Clinical Trials Methodology Conference 2011

Bristol, UK. 4-5 October 2011

Edited by Stephanie McAvoy

Published: 13 December 2011

These abstracts are available online at http://www.trialsjournal.com/supplements/12/S1

ADAPTIVE DESIGNS

A1
Adaptive designs: current status, future outlook
Andy Grieve
AptivSolutions, Robert-Perthel Strasse 77a, Cologne, Nordrhein-Westfalen, 50739, Germany
Trials 2011, 12(Suppl 1):A1

The world of pharmaceutical statistics has taken adaptive designs to its heart - at least in theory. However despite the very large number of methodological publications and presentations at conferences there are still very few examples of adaptive designs in the medical literature. In this talk I will discuss why this might be the case and will suggest areas where these approaches are likely to be of greatest value in the future.

A2
An adaptive seamless phase II/III clinical trial design incorporating short-term endpoint information
Nigel Stallard
Statistics and Epidemiology, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry, CV4 7AL, UK
E-mail: n.stallard@warwick.ac.uk
Trials 2011, 12(Suppl 1):A2

Adaptive seamless phase II/III designs enable a clinical trial to be conducted in stages with the most promising of a number of experimental treatments selected on the basis of data observed in the first stage to continue along with the control treatment to the second and any subsequent stages. The main statistical challenge in such a design is ensuring control of the type I error rate. Most methodology for such trials is based on the same endpoint being used for interim and final analyses. In some settings the primary endpoint can be observed only after long-term follow-up. In this case it may be desirable to use short-term endpoint data along with any long-term data available at the interim analysis to inform treatment selection. If short-term data are available for some patients for whom the primary endpoint is not available, basing treatment selection on these data may lead to inflation of the type I error rate.

This talk presents a new method [1] that allows the use of short-term endpoint information for treatment selection whilst controlling the overall type I error rate. The method builds on the work of Galbraith and Marschner [2], who proposed a method for using short-term endpoints in interim analyses for comparison of a single experimental treatment with a control. The method is based on adjustment of the usual group-sequential boundaries [3,4] to allow for monitoring of a test statistic obtained from fitting a bivariate normal model to the correlated short-term and long-term endpoints. The model can be fitted either using a linear mixed model or by combining the results from separate linear regression models for the short and long-term data [5].

Result will be presented from a simulation study to investigate the properties of the new method. In addition to verifying control of the type I error rate, these results show that the use of the short-term endpoint data can lead to an increase in power when the short and long-term endpoints are correlated.

References

A3
Flexible trial design in practice – dropping and adding arms in STAMPEDE: a multi-arm multi-stage randomised controlled trial
Matthew R Sydes 1*, Nicholas D James 2, Malcolm D Mason 2, Noel W Clarke 2, Claire Amos 3, John Anderson 4, Johann de Bono 5, David P Deeney 6, John Dwyer 2, Gordana Jovic 2, Alastair Ritchie 7, Martin Russell 3, Karen Sanders 7, George Thalmann 8, Mahesh KB Parmar 9
1MRC Clinical Trials Unit, London, UK; 2CRUK Institute for Cancer Studies, University of Birmingham, Birmingham, UK; 3School of Medicine, Cardiff University, Cardiff, UK; 4The Christie and Salford Royal Hospitals Foundation Trusts, Manchester, UK; 5The Royal Hallamshire Hospital, Sheffield, UK; 6The Institute for Cancer Research and Royal Marsden Hospital, Sutton, UK; 7Prostate Cancer Support Federation, Stockport, UK; 8Beatson West of Scotland Cancer Centre, Glasgow, UK; 9Inselspital, Bern, Switzerland
Trials 2011, 12(Suppl 1):A3

Objectives: STAMPEDE is a multi-centre, randomised controlled trial designed with novel multi-arm, multi-stage (MAMS) methods. Here we: (1) describe the methodological and practical issues arising with the stopping of recruitment to some trial arms following an intermediate analysis and (2) describe the issues surrounding the addition of new research arms during the trial.

Methods: The trial recruits men with locally advanced or metastatic prostate cancer starting standard long-term hormone therapy. There are 5 research arms and 1 control arm. The trial has a pilot stage assessing safety...
and feasibility, 3 intermediate “activity” stages (I-III) where the outcome measure is failure-free survival (FFS) and one final “efficacy” stage (IV) with overall survival as primary outcome measure. At the end of each stage, each research arm is formally compared pairwise to the control arm. Accrual of further patients is discontinued early for any research arm either not showing sufficient evidence of activity or with adverse safety considerations; accrual continues to arms showing activity with acceptable safety. The stopping guideline compares the treatment effect against a pre-defined cut-off value using the hazard ratio when the hazards are proportional and restricted-mean survival time otherwise. This interim hurdle becomes increasingly stringent stage-by-stage. The addition of new research arm(s) can be actively considered when sufficiently interesting agents emerge. New research arms are compared only to contemporaneously-recruited control arm patients using the same intermediate guidelines in a time-delayed manner. The addition of new research arms is independent of any of the original research arms stopping accrual early subject to adequate recruitment to support the overall trial aims.

**Results:** (1) After the second intermediate activity analysis (March-2011), the IDMC recommended and the Trial Steering Committee ratified discontinuation of recruitment to two research arms for lack-of-sufficient activity. Nearly 100 recruiting centres in UK and Switzerland had to promptly implement these changes. Detailed advanced preparation meant that activation was swift and recruitment continued seamlessly into Activity Stage III; MHRA and REC approval was not required because this was already included in the trial design. (2) An application to add a new research arm has been successfully approved by the CRUKs CTAC funding committee and accepted by the relevant industry partner. Details on the methodological and practical issues and implementation of these changes will be presented.

**Conclusions:** The STAMPEDE experiences shows that recruitment to MAMS trial is achievable and that mid-flow changes to trial design are practicable with good planning.

**Acknowledgements:** On behalf of the STAMPEDE Investigators.

---

**A5**

**Model selection and sample size adaptation: HYPAZ trial**

Simon Bond1,*, Adrian Mander2, Duncan Jodrell3, Joseph Cheryan4

1Cambridge Clinical Trials Unit, UK; 2MRC Biostatistics Unit, Cambridge, CB2 OSR, UK; 3Medical and Pharmaceutical Statistics Research Unit, Lancaster University, Lancaster, LA1 4YE, UK. E-mail: james.wason@mrc-bsu.cam.ac.uk

Trials 2011, 12(Suppl 1):A5

Pazopanib is a new cancer drug that works by limiting the growth of new blood vessels in tumours. About half of patients who take pazopanib develop high blood pressure (hypertension). This side effect can make patients have to reduce or stop their cancer treatment, and can cause other health problems.

Understanding how pazopanib causes high blood pressure will help us to advise doctors how to treat the high blood pressure effectively, so that patients can continue to take their cancer treatment safely. A proposed explanation is that hypertension is caused by endothelial dysfunction as a consequence of reduced nitric oxide (NO) bioavailability.

The aim of a study (HYPAZ) is to investigate the relationship between blood pressure and NO bioavailability in cancer patients treated with pazopanib, who develop hypertension. The predictor variable observed is the change in blood pressure from baseline. The response variable will be change in NO bioavailability measured through the change in forearm blood flow ratio (of infused versus control forearm) in response to intra-arterial acetylcholine infusion. These data will give a greater understanding to aspects of pazopanib’s mechanism and relationship to hypertension. The primary analysis for the study is to model and quantify the relationship between NO bioavailability and blood pressure. This patient study will be presented in the study there is little a priori evidence to suggest the shape, or to quantify the strength, of any relationship. Hence an adaptive design is proposed to assess the evidence for any relationship and judge possible sample size revisions at an interim analysis. To perform model selection at an interim analysis is a novel feature of this adaptive design. We consider some of the operating characteristics of the proposed design.

---

**A6**

**How Cancer Research UK is adapting to adaptive designs**

Julie Heam2, Nicola Keat, Kate Lav, Rowena Sharpe

Cancer Research UK Clinical Trials Team, UK

Trials 2011, 12(Suppl 1):A6

Clinical trials are central to Cancer Research UK’s purpose of carrying out world-class research to improve our understanding of how to prevent, diagnose and treat cancer.

The UK leads the world in the percentage of cancer patients entering clinical trials. In 2009, this figure reached an all time high of 16.8% of NHS cancer patients, of which ¾ entered a Cancer Research UK (CR-UK) supported trial (n=31,000 patients). Consequently, CR-UK has a major impact on the availability of treatment trials for UK cancer patients.

The strengths and limitations of randomised controlled trials have been discussed at length in the medical literature. In his 2008 Harveian Oration entitled ‘On the evidence for decisions about the use of therapeutic interventions’, Professor Sir Michael Rawlins made a plea for investigators to continue to develop and improve their methodologies and for decision makers to avoid adopting entrenched positions about the nature of evidence. The clinical trials community has responded by proposing innovative trial designs and funders are considering how to respond to the challenges of funding trials with novel designs.

Multi-arm, multi-stage trial designs, for example, can increase the chance of a single trial providing a positive results and saves time and potentially money compared to separate sequential trials. CR-UK already funds several trials of this design, including STAMPEDE in prostate cancer and ICON 6 in ovarian cancer. We also fund trial the AML16 trial in Acute Myeloid Leukemia which adopts the complex design to evaluate a number of agents concurrently dependent on the characteristics of patient subgroups and their response to treatment. The RATH-L trial in Non-Hodgkin’s Lymphoma adopts another innovative design, evaluating both the use of PET scans to determine treatment pathways and sub-randomisations to less or more intensive treatments. Trials such as these present issues that funders need to adapt to, including (1) educating our funding committees,
Our experiences with this phase II study emphasise the need for robust methods of analysis (e.g. Kaplan-Meier plots) for early and final outcomes, together with considerable effort for harms unacceptably high. Bayesian Hierarchical Models [2] may provide a better way to deal with the multiplicity of outcomes and secondary outcomes, though the use of composite outcomes and multiple "primary" outcomes, together with considerable numbers of secondary ones suggests that the issues are not resolved even for efficacy.

Pre-specification of adverse effects can rarely be done, at least not with any completeness. New ADRs that may be relatively rare may surprise investigators by their occurrence. Very much more statistical effort has been applied in analysis of efficacy than for analysis of harms, yet it is the absence of harm (safety) that is a primary concern of patients. Evidence both from surveys and even trials show that many patients have unrealistic expectations regarding benefits and the (total) absence of harms [1].

Conventional (Bonferroni) corrections for multiplicity may make the type II error for harms unacceptably high. Bayesian Hierarchical Models [2] have the potential to address these problems in RCTs as well as in observational studies and spontaneous reporting. Patients' needs must be met by applying the most effective methods of analysis (e.g. Kaplan-Meier plots) [3], clear reporting [4] and sensible interpretation to the ADRs seen in trials.
Eliciting harms data from trial participants: how perceptions of illness and treatment moderate recognition of relevant information to report

Elizabeth N Allen1, Karen I Barnes1, Adel Mushi2, Isolde Massawe2, Sarah G Staalke3, Ushma Mehta1, Lasse S Vestergaard4, Martha M Lemnge3, Clare I Chandler1

1Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa; 2National Institute of Medical Research, Dar es Salaam, Tanzania; 3National Institute of Medical Research, Tanga Centre, Tanga, Tanzania; 4Department of Clinical Research, London School of Hygiene and Tropical Medicine, London, UK; 5Independent Pharmacovigilance Consultant, Johannesburg, South Africa; 6Centre for Medical Parasitology at the Department of International Health, Immunology and Microbiology, University of Copenhagen, Denmark; 7Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK.

E-mail: elizabeth.allen@uct.ac.za

Trials 2011, 12(Suppl 1):A10

Background: There is no consensus on the ideal methodology for eliciting participant-reported harms, but question methods influence the extent and nature of data detected. This gives potential for measurement error and undermines meta-analyses of adverse effects. We undertook to identify barriers to accurate and complete reporting of harms data, by qualitatively exploring participants’ experiences of illness and treatment, and reporting behaviours; and compared the number and nature of data detected by three enquiry methods.

Methods: Participants within antiretroviral/antimalarial interaction trials in South Africa and Tanzania were asked about medical history, treatments and/or adverse events by general enquiries followed by checklists. Those reporting differently between these two question methods were invited to an in-depth interview and focus group discussion. Health narratives were analysed to investigate accuracy and completeness of case record form data and to understand reasons for differential reporting between question methods. Outcomes were the number and nature of data by question method, themes from qualitative analyses and a theoretical interpretation of participants’ experiences.

Results: We observed a cumulative increase in sensitivity of detection of all types of reports while progressing from general enquiry, through checklist, to in-depth interview. Questioning detail and terminology influenced participants’ recognition of health issues and treatments. Reporting patterns and vocabulary suggest influence from the relative importance that illnesses and treatments have for participants. Perceptions were often dichotomised (e.g. ‘street’ versus clinic treatments, symptoms experienced versus tests and examinations performed, chronic versus acute illness, persistent versus intermittent symptoms, activity- versus malaria-related symptoms) and this differentiation extended to ideas of relevance to report. South African participants displayed a ‘trial citizenship’, taking responsibility for the impact of their reporting on trial results, and even reaching reporting decisions by consensus. In contrast, Tanzanians perceived their role more as patients than participants; the locus of responsibility for knowing information relevant to the trial fell with trial staff as doctors rather than with themselves.

Conclusions: Our observations of how reporting relates to participant perceptions inside and outside trials could help optimise how harms data are elicited. Questions reflecting the different ways that biomedically defined illness and treatment data are perceived by participants may help them understand relevance for reporting. We will theorise how these two disparate trial environments may have influenced how participants understood their role, as this could help researchers advocate empowered participation in similar trials.

Acknowledgements: The authors would like to thank the staff and participants of the SEACAT and InterACT clinical trials. This study was supported by the Division of Clinical Pharmacology, University of Cape Town and the ACT Consortium which is funded through a grant from the Bill and Melinda Gates Foundation to the London School of Hygiene and Tropical Medicine.

