Promoting excellence in clinical research: from idea to publication

Stratis Kolibianakis MD MSc PhD

ESHRE Campus symposium Thessaloniki, Greece 5-6 November 2010

Disclosures

Consulting work for MSD, Merck Serono, Ferring

Stratis Kolibianakis MD MSc PhD

ESHRE Campus symposium Thessaloniki, Greece 5-6 November 2010

What to study:

patients, records or publications?

Goals of sampling criteria:

internal and external validity

Outcomes:
primary and secondary outcomes

Sample size assumptions:

how “large” is large enough?

A question has been set

A literature search has been performed

We are willing to do a study!

Is addition of recombinant LH necessary in patients undergoing ovarian stimulation for IVF?

Literature search controversial results

Records, patients, publications?

Perhaps we have given already LH in our unit.

Can we study our patient records?

Retrospective study

We have never given LH to our patients!

No records to examine

We need to study patients!

Prospective study

We have never given LH to our patients!

We cannot perform such a study!

(time, cost, the medication is not available)

The literature provides publications on the question of interest

Others have done the studies before us!

Perform a systematic review /meta-analysis

Study patient records

A retrospective study uses existing data

that have been recorded for reasons other than research

Some doctors in our unit

have added recLH to the stimulation scheme

in some patients during the past

Case report – Case series

Study patient records

Case report - Case series

A *case report* is a report of one unusual and/or instructive case

A *case series* is a report of multiple similar unusual or instructive cases

(We were the first to administer LH in our patients worldwide and the readers are dying to learn from our experience – Case series)

(We were the first to achieve live birth after addition of rec LH to FSH worldwide- Case report)

Study patient records

Case report Case series

cons

The investigator depends

on the availability and accuracy of the medical record

Subject to selection bias

because the investigator self-selects the cases

Uncontrolled!

Study patient records

Case control study

Superior to a case series because of the presence of a control group

Cases with and without the condition of interest are identified

The degree of exposure to a possible risk factor

is then compared between the 2 groups

(We identify the pregnant patients who suffered miscarriage

and investigate whether

they have been exposed to recLH in addition to FSH or not)

Study patient records

Case control study

The case-control study design assumes that

- **cases differ from controls only in having the disease**

- **exposure should be equally distributed between cases and controls if the exposure does not cause the disease**

- **greater exposure among cases would indicate that exposure increases the risk of the disease**

Study patient records

Case control study

The exposure is determined retrospectively

The data collectors are unaware of whether

a subject is a case or a control

Data collectors should be unaware of the study hypothesis

The cases and the controls

must be assessed for exposure in the same way

Study patient records

Case control study

Strengths

Fewer constraints by the frequency of the disease

Shorter waiting time than a prospective cohort study

Case-control studies are sometimes feasible

when randomized controlled trials are not

Case control studies cost less and have fewer practical restrictions

Study patient records

Case control study

Drawbacks

A less well defined target population

Risk of selection bias

Difficult or impossible to ascertain cause-and-effect, because of confounding factors

Study patient records

Retrospective study designs

are generally considered inferior to prospective study designs

A retrospective study design should never be used

when a prospective design is feasible

Study patient records

Many times investigators view retrospective studies as

"quick and dirty"

because the data are quickly gleaned

from existing records to answer a question

A well done retrospective study

may not be quick

is definitely not "dirty"

Study patient records

A retrospective study can serve a useful purpose

Focus the study question

Clarify the hypothesis

Determine an appropriate sample size

Identify feasibility issues

in a subsequent prospective study

Study patient records

Advantages

Inexpensive

Uses existing records

Allows study of rare occurrences

Easier to assess conditions where there is a long latency between exposure and disease

Can generate hypothesis that is then tested prospectively

Disadvantages

Relies on accuracy of written record or recall of individuals (recall bias): **garbage in garbage out**

Important data may not be available: **nothing in nothing out**

Difficult to control bias and confounders: no randomization, no blinding

May be impossible to access important information (restricted by statute or institutional regulations)

