

12th CardioVascular Clinical Trialist (CVCT) Workshops

Workshop on Cardiovascular Clinical trials

Luxembourg, Sunday 6th – Monday 7th December 2009

“Scientific Programme endorsed by”



Workshop Chairmen : David Gordon (Bethesda, MD, USA)
Desmond Julian (London, UK)
Bertram Pitt (Ann Arbor, MI, USA)
Faiez Zannad (Nancy, France)

EXECUTIVE SUMMARY

Topic I: Development of new agents for the treatment of heart failure

Unmet needs in heart failure

- The level of poor post-discharge outcome in Acute Heart Failure Syndromes (AHFS), including in HF patients with preserved ejection fraction (EF), remains unacceptably high despite use of standard therapies, including concomitant use of multiple drugs.
- In no other area of medicine does the patient appear to respond well to initial therapy only to succumb within 60-90 days to a subsequent catastrophic event. Consequently, solutions to this problem are being sought.
- Treating an acute effect might not be expected to have long-lasting benefits, as drugs given during the acute phase (e.g. β adrenergic agonist agents), increase the oxygen demand on the heart or reduce blood pressure – both effects that are deleterious to the heart.
- A better understanding of the pathophysiology of acute HF is needed. Furthermore, at present, physicians treat isolated aspects of the condition, but not the whole system. Little is known about the metabolic aspects of HF; a better understanding of these and their management is desirable, while a return to consideration of the heart as the main target for therapy may be warranted.
- The implementation of amended treatment guidelines may improve AHF performance measures since current targets for treatment success are set too low.
- There is a need to reassess the methodology to develop drugs and protocols for the management of acute HF, as all clinical trials conducted to date in this area have been unsuccessful. In future studies, the main pathophysiology, study population, end-points and study execution should be appropriate for the properties of the drug being investigated..
- Digoxin, unlike newer agents for AHFS, has been shown to improve haemodynamics in HF. Therefore, reassessing 'old' drugs with potential benefits in acute HF may be of value.
- Few studies are conducted sufficiently earlier in HF, e.g., in stage B, such that there is a disconnect between clinical trials and real populations.
- Systolic and diastolic heart failure are considered by some as different phenotypes of the same disease, but this remains controversial. HF patients may have a continuum of disease reflected in EF values; clinical trials, to date, have concentrated on systolic dysfunction and use powerful medications in the most severely affected patients to enhance end-point numbers. An expansion in focus to diastolic disease is now needed.
- Prevention of HF through management of hypertension, metabolic syndrome and obesity may be important. For example, if hypertension in the elderly (>80 years of age) is managed effectively, a 64% reduction in risk of HF can be attained.
- Studies are underway to investigate whether better management of diabetes will reduce the incidence of serious cardiovascular (CV) outcomes. Some glucose-lowering drugs are known to cause HF. Considering risk/benefits rather than overall deleterious effect in HF may be of value when assessing glucose-lowering drugs.

Are trials with “add on therapy design” the only way to go?

- The expectations of regulatory authorities’ in the design of clinical studies to investigate drugs in HF are based on *The 1999 Guideline on Medicinal Products* for the treatment of (chronic and acute) HF.
- The introduction to the guidelines states that:
 - HF is a heterogeneous syndrome
 - most chronic compensatory processes are detrimental and should be targeted by current treatments.
 - specific claims are envisioned (e.g. improvement in symptoms and CV morbidity; or reduction in mortality).
- Primary endpoints for clinical studies in chronic HF are suggested; an addendum to the guidelines also lists suitable endpoints for acute HF.
- In specific cases, depending on the nature of the product, a “partial add-on” may be acceptable (e.g. vs angiotensin antagonists, vs β -blockers, etc.)
- It is difficult to consider designs that differ from true add-on therapies. For example, added survival benefit is seen in adding a β -blocker to an ACE inhibitor (ACEI)/angiotensin II receptor blocker (ARB). Further benefit is gained by the addition of an agent like spironolactone to the combination, making the cessation or removal of any of these agents difficult to justify. Unequivocal evidence of comparable mechanism-of-action in early clinical trials of a novel compound, such as evidence of biomarker changes or a novel mechanism that combines the benefit of more than one existing therapy, could justify consideration of drug withdrawal in the study design.
- However, studies into the effects of the varied combinations being given have not been done and even drugs within the same class may have different pharmacological effects. Therefore, patients’ best interests might not be optimally served by simply adding drugs on to existing regimens. Limiting the ‘cocktail’ of drugs or use of devices might be more effective and/or safe than the add-on regimen being assessed.
- Genetic markers might allow differentiation of those who could benefit from each drug.
- It should be remembered that any effect noted in a clinical trial may not be maintained in reality and that drug benefits change with time and may need to be reviewed.
- Standardisation of background therapy in international trials, although desirable, is likely to prove difficult to achieve due to regional, local and even study centre differences.
- The management of acute decompensated HF should be investigated, because it may represent a paradigm shift in the treatment of HF; currently there are problems with disease definition and there is no specific therapy for its management beyond the empiric approach.