A11

Anti-epileptic drug harms: issues for meta-analysis

Catin Tudur Smith1, Arif Shukkalla2, Sarah Donegan3, Karla Hemming3, Graham A Powell4, Paula Williamson5, Anthony Manson6

1North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, Liverpool, UK; 2Clinical and Molecular Pharmacology, University of Liverpool, Liverpool, UK; 3Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK

Trials 2011, 12(Suppl 1):A11

Objectives: Decisions regarding choice and dose of anti-epileptic drug (AED) are driven by considering the potential benefits of reducing seizure frequency against the potential harms of alternative AEDs. Such decisions should be made using the best available evidence, which often requires a quantitative synthesis of data from multiple randomised controlled trials (RCT). However, the systematic review and meta-analysis of harms data is hindered by problems such as inadequate reporting, heterogeneity of harms definitions, and selective reporting bias. Here we will evaluate the quality of reporting of harms data in epilepsy trials, and assess the potential added value of incorporating harms data beyond the clinical indication of epilepsy.

Methods: To evaluate the quality of reporting of harms data in RCTs of AEDs in patients with epilepsy we have undertaken a systematic review [1]. We searched MEDLINE, the Cochrane Library and the Epilepsy Group register for published trials comparing AEDs in patients with epilepsy. Each trial was assessed according to a 23 item checklist developed from the CONSORT statement for the reporting of harms in clinical trials [2]. In a separate analysis, Bayesian panstratic meta-analysis models [3] were used to pool estimates of harm across studies and across indications of epilepsy, neuropathy and headache, allowing for variation between both study and indication.

Results: For the reporting quality review we identified 152 RCTs that met the eligibility criteria. None of the trials satisfied all criteria. The mean number of criteria per trial was 11.3 (standard deviation 4.3, range 0 to 21). No improvement could be detected following publication of the CONSORT statement for harms (difference in means: 0.6 with 95% CI (-0.9 to 1.8) p=0.53). Items that were not frequently reported were: definition of adverse events (36.2% of trials), use of a validated dictionary (21.7% of trials), use of a validated instrument (15.8% of trials), reporting of both number of patients and number of adverse events (19.1% of trials) and methods for handling of recurrent events (7.2% of trials). In the summary of harms data, borrowing strength from other indications resulted in a more precise effect estimate, and indicate that there is evidence for some adverse events across the range of indications.

Conclusion: Reporting of harms in RCTs of AEDs is poor and has not improved since the publication of the CONSORT guidelines on the reporting of harms. To allow reliable meta-analyses of harms data, improvements to reporting quality are essential. Preliminary results suggest that harms data from AEDs prescribed for headache and neuropathy may be useful to inform the harms profile of AEDs prescribed for epilepsy.

References
Background: Monitoring the safety of therapies is of paramount importance in protecting patients from harm and enabling risk-benefit assessment. The recording and reporting of measures of efficacy has received considerable attention and while by no means perfect, has advanced further than the parallel assessment of harm. The stimulus for this study came from a commissioned effectiveness and cost-effectiveness review of treatments for neuropathic pain in patients (the CEAN study) [1]. CEAN noted that the completeness of adverse event (AE) reporting varied between trials and some expert opinion was required where primary data were insufficient for modeling cost-effectiveness. Further, clinicians indicated that trials sometimes failed to provide adequate information for clinical decision-making and informing patients.

Objectives: To describe how AE data are collected and reported. To explore results post-2004 (when regulatory requirements regarding collection and reporting of AEs for RCTs were in place).

Methods: Relevant CEAN study publications (RCTs of anticonvulsants and antidepressants for post herpetic neuralgia and painful diabetic neuropathy) with separate primary data on impact on pain. Items for data extraction were generated using recommendations set out in CONSORT 2004 [2]. Additional information extracted sought to determine the criteria used by authors to select AEs for reporting (e.g. significant differences), the mode of collection (e.g. observation or questionnaire) and how they were collected (e.g. passively, actively (prompted). Double data extraction was performed.

Results: 53 publications were included, 12 were published post 2004. Key results are presented in Tables 1 and 2. A subset of the recommendations laid out in CONSORT 2004 were not adhered to by any of the publications. The collection method impacts on the number of AEs reported by patients and this was poorly reported by the majority of trials. The criteria used by authors for reporting AEs varied substantially across publications.

Conclusion: Synthesis of AE data across studies is hampered by the lack of information on collection methods and by arbitrary heterogeneous criteria used by authors for selecting AEs to be reported. In order to improve the usefulness of AE data reported by publications, validated methods for collection need to be developed, and core AEs for reporting need to be agreed. Online journal supplements can be utilised to overcome journal space limitations. The issues highlighted by this case study are likely to be relevant for AE reporting in RCTs in general, although solutions will likely need to be tailored to specific therapeutic-disease areas.

### Table 1 (abstract A12)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>All (N=53) n (%)</th>
<th>Post-2004 (N=12) n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reported total number who withdrew &amp; withdrew due to AE</td>
<td>48 (91)</td>
<td>12 (100%)</td>
</tr>
<tr>
<td>Reported that grading for AEs were assigned</td>
<td>30 (57)</td>
<td>9 (75)</td>
</tr>
<tr>
<td>Reported mode of collection (e.g. questionnaire, patient reported, observation)</td>
<td>23 (43)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>Distinction between severe/life threatening AEs and those that were not</td>
<td>29 (55)</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Reported the dictionary used for coding AEs</td>
<td>1 (19)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

### Table 2 (abstract A12) Criteria used to select AEs for reporting

<table>
<thead>
<tr>
<th>Criteria</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most frequent AEs</td>
<td>17 (32)</td>
</tr>
<tr>
<td>All AEs that occurred</td>
<td>17 (32)</td>
</tr>
<tr>
<td>Other</td>
<td>6 (11)</td>
</tr>
<tr>
<td>A pre-specified list of AEs</td>
<td>5 (9.4)</td>
</tr>
<tr>
<td>Unclear</td>
<td>3 (5.7)</td>
</tr>
<tr>
<td>Not applicable as no AE reported</td>
<td>3 (5.7)</td>
</tr>
<tr>
<td>Any AE with sig diff between treatment groups</td>
<td>2 (3.8)</td>
</tr>
</tbody>
</table>

### References

### A13

**Objectives:** IVAN and TANDEM are factorial randomised controlled trials (RCTs) of different treatment regimens for wet AMD, involving off-label use of bevacizumab. Safety data are being collected but, given follow-up visits every 4-8 weeks for up to 3 years and bevacizumab reporting, are difficult to collate. We have developed a database solution for the TANDEM trial to minimise duplicate adverse event or events or events without resolution dates, based on lessons learnt in the IVAN trial.

**Methods:** Duplicate reporting of events can arise over multiple visits for several reasons, for example:
- different verbatim reporting (or typographic errors) of the same event;
- evolution of an event through stages of symptom, investigation, diagnosis, treatment;
- ongoing events may not be reviewed at every visit and resolution dates may be missed; if the data record for such an event is incomplete over time, it is impossible to know whether a subsequent report describes the original event or a recurrence.

**Results:** The database solution has the following features:
- coded categories to match most common adverse events and mapped to MedDRA preferred terms, with an option to classify as ‘serious’, with reason;
- ‘other’ category;
- optional free-text qualification field for any event.

‘Other’ events will be reviewed regularly and additional categories added for common events not covered by the existing codes. After the first trial visit, a customised adverse event data form is printed from the trial database in advance of interviewing the participant. The top half of the form requires information about ongoing events to be updated, allowing evolution of an event to be described. The bottom half of the form allows new events to be documented using the coded categories.
and qualification field. Failure to update an ongoing event generates an automatic data validation query.

Conclusions: This solution provides complete transparency of adverse event reporting and real-time coding. It requires ongoing events to be updated and minimises the problems created by free-text reporting.

Acknowledgements: For the IVANI and TANCEM study groups.

BIOMARKERS

A14
Using biomarkers prospectively in adaptive clinical trials
Donald A Berry
M.D. Anderson Cancer Center, Houston, TX, USA
Trials 2011, 12(Suppl 1):A14

A standard approach to biomarker research is to retrospectively examine interactions between biomarkers and treatment effects. An alternative approach of examining such interactions prospectively has advantages. First, if the biomarker information is required for randomization then eliminates missing biomarker information. Another is that nonresponder biomarker subtypes in a multiarmed trial that do not respond to a particular arm can be excluded from that arm, perhaps gradually, using adaptive randomization. A consequence of adaptively excluding nonresponders is the potential to have a smaller, more focused trial.

There are limitations to such an approach. One is that the biomarker has to be available to enable adapting to the accumulating evidence. Another is that the outcome database must be updated in a reasonably timely fashion, and it must be connected to the patient assignment algorithm. In addition, the design, although fully prospective, is complicated to convey to investigators, patients, and IRBs, IECs, and other regulators.

A third possible design is to restrict trial eligibility to the population that is the drug’s target. This is efficient, but it relies on knowing the target. This design gives no information about biomarker by treatment interactions. And it inhibits learning about the roles of other biomarkers.

The adaptive approach is a compromise between the first and third approaches: Start with all-comers but restrict to the responding population as the trial results accumulate.

The approaches I describe employ randomization. All have an adaptive aspect in which accumulating trial results are analyzed frequently with the possibility of modifying the trial’s future course. Many treatment arms are possible, including combination therapies. So it is possible to learn about the way treatments interact with each other as well as the way they interact with biomarkers.

Having multiple biomarkers and multiple treatment arms increases the false-positive rate. Therefore it is essential to prospectively build some level of confirmation into the design. False-positive rates and statistical power can be evaluated by simulation and controlled.

Taking an adaptive approach is fruitless without information to which to adapt. There is little information available when the endpoints are long-term. However, early markers of therapeutic effect (longitudinal biomarkers, measurements of tumor burden, etc.) can be correlated with long-term clinical outcome.

I will give an example (called I-SPY 2) of an adaptive biomarker-driven trial in neoadjuvant breast cancer. The goal is to efficiently identify biomarker signatures for a variety of agents and combinations being considered simultaneously.

A15
Data modeling methods in clinical trials: experiences from the clinical trial methods in neurodegenerative diseases project
Athanasios Anastasiou1, Emmanuel Meachor2, John Hajcak2
1Signal Processing and Multimedia Communications Research Group, University of Plymouth, Drake Circus, Plymouth, UK; 2Institute of Health Service Research Peninsula College of Medicine and Dentistry
E-mail: athanasios.anastasiou@plymouth.ac.uk
Trials 2011, 12(Suppl 1):A15

Objectives: Clinical trials often generate large and diverse datasets. Data models are used to capture and organise the elements of the data in a meaningful way so that they can be stored and utilised by computer systems and support clinical decision making. This paper presents the data modeling considerations within the ‘Clinical Trial Methods in Neurodegenerative Diseases’ (CTMND) project funded by the NIHR [http://www.ctmnd.org].

The project adopts a holistic approach for the investigation of the suitability and efficiency of clinical observations in neurodegenerative diseases clinical studies. This ongoing research in novel clinical and surrogate outcome measures will be incorporated in an online data collection and analysis system to facilitate clinical trials and relevant research, taking into account, wherever possible, routinely collected NHS data.

This paper presents ongoing research in the project’s data modeling aspects with the following objectives:

1. To review the current state of the art data models for capturing clinical information from the available literature.
2. To compare and contrast their features against the data management requirements of the project and outline the key factors that affected the adoption of a specific model for the CTMND project’s information system.

Methods: A set of key papers and past reviews were collected from the currently available literature detailing the characteristics of standard data models used in healthcare such as CDISC’s ODM, Health Level 7 and others. The data models and associated approaches were compared and contrasted with each other by taking into account best practices and guidelines emerging from organizations such as the Object Management Group (OMG).

Finally, having concluded in a specific modeling approach we were also able to look forward at the possibilities that a particular solution enables and propose a flexible way to model clinical trial data.

Results: This review highlights a number of key data management and organization considerations that affect the adoption of a specific data model given the project specifications and resource constraints. The key factors were Current Resources, Interoperability (with current and future systems), Documentation and Reference Implementation availability.

Conclusions: Given the dynamic environment of clinical trials as well as the project’s objectives to propose novel outcome measures, we comment on the suitability of the Dual Model approach for the efficient organization of clinical study data but more importantly for its flexibility in modeling novel outcomes with minimal software maintenance. According to this approach, a handful of elementary data structures (numbers, character sequences, lists, trees and others) are made available to a higher level model that is responsible for their ordering and semantics. Additionally we provide specific details regarding system implementation.

Acknowledgements: The CTMND Consortium.

A16
Stratified medicine in practice: review of predictive biomarkers in European Medicines Agency (EMA) indications
Kinga Malottki1, Mousumi Biswas2, Jon Deeks1,3, Richard Riley1,3, Charles Caddick4, Lucinda Billingham1,2
1MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham, UK; 2Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK; 3Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham, UK; 4Centre for Clinical Haematology, Queen Elizabeth Hospital, Birmingham, UK
E-mail: k.malottki@bham.ac.uk
Trials 2011, 12(Suppl 1):A16

Objectives: Stratified medicine has been defined as using a biomarker to match a patient to a cohort that has exhibited a differential response to a treatment. This is important where the proportion of patients benefiting from treatment is low and possible adverse events can be serious. To maximise patient benefit, valid predictive biomarkers need to be used.

Several trial designs have been proposed to evaluate the use of predictive biomarkers in clinical practice including: enrichment, stratified and biomarker-based strategy designs.

Our aim was to review the EMA indications that include a predictive biomarker in order to investigate the type and strength of evidence considered sufficient for such decisions.

