Session 4:
Statistical considerations in confirmatory clinical trials II

Agenda

- Interim analysis
  - data monitoring committees
  - group sequential designs
- Adaptive designs
  - sample size re-estimation
  - Phase II/III trials
- Subgroup analyses
  - exploratory and confirmatory
- Missing data

Interim Analysis

Trial design with an interim analysis

- Unblinded interim analysis: Any review of data requiring patients to be **grouped according to the randomisation** before the database is frozen
- Unblinded interim analysis conducted to:
  - Assess whether to stop study early due to…
    - Safety concerns
    - Efficacy (overwhelmingly positive results)
    - Futility
  - Adapt the study design (e.g. choose between doses)
  - Planning other studies (not recommended for confirmatory studies)
- Blinded interim analysis: no grouping of treatments according to randomisation
  - Monitor total number of clinical events
  - Review ongoing safety data

Maintain study blind

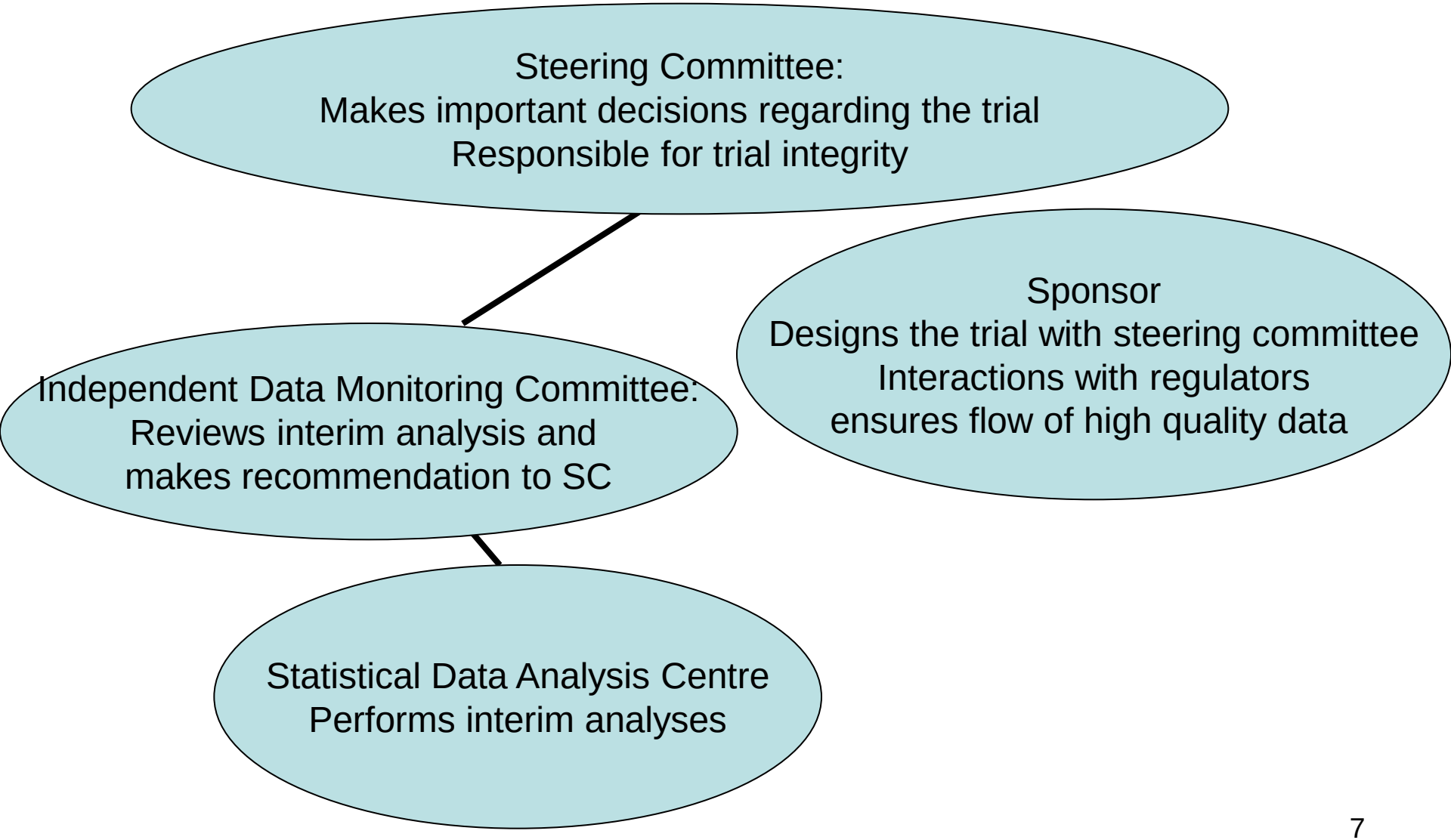
- Need to maintain blind among people directly involved in the study
  - Study staff
  - Investigators
  - Sponsor staff directly involved in the trial

