

Clinical Trials and Study Development

Laura Lee Johnson, Ph.D.

Statistician

National Center for Complementary and Alternative Medicine

johnslau@mail.nih.gov

Fall 2008

Your Question Comes First

May need to rewrite

If you change your question later

- May not have the power

- May not have the data

Need to know something about the population

COME TO THE STATISTICIAN EARLY AND COME OFTEN

Analysis Follows Design

Questions -> Hypotheses ->

Experimental Design -> Samples ->

Data -> Analyses -> Conclusions

Take all of your design information to a statistician early and often

Guidance

Assumptions

Objectives:
Study Development

Choosing a study design

Control groups (or lack thereof)

*Outlines of Study Protocols, Manual of Procedures (MOP),
and Components of Good Clinical Trials Report*

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Phase I to Phase IV Trials

National Cancer Institute: Dictionary of Cancer Terms

Phase I - The first step in testing a new treatment in humans
Test the best way to give a new treatment (for example, by mouth, intravenous infusion, or injection) and the best dose

Phase I

Dose increased little at a time

Find the highest dose that does not cause harmful side effects

Little known about possible risks and benefits of the treatments being tested

Trials usually include only a small number of patients who have not been helped by other treatments

Phase II

A study to test whether a new treatment has an anticancer effect (for example, whether it shrinks a tumor or improves blood test results) and whether it works against a certain type of cancer

Surrogate endpoints

Phase II Designs

Screening of new therapies

Not to prove 'final' efficacy, usually

 Efficacy based on surrogate outcome

Sufficient activity to be tested in a randomized study

Issues of safety still important

Small number of patients

Phase III

A study to compare the results of people taking a new treatment with the results of people taking the standard treatment (for example, which group has better survival rates or fewer side effects)

Phase IV

After a treatment has been approved and is being marketed, it is studied in a phase IV trial to evaluate side effects that were not apparent in the phase III trial.

Thousands of people are involved in a phase IV trial

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Study Design Taxonomy

Randomized vs. Non-Randomized

Blinded/Masked or Not

 Single-blind, Double blind, Unblinded

Treatment vs. Observational

Prospective vs. Retrospective

Longitudinal vs. Cross-sectional

Aspirin and Mortality

What is the best way to design a study to test if aspirin use reduces all cause mortality?

Aspirin and Mortality

Several ways to design the study

Two of them are 'best'

Aspirin and Mortality

ID a large group of people from a population at Time 0

Give them aspirin

Wait 5 years and observe all-cause mortality

Aspirin and Mortality

Picture of futuristic car

Aspirin and Mortality

Turn back time

Start over with the same large group of people at Time 0

Give them 'no aspirin'

Wait 5 years observing all-cause mortality

Compare outcomes aspirin vs. no aspirin

Estimate average aspirin effect

Aspirin and Mortality – Try 2

ID a large group of people from a population at Time 0

Clone them perfectly

Give one in clone pair aspirin, other no aspirin

Wait 5 years and observe all-cause mortality in each pair

Compare outcomes aspirin vs. no aspirin

Estimate average aspirin effect

Aspirin and Mortality – Try 2

Clones are a great idea!

Identical in all ways we can, and cannot, measure

But since we are not in Hollywood features what to we do?

Aspirin and Mortality – RCT

Identify a large group of people from a population at Time 0

Divide into 2 groups, at random

Give one group aspirin, one no aspirin

Wait 5 years and observe outcomes in each group

Compare outcomes aspirin vs. no aspirin

Estimate average aspirin effect

Aspirin and Mortality – RCT

Aspirin and no Aspirin groups should be comparable since randomizing yields groups with similar baseline characteristics

Except when randomization does not

Next lecture

Aspirin and Mortality – Observational Study

Prospective

Identify a large group of people from a population at Time 0

Some take aspirin, some do not

Wait 5 years and observe outcomes in each group

Compare outcomes aspirin vs. no aspirin

Estimate average aspirin effect

Aspirin and Mortality – Observational Study

Aspirin and no aspirin groups may not be comparable since they may not have similar baseline characteristics