Methods: We have undertaken a review to identify predictive biomarkers included in EMA indications, together with the supporting study designs and strength of evidence. The authorised, refused, withdrawn and pending
Fifteen predictive biomarkers were included in the indications of 18 drugs. For one biomarker the license was refused and for one withdrawn. Only three biomarkers were included in an indication before 2004. Five of the 18 drugs had an orphan designation. Thirteen biomarkers for 10 drugs were included in indications for treatment of various cancers (including a range of haematological diseases, breast, colorectal, gastric and lung cancer). Two biomarkers were included in the indications of four drugs for the treatment of HIV infections. The majority of identified studies were enrichment design or used a subgroup analysis (sometimes post hoc) to evaluate the predictive biomarker. One stratified and one marker-based strategy design study was identified.

Conclusions: The specialties where predictive biomarkers were identified were limited to cancer and HIV infection. No predictive biomarkers have been identified in other specialties where treatments are often effective only in a relatively small subgroup of patients (such as mental health). Our review found that evidence from subgroup analyses and studies with an enrichment design has been often considered sufficient to grant marketing authorisation. Such results would ordinarily be interpreted with caution due to underlying methodological limitations.

A17
Designing a preliminary adaptive study to inform a biomarker trial in Psoriasis
Andrew T Prevost1*, Jack Bowden2
1Department of Primary Care and Public Health Sciences, King’s College London, SE1 3QD, UK; 2Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, CB2 0SR, UK
E-mail: toby.prevost@kcl.ac.uk
Trials 2011, 12(Suppl 1):A17

Background: Biomarkers play different roles in trials, being accordingly classified into ‘prognostic’, ‘predictive’, ‘surrogate’, or combinations thereof. Knowledge of a biomarker’s role enables focused testing in late phase trials through a better choice from available designs (e.g. ‘stratified’, ‘strategy’, and ‘enrichment’). Preliminary studies can inform a biomarker’s role and the relative value of multiple biomarkers.

Motivation: Here we consider the design of a preliminary study of potentially predictive biomarkers in patients treated for Psoriasis. A clinical researcher came for methodological advice, bringing a related published study with exciting results but of highly dubious quality. The objective was to provide a design with better properties (less bias, high power, low cost), allowing multiple biomarkers and their combination to be assessed to inform any subsequent trial.

Methods: Prior preferences, agreed by the researcher, were for a prospective design, control groups, and power-based sample size calculation. A formal 10-minute presentation to the full team was required to explain pros and cons of an adaptive element (two recruitment stages) to the design. Power was assessed through simulation in R-software using Fisher’s method, involving the product of stage p-values.

Results: Effect size was defined in terms of the correlation between treatment response over time and a biomarker. Under a non-adaptive design, an R-squared of 20% could be detected with 90% power, 5% significance level, with 49 patients, with all 17 expensive biomarkers measured. The adaptive design offered an interesting alternative, employing p>0.3 to discontinue with biomarkers quarter-way through recruitment, requiring 24+72=96 patients. This offers more patients to develop a combination from an enriched biomarker set guaranteed to include the best five performers from stage one. The proportion of biomarkers expected to discontinue, conditional on underlying effect size, was considered graphically.

Conclusions: Incorporating methodological improvements into study designs requires understanding of methodology and collaborators. The cost-efficient two-stage design is an improvement on the related published study, and we outline further analysis-stage developments: reducing bias in estimates and providing valid confidence intervals and error rates [1,2]. The study proposal is currently going through ethics.

A18
Selection of subjects for clinical trials in Alzheimer’s disease and mild cognitive impairment with machine learning analysis of MRI and CSF biomarkers
Javier Escudero1*, John P Zajicek2, Emmanuel Ifeachor1
1Signal Processing and Multimedia Communications Research Group, School of Computing and Mathematics, Plymouth University, Plymouth, PL4 8AA, UK; 2Clinical Neurology Research Group, Peninsula College of Medicine and Dentistry, Plymouth University, Derriford, Plymouth, PL6 8BX, UK
E-mail: javier.escudero@ieee.org
Trials 2011, 12(Suppl 1):A18

Objectives: There is a need for techniques to conduct Clinical Trials (CTs) in Alzheimer’s Disease (AD) and Mild Cognitive Impairment (MCI) more efficiently to reduce their duration and cost. However, large variability in the rating scales increases the number of subjects required to obtain significant results in the CTs [1,2]. Additionally, Machine Learning can detect biomarker patterns to characterize AD and MCI. In this study, we assessed the usefulness of Machine Learning to select subjects with the clearest signs of the disease for inclusion in more efficient CTs [3,4].

Methods: We tested three Machine Learning classifiers: Logistic Regression (LR), Support Vector Machine (SVM) and Radial Basis Function (RBF) [4]. These techniques were trained to recognise disease patterns in 91 AD, 178 MCI and 106 cognitive normal (CN) subjects from ADNI [1] for whom baseline age, MRI hippocampal volume, MRI entorhinal cortical thickness, CSF Aβ42 and CSF phosphorylated Tau181 levels were measured. From the classifiers, we obtained a likelihood value that each subject was AD or MCI and not CN. Then, the patients with higher likelihood (i.e., clearer signs) of the disease were first selected for inclusion in hypothetical CTs. This approach was evaluated in the terms of reduction in the number of patients needed in the CTs to detect a 25% reduction in the hippocampal volume after one year (80% power, two-sided test, p-value=0.05) [3,4].

Results: Without the selection of subjects based on the classifiers, the hypothetical CTs required 109 patients for AD and 183 subjects for MCI per group (treatment vs. placebo). In contrast, the sample sizes decreased considerably when the classifiers based on the biomarkers were used to select one third of the subjects with the highest likelihood (clearest signs) of the disease, as shown in Table 1. All these group sizes were at least eight times smaller than those estimated when ADAS-cog, instead of the hippocampus, was the outcome measure in the CT (Table 2).

Conclusion: The results highlighted the potential of CSF and MRI biomarkers and Machine Learning classifiers (particularly LR and RBF) as objective tools to select subjects for more efficient CTs in AD and MCI.

Table 1 (abstract A18) Minimum number of subjects required per arm for a hypothetical CT with the hippocampal volume as outcome measure in AD and MCI. Two cases are considered: when all subjects are included in the CT, and when only the 33% of the subjects with the clearest signs of AD are selected for the CT

<table>
<thead>
<tr>
<th>Condition</th>
<th>Subset</th>
<th>LR</th>
<th>SVM</th>
<th>RBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>All subjects</td>
<td>109</td>
<td>109</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>33% of subjects with the clearest signs of AD</td>
<td>48</td>
<td>93</td>
<td>30</td>
</tr>
<tr>
<td>MCI</td>
<td>All subjects</td>
<td>183</td>
<td>183</td>
<td>183</td>
</tr>
<tr>
<td></td>
<td>33% of subjects with the clearest signs of AD</td>
<td>95</td>
<td>139</td>
<td>104</td>
</tr>
</tbody>
</table>
### Table 2 (abstract A18) Minimum number of subjects required per arm for a hypothetical CT with the ADAS-cog as outcome measure in AD and MCI. Two cases are considered: when all subjects are included in the CT, and when only the 33% of the subjects with the clearest signs of AD are selected for the CT.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Selection</th>
<th>LR</th>
<th>SVM</th>
<th>RBF</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>All subjects</td>
<td>1330</td>
<td>1330</td>
<td>1330</td>
</tr>
<tr>
<td></td>
<td>33% of subjects with the clearest signs of AD</td>
<td>1675</td>
<td>1605</td>
<td>259</td>
</tr>
<tr>
<td>MCI</td>
<td>All subjects</td>
<td>8878</td>
<td>8878</td>
<td>8878</td>
</tr>
<tr>
<td></td>
<td>33% of subjects with the clearest signs of AD</td>
<td>3098</td>
<td>1148</td>
<td>2119</td>
</tr>
</tbody>
</table>

However, further analyses are needed to corroborate these results and extend this approach to other biomarkers and classifiers.

**Acknowledgements:** This abstract presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10124). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Data used in the preparation of this abstract were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://www.loni.ucla.edu/ADNI/).

### References
2. Cummings JL. Integrating ADNI results into Alzheimer’s Disease drug development programs. Neurobiology of Aging 2010; 31:1481-1492.

### A19
Comparing diagnostic tests and biomarkers: trials in people with discordant test results

Richard Hooper, Karla Díaz-Ordaz, Andrea Takeda, Khalid Khan
Centre for Primary Care & Public Health, Blizard Institute, Queen Mary University of London, London E1 2AB, UK
E-mail: r.hooper@qmul.ac.uk

**Trials 2011, 12(Suppl 1):A19**

**Objectives:** Diagnostic tests are traditionally compared for accuracy against a gold standard, but there is growing interest in tests (or biomarkers) used to guide treatment choices rather than specifically to diagnose. The question of whether health outcomes are better using one test or another in the same population can be answered with a randomised trial. It has been suggested that an efficient approach to trialling is to give both tests to all participants, and randomise and follow up those with discordant results. We describe how to plan and analyse such a trial, and consider its efficiency compared with a conventional trial design.

**Methods:** We investigated two estimates of risk difference for a binary outcome: one based on analysing outcomes as if from a conventional trial (the trial estimate), and one which combined estimates of different parameters in the manner of a decision analysis (the decision analysis estimate). From theory we derived the bias and standard error of each. We also considered the impact of randomising participants to a testing strategy before the tests are administered rather than after.

**Results:** The trial estimate and decision analysis estimate are both unbiased. Using the decision analysis estimate (but not the trial estimate) the same precision is achieved by randomising before testing as by randomising after. Giving both tests to all participants means fewer need to be recruited: in one example from the literature the proposed design was nearly four times more efficient in this sense than a conventional trial design.

**Conclusions:** We suggest the term ‘randomised discordance trial’ for the design we have described. A discordance trial offers an efficient way to compare two diagnostic tests or biomarkers. Randomising before testing avoids selection bias. We have derived formulae for calculating sample size and for estimating risk difference and standard error using this design.

### A20
Choice of transformation for modelling non-linear continuous biomarkers

Deborah D Stocken 1, Lucinda J Billingham 1, Philip J Johnson 1, Nick Freemantle 1
1Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK; 2Primary Care and Population Health, University College London, London, UK

**Trials 2011, 12(Suppl 1):A20**

Identification of prognostic and predictive biomarkers is important for targeting treatments to patients and for the design and analysis of randomised controlled trials. Cox proportional hazards modelling is a standard method for assessing prognostic value of clinical biomarkers where time to occurrence of an event is the primary outcome of interest. An important issue in the analysis of prognostic factors is the functional form of the relationship between the factor and outcome specifically for continuous covariates. Continuous covariates are often simplified assuming a linear relationship with log-hazard or dichotomisation which may be inappropriate leading to loss of efficiency in statistical tests, bias and incorrect conclusions. The effects of important prognostic biomarkers may go unrecognised due to simplistic assumptions made in statistical modelling. Two polynomial based strategies, restricted cubic splines and fractional polynomials, are compared directly for determining the functional form of non-linear relationships between prognostic biomarkers and survival using two real datasets from randomised controlled trials in advanced pancreatic cancer and cardiac surgery. Fractional polynomials are an extended family of curves including non-integer and negative power terms. Spline functions are piecewise polynomials connected across intervals of a variable constrained to meet at the ‘knots’.

Multivariable models were constructed based on Cox proportional hazards regression using backward elimination. Internal validation to directly compare the fit of the restricted cubic spline and fractional polynomial strategies was carried out by calculating the sampling distribution of the difference in AIC between the models using nonparametric bootstrap analyses. Further analysis recalculated the univariate fractional polynomial transformation within each bootstrap resample to compare directly against a 5-knot restricted cubic spline. The influence of the size of the bootstrap samples was also investigated. The fitted functions generated by splines and fractional polynomials were similar resulting in comparable models. The methods are generally different in the extremities where there is often a paucity of data. Larger differences were seen between the two methods when sample sizes were reduced due to the reduced power to detect small effects but also to detect nonlinearity. Multivariable fractional polynomial transformations are an alternative approach to restricted cubic spline transformations for multivariable model building of continuous biomarkers with non-linear relationships with outcome.

### CLUSTERING AND CLUSTERS RCTS

### A21
Statistical issues in the analysis of non-pharmacological therapy trials with clustering by care-provider or therapy group

Chris Roberts
University of Manchester, Manchester, M13 9PL, UK

**Trials 2011, 12(Suppl 1):A21**

In trials of physical and talking therapies or group administered treatments, clustering of patients within care-provider or treatment group has
implications for sample size and statistical analysis analogous to those found in cluster randomised trials [1]. Between-cluster variation reduces precision and power when estimating treatment effects. Statistical analyses that fail to take account of this form clustering rest on the assumption of no clustering effect and may therefore lack conclusion validity. In trials with treatment related clustering, the cluster size may differ systematically between treatment arms, due to differing number of therapists in each arm or differences in the size of therapy groups between arms. The intra-cluster correlation due to therapist or therapy group may also differ between treatment arms. An extreme case of such heterogeneity is the partially nested design in which the clustering effect is absent in one treatment arm. Where both cluster size and the underlying variance components differ between treatments, failure to correctly model heterogeneity can bias estimates of the intra-cluster correlation coefficient and test size [1]. This contrasts with cluster-randomised trials where the distribution of cluster sizes in each treatment arm will be similar due to randomisation making cluster randomised trials robust to heteroscedasticity. In a cluster randomised trial, it is generally assumed that each subject belongs to just a single unit of randomisation so that the design is hierarchical. In non-pharmacological therapy trials patients may receive treatment from more than one therapist or in more than one therapy group so the multilevel model is no longer strictly hierarchical. Statistical analysis will therefore require simplifying assumptions regarding the pattern of clustering such as use of a primary therapist or primary group for each patient or the application of a multiple membership model [2].

In conclusion, analysis of trials with treatment related clustering may therefore require more complex methods of analysis than cluster randomised trials.