The use of biomarkers in heart failure trials (for patient selection and/or surrogates)

- Biomarkers may aid in the selection of patients for a specific therapy, may be the target for therapy and/or may measure the response to therapy. As surrogate endpoints, they are unlikely ever to be suitable for Phase III studies but can help to facilitate drug development provided the following criteria are met:
 - reliable and available analytic method(s) allow their measurement
 - their underlying pathobiology is well-defined, plausible, and ideally part of the causal pathway
 - minimal biologic variability; any that does exist should be unrelated to the outcome/pathobiology of interest
 - strong evidence exists for an association between biomarker and outcome – this relationship can be seen under the study conditions in which modifying the outcome may be associated with a directionally similar modification of the biomarker measure
- Predictive biomarkers, as opposed to prognostic markers, are markers of a more profound therapeutic effect than simple improvement in symptoms; preliminary evidence for the efficacy or value of a predictive biomarker requires a test of interaction between treatment and marker - evidence that a marker is prognostic is not sufficient.
- Well-designed, randomised clinical trials (RCT) provide the strongest evidence that a marker is predictive, and it is recommended that RCTs should include both marker (+) and (-) subjects. Evidence that a biomarker is predictive should be stronger than results of a single retrospective analysis before proceeding to a confirmatory Phase III trial.
- Some biomarker-based endpoints are more plausible than others (e.g. myocardial infarction (MI), glomerular filtration rate (GFR), haemodynamics, and weight). Investigators should be willing to accept biomarkers such as GFR that can be proven to reflect organ function and not non-haemodynamic effects; whenever Quality of Life (QoL) is used as a marker, the scale used should be formally validated for the subject population using the appropriate statistical tests for internal consistency, etc., that now exist.
- Multiple markers have been used in HF trials to assess the effects of drugs that target the various pathways and processes affected by HF. For example, increased concentrations of brain natriuretic peptide (BNP) have been associated with left ventricular (LV) systolic dysfunction and correlate with increased mortality in chronic HF patients. Although assessment of BNP in HF clinical trials has been undertaken, only two of the six trials conducted have found that BNP-guided assessment of care made a significant difference to the study outcome.
- Heterogeneity may be a factor limiting the utility of BNP as a marker, as increases in its level are found in conditions unrelated to HF and mortality. Unfortunately, investigators are becoming

increasingly reliant on BNP as a marker. Ideally, biomarker status should be assessed after the clinical diagnosis is made, to better understand its role.

- All those undertaking clinical studies should collect and store blood samples to facilitate future biomarker development.

Endpoint related issues: Unconventional and patient related outcomes vs. survival