Aspirin and Mortality

Obs. Study *Retrospective*

ID a large group of people from a population at Time *Today*
(*but if all-cause mortality = outcome?*)

In the past 5 years (prior to death) some took aspirin, some did not

Today observe outcomes over the past 5 years in each group

Compare outcomes aspirin vs. no aspirin

Estimate average aspirin effect

Aspirin and Mortality Case Control Study

Aspirin is an 'exposure', BUT

Take people after ?first heart attack? and a group of controls

Match some risk factors?

Assess prior aspirin use

Dose

Frequency

Plus assess a million other risk factors

Aspirin and Mortality Case Series and Survey

May have started with case series

Now might be late

Unless you have seen something interesting

May have done a survey next (late now)

If interested in preliminary duration of use information could be useful

Likely to incorrectly evaluate temporal

Dead do not answer surveys

Surrogate responders may not help

Analyses

Fancy methods

Bread and butter

- Chi-square, T-tests, Wilcoxon tests

- Linear or logistic regression

- Basic survival (K-M, Cox PH)

Extensive Exploratory Data Analysis

Plots to match analysis

Banish ANOVA

ANOVA has its place

Rarely in the study of humans

Great when computers less memory than a cell phone

More robust methods will run on modern computers

Exciting methods run on PS3 game console

Repeated measures ANOVA is worse

What Do We Test? **Effect or Difference**

Difference in Means or Proportions

Odds Ratio (OR)

Relative Risk (RR)

Hazard Ratio (HR)

Correlation Coefficient

And many other things.....

Risk: Difference vs. Ratio (new)

Difference in the absolute risks

- Attributable risk

- Excess risk attributable to exposure

Relative Risk (RR)

- Ratio of two absolute risks

Hazard Ratio (HR)

- Ratio between predicted risk of an event for member of A and that of a member of B, holding everything else constant

Is ratio the best to talk to people?

Difference vs. Ratio (new)

Invasive breast cancer WHI (JAMA 288[3]:321-33)

Increase observed estrogen+progestin group

Difference in risk

38 vs 30 per 10 000 person years

Hazard Ratio (HR)

26%

Is your personal risk 26%? No

8 more invasive breast cancers per 10 000 person years? Yes

What Do We Test?

Clinically important difference

Smallest difference considered biologically or clinically relevant

Medicine: usually 2 group comparison of population means

Do Not Confuse

Association
Causality
Confounding

Correlation
Prediction

High OR Does Not a Good Test Make

Everyone loves prediction. BUT

Diagnostic tests need separation

Not logistic regression with high OR

Strong association between 2 variables does not mean good prediction of separation

Measure of evidence should match the Question

Ideal Study - Gold Standard

Randomized

Double blind / masked

Treatment

Prospective

Parallel groups

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Observational Randomized

Can ONLY show Association

You will never know all the possible confounders!

Can show Association AND Causality

Well done non-adaptive randomization → unknown confounders
should not create problems

Observational Studies

Case Reports/Case Series

Cross-sectional Survey

NHIS (National Health Interview Survey)

Case-Control Study

Groups with or without outcome

Determine who was exposed to risk factor

Cohort Study

Follow a group for a while

Cardiovascular Health Study

Experimental?

Are 'experimental' and 'randomized' interchangeable?

Depends. In text, yes.