- May require evaluation of interim analysis by independent data monitoring committee (IDMC)..

IDMC for confirmatory trials

- Independent of investigators, sponsor involvement discouraged
- Includes clinical experts in the therapeutic area and a statistician
- Safety monitoring primary responsibility, may monitor efficacy
- Makes recommendations that impact the future conduct of the trial,
  - include continuing, terminating or modifications to the trial
- Implementation of IDMC recommendation is responsibility of the sponsor
  - Possible to ignore recommendations

Committees for a large trial



Interim analysis for efficacy

- Allows trial to stop early for overwhelming efficacy
  - May be necessary for serious outcomes to avoid unnecessary placebo exposure
  - Can mean medicine available to patients earlier
- Risks with stopping early include:
  - Reduction in available safety database.
  - Increased variability in estimates of treatment effects.
  - Reduced information on secondary endpoints
  - Acceptance of study results is not only based on a statistically significant primary result
  - May need sufficient data to explore important subgroups

Consistency of results

- Regulators interested in assessing results before and after interim analysis
  - Substantial discrepancies with respect to the types of patients recruited and / or results obtained will raise concern
  - Difficult to interpret conclusions if it is suspected that the observed discrepancies are a consequence of dissemination of the interim results.
  - Difficult to convincingly demonstrate that no unblinded interim results have been released.
  - Differences between stages can occur by chance so Interim analyses always introduce this risk

P-value adjustment

- If the interim analysis can only stop the trial for safety or futility, no p-value adjustment required
  - Need to make this clear in the protocol
- If interim analysis can stop for efficacy, then need to adjust for more than one look at the data
  - If there is truly no difference between treatments, have more than one chance a false positive
  - Need to control overall probability of a false positive
- If study stops for efficacy at interim there is a sample size saving compared to a fixed sample size study
  - But if the trial continues to completion, sample size is larger because of p-value adjustment

Group-sequential design

- Conduct one or more interim analyses during the course of a study.
- Two possible decisions after each interim analysis:
  - Continue the trial as planned.
  - Terminate the trial
- Control overall Type I error rate.
  - Construct stopping boundaries that enable the trial to stop early if there is overwhelming evidence of efficacy,
  - Maximum sample size (sponsor commitment) is known up front
  - O'Brien/Fleming approach typical option as the penalty for conducting interim analyses is small.
- Generally well accepted by Regulatory authorities.

Benefits & limitations of group sequential

- Benefits
  - Very well established methodology.
  - Understood and accepted by regulators (ICH-E9).
  - Allows the flexibility to stop early for efficacy
  - Can vary timing and number of interim analyses
- Limitations
  - Interim analysis performed on the same endpoint at interim and final
  - Design focus is on maximum sample size, fixed in advance
  - Can’t amend the design e.g. to drop treatments or doses

TORCH trial

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 FEBRUARY 22, 2007 VOL. 356 NO. 8

Salmeterol and Fluticasone Propionate and Survival in Chronic Obstructive Pulmonary Disease

Peter M.A. Calverley, M.D., Julie A. Anderson, M.A., Bartolome Celli, M.D., Gary T. Ferguson, M.D., Christine Jenkins, M.D., Paul W. Jones, M.D., Julie C. Yates, B.S., and Jørgen Vestbo, M.D., for the TORCH investigators*