Difficult to establish cause and effect

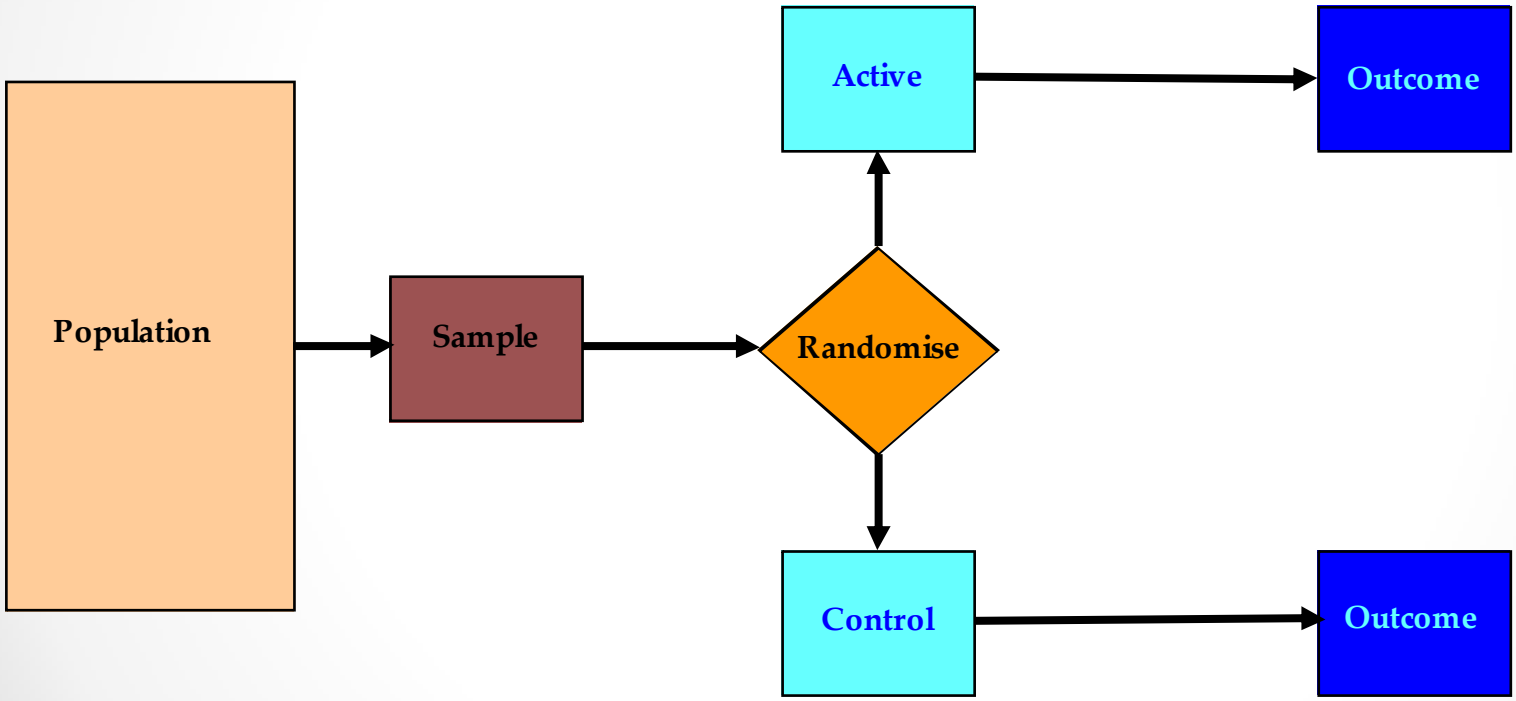
Results are, at best, hypothesis-generating

Study patients

A study in which people are divided into groups
that are exposed or not exposed to the intervention(s) of interest
before the outcomes have occurred

Randomised controlled trials are always prospective studies
case control studies never are
Concurrent cohort studies are prospective studies,
Historical cohort studies are not

RCT to assess patient outcomes



Study patients

Can demonstrate

an appropriate temporal sequence between exposure and outcome

Since exposure is determined first and

in a time period preceding assessment of outcome,

it is easier to ascribe the outcome to the exposure

than it is in studies

where the temporal sequence is more difficult to determine

Study patients

Permit the direct calculation of incidence rates

in both the exposed and unexposed groups

This makes it easy to calculate risk or rate ratios (or differences)

Permit multiple outcomes to be assessed in the same study

Study patients

cons

Large number of patients are needed

Lost of follow up (attrition)