- Survival is an important endpoint for the registration of new CV drugs for HF and for their gaining HF guideline acceptance.
- Composite endpoints containing outcomes such as hospitalization for HF or CV hospitalization are of value in reducing sample size but, due to variability (e.g. in the criteria for hospitalization), composite endpoints need to be standardised and controlled.
- There is no reliable or generally acceptable surrogate endpoint for new drugs for HF. Patient-related outcomes (e.g. dyspnoea, exercise tolerance, peak VO_2 , QoL) are limited by patient variability and the lack of standardization of the evaluating instruments, and these outcomes generally have little clinical influence in HF medicine.
- Nevertheless, unconventional and patient-related outcomes are becoming increasingly important in the early evaluation of new HF agents, although they do not yet reliably predict the outcomes of large scale RCT.
- The increased use of such surrogate markers in Phase II studies, and the difficulties in interpreting their meaning, calls into question the value of Phase II study data in predicting the outcome of Phase III studies. More discrimination and scepticism in the assessment of Phase II data are needed when making decisions to move to Phase III, as even weak signals from Phase II studies are psychologically attractive.
- Biomarkers may have a role to play in the future, but further investigation of their value is needed. It might be that studying sufficient numbers of biomarkers will inspire confidence in their value, particularly in the identification of patients likely (or unlikely) to benefit from therapy.
- Dyspnoea as an endpoint in HF trials is considered to be a rapid means of assessing whether the treatment is having a symptomatic effect, and it is likely to continue as a required endpoint until more reliable short-term indicators are identified.

Topic 2: Risk-benefit decisions for anti-thrombotic therapies: Clinical and regulatory challenges

Selecting the appropriate study design: Positive control vs. non-inferiority comparative trials. Which way to go?

- Non-inferiority (NI) (equivalence) trials are increasingly common and enable new treatments to be assessed against active controls provided a pre-specified NI margin (Δ) falls outside the 95% CI of treatment difference (usually defined by the upper CI margin, M_1). Thus, choice of Δ is a key issue – if it is too generous, the NI outcome is questionable; if it is too small, very large numbers of patients are needed in the study.
- The patients, the active control treatment, the primary outcome to be assessed and the length of follow-up must all be chosen appropriately to ensure meaningful results are obtained from NI studies.
- As NI studies usually use an intent-to-treat (ITT) analysis, patient losses can create problems and can diminish real differences between treatment groups. Regulatory authorities are increasingly asking for both ITT and per protocol (PP) analyses, although the data set which is used as the foundation of their ruling is not clear.
- Unless the effect size of the active control is large the design of NI studies is challenging because the active control effect is an assumed value and trialists are conservative in their assumptions.
- Also, because NI assumptions are based on outcomes of previous RCTs, trialists need to ensure that clinical practice as well as study endpoints and their measurement have remained constant. For example, clinical practice in HF has changed over time, so it would be impossible to conduct NI trials of new vs 'old' HF drugs.
- Additionally, the control against which a therapy is assessed for non-inferiority may be queried, as happened with the ON-TARGET NI study. This study used the HOPE RCT as its control. However, the outcome from the TRANSCEND RCT, which showed a smaller benefit of the therapies being assessed compared to those seen in HOPE, potentially undermined the statistical basis on which ON-TARGET was based.
- No such assumptions are needed in positive control superiority studies that aim to show a difference between therapies, although use of suitable and current comparators is just as necessary.
- Superiority of an agent can be tested against placebo, against placebo in add-on studies, in dose-response studies or against an active control; the results of each of these studies are easily interpreted.
- Superiority studies compare different pharmacological effects, the potency of the same pharmacological mechanism, and the suitability of the study drug in patients intolerant of the standard therapy.

- Even where the advantage of a new drug is safety, efficacy also has to be demonstrated and persuasive NI or superiority data are still required.
- A critical distinction needs to be made between safety studies, where an arbitrary limit is placed on the level of similarity between agents, and NI studies. Unlike safety studies, NI studies require that the loss of effect of the active control must be excluded in order to show non-inferiority,. Determining this value accurately is therefore crucial. One means of doing this is use of the 95-95 rule (95% CI of the lower bound x 95% CI [control – test drug]). It is widely accepted that assessment of loss of effect of the active control must be considered in NI studies; however, due to its recognised limitations, use of the 95-95 rule remains controversial.

Risk benefit issues in progressing to Phase III. Do surrogates and/or adaptive design help?