Quasi experimental

Experimenting

No control

Not randomized

Quasi Experimental or Non-Randomized Experimental Studies

No control group

Early in investigation

Concurrent control “group”

Treatment assignment not by randomization

Historically controlled

Missing/poor data

Non-comparability of groups

No placebo/control = problems

Patients tend to do better by receiving some treatment, even placebo or standard of care (soc)

Comparing a patient on treatment to baseline does not take this into account

Additional Problems

Researchers tend to interpret findings in favor of the new treatment

- Investigator/participant bias

Impossible to distinguish the effect of time from treatment effects

- Confounding

Human Assumptions and Concurrent Control Groups

Newer equals better

Systematic allocation is unreliable and many times NOT systematic

- Bias

- Manipulation

No randomization -> impossible to establish if comparable groups

Historical Control Study

Small patient pool

Pediatrics

Cancer research

Responses compared to controls from previous studies

Only half the patients

No “placebo exposure”

Historical Control Problems

Serious bias for assessing treatment efficacy

Controls not a good comparison group

Historical Controls and Time

Treatments, technology, patient care changed over time
Patient population characteristics have changed over time

Non-randomized Phase II design problems

Placebo effect

Investigator bias

Unblinded treatment/assessment

Regression to the mean

 Natural reduction in disease activity over time

Observational Studies

Why can observational studies only find a weaker degree of connection?

- Subject to confounding

- Can correct for what you know, but nothing to be done about the unknown

Sometimes it is unethical to do a randomized trial (e.g. smoking)

Causation vs. Association

Causation

Established by randomized experimental studies and clinical trials

Association

Observational studies can merely find association between a risk factor and an response

Example

Example

JAMA 2004 recommendations for adult HIV

Optimal time to initiate HAART to maximize survival/AIDS-free survival

CD4 cells/ μ L: > 200 but ≤ 350

Dynamic treatment regime

Waiting on a person's CD4 count, which changes over time

Study Needs

Cohort recently diagnosed/infected with HIV
Antiretroviral naïve

Nonrandomized Observational

System to ID treatment status over time

Time t changed treatment status

Record data on confounders until the end of study, AIDS, or death

All time-varying risk factors used to decide to change treatment status at time t

Survival analysis

Comparing what groups?

HIV OutPatient study (HOPS)

Distinguish

“Observational studies are often analyzed as if they had resulted from a controlled study, and yet the tacit assumption of randomness can be crucial for the validity of inference.”

Copas, J.B. and Li, H.G. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society*, 59: 55–95.

Another Design?

Randomized Study

Always treat/never treat

Regardless of CD4

Use intent to treat (ITT) analysis

Follow until end of study/death/AIDS

Survival analysis

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Books

Statistical Rules of Thumb by Gerald van Belle (vanbelle.org
for updates & monthly rule)

Hosmer and Lemeshow books

Epidemiology by Leon Gordis

More Books

*Statistical Reasoning in Medicine: The Intuitive P-Value
Primer* by Lemuel Moye

Designing Clinical Research: An Epidemiologic Approach,
edited by Stephen Hulley

And More Books

Data Monitoring Committees in Clinical Trials: A Practical Perspective
by Ellenberg, Fleming, DeMets.

Fundamentals of Clinical Trials by Friedman, Furberg, DeMets

The Statistical Evaluation of Medical Tests for Classification and Prediction by Margaret Sullivan Pepe

Articles

British Medical Journal: Statistics Notes

Link broke; search in BJM or via your favorite
search engine

Statistics in Medicine

NEJM: Equivalence trials

October 16, 1997

FDA Guidance

ICH E9 Statistical principles

ICH E10: Choice of control group and related issues

ICH E4: Dose response

ICH E8: General considerations

FDA draft guidance on drug interaction study designs (and analyses), Bayesian methods, etc.

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Conducting a Clinical Study: Study Protocol

Road map for performance of study

Anticipate problems

Facilitates communication with potential collaborators, employers, funding agencies

Assists in manuscript preparation

Protocol Components

Background and rationale

Specific objectives (3-5 aims of study)

Clinical trial should include specific hypothesis regarding primary outcome

Concise statement of design

Methods and analysis

Responsibility and authorship

Statement of Design

"An observational study of decline in pulmonary function in persons living in heavily industrialized areas compared to persons in non-industrial areas."

"A prospective, non-concurrent study of postoperative pneumonia in patients receiving regional vs. general anesthesia for peripheral vascular grafting."