TORCH trial

- Trial comparing mortality in COPD
- Independent IDMC
  - Interim analysis for safety every 6 months
  - Two formal efficacy interim analyses
- Final analysis
  - Unadjusted p-value 0.041
  - Adjusted p-value 0.052

Adaptive Designs

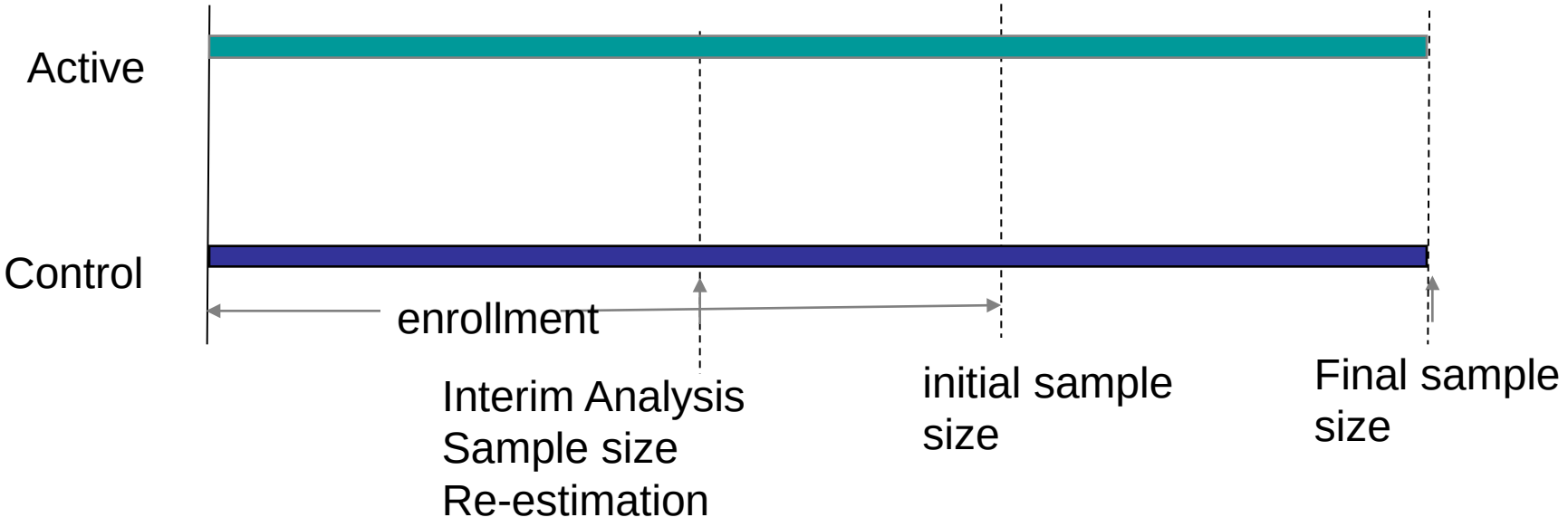
Definition

- **Adaptive Design** – any design which uses an interim analysis to modify aspects of the design (e.g. sample-size, number of treatment arms)
  - Type of design modification has to be pre-specified in the protocol
- Requires control of the type I error for regulatory purposes
- Requires assessment of homogeneity of results from different stages
  - Need to justify combining results from different stages

Sample size re-estimation

- Uncertainty about sample size assumptions. E.g. size of placebo effect
- Whenever possible, use blinded sample size reassessment e.g. total number of events
- Need to pre-specify size of treatment effect to be detected
- If based on unblinded analysis, need to show control of type I error