- Studying multiple doses in a Phase III trial is expensive and may lead to more complex instructions for use. However, choice of doses is a cause of concern, particularly in anti-thrombotic medications.
- While there is no clear relationship between anti-coagulant or anti-platelet activity and clinical endpoint event rates, the doses selected in Phase II studies aim to minimize bleeding rather than to assess clinical event reduction. Usually, any efficacy data are of limited value as studies tend to be underpowered.
- In the effort to balance efficacy and safety, Phase II studies of anti-thrombotic drugs have essentially become safety studies. Consequently, the single dose that is taken into phase III studies provides inadequate evidence of the optimal use of the drug.
- In studies conducted to assess new oral anti-thrombotic agents, these agents were tested as add-on therapy to anti-platelet therapy (aspirin and clopidogrel). Although there was a trend towards a reduction in ischaemic events and fewer bleeds with use of oral anti-thrombotic agents, this was compromised by the presence of clopidogrel. Thus, due to ethical considerations in the trial design, safety appears compromised and the likely benefit diminished.
- Clearly, there is a need to amend the clinical trials design. The FDA preferred approach is:
 - to use phase II to confirm the operating range of activity and the limits to dose (death, stroke)
 - to take multiple doses into phase III
 - to preserve strong α control on whether the drug is effective
 - analysis by data-regression or pooling some or all of the data
 - strong α control by dose is not essential
 - avoid over-valuing bleeding, especially by the Data Management Committee (DMC).
- Such adaptive study design is needed to preserve the integrity of the main success criterion. Therefore, it is highly desirable that algorithm-driven sample size adjustments are made throughout the first-phase of the Phase III study in response to any event that occurs during this time. An interim evaluation of variability would allow the criteria established in the first-phase to be assessed. Such seamless designs may have logistical usefulness, even if data derived from the first-phase clinical outcomes cannot be used in the final outcomes assessment.
- Conducting seamless Phase II/III studies may be a way forward in the assessment of anti-thrombotic agents. However, without cardiovascular outcomes data, determining where particular doses might lie on a safety/efficacy curve could be problematic, thus making dose selection for Phase III study more difficult.
- Pooling data for analysis could offer a possible solution, although it also appears controversial, as this might affect the α value. As this would be done purely to assess safety, there is no inherent

disadvantage in combining doses. However, pooling of doses could interfere with the potential to identify the optimal dose for efficacy and with the control of type 1 error. Possible solutions could be to conduct an interim analysis to assess the various drug doses being used and to use statistical methods to compensate or protect α , but these are not currently endorsed.

Bleeding outcomes: Definitions and adjudication issues

- Different definitions of bleeding across trials and variations in the way that bleeding data are captured make comparisons between studies difficult.
- Harmonization of collection and reporting of bleeding data in trials of anti-thrombotic drugs would be welcomed not only to facilitate comparisons between trials, but also to standardise the role of the DMC in recommendations for dosage alteration or study cessation.
- It was recommended that bleeding results should be presented in grid-like tables, by location and amount of blood loss, such that different definitions may be applied; this format is similar to that adopted by the ESC, the ACC, the AHA and the WHO in their bleeding guidelines.
- Such standardised reporting may facilitate modelling of the bleeding events that occur during anti-thrombotic trials, assessment of whether small bleeds precede major bleeds, and separation of the effect of interventions on total bleeding. A detailed grid would also allow events to be followed more closely, thereby sensitising DMCs to trends that might need attention.
- Prospective guidelines on what constitutes excessive bleeding should be created and rules for stopping studies should be in place at the study outset. The role of the DMC is to alert the Steering Committee should this threshold be reached; they should not make *ad hoc* decisions on what level of bleeding is excessive. Any bleeding without perceived benefit from the drug is considered unacceptable.
- Assessment of the bleeds by independent adjudicators would assist in their classification. This is of particular value where trialists are non-cardiologists. This in turn can reduce the rate of false positives, so that relative risks can be estimated without bias; it will have no bearing on the number of false negatives, however.
- Establishing clear guidelines on bleeding events for trialists would also correct for the currently-suspected systematic under-reporting of safety events.