Methods: Inclusion Criteria

Definition of patient population

Specific as possible (but not too restrictive)

Inclusion criteria

Disease or condition under study

Prior myocardial infarction, smokers

Other information

Age

Sex

Area of residence or hospitalization

Methods: Exclusion Criteria

Participants must not have *any* specified criterion

Generally include conditions making study difficult or impossible

Patients in whom one treatment or other is inappropriate or unethical

Coronary Artery Surgery Study excluded patients with left main coronary artery disease

Methods: Exclusion Criteria

"Logistic" concerns

- Aged under 18

- Critically ill

Circumstances making determination of outcome difficult or impossible

- Expected to leave area

- Unable to communicate in language study team uses

- Pregnancy

Common Mistakes

Unnecessary exclusion/inclusion criteria

Plans for the trial made without any reliable data on participant availability

Pilot recruitment

Unrealistic timetable for recruitment or no recruitment goals

Revision of sample size calculations to make them consistent with recruitment realities

Outcome Definitions

Be specific and as clear as possible

Primary vs. secondary outcomes

Standard clinical definitions

Textbook: usually not specific enough

Consensus conference

Definition of hypertension

Recognized expert body (WHO, AHA)

Outcome Definitions

Appointed panel of experts

Previously widely-recognized study (SHEP, WHI, SOLVD)

Adjudication: submit to panel of masked, unbiased "experts"

Treatment Definition

Specify as much as possible without interfering with patient management

Realize that generalizability often lost in quest for specificity

Specify criteria for withdrawal from study or deviation from protocol

List concurrent medications, procedures, etc. that are prohibited or permitted

Masking/Blinding

Specify whom to be masked, why, how, and to what

Assess effectiveness of masking

Specify criteria for unmasking, whom to be unmasked

Mask determination of outcome so that reviewers are unaware of treatment assignment; provide information on "need to know" basis

Study Design Taxonomy (new/old)

Randomized vs. Non-Randomized
Blinded/Masked or Not
 Single-blind, Double blind, Unblinded
Treatment vs. Observational
Prospective vs. Retrospective
Longitudinal vs. Cross-sectional

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

MOP: Manual of Procedures or Manual of Operations

Can another investigator step into the study (or reproduce it) at any time?

Hope so!

Which data to be collected, how

Timetable for follow-up

Examples of Chapters in MOP

Overview	Recruitment
Eligibility	General procedures
Informed consent	Screening
Randomization	Follow-up visits
Retention	Intervention
ECG	Blood collection
Physical assessment	Fitness testing
DXA scanning	Health events
Participant safety	Adverse events
Data management	Quality control
Interviewing	Study organization
Study website	

Subject Recruitment aka The Facts of Life

Early estimates unrealistically high

Takes a major effort

People presumed eligible for study during planning will disappear mysteriously as soon as the study starts

Recruitment will be more difficult, cost more, and take longer than planned

Chart showing data for study on life – cumulative number randomized versus study goals for all clinics.

Prepare!

- Collect reliable data to estimate participant availability
- If matching, allow for screening twice as many controls per discrete variable matched upon
- Decide on general recruitment approach
- Outline steps in recruitment process
- Establish necessary recruitment contacts

Recruitment Mistakes/Problems

Competing with private physicians for patients

Providing basic care rather than referring patient back to primary care physician

Failure to maintain adequate contact with referring physician

Recruitment Mistakes/Problems

Attempting recruitment without the support of colleagues

Taking access to medical records for granted

Failing to secure enthusiasm and commitment of staff

Inadequate publicity

Protection of Human Subjects

Monitoring for adverse effects

Informing patient, physician of complications or abnormalities

Interim analyses (*pre planned*)

Data Safety Monitoring Board (DSMB)

Informed Consent

Written informed consent

Institutional review board (IRB): independent review and monitoring by panel including members outside institution

Informed Consent: Approach

Find proper setting: quiet, private

Provide adequate time

Encourage *potential participant* to discuss with others (family members, physician), ask questions

Ensure participant's competence to give consent

Provide copy of signed consent

In unblinded studies, must be willing to participate regardless of random assignment

Mistakes in Consent Process

Inadequate time

Failure to specify required procedures

Inadequate documentation

Vague or inaccurate statements

Making commitments that cannot be met

Use of untruths to protect study design

Consent after the fact

Speaking for the patient ("I understand that...")