References

A22 Stratified randomisation: a hidden form of clustering?
Brennan C Kahan1, Tim P Morris
MRC Clinical Trials Unit, 125 Kingsway, London, WC2B 6NH, UK
Trials 2011, 12(Suppl 1):A22

Objectives: Many randomised trials use stratified permuted blocks or minimisation to balance key prognostic variables between treatment groups. It is widely argued in the statistical literature that any balancing variables should be adjusted for in the analysis, however a review of major medical journals shows that this is not commonly done. Our objective was to determine the effects of an unadjusted analysis after balancing.

Methods: The statistical properties of an unadjusted analysis after balancing are explored using theoretical results. A major simulation study using data from 5 trials is performed to determine the potential impact in real life situations.

Results: We show that balancing on baseline covariates leads to correlation between the treatment groups (similarly, cluster randomised trials lead to correlation within treatment groups). If this correlation is ignored, and an unadjusted analysis is performed, the estimated variance of the treatment effect will be biased upwards, resulting in type I error rates that are too low, and a reduction in power. Conversely, an adjusted analysis results in nominal type I error rates, and optimal power.

Conclusions: Prognostic variables that have been balanced between treatment groups in the randomisation process should be adjusted for in the analysis. Unadjusted analyses lead to invalid results, whereas adjusted analyses maintain nominal properties.

A23 A review of methodology for sample size calculations in cluster randomised trials
Clare Rutterford1, Sandra Eldridge1, Andrew Copas2
1Centre for Primary Care and Public Health, Queen Mary University of London, London, E1 2AB, UK; 2Hub for Trials Methodology Research, MRC Clinical Trials Unit, London, UK
Trials 2011, 12(Suppl 1):A23

Objectives: To produce a thorough review of the existing state of knowledge on sample size calculations for cluster randomised trials (CRTs) and to identify gaps in the knowledge.

Methods: A systematic review is being conducted of sample size methodology for cluster randomised trials. The sources for the search include electronic databases PubMed and Web of Science, key text books on cluster randomised trials and discussions with experts in the field. The search strategy involves a compliment of Medical Subject Headings and free text terms to aid a comprehensive search. The references of papers eligible for the review will also be searched and a search on the first author conducted. This process will continue until no more additional papers are located.

This work forms the beginning of a PhD research project.

Results: Of 8697 citations obtained from PubMed and Web of Science, the majority have currently been assessed for eligibility into the review and 57 papers so far identified for inclusion. The majority of papers discuss sample size for continuous or binary outcomes, with four papers discussing time to event outcomes. In terms of the analysis method used, most assume a random effects analysis (cluster specific approach) or a cluster level analysis, with fewer papers assuming a generalized estimating Equation (population averaged approach) methodology.

An emerging theme, discussed in six papers, is sample size methodology for 3-level cluster randomised trials, where we may randomise clinics (level 3) and each clinic will treat multiple subjects (level 2) who in turn are measured on repeated occasions (level 1 units). Eight papers consider sample size calculations for trials with varying cluster sizes. These papers account for the loss in power due to varying cluster sizes through an examination of the relative efficiency of unequal versus equal cluster sizes or by proposing an appropriate design effect to account for this loss for both continuous and binary outcomes. Sample size for alternative trial designs such as cross-over trials, stepped wedge designs, testing for non-inferiority, stratified, and matched designs were identified. Papers covering adjustments to sample size for dealing with non-compliance or attrition, accounting for the use of cluster or person level covariates and dealing with imprecision in the estimate of the intraclass correlation coefficient (ICC) were identified.

Conclusion: We will provide the results of the search and preliminary insight into potential gaps in the knowledge.

A24 Clustering in surgical trials – database of intra-cluster correlations
Jonathan A Cook1, Thomas Bruckner2, Graeme S MacLennan1, Christoph M Seiler3
1Health Services Research Unit, University of Aberdeen, Aberdeen, AB25 2ZD, UK; 2Institute of Medical Biometry and Informatics, University of Heidelberg, 69120, Heidelberg, Germany; 3Department of General, Visceral and Trauma Surgery, University of Heidelberg, Heidelberg, 69120, Germany
Trials 2011, 12(Suppl 1):A24

Background: Randomised trials evaluating surgical interventions are often design and analysed as if the outcome of individual patients is independent of the surgeon providing the intervention. There is reason to expect outcomes for patients treated by the same surgeon to be more similar than those under the care of another surgeon due to previous experience, individual practice, training, and infrastructure. Such a phenomenon is referred to as the clustering effect and potentially impacts on the required sample size. This depends upon the design and analysis adopted. However, trialsists have little data upon which to assess the impact and base trial design. The aim of this study was to quantify the clustering effect by producing a database of surgical trial ICCs.

Methods: Intra-cluster correlation coefficients (ICCs) were calculated for outcomes from a set of 10 multicentre surgical trials for a range of outcomes and different time points for clustering at both the centre and surgeon level.

Results: ICCs were calculated for 198 outcomes across the 10 trials at both centre and surgeon cluster levels. The number of cases varied from 138 to 1370 across the trials. The median (range) average cluster size was 32 (9 to 51) and 6 (3 to 30) for centre and surgeon levels respectively. ICC estimates varied substantially between outcomes though uncertainty around
individual ICC estimates was substantial. Full details are available online [http://www.abdn.ac.uk/hsru/research/research-tools/study-design].

**Conclusions:** Our data for multicentre trials of surgical interventions suggests clustering of outcome is more of an issue than has been previously acknowledged. This database provides trialists with valuable information to aid the design of surgical trials. We anticipate that over time the addition of ICCs from further surgical trial datasets will enhance the usefulness of the database.

**Acknowledgements:** The authors would like to thank the trial groups for access to the trial data and help preparing the data for analysis. The Health Services Research Unit is core funded by the Chief Scientist Office of the Scottish Government Health Directorates. Jonathan Cook holds a Medical Research Council UK fellowship (G0601938). Views expressed are those of the authors and do not necessarily reflect the view of Chief Scientist Office.

---

**A25 Sample size in cluster randomised trials with unequal clusters**

Ivana Holloway*, Amanda Farrin

Clinical Trials Research Unit (CTRU), University of Leeds, Leeds, UK

**Methods and results:** Unbalanced cluster size decreases the statistical power in CRTs. Even if the original sample size calculation considers clustering, this sample size is underestimated in the case of unequal cluster size. A common reason for ignoring variability of cluster size is the lack of appropriate, easily usable sample size calculation formulae.

In the TRACS trial, all centres were randomised at the same time and unequal cluster size was not anticipated. However, differences in recruitment rate, a higher than expected loss to follow-up and varying by centre occurred. Theoretically, including more clusters, each recruiting the same number of participants, would be an optimal solution. In practice, due to time, logistics and budget constraints, the number of centres was fixed, so overall more participants were recruited and the maximum cluster size was capped. In the LoTS Care trial, imbalances were expected, because centres were randomised in two phases and the overall recruitment period was fixed.

In both trials, we re-assessed sample size calculations and studied the effect of conservative, typical and extreme scenarios in terms of cluster size on the statistical power. For both calculations, various values of drop-out rate, design effect and coefficient of variation were considered and the effect on statistical power was calculated. Using the most conservative estimates, the overall power dropped by 2.3% when compared to calculation of the power based on equal cluster size. For both trials, statistical power based on equal cluster size was estimated to be 90%; so in the presence of unequal cluster size, power above 80% was preserved.

**Conclusions:** Using TRACS and LoTS Care trials as examples, we have demonstrated the importance of incorporating unequal cluster sizes into calculations of robust sample size for CRTs.

---

**A26 Sample size determination through power simulation; practical lessons from a stepped wedge randomised control trial (SW CRT)**

Munyaradzi Dimairo1-4*, Mike Bradburn5, Stephen J Walters1,2

1CTRU, SchARR, University of Sheffield, Sheffield, 51 4DA, UK, 2M~SG, SchARR, University of Sheffield, Sheffield, S1 4DA, UK, 3NIHR RDS for York & Humberside, Sheffield, S1 4DA, UK

**Trials 2011, 12(Suppl 1): A26**

**Objectives:** To describe the design of a stepped wedge randomised control trial (SW RCT) and demonstrate sample size estimation (total number of clusters and time periods) through power simulation when standard formulae are unavailable.

**Methods:** We describe the robustness of a stepped wedge design compared to other designs with respect to the referenced trial. We highlight difficulties faced during sample size estimation for a stepped wedge design in the presence of uncertainty of parameters and design constraints, and demonstrate how assessing the sensitivity can be achieved through power simulation.

**Results:** Assuming the statistical power needs to be 80% or higher, a total of 6400 simulations will estimate the power to a standard error of 0.5%. In the present study, a total of 24 clusters and 5 time periods were required to give a study power of approximately 85% assuming a type I error rate of 5% and a 20% [RR=1.20] increase in lung cancer diagnosis rates attributable to the intervention within 6 months. Hence, 6 clusters will receive the intervention per each time period in a delayed intervention fashion and all clusters will act as controls during the first time period. Study power was insensitive to distributional assumptions under consideration in this case.

**Conclusions:** The stepped wedge CRT design is robust, flexible, offers a useful practical alternative in evaluating intervention in routine practice than other designs and provides an opportunity to model time trends. Despite this, its implementation in routine practice is being hampered by complex statistical issues including sample size estimation. However, we have demonstrated that sample size calculation through power simulation is a simple and efficient approach when the use of routine formulae is impossible. Moreover, it is straightforward to assess sensitivity to issues such as heterogeneity across clusters, including the impact of design constraints such as centre or learning effects. We propose simulation-based power calculations should be considered routine practice in such circumstances. Finally, using a computer is easier than finding a book with the right formula.

---

**A27 Coping with clustering in sample size calculations**

Rebecca Playle1,2, Mark Kelly3, Robert Newcombe1,4, Kenny Hood1

1SEWTO, Cardiff University School of Medicine, Cardiff CF14 4YS, UK, 2Dept of Primary Care and Public Health, Cardiff University School of Medicine, Cardiff CF14 4YS, UK

**Trials 2011, 12(Suppl 1): A27**

**Objectives:** The objective of this study is to discuss some of the issues and give some examples of dealing with clustering in individually and cluster randomised trial sample size calculations.

**Background:** Clustering often occurs in individually randomised trials and is sometimes ignored when calculating sample sizes. Examples include practitioner effects in individually randomised trials with more than one practitioner or therapist effects from group interventions where clustering may be in one arm only. Additional levels of clustering may also occur in cluster randomised trials and their effects can be difficult.

**Methods:** Sample size calculations for several recent grant applications submitted by the South East Wales Trials Unit (SEWTO) have attempted to account for clustering or reduce inter-practitioner variation in different ways.

**Results:** Example 1: Clustering effects in an individually randomised trial of a group intervention for drug and alcohol detoxification based in prison were estimated for the intervention arm and the sample size of both arms inflated. Example 2: Clustering effects of anaesthetist were not estimated for a multi arm trial of 4 different airways devices but sample sizes were inflated to allow for small unknown effects. Sample size estimation used comparative and equivalence methods in three parallel multi arm trials. Example 3: Possible clustering in a trial of fissure sealant vs fluoride varnish at the school or family level was not accounted for in sample size calculations. All schools in the trial are Communities First schools and likely to be fairly homogeneous. Example 4: Unknown clustering effects of weight loss slimming groups were not accounted for in the intervention arm of a cluster randomised trials of a healthy lifestyle programme in obese pregnant women. Training has been implemented to reduce variability among group leaders.

**Conclusions:** There appears to be no consensus regarding dealing with additional clustering effects and best efforts are made on a trial by trial basis for sample size estimation. Access to data for the estimation of intraclass correlation coefficients and/or dissemination of the results will benefit all researchers designing trials with clustering issues.
EQUIVALENCE AND NON-INFERIORITY TRIALS

A28 Equivalence and non-inferiority trials
Ralph B D'Agostino Sr
Trials 2011, 12(Suppl 1):A28

Many effective drugs, biologics and devices exist. Because of this, it is often considered unethical to undertake placebo controlled clinical trials to evaluate new treatments. Rather, randomized active control non-inferiority trials have become the norm. These trials, however, are accompanied by serious issues including selecting an active comparator, identifying previously run active comparator placebo controlled trials, deciding upon the form of comparison (e.g., differences or ratios), setting a non-inferiority margin, designing a new study where consistency with past trials hold, finding study sites where the trial can be performed, selecting the appropriate analysis set (Intent-to-treat or per protocol) and determining sample size. These issues have proven to be more problematic than naively anticipated. Further with the accumulation and advancement of these studies and the attempt to expand their use, new problems arise and others can be anticipated to arise. Some common problems are the realization that consistency with previous active comparator placebo trials does not hold, that the multiple available ad hoc (non-theory driven) tests present confusion rather than help and that the outcome data for the active comparator often do not match the anticipated results. Newer problems are the attempts of switching from non-inferiority to superiority testing and superiority to non-inferiority testing, the use of interim analyses, multiple treatments and dealing with recurrent events as the trial outcomes. In this talk we present a brief overview of the above beginning with the migration from equivalency test to non-inferiority test, moving through the issues and with emphasis on the major problems.

A29 Is there a danger of “biocreep” with non-inferiority trials?
Primrose Beryl, Werner Vach
Clinical Epidemiology Group, Department of Medical Biometry and Statistics, University Medical Center Freiburg, Germany
E-mail: beryl@imbi.uni-freiburg.de
Trials 2011, 12(Suppl 1):A29

Background: Non-inferiority (NI) trials test a hypothesis that a new treatment is inferior to standard treatment only to a negligible degree. Biocreep basically refers to the cyclical phenomenon where a slightly inferior treatment becomes the active control for the next generation of NI trials which over time leads to degradation of the efficacy of the investigational treatment [1,2]. We studied the effect estimates from an unselected set of all the registered non-inferiority trials conducted within a seven-year period.