Combining efficacy and safety in composite endpoints

- A review of several clinical trials investigating the CV benefits of anti-thrombotic therapy using a combination of efficacy and safety endpoints yielded several interesting findings.
- 'Net clinical benefit' is not a substitute for benefit-risk.
 - In SPORTIF III, a study of stroke prevention that compared ximelagatran and warfarin in 3,407 patients with atrial fibrillation (AF), the primary endpoint was stroke and systemic embolic events. Results for the primary endpoint were not significantly different between therapies, but combining the primary endpoint with major bleeds and death generated a statistically significant benefit for ximelagatran compared to warfarin. However, there were 24 MI events in the ximelagatran group compared to 13 in the warfarin group; a component not included in the combined endpoint.
- Time has an effect.
 - The RITA 3 study compared early intervention and conservative treatment in 1,810 patients with acute coronary syndromes (ACS). This study had two primary endpoints; death, non-fatal MI and refractory angina at 4 months and death and non-fatal MI at 1 year. Despite benefit being noted in the early intervention group at 4 months, this difference was no longer present at 1 year.
- One component can drive the results.
 - The ACTIVE study to prevent vascular events in 7,554 patients with AF compared clopidogrel plus aspirin with aspirin alone. The primary endpoint (stroke, MI, non-CNS systemic embolism, death from vascular causes) was statistically significant in favour of clopidogrel. However, this result was driven principally by the incidence of ischaemic stroke.
- Using more endpoints is not always better and results vary depending on the combination of endpoints.
 - In the PROactive study, a study to determine whether there were treatment effects on the complications in patients with type 2 diabetes, the primary endpoint had eight components. No significant difference was noted between treatment groups when the combined eight-component primary endpoint was assessed. However, a statistically significant difference was noted when just three components of the primary endpoint were combine for assessment.
- Thus, combining safety and efficacy endpoints can make interpretation of study outcomes problematic, potentially diluting the actual benefit as components may respond differently to the test drug at different times and may have opposing pharmacodynamic effects.
- Composite endpoints (efficacy, safety and efficacy, and safety) should be used cautiously. Trading efficacy for safety may be acceptable in some circumstances. It is recognised that such endpoints will enable a trialist to 'gain' on safety and to increase the number of events reported.
- Such combinations should be discouraged by regulatory authorities and, at the very least, results for individual components should be reported separately.

- Current assessments of safety events are highly subjective. However, if the magnitude of safety events was to be graded according to prospectively-agreed criteria, outcomes analysis would become difficult.

Topic 3: Need for outcomes in the development of anti-diabetic medicinal products

Cardiovascular efficacy outcomes in diabetes trials

- To date, no benefits on macrovascular outcomes have been demonstrated in patients with type 2 diabetes.
 - Three large outcome studies (UKPDS, ADVANCE and ACCORD) demonstrated that controlling blood glucose levels in patients with type 2 diabetes produced no mortality benefit.
 - In the ACCORD study, intensive blood glucose-lowering therapy was associated with increased CV risk compared with standard blood glucose-lowering treatments. (In a sub-analysis, however, patients with low baseline HbA_{1c} values and no history of CV disease were at a lower risk of CV events than those patients with high baseline HbA_{1c} levels and a history of CV disease.)
- In the Diabetes Control and Complications Trials (DCCT) study in type 1 diabetes, it was 8 years before any macrovascular benefit was demonstrated. However, the UKPDS 10-year follow-up clearly demonstrated that tight glucose control reduced microvascular complications but had no effect on macrovascular CV disease. Thus, far longer study durations than are currently undertaken may be required to obtain significant improvements in macrovascular outcomes in type 2 diabetes. .
- It is questionable whether pharmaceutical or other sponsors will embark on such long-term studies.
- Surrogate markers for macrovascular disease are needed. Carotid intimal thickness correlates with coronary ischaemia and is a potential candidate that should be investigated further.
- Relevant efficacy endpoints must be used in non-inferiority studies that investigate diabetes and CV events. Endpoints that are not directly related to the actual mechanism of the disease can affect the study's outcome – and, in diabetes, the mechanism underlying CV outcomes remains unknown.
- To establish superiority, efficacy endpoints need to be comprehensive. Considering those used in the ACCORD study (death, stroke, and MI), approximately half the deaths were not CV in nature, therefore including this in a primary composite efficacy endpoint would skew the results.
- Indeed, composite endpoints can become meaningless in terms of superiority. Those conducting diabetes trials should recall that diabetes is not a purely CV disease and should be cautious when emulating CV study designs.
- Ultimately, as at least 50% of the complications associated with type 2 diabetes are microvascular in nature, it may be more appropriate to focus on microvascular outcomes as endpoints in clinical studies assessing anti-diabetic medicines.