Specifics of Laboratory Methods

Enzyme determinations: laboratory methods?

Chest x-rays

PA and/or lateral

Supine or erect

Clinical measurements

Blood Pressure (BP)

Supine or standing

How many times? Average?

Rest periods between measurements?

Feet on floor? Arm is where? Cuff size?

Heart sounds in left lateral decubitus

Data Management and the Subject Record

Each participant should have his or her own study record stored in locked area when not in use

Each participant should have a study number for use as identifier name should not be in data base, coding forms, etc.

Data Management

If multiple data sources are needed, use separate forms and system to keep track of progress in data collection (e.g., colored-dots, transmittal forms, etc.)

Subgroup Analysis

Often performed when no overall effect found

Used to look for high-risk or peculiar groups with marked treatment effect

Beware of "data-dredging"- looking at many, many subgroups until one "significant" effect found

Subgroup Analyses

Limit number of subgrouping variables

Look at all members of the subgroup

a priori and *a posteriori*

Choose cut points independent of treatment differences

Blood pressure treated to goal of 140/90

Cut blood pressure at 140 vs. >140 will introduce bias of
successful vs. unsuccessful treatment

Subgroup Analyses

Stringent significance testing, especially if number of hypotheses tested is large

When possible, validate findings before reporting on an *a posteriori* (data-driven) subgrouping variable

Report methods and procedures

Be cautious regarding conclusions

Other Problems

Changes in procedures necessary

Changes in inclusion or exclusion

Changes in data collection procedures

Revisions as needed, dated, with replacement pages in
MOP

Drift in measurements

Change in health and treatment patterns or practices within
the community

Write the MOP

So anyone you might hire for any position can follow the entire document and run the study

Undergrad, post doc, statistician, anyone

Outline

Phase I-IV studies

Build-A-Study

Analyses (small)

Control groups (or lack thereof)

Books and Articles (1)

Outline Study Protocol

Outline Manual of Procedures (MOP)

Questions

Components of Good Clinical Trials Report

Questions?

Components of Good Clinical Trials Report: Design Specifies

**(after Dr. Curtis L. Meiner, Professor of
Epidemiology, Johns Hopkins School of Hygiene
and Public Health)**

Purpose of study

Primary outcome measure

Test and control treatments

Level of treatment masking: unmasked, single-or
double-masked

Planned recruitment goal

Eligibility and exclusion criteria

Method of patient recruitment

Continued

Number of patients enrolled

Number of patients in analyses

Equal to number allocated to treatment, or explanation should be given

Method of treatment allocation

Stratification variables

Methods measuring treatment adherence

Planned and actual length of patient follow-up

Patient Safeguards

Informed consent, Institutional Review Board (IRB) approval

Measures taken to protect patient confidentiality

Procedures to monitor study results for evidence of treatment effects

Data Collection Schedule

Frequency of baseline visits

Frequency of follow-up visits

Definition of dropouts

Results

Number of patients enrolled by treatment group

Number of deaths observed

Comparison of treatment groups for the primary outcome measure

Results: Completeness of Follow-up

Number of missed examinations

Number of dropouts and withdrawals

Number of participants lost to follow-up

More Results

Selected baseline characteristics

Multiple regression analyses using baseline characteristics
to provide adjusted treatment comparisons

Treatment comparisons by level of adherence

Conclusions

Test of primary hypothesis/outcome

Test of secondary hypotheses as applicable

Limits on generalization of the results indicated

Discussion of statistical power if no treatment difference is detected

Although I am not a big fan of this depending...

Thank you!