Methods: We did a search for all NI trials registered in the National Library of Medicine (NLM)’s Clinical trials register (https://clinicaltrials.gov) which was carried out between January 2000 and December 2007. Trials studying non-inferiority of efficacy as the primary objective were only included. We did a search for information regarding the primary results from these trials in the following steps: the NLM website [3], The Pharmaceutical Research and Manufacturers of America (PhRMA) [4], the International Federation of Pharmaceutical Manufacturers and Associations (IFPMA) website [5] and Pubmed. Web-based search engines and personal communication were also used. Using the retrieved study results, a descriptive and exploratory analysis of the study characteristics and a meta-analysis of the effect estimates were performed using STATA 11 [6].

Results: Of the 113 registered NI trials, 83 met the inclusion and exclusion criteria. The final results were available for 63 of the 74 completed studies with result estimates with the help of NLM website, 44, PhRMA-2, Pharmaceutical websites-7, Pubmed-18 and others-3. The final results were available for 63 of the 74 completed studies. The source of the study results and effect estimates was 53 scientific journal articles, 13 clinical study reports, 4 press releases and 4 reported on the register records. We intend to present the distribution of true effect of NI trials derived based on the above estimates.

Conclusion: We found a very high likelihood of retrieving results from registered clinical trials making it possible to calculate the pre-study distribution of the true effect in non-inferiority trials. The unanticipated finding of a positive average effect estimate suggests that a decline in standard treatment effect (biocreep) is not imminent, at least on average. However, the intimidating risk of approval of treatments with true negative effects reiterates the need for a careful choice of the margin in NI trials.

References
5. IFPMA: Joint position on the disclosure of clinical trial information via clinical trials registries and databases. 2008 [http://clinicaltrials.ifpma.org/]
6. Stata Corp: Stata Statistical Software. College Station, TX: StataCorp LP 2009.

A30 Choice of randomisation time-point in non-inferiority studies of reduced treatment duration: experience from the SCOT study
Jim Paul,1 Tim Iveson,2 Rachel Midgley3, Andrea Harkin1, Michelle Masterton4, Laura Alexander5, Jim Cassidy6
1Cancer Research UK Clinical Trials Unit Glasgow, Glasgow, G12 0YT, UK
2Southampton University Hospitals NHS Trust, Southampton, SO16 6YD, UK
3Department of Clinical Pharmacology, Oxford, OX3 7DQ, UK
4Oxford Clinical Trials Office, Oxford, OX3 7DQ, UK
Trials 2011, 12(Suppl 1):A30

Background: In a non-inferiority study where the aim is to compare a reduced treatment duration there is a choice of two randomisation time-points. Patients can be randomised at the end of the reduced treatment period to continue or not to the standard duration. This is optimal for maximising compliance, but has the disadvantage for the patient that treatment duration is not known from the outset. It is also more difficult to implement and, because of patients dropping out prior to randomisation, may result in an unrepresentative patient group. The alternative is to randomise patients prior to starting treatment.

Methods: SCOT is a large (9500) phase III randomised non-inferiority study comparing 6 versus 3 months adjuvant treatment in colorectal cancer. In the first year of recruitment centres were randomised to either randomise patients prior to treatment (Up-front:U) or after they had completed 3 months of treatment (Delayed:D). In June 2009 the performance of the two approaches was reviewed in terms of recruitment rate, compliance and drop-out by the Trial Steering Committee (TSC) and the recommendation was made to change all sites to U. Using data to the end of 2010 we have looked at recruitment before/after the change, restricted to centres open >3 months prior to the change date. Updated drop-out information is also provided.

Results: 215 patients were registered for D and of these 159 were randomised; a drop-out rate of 26% (95% confidence interval [CI] 21-32%). This drop-out rate is higher than the general rate of patients stopping treatment at 3 months 18% (95% CI 16-21%). For the 41 centres allocated to D the median [interquartile-range] randomisation rate prior to changing to U was 4.09 [1.29-7.09] patients/centre/year; after changing to U the equivalent figures were 8.87 [4.77 – 15.34]. This increase is statistically significant (p<0.001). For the 36 centres allocated to U the median [interquartile-range] randomisation rate prior to the date the D centres changed to U was 5.21 [3.56-11.55] patients/centre/year; post this date the equivalent figures were 7.52 [2.22 – 14.31]. This increase was not statistically significant (p=.121).
The increase in recruitment rate in the D centres is much higher than in the U centres post the randomisation change date (p=.001, test for interaction).

Conclusions: In 2009 the TSC recommended that SCOT continue with up-front randomisation only. This was based on the higher randomisation rate with U and a high drop-out rate with D. Further data has endorsed that recommendation. The SCOT study is supported by a trial grant from the MRC. Cancer Research UK supports the work of the Glasgow Clinical Trials via a programme grant.

A31
The evolution of and challenges in defining the clinical endpoint in tuberculosis treatment trials with non-inferiority designs
Patrick PJ Phillips, Angela Crook, Andrew J Nunin MRC Clinical Trials Unit, London, WC2B 6NH, UK E-mail: patrick.phillips@ctu.mrc.ac.uk Trials 2011, 12(Suppl 1):A31

Background: The first clinical trial to evaluate an effective drug for the treatment of tuberculosis (TB) conducted in 1946 by the Medical Research Council (MRC) under the direction of Bradford Hill is also widely regarded as the first properly conducted randomised clinical trial in any disease area. This new treatment was shown to significantly reduce the short term risk of death in patients with TB. As treatment improved, death became an increasingly uncommon outcome and change in clinical symptoms, radiographic appearance and bacteriological response were the commonly reported endpoints. Eventually, bacteriological failure and relapse became the primary endpoints of interest. Patients lost to follow-up were excluded from the primary analysis, as were patients not completing treatment or changing treatment for adverse events in what would today be regarded as a per protocol (PP) analysis.

In recent years the first phase III trials for new TB regimens for several decades have been initiated. Due to the high efficacy of the standard regimen in trial settings, these trials are designed as non-inferiority trials. Guidelines recommend both an intention-to-treat (ITT) and a PP analysis for non-inferiority trials. Classifying all patients with missing endpoints as unfavourable, as is recommended by the US FDA for other infectious diseases, is not appropriate in the context where losses to follow-up in the 12-18 months after treatment (when signs and symptoms have usually ceased) could easily exceed the true bacteriological unfavourable outcome (approximately 5%).

Objective: In this paper, we will discuss the complexities in (1) defining the composition of the clinical endpoint used in a TB treatment trial with a non-inferiority design and (2) performing the appropriate analyses. The relative roles of bacteriology and clinical signs and symptoms as well as the classification of patients with missing endpoints are important issues which can radically affect the overall trial results.

Case study: We will illustrate the impact of classifying unassessable patients and unrelated mortality in different ways from recent trials including a regulatory submission to the FDA.

Conclusions: There is no conservative or ‘best’ approach to defining the endpoints for non-inferiority TB treatment trials and considerable care needs to be taken in formulating a precise definition. Both ITT and PP and other sensitivity analyses are essential to eliminate bias and give robust results. Merely classifying all losses to follow-up as unfavourable leads to meaningless results, since true events will be swamped by random noise.

A32
Challenges of defining a non-inferiority margin: a case study of non-inferiority randomized controlled trials of oral anti-thrombolytic agents for prophylaxis of venous thromboembolic events after orthopedic surgery
Gracie Wangge1, Olaf H Klungel1, Kit CB Roes1, Antonius de Boer1, Ama W Hoes1, Mirjam J Knol1,2
1Department of Pharmacoepidemiology & Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences (UIPS), University of Utrecht, The Netherlands; 2Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands
Trials 2011, 12(Suppl 1):A32

Objective: To identify problems and difficulties in determining a non-inferiority (NI) margin using the case of NI randomized controlled trials (RCTs) of oral anti-thrombolytic agents for prophylaxis of venous thromboembolic events (VTE) after orthopedic surgery.

Methods: We searched in Pubmed and Cochrane-central-register-for-controlled-trials for all NI RCTs of direct thrombin inhibitors (DTI) and direct inhibitors of factor Xa (DXAI) for prophylaxis of VTE. All NI trials had exoanoparin as their active comparator. Using the draft FDA guidelines for NI trials, we determined an NI margin, referred to as the reference NI margin, based on all published placebo-controlled trials on exoanoparin for the same indication, identified in PubMed and Cochrane-central-register-for-controlled-trials. We used preserved-effects of 50% and 67% to calculate the reference NI margin.

Results: We identified 12 NI trials and 4 placebo-controlled trials of exoanoparin from our searches. All NI trials studied oral drugs. Trials in DTI used the risk difference (RD) to define their NI margin, and it ranged from 0.02 to 0.092. Trials in DXAI used the RD (ranging from 0.035 to 0.056) or risk ratio (RR) (1.25) or both to define their NI margin. Furthermore, the NI margins using the RD were stricter than the 50% preserved-effects reference NI margin (0.02 to 0.092) vs. 0.115. The NI margins in the trials using the RR were stricter than the 50 and 67% preserved-effects reference NI margin (1.25 vs. 1.46 and 1.28). In one trial, the test drug might have been concluded as non-inferior to exoanoparin if the 50% preserved-effects reference NI margin of RR were used.

Conclusions: Although a same comparator was used, a large variation in NI margins among NI RCTs of oral anti-thrombolytic agents for prophylaxis of VTE after orthopedic surgery exists. Using different NI margins could lead to different conclusions of the drug’s efficacy. Challenges that became apparent during determination of an NI margin were 1) missing unpublished results of placebo-controlled trials, 2) how similar should placebo-controlled trials and NI trials be to maintain the constancy assumption, 3) whether fixed or random effects analysis should be used in the meta-analysis, 4) whether to calculate the NI margin on an absolute or relative scale, 5) which preserved-effects to use, and 6) whether further clinical judgment is needed.
strategy failure). It also enables the QALY data from the initial phase II part of the trial to be used to inform and verify the powering of the overall phase III trial, as minimal data was available during initial trial design to power on a QALY outcome. Other novel outcome measures (time to strategy failure and summative progression free survival interval) have also been required, due to the intermittent nature of the DFS reducing the utility of the standard progression free survival endpoint. With increasing use of targeted therapies and interest in DFS due to potential QoL and cost effectiveness benefits, these novel endpoints will be increasing required in future clinical trial designs.

The STAR trial will be an exemplar trial in the evaluation of optimal treatment strategies for targeted therapies in other diseases. We will discuss issues relating to the trial design and the methodology relating to the novel endpoints in this study, as well as comment on the implications for future trial design.

A34

The ABC of non-inferiority margin setting: an investigation of approaches

Steven A Julious

Trials 2011, 12(Suppl 1):A34

Objectives: To assess the efficacy of a new investigative treatment a non-inferiority study is undertaken when it is no longer ethical to have a placebo control. Instead an active controlled trial is undertaken. The objective is thus to show that the new treatment is no worse than the active control.

In analysing a non-inferiority trial, the following ABC needs to be considered [1]:

1. The Assay sensitivity of the active control in both the placebo-controlled trials and in the active controlled non-inferiority trial is the same.
2. Bias is minimised through steps such as ensuring that the patient population and the primary efficacy endpoint are essentially the same for the placebo-controlled trial and the active-controlled trial.
3. Constancy assumption of the effect of the common comparator. Such that for two trials in sequence: Trial 1 and Trial 2 the control effect of Treatment B vs. Placebo in Trial 1 is assumed to be the same as the control effect of Treatment B vs. ‘Placebo’ in Trial 2.

This presentation will describe how this ABC can be considered.

Methods

A major issue in designing a non-inferiority study is the setting of the non-inferiority limit. The Food and Drug Administration (FDA) discuss setting a limit so it would be possible to demonstrate superiority over placebo. This comparison would need to be done indirectly as placebo is not given concurrently. A margin could be set for non-inferiority therefore which will enable superiority to placebo to be demonstrated. The European regulators discuss surveying experts to quantify the margin. An approach that uses both the objective observed data and subjective opinion to set a non-inferiority margin would useful which lends to Bayesian approaches.

Results: There is an issue with indirect comparisons if they are done retrospectively as the effect over placebo may not be as great today as when a placebo controlled trial was last undertaken [2]. The presentation will give a number of examples where the effect of treatment has fallen over time. This is known as placebo creep and could bias the estimate of effect over placebo. Comparison will be made of simple indirect and Bayesian approaches to set non-inferiority margins.

Conclusions: When making indirect comparison over time to determine a non-inferiority margin, if there is a suspicion of placebo creep then the simple ABC for setting a margin may fail and approaches such as a simple Bayesian approach may need to be considered.

References


GOVERNANCE AND REGULATORY ISSUES

A35

Streamlining regulation of clinical trials – update on the Academy of Medical Sciences review and government’s response

Gary A Ford

Newcastle University, NE1 4LP, UK

Trials 2011, 12(Suppl 1):A35

The UK has a world leading reputation for clinical research. However this position is being undermined by an overly complex regulatory and governance environment that has had a very negative effect on undertaking clinical trials in an efficient way. The European Clinical Trials Directive (EUCRD) was implemented into UK law in May 2004 and was intended to provide greater protection to patients and volunteers participating in clinical trials and increase the quality of trial conduct. At the same time the Research Governance Framework (RGF) was introduced into the NHS. Increasingly the research community has viewed implementation of both the EUCRD and RGF as being disproportionate and ‘gold plated’.

In 2010 the Academy of Medical Sciences was invited by Government to review the regulation and governance of health research and make recommendations to improve current processes. Despite recent attempts to improve parts of the regulation pathway, the review identified that significant challenges remain, including: delays and duplication in obtaining NHS research permissions; a lack of proportionality in the regulation of clinical trials; and a healthcare culture that fails to fully support the value and benefits of health research. The Academy’s report recommended the creation of a new health regulator to rationalise the regulation and governance of health research; steps to streamline the approval of research studies by NHS Trusts; and changes to improve the UK environment for clinical trials.