Cardiovascular safety outcomes in diabetes trials

- A meta-analysis of rosiglitazone and CV risk indicated an apparent excess of MI in patients treated with this drug. Subsequent publication of the interim and final results of the RESULT study identified HF as the real problem. Consequently, the glitazones are not recommended in high-risk patients.
- To avoid such a safety issue arising in future, the FDA produced proscriptive industry guidelines requesting the evaluation of CV risk in all new anti-diabetic therapies. In short, these guidelines require that, prior to submission, studies demonstrate non-inferiority to a standard therapy or comparator with a risk ratio (RR) margin of 1.8 (i.e. 95% CI excludes 1.8).
 - if this cannot be shown, a large safety trial should be conducted
 - if the upper CI boundary is between 1.3 and 1.8, then a large post-marketing trial is required to show non-inferiority with a RR margin of 1.3
 - if the upper CI boundary <1.3 , no such trial is required.
- Two approaches can be taken to acquire these data: a single large trial or a meta-analysis. Regardless of approach, safety trials are likely to dominate research into anti-diabetic medicines, potentially at the expense of efficacy data.
- A number of statistical issues arise due to the complicated set of choices that have to be made, particularly in relation to any corrections being made to the CI generating α . Some think it is unlikely that there will be an α penalty to test the RR=1.8 hypothesis, unless multiple looks at the data are taken; for RR=1.3 a penalty should be paid for looking, however this is by no means assured. Moving to claim for superiority, if a study is adapted once results look promising, this should generate no α penalty. However, as these are essentially safety studies, there may be no α penalties at any stage. Clearly, the need for consensus on the statistical aspects of such studies is needed.
- Furthermore, there is huge debate about whether it is permissible to add studies to a meta-analysis in order to attain targets, if these should fall just short of the stipulated RR. Whether such decisions should be based on firm guidelines or should follow consultation with regulatory agencies remains an unanswered question.
- Additionally, a purely safety trial may raise issues about the ethics of recruiting patients into a trial that does not offer an efficacy benefit.
- If such a trial is acceptable, there is considerable debate about the type of patients to be selected; e.g. high-risk patients (for maximum events) or those in the earliest stages of diabetes (to potentially shorten the length of the safety study). Consensus on this matter is needed.
- Concerns were expressed about prolonged drug use and its effect on CV risk. To avoid CV 'surprises', it was suggested that every CV drug should have a HR <1.5 before approval.
- CV events in patients with diabetes represent a multi-dimensional problem, and factors other than the type of anti-diabetic drug given need to be considered when designing clinical studies.

- In creating the guidelines for anti-diabetic agents, the FDA may have been overly cautious. Such measures will create unnecessary anxiety among patients and physicians, and are likely to inhibit drug development in this field in the future. For every licensed product, risk management should already be stressed and creation of good quality post-marketing surveillance should be a priority.