Expensive

Administrative problems

Studying publications

Systematic Review

- Focuses on a clinical topic and answers a specific question
- An extensive literature search is conducted

to identify all studies with sound methodology

- The studies are reviewed, assessed, and the results summarized

according to the predetermined criteria of the review question

Meta-analysis

Following a systematic review

combination of the results

using accepted statistical methodology

as if they were from one large study

Studying publications

elements of a systematic review

Question

Search strategy

Inclusion criteria

Validity assessment

Data extraction

Meta-analysis when appropriate

Studying publications

Narrative review	**vs.**	**Systematic review**
Question vague		Question explicit
Methods not explicit		Methods explicit: search, inclusion, extraction, stats
No focus on study quality		Validity assessment
No quantitative summary of results		Meta-analysis when appropriate

Study publications
Advantages of systematic reviews

Explicit methods limit bias in identifying and rejecting studies

Conclusions are more reliable and accurate because of methods used

Large amounts of information can be assimilated quickly by healthcare providers, researchers, and policymakers

Delay between research discoveries and implementation of effective diagnostic and therapeutic strategies may be reduced

Results of different studies can be formally compared to establish generalisability of findings and consistency (lack of heterogeneity) of results

Reasons for heterogeneity (inconsistency in results across studies) can be identified and new hypotheses generated about particular subgroups

Quantitative systematic reviews (meta-analyses) increase the precision of the overall result

Study publications

Meta-analysis

Pros

Statistical power

Applicability

Precision of results

Objectivity

Quality control

What type of study?

Goals of sampling criteria:

Internal and external validity

Population

a group of individuals persons, objects, or items
from which samples are taken for measurement

Sample

a finite part of a statistical population
whose properties are studied
to gain information about the whole (Webster, 1985)

When dealing with people, it can be defined as a set of respondents(people) selected from a larger population for the purpose of a survey

What is sampling?

Sampling

act, process, or technique

of selecting

a suitable sample, or a representative part of a population

for the purpose of determining parameters or characteristics

of the whole population.

What is sampling?

To draw conclusions about populations from samples,

we must use inferential statistics

which enables us to determine a population`s characteristics

by directly observing only a portion (or sample) of the population.

Census vs. sample

A census rather than a sample

should always be used to obtain information about populations.

But there are many reasons why a census is not used

Economy

Timeliness

The large size of many populations

Inaccessibility of some of the population

Destructiveness of the observation

Accuracy

BIAS AND ERROR IN SAMPLING

A sample

is expected

to mirror the population from which it comes

Unfortunately, there is no guarantee

that any sample

will be precisely representative

of the population from which it originates

Chance may dictate

that a disproportionate number of untypical observations will be made

Pitfalls in sampling

What can make a sample unrepresentative of its population?

Sampling error

comprises the differences
between the sample and the population
that are due solely to the particular units
that happen to have been selected

Causes for sampling error

Chance:

That is the error that occurs just because of bad luck

This may result in untypical choices

Unusual units in a population do exist

and there is always a possibility

that an abnormally large number of them will be chosen

Causes for sampling error

Sampling bias

tendency to favor

the selection of units

that have particular characteristics

Sampling bias is usually the result of a poor sampling plan

The most notable is the bias of non response

when for some reason some units

have no chance of appearing in the sample

Non sampling error (measurement error)

An error that results solely

from the manner in which the observations are made

e.g. inaccurate measurements

due to malfunctioning instruments or poor procedures

Types of samples

The convenient sample

A convenience sample results when the more convenient elementary units are chosen from a population for observation

The judgment sample

A judgment sample is obtained according to the discretion of someone who is familiar with the relevant characteristics of the population

The random sample

This may be the most important type of sample

A random sample allows a known probability that each elementary unit will be chosen

For this reason, it is sometimes referred to as a probability sample

Types of random samples

A simple random sample

A systematic random sample

A stratified sample

A cluster sample

Internal Validity: Definition

Internal validity refers to the extent to which we can accurately state that the independent variable produced the observed effect

If

the effect on dependent variable

is only due to variation in the independent variable(s)

then

internal validity achieved

Internal validity

All that internal validity means is that we have evidence that what we did in the study caused what we observed to happen

It doesn't tell us

whether what we did for the study was what we wanted to do

or

whether what we observed was what we wanted to observe

those are construct validity concerns

It is possible to have internal validity in a study and not have construct validity

External Validity: Definition

Relates to generalizing findings

to or **across** target populations

to or **across** tasks

to or **across** environments

External validity involves the extent to which the results of a study can be generalized (applied) beyond the sample