Sample size re-estimation



Group sequential vs. adaptive

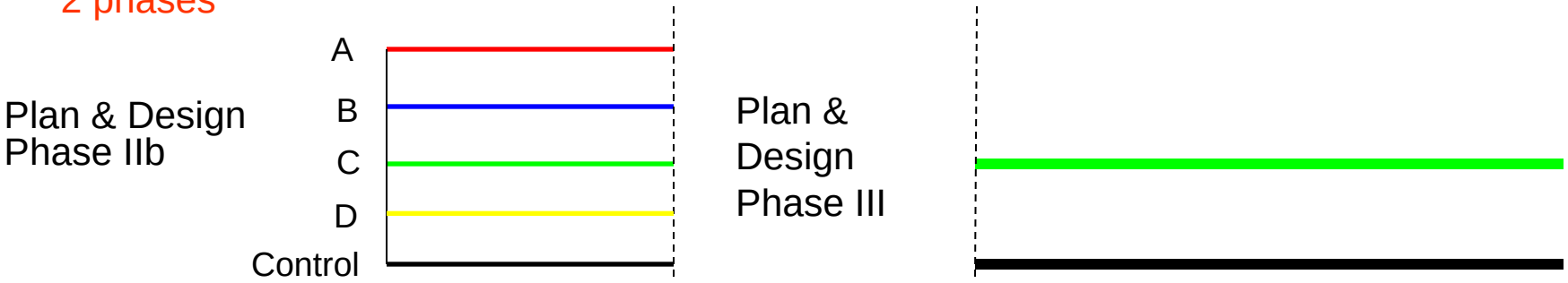
- Group sequential design: focus is on maximum sample size
  - Plan larger trial, stop early if unexpected large efficacy
  - More statistically efficient
- Adaptive design: focus is on initial sample size
  - Start smaller, expand if need to
  - More complex analysis may be required

Phase II / III trials

Standard
2 phases

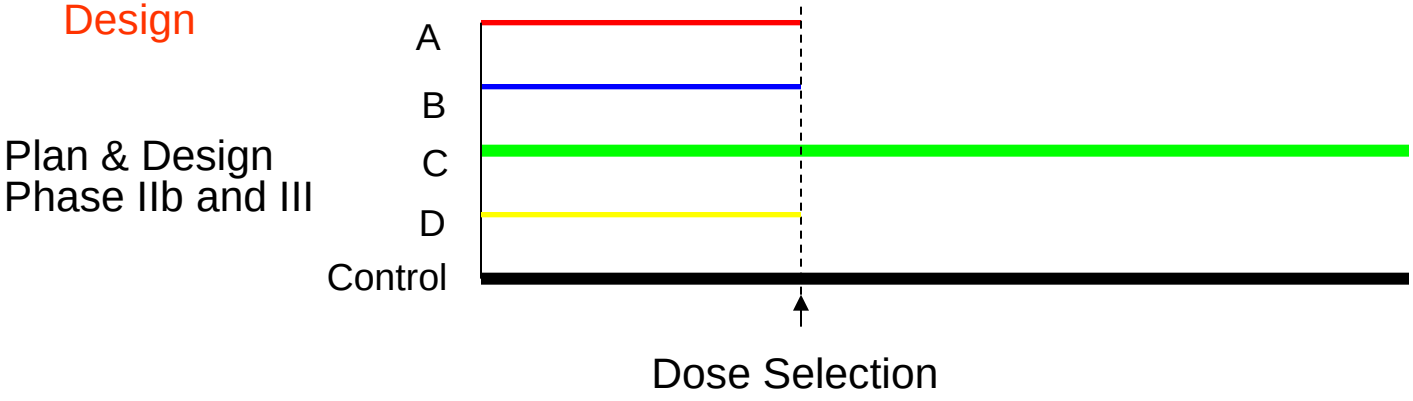
Learning

Confirming



Adaptive
Seamless
Design

Learning, Selecting and Confirming



Phase II / III trials

- Initially investigate multiple doses of experimental treatment
- Select dose to take forward based on interim analysis
- Only continue this dose and placebo for rest of study
- Requires careful control of type I error
- Can use short term endpoint for dose selection, longer term endpoint for confirmatory part of the trial

ELSEVIER

Contents lists available at ScienceDirect

Pulmonary Pharmacology & Therapeutics

journal homepage: www.elsevier.com/locate/ypupt

Integrating indacaterol dose selection in a clinical study in COPD using an adaptive seamless design

