and Disease Free Survival are only for patients that meet particular response criteria

38

Types of Time Dependent Endpoints

Variable	Population	Comment
Pharmacokinetics	All patients	A series of parameters that describe the absorption, distribution, metabolism and elimination of a drug as a function of time and exposure
Overall Survival	All patients	Typically time between study entry and death. Measurement is usually unambiguous. Cause of death may be difficult to determine. Effective therapies can result in long follow up times for completing studies.

39

Types of Time Dependent Endpoints

Variable	Population	Comment
Progression Free Survival / Time to Progression	All patients that progress	Typically time between study entry and first date of disease progression with death being considered as progression. The parameters for progression must be reliably defined and the assay validated- may be symptom based, imaging study, biomarker, or patient reported outcome.
Disease Free Survival	Only complete responders	Parameter for progression must be reliably defined and assay validated- may be symptom based, imaging study, biomarker, or patient reported outcome.

40

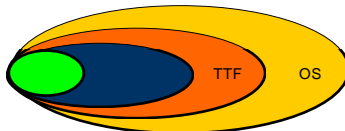
Types of Time Dependent Endpoints

Variable	Population	Comment
Time to Treatment Failure	All patients that change treatment	Treatment failure must be reliably defined and may include disease progression or unacceptable toxicity. Unacceptable toxicity can be a highly individual assessment and treatment failure can be due to multiple factors, not all of which are objective. Time to treatment failure is particularly difficult to interpret.
Duration of Response	Only responders	Typically time between first date of response and first date of disease progression. Response is usually defined as having a minimum duration (typically 4 weeks) to be considered a response.
Time to Response	Only responders	Typically time between date of study entry and first date of response. Response is usually defined as having a minimum duration (typically 4 weeks) to be considered a response.

41

Proportion of Population

Duration Endpoints (Variable Time)



DFS= Disease Free Survival = Complete Responders
PFS= Progression Free Survival = Non-Progressors
TTF= Time to Treatment Failure= Non-Progressors plus Not Tolerated Toxicity
OS= Overall Survival = All Survivors

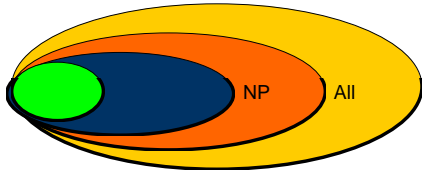
42

Types of Time Independent Endpoints

Variable	Population	Comment
Pharmacokinetic/Pharmacodynamic Relationships including Patient Reported Outcomes	All patients where measurements are taken	Description of relationship between drug exposure and a clinical or biochemical effect
Response	% of All Patients or a continuous variable such as a drug level	Criteria are extremely variable- may be drug levels, symptom based, imaging study, biomarker, or patient reported outcome. Response is often subdivided into categories that may be ordered (for example Complete Response, Partial Response, Stable Disease, Progression). Ordered categories require additional analyses and in some cases are combined.
Adverse Events	Usually % of All Patients	Standard reporting criteria are available from several sources
Landmark	% of All Patients	Highly variable- paradigm -an example would be % of patients alive at 2 years for a life threatening illness, but must be meaningful with regard to disease and patient population

43

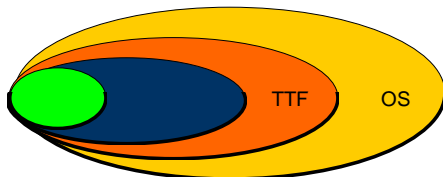
Rate Endpoints (Fixed Time)



CR= Complete Responders
 OR= Overall Response
 NP= Non-Progressors
 All = All patients intended to treat

44

Landmark Endpoints (Fixed Time)



DFS= Disease Free Survival = Complete Responders
 PFS= Progression Free Survival = Non-Progressors
 TTF= Time to Treatment Failure= Non-Progressors plus Not Tolerated Toxicity
 OS= Overall Survival = All Survivors

45

Composite Endpoints and Scales

- Some outcomes are a composite of multiple components such as a list of signs and symptoms
- The elements should be of approximately equal significance for the scale to be useful
- Each component should be of similar clinical significance and potentially observable in each patient

46

Composite Endpoints

- Determined by prospective algorithm
- Categories should not be confused with values, for example, if categories are numbered or scale scores add to a total, there is not a justification to perform arithmetic operations on those labels
- Analysis may include analysis of each component as well as the composite
- Must include guidelines for interpretation
- If driven by one domain or topic, then claim may be restricted to that domain

47

What about endpoints not in the study protocol?

In no circumstances should an endpoint be added to the analysis after the study is completed. It is not possible and not valid to interpret such an endpoint. If an apparent difference appears between the groups for an endpoint not prospectively intended to be analyzed, a new study should be designed using that endpoint as a primary endpoint.

48

Product Development Plan

- The product development plan should be considered as an integration of all of the data that would support a claim
- The product package insert should be a general guide to collect specific data for each section
- Outcome measures should be integrated across studies so that the preferred efficacy variable is assessed as early as feasible in the development plan and the confidence about the outcome continues to increase in precision

49


Predictive Value of Endpoints

- Can results in early phase studies predict outcomes for later phase studies?
 - Results are variable and approaching random
 - Different studies can have different outcomes using the same endpoints
 - Reflection of lack of precision of endpoints
 - For example, response rate is less precise than magnitude and duration of response
 - New assessments needed

50

Overlapping Outcome Variables

Development Plan Sequence



	Study 1	Study 2	Study 3	Study 4
Study	1	2	3	4
Outcomes				
CI				

51

Outcome Measures

- All measurements have associated confidence intervals, which are calculated from statistical tables.
 - The smaller or narrower the confidence intervals, the greater is the certainty of the result.
 - This can be achieved through either a large study population size or a large therapeutic effect.
 - The most difficult results to interpret are from a small population size with a small effect
- Convention is symmetric confidence intervals but asymmetric may be relevant and applicable

52

Endpoint Summary

- Precision, reproducibility, and predictive value are important criteria for endpoint selection
- Different endpoints have different predictive value
- The integration of all evidence is required to understand potential benefit and risk
- Make no assumptions- if something is not measured well, its status is unknown

53

Resources

- Multiple Guidance for Industry documents may be found on the FDA website
- Specific comments and policies may be found on the FDA and OHRP websites
- Contact information
 - Steven Hirschfeld, MD PhD
 - hirschfs@mail.nih.gov

54