Macro and microvascular endpoint definitions

- Macrovascular outcomes used in clinical studies include CVD death, non-fatal MI, non-fatal stroke, non-fatal HF, resuscitated cardiac arrest, revascularization, angina (unstable, worsening, new) and transient ischaemic attack (TIA).
- The microvascular outcomes usually assessed include retinopathy, microalbuminuria (nephropathy), neuropathy and, more recently, vascular dementia.
- Because these endpoints possibly share common pathologies, it is tempting to combine them. However, it is likely that macrovascular complications precede microvascular effects, and tight glycaemic and blood pressure control have only been shown to be beneficial for microvascular complications. . Use of CV drugs has been of limited benefit, however, in modifying macrovascular endpoints.
- Retinopathy is used a marker for systemic micro- and macrovascular diseases including CHD, stroke and HF. However, retinopathy may be due not only to diabetes but also to raised blood pressure, and it can also occur in 10% of people with normal glucose and blood pressure levels.
- Results from clinical trials investigating the role of anti-diabetes drugs suggest that retinopathy is a useful endpoint, although the different methods used in its assessment make comparisons between studies difficult, and its precise definition is complicated.
 - In the Diabetes Retinopathy Candesartan Trial (DIRECT), candesartan reduced the incidence of retinopathy but had no effect in its progression in type 1 diabetes; in type 2 diabetes candesartan reduced the risk of progression and increased regression of retinopathy.
- Tight glucose control is beneficial in reducing microvascular disease outcomes in diabetes whatever the drug used (sulphonylurea, metformin, insulin, etc.)
 - In the DCCT study, tight glucose control had an impact only in those patients entering the study with no signs of retinopathy: it reduced incident retinopathy by 75%, but reduced progression by 50%.
- The time-course of microvascular benefit is protracted and can take 7 years or more to become apparent.
- In nephropathy, the results from the MICRO-HOPE study suggest that use of an ACEI improves CV and renal outcomes in patients with diabetes.
- Extremely large studies are required to detect small improvements in macrovascular outcomes . Therefore, it is important to develop, define and use new endpoints (e.g. endothelial function, FDG-PET scanning, dynamic contrast-enhanced MRI, MSCT-Scan, IVUS imaging (Palpography, Radiofrequency analysis), to assist the selection of new compounds.

- The endpoints used to study diabetes may need to be different from those used in CV studies. It would be interesting to determine whether different types of HF are associated with different pathophysiological aspects of diabetes and hypertension.
- Dementia as an endpoint in CV studies is gaining interest among clinicians, as it is recognised that stroke patients are at increased risk of dementia – thus, vascular dementia could serve as a macrovascular endpoint in CV studies. However, white matter disease, which also produces dementia, is a compelling microvascular endpoint. Assessment of dementia progression is difficult at present, but tremendous activity is ongoing in the search for surrogates.

Topic 4. Current limitations of the registration data package

- To take account of the fact that clinical trials do not truly reflect day to day clinical practice, and because the data package at submission is based on clinical trial data, risk management plans are put in place to assess and monitor risk and benefit, and to assess how the drug is incorporated in medical practice.
- Clinical practice guidelines also assess a product's utility. Such guidelines make recommendations based on level of evidence and expert opinion, and there is often a lack of direct data supporting their recommendations on use of drug combinations.
- Occasionally, regulatory authorities and clinical practice guidelines make different or contradictory recommendations. For example, an unlicensed product may be used in clinical practice either because physicians have found that it offers benefits outside those listed in the SPC and include it in treatment guidelines, or because pharmaceutical companies seek to influence support of a licensing application through inclusion of its product in a set of guidelines.
- Data packages are highly variable. In the CV area generally, packages comprise large clinical trials as part of the initial submission.
- Data packages frequently lack data on older people, on high-risk patients and on women, and as such are not representative of the population as a whole. There is no systematic insistence on inclusion of these patients at present, although it is known, for example, that antihypertensive medications behave very differently in the elderly or that patients with concurrent kidney disease might benefit from therapy.
- Good quality comparisons of multiple drugs of the same class are rarely submitted to regulatory authorities. The FDA has requested the NIH to conduct such trials.
- Similarly, dose-response data are generally poor and doses can seldom be compared across studies.
- Combinations of antihypertensive therapies will only be approved as first line therapy if a definite benefit can be proven in a specific population.
- Limited experience of the value of pharmacovigilance studies exists in Europe at present; in the US these have facilitated the recognition of a safety issue and have led to requests for further studies to be undertaken.

Topic 5. Comparative effectiveness. Pragmatic trials of alternative intervention strategies (eg SYNTAX, RITA 3, COURAGE...)