Peter J. Barnes [a,*], Stuart J. Pocock [b], Helgo Magnussen [c], Amir Iqbal [e], Benjamin Kramer [d], Mark Higgins [e], David Lawrence [e,1]

Indacaterol trial

- Stage I (N = 115 per group, 7 groups)
- 75, 150, 300, 600 mg indacaterol
  - vs placebo vs formoterol vs tiotropium
- Interim based on 2 week efficacy outcome
- two doses selected for to Stage 2
  - lowest dose meeting pre-defined efficacy criterion + next dose
- Final analysis performed after 26 weeks
- Careful control of type I error
- Second conventional phase III trial started in parallel after interim analysis

Phase II / III trials

- Other option, “non-inferentially seamless”
  - Two part protocol, Part A decides dose
  - Part B is confirmatory study but doesn’t use data from Part A in analysis
  - Avoids need for unblinded interim and alpha adjustment

Phase II/III trials

- Advantages of adaptive seamless designs
  - Increase of information value per patient
  - Shorter overall development time
- Issues
  - Number of treatment groups can change during trial with resulting implications in drug supply
  - Careful consideration of trial integrity issues (unblinding, consistency between stages)
  - Use of phase II/III designs misses opportunity to discuss/agree dose with regulatory authorities e.g. end-of-phase II or CHMP advice

Subgroup Analysis

Confirmatory subgroup analysis

- Generally requires pre-specification that a subgroup is expected to have larger effect
- Usually expected in the context of an overall positive trial
- Not usually possible to rescue a trial with overall non-positive result

Subgroup analysis

- Overall concern that the response of the “average” patient may not be the response of the all patients in the study
- Routine requirement for analysis by subgroup
- Aim
  - Identify patient groups with differential treatment effects
  - Assessment of internal consistency
  - Licence can be restricted if not sufficient evidence of a positive risk-benefit in the subgroup

Typical list of subgroups for analysis

- Sex
- Age
- Race
- Region
- Baseline severity measure 1
- Baseline severity measure 2
- Clinical events in the previous year
- Baseline medication
- Baseline blood biomarker

Multiplicity

- Results from analyses are interpreted as the true results for that group of patients
- Subgroup differences in treatment effect can arise by chance
  - Hard to identify what is a true difference
- Single subgroup with 5 levels, equal n, 90% power to detect overall effect*
- No true difference among subgroups
- Probability of observing at least one negative subgroup result = 32%

* Li Z, Chuang-Stein C, Hoseyni C. Drug Inf J. 2007;41(1):47–56

Classic example of dangers

- ISIS-2 trial aspirin vs placebo for vascular deaths
- Overall trial extremely positive for reduction in mortality
- Subgroup analysis by star sign
  - Gemini or Libra: adverse effect of aspirin on mortality
  - Remaining star signs: highly significant effect of aspirin on mortality

ISIS-2. Lancet 1988; 332:349-360

Multiplicity: is the difference real?

- Biological plausibility
  - Pre-definition
    - Differential effect anticipated
    - Plausible but not anticipated
    - Not plausible, hypothesis generating
- Consistency across endpoints
- Replication across two trials
  - But meta-analysis can still have subgroup problems

Design assumption

- Frequent assumption (by sponsors): patient population is homogeneous
  - Pragmatic approach for sample size determination
  - Should expect a consistent treatment effect
  - Anything else due to chance

- Alternative assumption (by regulators): treatment effect will vary between subgroups
  - Burden of proof to establish an effect in each heterogeneous subgroup is with the trial sponsor

Can we limit the number of subgroups?

- Design stage, pre-specification
  - Scientific rationale for heterogeneous effects?
  - Should separate trials be performed?
  - Pre-agreement with regulatory authorities on important subgroups may be helpful
- Need for subgroup analysis is related to the overall patient population
  - Sponsors may identify targeted populations
  - The more homogeneous the population studied, the fewer requirements there should be for subgroup analyses

How to assess results?