Since the publication of the Academy’s report a number of positive changes have been made. The 2011 Plan for Growth outlined a package of government measures to foster growth in the healthcare and life sciences and took forward a number of recommendations from the Academy report, including establishment of a new Health Research Authority to streamline regulation and improve the cost effectiveness of clinical trials, and making future funding by the National Institute for Health Research to NHS Trusts conditional on meeting benchmarks. The Government has created a new duty for clinical commissioning groups to promote research innovation and for the NHS Commissioning Board to ensure provision of treatment costs for patients who are taking part in research. The MHRA are introducing a risk based approach to management of clinical trials based on the marketing status of an investigational medicinal produce and the standard medical care that would facilitate a proportionate approach to trial activities. The ongoing review of the EUCRD may offer a further opportunity to reduce the regulatory burden on academic trials.

The above changes are welcome and should result in reductions in the time to obtain approvals and permissions for clinical trials and some reduction in trial management costs. However it remains to be seen whether the health research regulatory agency will have sufficient authority to achieve more efficient working between regulators, researchers and the NHS to maintain the UKs historical position as a world leader in designing and delivering high quality clinical trials.

A36

Interim reports for data monitoring committee review vs final reports for regulatory filing

KyungMann Kim

University of Wisconsin-Madison, Madison, WI 53792, USA

E-mail: kmkim@biostat.wisc.edu

Trials 2011, 12(Suppl 1):A36

Independent data monitoring committee is responsible for review of the ongoing safety of participants and the validity and integrity of data in clinical trials and for making recommendations to the sponsor whether the trial should continue as planned, be modified or be terminated because of safety concern, futility of the trial, or treatment benefit. Use of DMCs by government and industry sponsors has been growing steadily ever since the announcement by the US National Institutes of Health in
A37

A practical solution to ‘Continuing Care Site’ issues in neonatal clinical trials – a pragmatic approach to regulatory and research governance review

Kayleigh Morgan¹, Ed Juszczak, Ursula Bowler
NPEU Clinical Trials Unit, University of Oxford, Oxford, UK
E-mail: kayleigh.morgan@npeu.ox.ac.uk

Trials 2011, 12(Suppl 1):A37

I2S2 is a randomised controlled trial of iodine supplementation in preterm infants examining whether iodine supplementation can improve neurodevelopmental outcome at two years of age. 20 Neonatal Units in the UK will recruit around 1400 infants <31 weeks of gestation into the trial; infants will receive a daily dose of sodium iodide or sodium chloride placebo until 34 week's corrected age. Due to the nature of neonatal care, approximately 50% of infants participating in I2S2 are likely to be transferred from the 'Recruiting Site' to another hospital for continuation of their clinical care. Hospitals which are not the primary research site may receive infants with little warning and be required to continue treatment under the trial protocol. Hospitals which are research active but have a reduced level of involvement have been defined as 'Continuing Care Sites' and 'Data Collection Sites'; such sites require Research Management and Governance review proportional to their level of involvement in the clinical trial. This three-tiered approach was negotiated between the NPEU Clinical Trials Unit, R&D Forum and the National Institute for Health Research. It was approved by the Medicines and Healthcare products Regulatory Agency and Research Ethics Service as part of a substantial amendment. The use of a generic Site-Specific Information (SSI) and supporting document entitled 'Statement of Responsibilities' clearly defines site involvement, NHS permission and funding; both documents negate the need for site agreements with the sponsor. NHS permissions are granted for each Neonatal Unit across the UK; those sites without identified Principal Investigators issue provisional NHS Trust approval with confirmation of Principal Investigator on point of transfer. Appropriate trial specific training materials are provided to all sites including training in the reporting of SAE's and SUSAR's. This approach allows NHS permissions to be issued in advance of any transfers of study infants recruited to I2S2 and can continue trial procedures at different sites. Withdrawing an infant from the trial because regulatory and research governance paperwork is not in place, seems a wasted effort for all individuals involved, and more importantly unethical to parents. This streamlined, pragmatic approach to governance review has proved a successful strategy from which many can learn.

Trial registry: [http://www.clinicaltrials.gov; Identifier: NCT00638092]. Funded by UK Medical Research Council.

Acknowledgements: Presented on behalf of the I2S2 collaborative group. The authors would like to acknowledge contributions and support from the R&D Forum, NIHR CRN CC, BBC CLRN and the University of Oxford.

A38

Exploring the ethical and practical challenges of conducting clinical trials in care home settings

Fiona Wood¹, Hayley Prout², Arun Acharjya³, Jacqueline Nuttall⁴, Kenny Hood⁵, Christopher Butler⁶
¹Department of Primary Care and Public Health, School of Medicine, Cardiff University. Heath Park, Cardiff, CF14 4XN, UK; ²South East Wales Trials Unit, Department of Primary Care and Public Health, School of Medicine, Cardiff University. Heath Park, Cardiff, CF14 4XN, UK.

Trials 2011, 12(Suppl 1):A38

Background: The PAAD Study (Probiotics for Antibiotic Associated Diarrhoea in care homes) involves 2 stages. The first stage is a prospective 12 month observational study to collect data on the amount and type of antibiotic prescribed episodes of Antibiotic Associated Diarrhoea and C.difficile Associated Diarrhoea and outcome in a randomly selected sample of 9 care homes. The second stage is currently being designed but will take the form of a RCT of probiotic vs. placebo administered alongside the antibiotic. These two studies are governed by separate laws and regulations in relation to mental capacity. PAAD stage 1 is covered by the Mental Capacity Act and, for those service users who lack capacity, personal consultants are able to give agreement to the service user’s participation. PAAD stage 2 is covered by the Medicines for Human Use (Clinical Trials) Regulations and consent is given by a personal legal representative or a professional legal representative of the participant. The PAAD study also presents novel challenges in relation to advanced consent. We propose to explore some of the ethical and practical challenges of conducting these studies (and others like it) within the care home setting using qualitative methods. Our purpose is to optimise the informed consent process in a vulnerable population in preparation for stage two of PAAD.

Methods: We are conducting focus groups with staff within the 9 care homes who are participating in PAAD stage 1. We are also conducting face-to-face interviews with service users who have capacity, with relatives of service users, and with GPs who provide primary care to the care homes. Focus group and interview schedules focus on issues such as: the merits and problems associated with a number of models of consent for both stage one and stage two of PAAD, views about the feasibility and acceptability of taking advance consent/assent for research trial procedures, and views about potentially raising concerns about the possibility of c.difficile circulating within the care home.

Results: Data collection is in early stages, and we will have data analysed by Autumn 2011.

Conclusion: We anticipate that our data will not only optimise the design of our informed consent process for stage 2 of PAAD, but will also provide generalisable insights to guide other studies which plan to conduct clinical trials amongst vulnerable populations within care home settings.

A39

Risk-adapted approaches to the management of clinical trials: guidance from the Department of Health (DH)/Medical Research Council(MRC)/Medicines and Healthcare Products Regulatory Agency (MHRA) Clinical Trials Working Group

Sarah Meredith¹, Martyn Ward², Gillian Booth³, Andrew Fisher⁴, Carrol Gamble⁵, Heather House⁶, Martin Landray⁷
¹MRC Clinical Trials Unit, London, UK; ²Clinical Trials Unit, Medicines and Healthcare Products Regulatory Agency, UK; ³MRC Clinical Trials Unit, Liverpool University, Liverpool, UK; ⁴Clinical Trials and Research Governance, Oxford University & John Radcliffe Hospital, Oxford, UK; ⁵Clinical Trials Service Unit, Oxford University, Oxford, UK.

Trials 2011, 12(Suppl 1):A39

In 2009, the Department of Health asked the MRC and MHRA to identify major obstacles to non-commercial clinical trials research in the UK and
suggest remedial actions. Risk-proportionality in trial management and monitoring was identified as a key area and a sub-group formed to:
- Develop a process to facilitate the agreement of key stakeholders on the level of risk associated with a clinical trial of an investigational medicinal product (CTIMP).
- Identify how risk-adapted approaches for CTIMPs can be achieved within the current regulatory framework.
- Develop guidance on risk assessment and the risk-proportionate management of clinical trials.

The resulting guidance focuses on the risks inherent in a trial protocol which impact on participant safety and rights, and the reliability of the results. A two-part assessment is suggested: 1) a simple IMP risk categorisation based on marketing status and standard medical care, and 2) assessment of the trial design, population and procedures to identify specific areas of vulnerability.

The first part, IMP risk category, has implications for simplifications of initiation and conduct of a CTIMP that may be possible within the current regulatory framework. Possible risk-adaptations include: the need for competent authority authorisation; content of the Clinical Trials Authorisation (CTA) application; IMP management; safety surveillance; trial documentation; and GCP Inspection. The risks associated with the IMP also determine trial procedures for monitoring participant safety.

The second part of the risk assessment addresses other aspects of clinical trial design and methods: safety risks from clinical procedures; risks related to participant rights; and risks to the reliability of trial results. It is designed to help trialists identify potential vulnerabilities and to prepare tailored trial management and monitoring plans to minimise risks which may be reviewed and modified throughout the life of a trial. The IMP risk category and safety monitoring plan may be submitted to the MHRA with the CTA application to ensure that there is shared understanding on this key aspect of a trial. We hope that the entire risk assessment and associated plans will provide the basis for a common understanding of stakeholders of the risks for that trial, and facilitate a risk-proportionate approach to trial activities.

The guidance is available on the MHRA and NETSCC websites.

Acknowledgements: We would like to thank Louise Mawer and Wilma van Riel and the members of the Clinical Trials Working Group for their contribution to the development of this guidance.

HEALTH ECONOMICS AND TRIALS

A40

Using economic modelling to contribute to the prioritisation and design of clinical trials: ready for prime time

Mark Sculpher
Centre for Health Economics, University of York, York, YO10 5DD, UK
E-mail: mark.sculpher@york.ac.uk
Trials 2011, 12(Suppl 1):A40

It is now generally recognised that health systems internationally have resource constraints and need to set priorities in selecting the interventions they fund. Hence considerations of value for money are central in health policy work. Cost-effectiveness analysis is a formal assessment of value, and is central to how many health care systems make resource allocation decisions. The purpose of clinical trials is to generate evidence to support various types of decision making, and an increasing number of trials are designed primarily to inform health system decisions (e.g. the NIHR Health Technology Assessment Programme in the UK). Consequently, the process of designing a clinical trial and determining whether it is a priority for research funding needs an explicit consideration of whether it can contribute to better decisions in the future. This need to tie trials to ultimate decisions about cost effectiveness is at odds with the role that economic analysis has generally assumed in clinical trials over the last 20 years - i.e. trials being designed largely to address clinical questions but offering some add-on economic data to facilitate some form of cost effectiveness analysis when the trial reports.

Analytical methods exist which assess the value of trials in terms of the likelihood of their improving resource allocation decisions. Based on Bayesian decision theory, these methods quantify the uncertainty relating to the most cost effective approach to managing a specific patient group based on existing evidence, couple this with factors such as the size of the patient population to estimate the expected value of perfect information which can be used to begin prioritising trials. Extensions to these methods consider the appropriateness of specific trial designs by assessing their marginal costs and benefits in terms of reduction in the cost of decision uncertainty. These methods have been used within the HTA Programme and there are examples of impact in trial funding decisions. They are, however, perceived as being complex and requiring significant analysis. Although this perception can be challenged, there is undoubtedly a need for methods that can be applied routinely to assess the potential trial value. These might include the need for all trial proposals to include modelling to establish a plausible effect size which is sufficient to demonstrate an intervention’s cost-effectiveness in the context of all existing evidence. This modelling would also be used to justify design features such as the choice of comparators, endpoints and follow-up periods.

A41

Analysis of adverse events and quality of life data for an economic evaluation of adjuvant chemotherapy in colorectal cancer: when can we stop collecting?

Kathleen A Boyd, Andrew H Briggs, Jim Paul, Tim Iveson, Rachel Midgley, Andrea Harker, Gaynor Bates, Laura Alexander, Jim Cassidy

1Health Economics & Health Technology Assessment, University of Glasgow, Glasgow, G12 8RZ, UK; 2Cancer Research UK Clinical Trials Unit Glasgow, Glasgow, G12 0YN, UK; 3Southampton University Hospitals NHS Trust, Southampton, SO16 6YD, UK; 4Department of Clinical Pharmacology, Oxford, OX3 7DQ, UK; 5Oxford Clinical Trials Office, Oxford, OX3 7DQ, UK
E-mail: Kathleen.Boyd@glasgow.ac.uk
Trials 2011, 12(Suppl 1):A41

Background: The SCOT colorectal cancer trial is a phase III randomised controlled, multi-centre, non-inferiority trial comparing the efficacy of 12 weeks of chemotherapy versus 24 weeks and the associated toxicity, along with a lifetime economic analysis. The economic analysis will use prospective trial data to inform a lifetime cost-effectiveness model. Quality of life and toxicity data are only collected in a proportion of patients as the sample size required for these comparisons is much smaller than that required to show a significant difference in efficacy. This paper reports on an analysis of adverse events (AEs) and quality of life data, undertaken to determine whether it was appropriate to stop collecting quality of life data from new recruits.

Results: Univariate analysis showed that the majority of AEs significantly reduce quality of life (significant at 3% or less); however, multivariate analysis was inconclusive as AEs were highly correlated with one another. Univar and multivariate regressions were re-run using indicator variables for AE grade groups for each specific adverse event type, and showed a highly significant (1% level) negative impact on quality of life as severity of grade increased, particularly for the adverse events Diarrhoea (-0.04 grade 1/2, -0.09 grade 3/4), Fatigue (-0.02 grade 1/2), Nausea (-0.05 grade 1/2, -0.14 grade 3/4), Neuropathy sensory (-0.02 grade 1/2, -0.19 grade 3/4), and Vomiting (-0.05 grade 1/2). Quality of life had a significant negative relationship with severity of AE grade, regardless of the specific event (0.017 grade 1, -0.058 grade 2, -0.062 grade 3, -0.145 grade 4).

Conclusions: Adverse events impact negatively on quality of life; however, multivariate regression was inconclusive due to multicollinearity between the adverse events. There was strong evidence to show that the AEs and AE grades significantly reduced quality of life using univariate analysis, and therefore we can be confident that the sample from June 2008-Dec 2010 is sufficient and can stop collecting such data from new recruits.