- The Congressional Budget Office defines Comparative Effectiveness Research (CER) as ‘a rigorous evaluation of the impact of the different options available for treating a medical condition’.
- Many RCTs do not treat patients as they would be treated in real-life. Thus, there is a need for CER in the form of pragmatic trials that assess drugs in a real-life setting, which means there should be broad eligibility of patients entered, flexible interventions, typical practitioners, no follow-up visits, objective clinical outcomes, usual compliance and ITT population assessment.
- CER generally tests strategies rather than specific interventions, which means that findings may be useful for extrapolation to common clinical situations. However, while strategies tested may be clinically relevant, crossovers may be so prevalent and uncontrolled that results are not interpretable. Furthermore, any difference in intervention at the outset may limit the strength of the conclusions.
- CER study populations are relatively representative of real-life because they minimise exclusions, but confounding variables may limit the interpretability of any findings.
- RITA, COURAGE and SYNTAX are pragmatic trials assessing PCI/CABG. While it was agreed that these studies represent the real world and make physicians rethink their approaches to patient selection, it is difficult to establish the benefit or risk of any therapy due to factors such as the levels of heterogeneity either in the study populations or in the therapies utilised, frequency of cross-overs within a study, lack of blinding and ITT analyses.
- Thus, it is difficult to envisage how these studies could drive cost-effectiveness or budgetary decisions.

Topic 6. Alternative to Randomised Clinical trials?

Registries vs. trials

- Consistency between observational studies and clinical trials supports extrapolation of hypotheses or findings from one to the other.
- Occasionally, because real world and clinical study settings are different, the two may produce disparate or misleading findings. For example, a small clinical trial suggested there might be increased risk of cardiac death and MI with drug-eluting stents (DES) compared with use of bare-metal stents. These findings were supported by two meta-analyses and a registry study – and then contradicted by the results from a pragmatic study and another registry study. Thereafter, a larger RCT showed there was no difference in risk between the two types of stent.
- A meta-analysis of data from 21 RCTs vs. 31 registries was undertaken to try and understand why such conflicting results were being generated. The combined RCTs meta-analysis indicated no difference existed between the stents, but the registries meta-analysis showed that DES were associated with a lower mortality than bare-metal stents. Thus, when comparing stents, something fundamental is awry in the assumptions being made about their use, most likely that mortality risk depends on factors related to PCI.
- When such discrepancies arise, only randomised treatment assignment can provide a reliable unbiased estimate of a treatment effect. With observational studies, the design is not experimental, treatment is not randomly assigned and, while adjustments can be made for identifiable differences, it is never certain that these are adequate.
- Observational studies may be of utility:
 - in assessing benefits when outcome is particularly poor and a large proportion of patients is likely to derive a benefit from the treatment
 - in assessing safety when the outcome of interest is rare
 - when ethical considerations, practicality, clinical judgment, doctors' or patients' unwillingness to participate make RCTs unrealistic
 - in assessing the long term transferability of chronic treatments tested in RCTs separately and for a relatively brief period of time
 - to provide more generalisable evidence about the effects of treatment involving populations of patients or clinicians that are more representative of clinical practice than those involved in trials.
- Observational research and RCTs are not competitive methods of assessing treatment effects on mortality and major morbidity.

Observational studies may provide complementary evidence to RCTs.

Bias vs. Real world representation

- Bias is often discussed as an increase of type I error rate, but it can be an artificial decrease or increase in type II error rate.
- A number of sources of bias exist:
 - trial design - randomisation improves comparability of treatment groups at baseline and blinding reduces many sources of bias
 - trial conduct, logistics - induce biases that are hard, if not impossible, to quantify
 - the statistical analysis method - induces bias associated with the probability of making an incorrect conclusion (e.g. inflate type I error rate) or with effect estimation (e.g. underestimate or overestimate a treatment effect)
 - inference generalisation - the direct application of the result of a RCT to the real world setting can be misleading due to difference between the settings
 - regional differences.
- Selection bias is of concern in both superiority and in NI trials.
- Both per protocol superiority and non-inferiority trials designed to compare responses in *proportions* of patients will experience bias if patients are lost to the study after study initiation, particularly if these losses occur more frequently in just one of the comparator groups.
- This presents difficulties for the study statisticians who would like to capture all patients enrolled but do not know how to deal with the problem of patient withdrawal, as this too would create a bias.
- As biases can easily and unknowingly be introduced into clinical trials, more attention needs to be paid to their avoidance.