Can we apply what we found in our study

to other people (**population validity**)

or settings (**ecological validity**)

Threats to external validity

- Treatment-Attribute Interaction
- Treatment-Setting Interaction
- Multiple-Treatment Interference
- Pre-test Sensitization
- Post-test Sensitization

“Trade-off"

between

internal validity and external validity

When measures are taken or procedures implemented
aiming at increasing the chance for higher degrees of internal validity,
these measures may also limit the generalizability of the findings

This situation has led many researchers call for "ecologically valid" experiments
By that they mean that experimental procedures
should resemble "real-world" conditions

“Trade-off"

between

internal validity and external validity

Effectiveness

relates to how well a treatment works in practice,

as opposed to

Efficacy

measures how well it works in clinical trials or laboratory studies

Outcomes:

primary and secondary outcomes

Primary outcome

is typically the clinical parameter of interest

provides the central justification for the trial

determines the study size

(Live birth after LH addition in ovarian stimulation)

Secondary outcomes

also motivated the trial,

but that by themselves would be unlikely to justify a full-scale intervention

(Number of COCs retrieved, E2 on the day of hCG administration)

Primary outcomes

Primary endpoints represent the axis
around which the trial's logistical machinery revolves

The findings for the primary endpoints of the study will determine
whether the study is
positive, negative, null, or uninformative,
thereby serving as the ruler against which the trial's results will be measured

The analyses of primary endpoints
are often described as **confirmatory analyses**,
because the analyses confirm
the answer to the scientific question which generated the clinical trial

Secondary outcomes

The endpoints of the clinical trial
that were prospectively selected during the trial's design phase,
but had no a priori alpha allocated to them,
are termed ***secondary end- points***

These endpoints, being prospectively selected,
produce trustworthy estimators of effect size, standard error, CIs, and *p*-values,
all of which measure the effect of the clinical trial's intervention

Secondary outcomes

However, drawing confirmatory conclusions
about the effectiveness of the intervention being studied by the clinical trial,
based on the results of secondary endpoints in general,
cannot be permitted,
since conclusions based on these secondary endpoints
will increase the familywise error level above acceptable levels

Secondary outcomes

The role of analyses carried out on secondary endpoints
is to provide support for the conclusions drawn from the trial's primary endpoints

Secondary endpoints can provide important information
about the nature of the biologic mechanism of action of the compound
that is being studied in the clinical trial

If they are endpoints that are related to the primary endpoint,
they can add additional persuasive force
to the argument for the beneficial effect of therapy

Typically, there are more secondary endpoints than primary endpoints

Surrogate outcomes:

Outcome measures that are not of direct practical importance
but are believed to reflect outcomes that are important

A laboratory measurement or a physical sign
used as a substitute for a clinically meaningful end-point
that measures directly how the patient feels, functions or survives.

Changes induced by a therapy on a surrogate end-point
are expected to reflect changes in a clinically meaningful end-point

Study of surrogate variables may

shorten the period of study

lower the sample size required

lower the costs of the study

Sample size assumptions:

how "large" is large enough?

A study must be of adequate size, relative to its goals

It must be "big enough"

that an effect of such magnitude as to be of scientific significance

will also be statistically significant

It must not be "too big,"

where an effect of little scientific importance

is nevertheless statistically detectable

Sample size assumptions:

how "large" is large enough?

Sample size is important for economic reasons:

an under-sized study

can be a waste of resources

for not having the capability to produce useful results

an over-sized study

uses more resources than are necessary

Sample size for percentages

Baseline proportion

Hypothesized difference – effect size (proportion in the intervention group)

Alpha

Beta (power:1-b)

Allocation ratio

One sided/two sided test

Where alpha is the probability of a type I error (rejection of a correct null hypothesis)

Beta is the probability of a type II error (acceptance of a false null hypothesis)

Two sided tests should be used unless there is a very good reason for doing otherwise

Effect size

Base it on substantive knowledge

Base it on previous research

Use conventions

Sample size for means

Baseline mean

Hypothesized difference (mean in the intervention group)

Alpha

Beta (power:1-b)

SD in the groups compared

Allocation ratio

One sided/two sided test

Power

The ability to reject the null hypothesis when it is false

The probability that the test will correctly detect a treatment effect

Sample size assumptions:

how "large" is large enough?