Acknowledgements: The SCOT study is supported by a grant from the MRC.
Cancer Research UK supports the Glasgow Clinical Trials via a programme grant.
A42

What is the value of collecting detailed costing data in clinical trials?
Helen Dakin, Giselle Abangma, Sarah Wordsworth
Health Economics Research Centre, University of Oxford, Oxford, OX3 7LF, UK
Trials 2011, 12(Suppl 1):A42

Objectives: Cost data for trial-based economic evaluation can be obtained through micro-costing (collecting resource use and unit cost data for each centre or patient), gross-costing (average costs based on top-line budgets) or provider tariffs (e.g. healthcare resource groups, HRGs). Most studies use a combination of approaches due to data availability, although there is little guidance on which is best. We report a systematic comparison of the three costing approaches in IVAN: a non-inferiority randomised controlled factorial trial of treatment regimens for age-related macular degeneration (AMD), where policy makers are interested in the efficacy and cost-effectiveness of two dosing regimens of bevacizumab (Avastin) and ranibizumab (Lucentis). We aimed to assess the extent to which micro-costing, gross-costing and HRGs differ, and to investigate resource use variation between UK hospitals and explore possible reasons for this variability.

Methods: We compared micro-costing, gross-costing and HRG estimates of consultation costs using IVAN data. Nineteen IVAN trial centres were sent questionnaires on the resources required to set up and run clinics. Resources were valued using national unit costs to give micro-costing estimates that are compared against Department of Health gross-costing estimates and the HRG for ophthalmology outpatient consultations. Regression analyses explore the variability between centres.

Results: Fourteen centres (74%) returned questionnaires. The mean cost of a follow-up ophthalmology outpatient consultation is £74 compared with an HRG cost of £53. Preliminary micro-costing suggests that both HRGs and gross-costs substantially underestimate the cost of consultations to administer treatment (excluding drug costs) or monitor outcomes. Micro-costing also highlighted substantial variation in consultation costs, facilities, organisation and resource use not captured within HRGs or gross-costs. Clinic size did not explain variations in consultation costs.

Conclusions: Although data analysis is ongoing, initial results suggest that micro-costing estimates for administration and monitoring of Avastin/Lucentis are higher than gross-costs or HRGs. HRG costs were lowest, suggesting that hospitals must cut costs substantially to break even on such consultations. Differences in costing methodology are likely to affect cost-effectiveness results; particularly in the context of a non-inferiority trial comparing different dosing regimens, where cost differences will drive conclusions about cost-effectiveness. Only micro-costing differentiated between consultations for monitoring and drug administration. Micro-costing (unlike other approaches) also showed how costs and patient management vary between UK centres, facilitating analysis of heterogeneity and identification of potential efficiency improvements. This demonstrates the value of collecting detailed resource use data in trials.

Acknowledgements: Prepared on behalf of the IVAN Trial Study Group.

The IVAN trial and its economic evaluation are funded by the UK National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Programme (project number 07/36/01, http://www.hta.ac.uk). The views and opinions expressed are those of the authors alone and do not reflect those of the NIHR or the UK Department of Health.

A43

Bayesian analysis of trial-based cost-effectiveness in the presence of missing data; effects of including covariates on efficiency of estimation in the ASTER trial
Linda Sharples1, Christopher Jackson1, Ella Wheaton2, Robert Rintoul1
1MRC Biostatistics Unit, Cambridge, UK; 2Cancer Research UK, London, UK;
Papworth Hospital NHS Foundation Trust, Cambridge, UK
Trials 2011, 12(Suppl 1):A43

Objectives: Bayesian estimation of missing resource use data and expected costs in the ASTER trial of endoscopic staging followed by surgical staging if negative (ES), compared with surgical staging alone (SS), in candidates for lung cancer surgery. To assess how covariates, that are included in the model to justify a ‘missing at random’ assumption, affect estimates of expected costs.

Methods: ASTER was a prospective, international, open-label, randomised-controlled study, with a trial-based economic analysis over 6 months. Due to delays in starting the health economic study, resource use data were collected prospectively for the second half of the study only. Although resource use data could be ascertained retrospectively, some items were difficult to ascertain once patients had been discharged from the trial centre. A Bayesian parametric model was developed to estimate missing resource use items and expected costs. Missing resource use data were modelled using Binomial, Poisson, over-dispersed equivalents of these, or using a hurdle count model if only a proportion of the patients had the event (e.g. chemotherapy). Covariates considered were randomisation group, centre, age, sex and stage of lung cancer. The total expected cost was calculated as the sum of the resource use component-specific expected costs for each randomisation group.

Results: ES was more sensitive, resulted in fewer futile thoracotomies and had better utility during staging than SS. All patients had initial diagnostic tests and management recorded but subsequent resource use components were missing for 10-20% of cases, and only 71% had complete resource use data. Using the complete cases only, the mean 6 month cost of ES was £10,614 (£8515, £13,073) per patient versus £11,788 (£9053, £15,321) for SS, mean difference £1174 (-£948, £3912), so that ES was cheaper but with considerable uncertainty in these estimates. The Bayesian model aimed to recapture power lost due to missing data and when randomisation group was the only covariate, point estimates were reduced by 5% and posterior standard deviations were reduced by 10% compared with complete case analysis. Inclusion of other covariates resulted in small subgroups, imprecise point estimates for covariates and a resulting increase in the posterior variance for expected costs.

Conclusions: Resource use with patient groups is highly variable and trials are rarely powered for secondary outcomes that drive costs, so that inclusion of many parameters in a Bayesian analysis may result in inefficient estimation. Covariate selection should consider both the missing data mechanism and efficiency.

Acknowledgements: We acknowledge to contribution of the ASTER trial investigators. The UK data collection and cost-effectiveness study were funded by the National Institute for Health Research Health Technology Assessment programme.

A44

HERALD (Health Economics using Routine Anonymised Linked Data)
Muhammad J Huq1, Sinead Brophy1, Steven Macley1, Leila M Pinder1, Mark D Atkinson1, Roxanne Cooksey1, Ceri J Phillips1, Stefan Siebert1
1College of Medicine, Swansea University, Swansea, SA2 8PP, Wales, UK;
2College of Human and Health Sciences, Swansea University, Swansea, SA2 8PP, Wales, UK
Trials 2011, 12(Suppl 1):A44

Background: Health economic analysis traditionally relies on patient derived questionnaire data, routine datasets, and outcomes data from experimental randomised control trials and other clinical studies, which are generally used as stand-alone datasets. Herein, we outline the potential implications of linking these datasets to give one single joined up data-resource for health economic analysis.

Method: The Health Information Research Unit (HIRU) at Swansea University has set up the Secure Anonymised Information Linkage (SAIL) database, which brings together and links a wide range of anonymous patient-level data [1,2]. The linkage of individual level data from questionnaires with routinely-captured health care data allows the entire patient journey to be mapped both retrospectively and prospectively. We illustrate this with examples from a population-based Ankylosing Spondylitis (PAS) cohort [3] by linking patient reported study dataset with the routinely collected general practitioner (GP) data, inpatient (IP) and outpatient (OP) datasets, and Accident and Emergency department data in Wales.

Potential benefits of data linkage: The linked data system allows: (1) retrospective and prospective tracking of patient pathways through multiple healthcare facilities; (2) validation and clarification of patient-reported recall data, complementing the questionnaire/routine data information; (3) obtaining objective measure of the costs of chronic conditions for a longer time horizon, and during the pre-diagnosis period; (4) assessment of health service usage, referral histories, prescribed drugs
and co-morbidities; and (5) profiling and stratification of patients relating to disease manifestation, lifestyles, co-morbidities, and associated costs.

**Results:** Using the GP data system we tracked 183 AS patients retrospectively and prospectively from the date of questionnaire completion to gather the following information: (a) number of GP events; (b) presence of a GP ‘drug’ read codes; and (c) the presence of GP ‘diagnostic’ read codes. We tracked 236 and 296 AS patients through the OP and IP data systems respectively to count the number of OP visits; and IP admissions and duration. The results are presented under several patient stratification schemes based on disease severity, functions, age, sex, and the onset of disease symptoms.

**Conclusion:** The linked data system offers unique opportunities for enhanced longitudinal health economic analysis not possible through the use of traditional isolated datasets. Additionally, this linkage provides important information to improve diagnostic and referral pathways, and thus helps maximise clinical efficiency and efficiency in the use of resources.

**Acknowledgements:** Funding for the PAS cohort was provided by a Medical Research Council (MRC) and National Institute for Social Care and Health Research (NISCHR) grant (G080058).

**References**


**A46**

**How best to handle the evidence in a cost-effectiveness analysis of TAVI in the UK**

Aileen Murphy1,2, Elizabeth Fenwick1, Andrew Briggs3, Matthew Neilson2

1Department of Economics, University College Cork, Ireland; 2HEHTA, University of Glasgow, Glasgow, Scotland; 3Department of Health Sciences, School of Medicine, University of Bristol, Bristol, BS8 2PS, UK

**Aims:** Despite available literature and results from RCT the question of whether Transcatheter Aortic Valve Replacement (TAVI) is cost-effective, and for whom, persists in the UK. This paper investigates how evidence from published literature, RCT and registry data can be employed and best synthesised to determine the cost effectiveness of TAVI and establish for whom is TAVI most suitable.

**Methods:** A decision analytical model, incorporating a short-term decision tree and long-term Markov model were constructed. The use of Monte Carlo simulation and Value of Information analysis enables both the cost-effectiveness of TAVI to be estimated, based on the existing information, and the value of, and requirements for, further evidence collection to be determined. The model allows for the heterogeneous patient population by considering different three patient risk cohorts (low, medium and high). Initially the model was populated using the best available literature; then updated when RCT results became available. However, the trial data did not fully reflect UK practice as such the model was updated with UK Registry data. Probabilistic analysis is undertaken at each stage and results are presented for the cost-effectiveness and the value of undertaking further research.

**Results:** The initial model results revealed that TAVI is cost-effective for inoperable patients only compared to medical-management with an ICER of £22,603. The EVPI per high risk patient ranges from £462 to £1,277. When the results from the PARTNER trial were incorporated into the model, TAVI remained cost-effective for these patients compared to medical-management, ICER £25,875. The EVPI using the PARTNER evidence ranged from £69 to £1,170 per patient. At each of these stages, there is potential value in undertaking further research in this patient sub-group within the UK.

**Conclusions:** This paper reveals that a flexible model accommodating an evolving evidence base is a necessity when dealing with novel technologies for which data is scarce, such as TAVI. While the US RCT evolves the database, the suitability of this data is questionable for a UK cost-effectiveness model. Similarly the suitability of the UK TAVI registry is also undefined as it is not randomised. This paper examines how best to synthesise and incorporate these data into a decision analytical model to reveal for whom TAVI is cost-effective in the UK.

**LATE BREAKING SESSION / OTHER**

**A47**

**The influence of CONSORT on the quality of reporting of randomised controlled trials: an updated review**

Lucy Turner1,2, David Moher1, Larissa Shameer1, Laura Weeks3, Jodi Peters1, Amy Plint4, Douglas G Altman2, Kenneth F Schulz2

1Ottawa Hospital Research Institute, Ottawa, Canada; 2Children’s Hospital of Eastern Ontario, Ottawa, Canada; 3Centre for Statistics in Medicine, University of Oxford, Oxford, UK; 4FH360, North Carolina, USA

**Background:** The Consolidated Standards of Reporting Trials (CONSORT) Statement was developed in response to concerns about the quality of
reporting of randomized controlled trials (RCTs). The checklist is an evidence-based minimum set of recommendations for reporting RCTs, intended to facilitate the complete and transparent reporting of RCTs and aid in their critical appraisal and interpretation. In 2006, Plint and colleagues published a systematic review examining the effectiveness of CONSORT for improving the reporting of RCTs in journals that have formally endorsed the guidance (i.e. at minimum recommend that authors use CONSORT) [1]. Despite poor methodology of some included studies, use of CONSORT was found to be associated with improvement in the quality of reporting of RCTs.

Objective: To update Plint et al.’s systematic review assessing the influence of the CONSORT Statement’s checklist (2001) on the quality of reporting of RCTs.

Methods: Conventional systematic review methods employed in the original review by Plint et al. have been implemented. The search for new studies spanned August 2005 – March 2010. Two independent reviewers screened studies for eligibility; extraction and validity assessment of studies were conducted by a single reviewer and a second reviewer performed verification. Reporting quality was assessed by comparing the proportion of RCTs adhering to individual CONSORT items or a total sum score between comparison groups.

Results: Of 2896 possibly relevant studies, 53 reports of 50 quasi-experimental studies have been included, compared to 8 in the earlier review. In total these studies assessed adherence to CONSORT in 16,222 RCTs. When comparing reporting in RCTs of CONSORT endorsing journals with CONSORT non-endorsing journals; 25 of 27 outcomes yield higher relative reporting of these items in endorsing journals, of which 7 were statistically significant. The largest positive effect, across 16 studies, showed that reporting of allocation concealment was 81% greater in CONSORT endorsing journals (RR = 1.81, 95% CI 1.37 to 2.40). There is no evidence to suggest that CONSORT endorsement has a detrimental influence on the quality of reporting of RCTs.

Impact: Despite the questionable validity of the included studies, this updated review provides stronger evidence suggesting that CONSORT is associated with improved reporting of RCTs. This information is helpful to authors, peer-reviewers and journal editors when deciding whether to recommend or enforce the use of CONSORT.