- Tests for interaction of limited value when investigating subgroup differences
  - Low power to detect heterogeneity
  - Still have 5% or 10% false positive rate
  - Hypothesis testing not appropriate
- Estimates and CI of size of interaction can be helpful to show what differences a trial can reliably estimate

Consistency of effect

- Alternative to interaction tests is to look at effect size in each subgroup
- Formal requirements have been proposed
  - e.g. that effect size in each subgroup must at least be positive
- All requirements are problematic

Subgroup analysis - summary

- Subgroup analysis is major statistical challenge
  - Hard to identify true effects versus false positives
- Pre-identification of important subgroups helpful for interpretation
- Subgroup analysis should depend on heterogeneity of the population
  - Less requirement when population is targeted
- Difficult to define consistency of effect
  - Interaction tests are of limited value
  - Requirement for each subgroup to show given level of effect is problematic

Peto [2011]

- “The appropriate interpretation of apparently different results in different subgroups of trial results is still one of the most difficult matters of judgement in the interpretation of randomised evidence”
- At present, many clinicians and regulatory agencies pay far too much attention to irregularities between the apparent effects in different subgroups

Missing Data

Missing data analysis

- Increased regulatory focus on missing data
- All statistical analyses where data is missing rely on untestable assumptions about unobserved data
  - Best strategy is avoidance
- Missing data more problematic if imbalance in withdrawal rates across treatment arms or characteristics of withdrawals different to completers

ITT analysis (De Facto estimands)

Two separate aspects:

- Including all randomised patients and all available on-treatment data (ITT Population)
- Assessing outcome regardless of whether the patient remained on the assigned treatment

First principle almost universally agreed

Second principle less well-understood,

  - either requires follow-up off treatment
  - or an assumption regarding missing data

Collection of data after treatment discontinuation

- Treatment discontinuation should not necessarily mean withdrawal from study
  - May need to follow-up subjects post-withdrawal from study drug for safety and key efficacy
- Academic consensus is strongly in favour of continued data collection
- CHMP missing data guideline
  - "Continued collection of data after the patient's cessation of study treatment is strongly encouraged, in particular data on clinical outcome"
- FDA and Europe now often request this
  - Ongoing debate whether required in all cases e.g. for symptomatic endpoints where effective medication is available to those discontinuing randomised treatment

Why is subject retention so important

- Missing clinical trial data is a key focus for regulatory authorities
- High levels of missing data can raise questions about integrity of a trial in general
- May negatively impact interpretation of efficacy and safety data
- Multiple analysis typically required, may show sensitivity of conclusion to missing data assumptions
- Requires a particular focus in long term or outcome studies

Prevention of missing data

- Focus on efforts to retain patients in trials
- Informed consent can allow for further follow-up contact off randomised treatment
- Designs can allow for multiple types of follow up, even if a subject no longer wishes to take IP
  - Contingency plans for collecting data for patients not attending visits
- Avoid withdrawal criteria where possible
  - Not all protocol deviations warrant exclusion from treatment or from the study.
  - Subjects should remain in the study unless there is a safety concern (even if the deviation is considered to impact efficacy)
- Monitoring sites for level of missing data

ITT analysis for normal data

- Historically analysis performed using LOCF (last observation carried forward)
- May not be a reasonable assumption for what happens when a patient discontinues
- Artificially increases sample size, does not reflect true variability of the trial
- Now discouraged by academics, less favoured by regulators

ITT analysis for normal data

- De jure analysis estimates what would happen if patient continued treatment
- Alternative approaches (de facto analyses) make assumptions about what happens to withdrawals e.g.
  - Active treatment withdrawals have similar future changes to placebo
  - Active treatment withdrawals jump to placebo mean

Some less obvious consequences…

- Apparent efficacy of a treatment will tend to reduce over time as withdrawals only increase, regardless of pharmacological effect
- Apparent efficacy in a subgroup will depend on withdrawals rates in the subgroup

Missing data

- De facto analysis often now required for both FDA and Europe
  - Alternative ideas exist, no standard analysis approach yet
  - Lack of robustness may mean the trial is not viewed as positive
  - Methods for some types of data not well developed
- Field is moving quickly, advisable to proactively address the issue in regulatory advice
- Best solution is to minimise missing data as far as possible