In Reject-Support research:

The researcher wants to reject *H0*

Society wants to control Type I error

The researcher must be very concerned about Type II error

High sample size works for the researcher

If there is "too much power," trivial effects become "highly significant."

In Accept-Support research:

The researcher wants to accept *H0*

"Society" should be worrying about controlling Type II error

The researcher must be very careful to control Type I error

High sample size works against the researcher

If there is "too much power," the researcher's theory can be "rejected"

by a significance test even though it fits the data almost perfectly

Sample size assumptions:

how "large" is large enough?

There is a growing amount of software for sample-size determination,

nQuery Advisor (Elashoff, 2000),

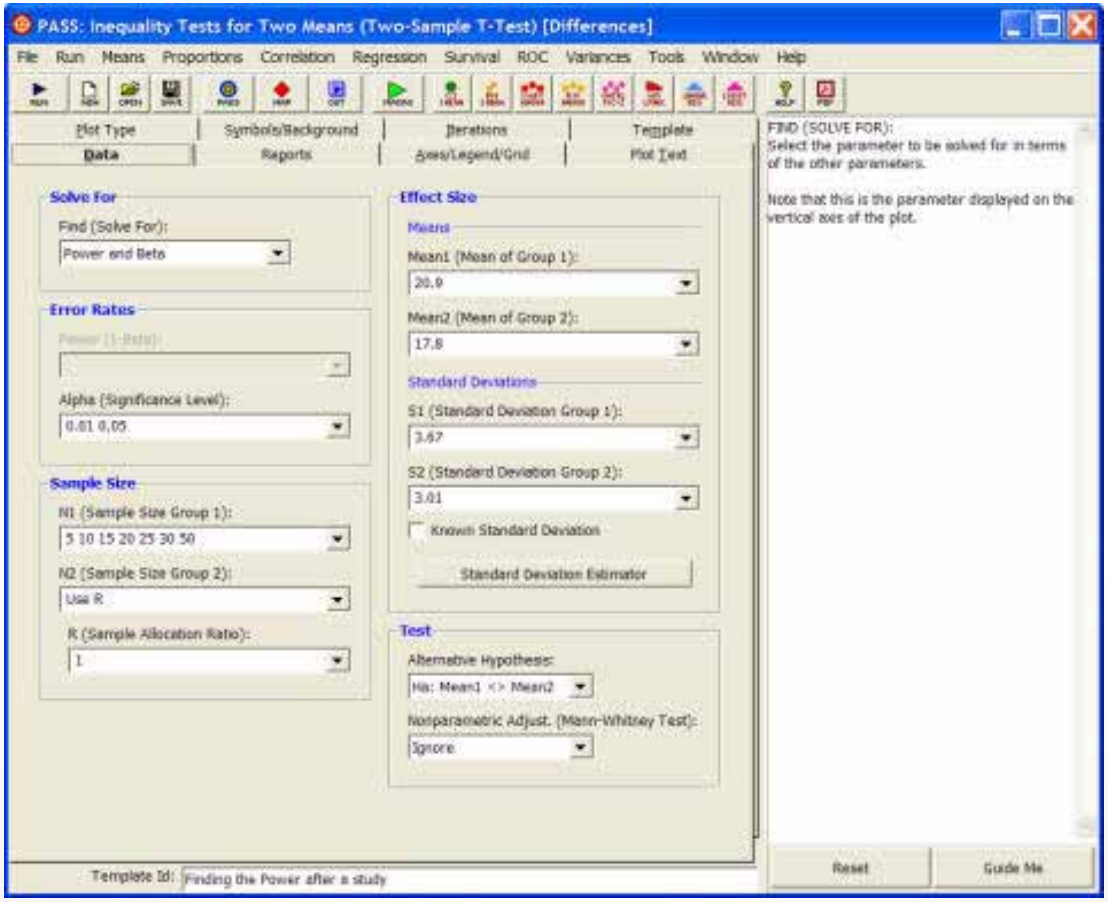
PASS (Hintze, 2000),

UnifyPow (O'Brien, 1998),

Power and Precision (Borenstein et al., 1997).

Sample size assumptions:

how “large” is large enough?



What to study:

patients, records or publications?

Goals of sampling criteria:

internal and external validity

Outcomes:
primary and secondary outcomes

Sample size assumptions:

how “large” is large enough?