Reference

A48
What can available randomised controlled trials evaluating monitoring strategies tell us about the design and analysis of future trials?
Jacqueline Dinnes1*, Jenny Hewison1, Doug Altman1, Jon Deeks1
1Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK; 2Leeds Institute for Health Sciences, University of Leeds, Charles Thackrah Building, 101 Clarendon Road, Leeds, LS2 9LT, UK; 3Centre for Statistics in Medicine, University of Oxford, Wolfson College Annexe, Linton Road, Oxford, OX2 6UD, UK
Trials 2011, 12(Suppl 1):A48

Background: Randomised controlled trials (RCTs) of monitoring regimes present unique challenges. Trials of monitoring evaluate a strategy, “a planned and organised system of repeated assessments and subsequent decisions about additional interventions, such as starting, stopping or modifying treatment” [1] all of which should be specified in advance, and ideally supported by previous research. The complexity of the intervention and consequent potential for “interactions between tests, repeated tests, test results and the decisions based on these results” [1] also necessitates large sample sizes in order to detect an effect on important patient outcomes.

Objective: To gain an insight into the extent to which monitoring schedules evaluated in RCTs have been informed by prior research and to highlight potential issues for design and analysis of future trials.

Methods: A review of available RCTs of monitoring regimes intended to allow earlier intervention in people with, or at high risk of, a condition that is likely to recur or progress was conducted. An electronic search of the CENTRAL database was carried out using the keywords (monitor* or serial* or surveill*) in the title or the same keywords in the abstract combined with (biomarker* or marker* or test* or examination* or measure*). RCTs comparing at least one formal monitoring strategy to no formal monitoring, to an alternative monitoring strategy, or to immediate treatment were included. Trials of monitoring where the main purpose was treatment titration, improvement in adherence to a treatment regimen, or evaluating ways of delivering monitoring were excluded. One author conducted the searches, screened the search results for relevant studies and carried out the data extraction. A narrative synthesis will be performed.

Results: Results will be presented for over 20 RCTs. In addition to information on trial quality, data will be presented on the extent to which the components of the monitoring schedule (testing frequency, threshold for intervention and actions consequent to crossing that threshold) have been specified. The MRC framework for evaluating complex interventions will be used to help inform the extraction of this information [2]. A further focus will be on the types of outcomes assessed by the trials and reporting of sample size calculations.

Conclusions: The review’s results will be used to inform guidance regarding the future evaluation of monitoring regimes.

References

A49
Cochrane systematic reviews as a source of information for practice and trials
Mike Clarke1*, Thomas T Clarke2, Lorcan Clarke2
1All-Ireland Hub for Trials Methodology Research, Queen’s University Belfast, Northern Ireland; 2UK Cochrane Centre, National Institute for Health Research, Oxford, UK
E-mail: m.clarke@qub.ac.uk
Trials 2011, 12(Suppl 1):A49

Background: Systematic reviews should be used to provide the ethical, scientific and environmental justification for a new randomised trial [1]. The Cochrane Collaboration is the world’s largest organisation dedicated to preparing and maintaining systematic reviews of the effects of healthcare interventions [2]. These are published in the Cochrane Database of Systematic Reviews (CDSR) and all Cochrane Reviews have a rigid structure, which includes the authors’ conclusions on the implications for practice. Therefore, this collection of reviews provides a readily available source of information on the evidence base for the use of a large number of treatments in health care.

Objectives: We assessed all new and updated reviews in the first twelve monthly issues of CDSR from February 2010 to January 2011 to identify their relevance to the evidence base for current practice and for future trials to resolve continuing uncertainties.

Methods: At least two authors independently examined the implications for practice in the authors’ conclusions of each new and updated review in issues 2-12 2010 and issue 1 2011 of CDSR. Each review was coded to indicate whether its authors concluded that a specific intervention should only be used in research, was not supported or refuted by the evidence in the review, had been shown to be effective, or should not be used in practice (or could not be recommended). Each review was also coded by area of health or health care. The final decisions on coding were taken by one author.

Results: These 12 monthly issues of CDSR contained 390 new and 462 updated reviews, many of which provide evidence on more than one intervention. The authors of 25 (2.9%) of these reviews concluded that an intervention should only be used in research and 494 (58.0%) concluded that the evidence for an intervention was insufficient to support or refute its use. At least one intervention in 413 (48.5%) of these reviews was judged to be effective by the review authors and the authors concluded that an intervention should not be used or could not be recommended in 139 (16.3%) reviews.

Conclusions: Cochrane Reviews identify many interventions for which the research evidence supports their effectiveness and many where the
Evidence is insufficient to assess benefits and harms. This latter group, along with reviews in which the authors recommend restricting an intervention to use in research, provide examples of uncertainties which could be resolved through randomised trials.

**References**


### A50

**Biases in clinical trials with sequential monitoring**

Manjula Schou, Ian C Manschner

Department of Statistics, Macquarie University, Sydney, NSW 2109, Australia

**Objectives:** It is well known that a sequentially monitored clinical trial that stops early for benefit has a crude treatment difference that overestimates the true treatment effect. This has led to extended debate in the literature, with some researchers arguing that early stopping is an important source of bias in meta-analyses of clinical trials. We therefore investigated the implications of excluding studies that stopped early, so-called truncated studies, from estimation of treatment effects.

**Methods:** The effect of excluding truncated studies was investigated by examining the statistical properties of sequentially monitored studies conditional on reaching the planned final analysis. Using theory and simulation we studied clinical trials with standard sequential rules for stopping early due to benefit. As well as estimation bias, we studied information bias measured as the difference between standard measures of the statistical information, such as sample size, and the actual information based on the conditional sampling distribution.

**Results:** We found exclusion of truncated studies leads to both estimation bias and information bias. Treatment differences are underestimated and information is overestimated. Most importantly, the magnitude of information bias is an increasing function of the magnitude of estimation bias. This has important implications for meta-analyses that typically weight by sample size. In particular, it means that studies with the most biased treatment effect are the most overweighted studies in a meta-analysis. The magnitudes of both estimation and information biases can be practically significant. When all studies were included in meta-analyses, both truncated and non-truncated, the estimation of treatment effects was unbiased.

**Conclusions:** Crude methods of analysis for sequentially monitored studies can lead to underestimation bias if truncated studies are excluded from estimation of treatment effects. Furthermore, information bias resulting from this exclusion leads to a double whammy effect, in which the most biased studies are the most overweighted studies in a meta-analysis. Since exclusion of truncated studies is problematic, we advocate wider reporting of adjusted estimates of treatment effects that take account of any interim monitoring, and recommend that all studies, both truncated and non-truncated, are included in meta-analyses.

### A51

**Making best use of existing evidence when planning trials**

Verena Roloff', JPT Higgins

MRC Biostatistics Unit, Cambridge, UK

**Objectives:** Although meta-analyses are typically viewed as retrospective activities, they are increasingly being applied prospectively to justify planned research by providing up-to-date evidence on specific research questions. Since healthcare policy makers often base decisions on systematic reviews of reliable evidence rather than on single clinical trials, trialists should consider planning additional research with regard to the meta-analytic result.

Since adding a new study to an existing meta-analysis creates a multiple testing scenario, which standard methods do not well address, nominal significance levels need to be adjusted to preserve the overall type-I error rate. Sequential approaches to meta-analysis have been proposed, although these do not lead directly to recommendations on planning new studies.

**Methods:** We propose to use the framework of adaptive clinical trial design methods to plan studies for addition to a meta-analysis.

**Results:** We discuss the implementation of the adaptive method proposed by Bauer and Köhne (1994) to the meta-analysis framework, and discuss the implications of heterogeneity among studies when applying them to a random-effects meta-analysis.

By additionally deriving the conditional power of a random-effects meta-analysis under different assumptions (about the number of additional studies, their information sizes and of the heterogeneity anticipated among them), we can assess the impact of new trials on the meta-analysis. For instance, if heterogeneity is anticipated, it might not be possible for a single study to reach the desirable power no matter how large it is. Simple graphs that summarize the conditional powers of possible design alternatives provide a convenient way to explore different strategies for planning research.

We illustrate this framework for designing new trials for different scenarios using data from the Cochrane Database of Systematic Reviews.

**Conclusions:** We provide a framework for designing new trials with regard to the meta-analytic result rather than the study in isolation.

### A52

**Protecting intellectual property associated with health technology trials – another barrier to multi-centre trials?**

Sue Ross',1 Laura Magee2, Stephen Wood1

1Departments of Obstetrics and Gynaecology and Community Health Sciences, University of Calgary, Calgary, T2N 2T9, Canada; 2Department of Medicine, University of British Columbia, Vancouver, V6H 3N1, Canada

E-mail: sue.ross@bertahealthservices.ca

**Objectives:** To examine the approaches to protection of intellectual property in multi-centre trials currently being conducted in Canada.

**Methods:** Two ongoing international multicentre perinatal trials, both funded by the Canadian Institutes of Health Research, were selected for study on the basis of their contrasting approaches to protecting intellectual property. These approaches were examined in detail to understand their motivation, and to estimate the impact of these approaches on centre recruitment.

**Results:** CHIPS (Control of Hypertension in Pregnancy Study, ISRCTN71416914) – is recruiting 1028 pregnant women in 4 countries. Women with hypertension are randomised to tight or less tight control of hypertension. Primary outcome: composite of pregnancy loss/neonatal intensive care.

Intellectual property is safeguarded by publishing the protocol online [1]. Positive consequences: possible/actual sites have easy access to full study design; potential for open discussion between collaborators; study investigators will be held to high standards of reporting. Negative consequences: details are available with potential for plagiarism.

FACT (Folic Acid Clinical Trial, ISRCTN23781770) – is recruiting 3656 pregnant women in 4 countries. Pregnant women are randomised to receive either 4 mg folic acid or placebo daily. Primary outcome: development of pre-eclampsia.

Intellectual property is safeguarded by requiring local investigators/institutions to sign non-disclosure agreements (NDAs) before the full protocol is provided. Positive consequences: details of the study not available unless legal agreement is signed. Negative consequences: may restrict academic openness; provide additional barriers to site recruitment; investigators may present selected results.

**Conclusions:** The two trials illustrate contrasting approaches to protecting intellectual property associated with study design. This issue is becoming more important for academic institutions whose reputations and wealth are influenced by ownership and management of the intellectual property generated by faculty members. Some institutions prefer to manage risk using legal measures. In the case of trials, institutions protect their intellectual property by introducing NDAs into the sub-site agreement process. NDAs between the lead institution and sub-sites may represent a legally responsible approach. Unfortunately there are potential disadvantages: adding an extra legal step into sub-site recruitment will make this process more difficult; this step may reduce academic openness and...
collegiality; and restricting the availability of the protocol could allow investigators to present selected results. The use of non-disclosure agreements is an increasing trend in Canada. This trend will impact on the work of clinical trialists, perhaps making site recruitment even more difficult.

Reference

A53
Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies
Mark Jitlal,1, Iftekhar Khan1, Sow Ming Lee2, Allan Hackshaw1
1Cancer Research UK & UCL Cancer Trials Centre, University College London, London, W1T 4J, UK; 2Department of Oncology, University College London Hospital, London, NW1 2PG, UK
E-mail: mj@ctc.ucl.ac.uk
Trials 2011, 12(Supp 1):A53

Background: Phase III clinical trials are generally large and expensive, so stopping early for futility is a potentially attractive approach. It could avoid using further patients and funds on an ineffective intervention. We aimed to see how well futility performs for clinical trials in practice.

Materials and methods: We retrospectively applied a futility method to ten cancer trials, in which the final hazard ratios (HR) showed a large, moderate, or no treatment benefit. The target sample size was reached in all. The conditional power (CP) is the probability of obtaining a HR of a specified magnitude, which is statistically significant. Futility analyses were performed after 25, 50 and 75% of patients were recruited, or events observed. Two methods were used, assuming future data are consistent with either (i) the target HR, or (ii) the observed interim HR. Low CP suggests futility.

Results: Futility analyses could stop some trials with no overall benefit, but not all, depending on the method used and timepoint. For example, after observing 50% of the target number of events, and assuming that future data is consistent with observed data, 4 out of 5 trials with no benefit could be stopped early (CP≤9%). Among 3 (of these 4) studies, trial duration could be reduced by 4-40 months, saving £44k-£385k. The other trial would have already finished accrual and hence no savings made. Of concern is that all 4 trials with moderate treatment effects could have stopped early at some point. For example, assuming that future data is consistent with the observed data, these trials could have stopped after 25% of patient accrual (CP≥9%). However, the final HRs for all four trials showed clinically worthwhile benefits.

Conclusions: Appropriate application of futility methods can substantially shorten trial duration and reduce costs for trials which ultimately show no benefit. However, studies with moderate treatment effects could be stopped early, whilst some studies with effect may not have been detected as such. Futility needs to be applied with great care to avoid missing a worthwhile treatment. We suggest several criteria for stopping a trial early: low CP (e.g. ≤15%); sufficient number of events; remaining patient accrual is likely to take several months; and lack of evidence of a benefit for important secondary endpoints and pre-defined subgroups.

A54
The use of systematic reviews in the design of randomised trials
Ashley P Jones1, Carrol Gamble2, Mike Clarke3, Paula R Williamson1
1Department of Biostatistics, University of Liverpool, L69 3GS, UK; 2Clinical Trial Service Unit, Oxford, OX3 7LF, UK
E-mail: apjones@liverpool.ac.uk
Trials 2011, 12(Supp 1):A54

New randomised trials should be planned and reported taking account of knowledge from a systematic review of the existing research, but there is little empirical evidence to show how systematic reviews are used in the planning stages of new trials.

A systematic review could be used to inform the design of a new trial in several ways: (1) to choose the most appropriate forms of the interventions for the experimental and control groups, (2) to inform the sample size calculation (e.g. an estimate of the standard deviation), (3) to aid the choice decision of outcomes to measure (or the definition of an outcome), (4) to identify potential problems with consent, treatment withdrawal or retention, (5) to predict likely adverse events that may not otherwise be expected.

Furthermore, a systematic review could be used to conduct a sample size calculation for an updated meta-analysis incorporating the existing evidence and the eventual findings of the trial being planned. Although there is debate as to whether a trial should be powered in its own right or as part of an updated meta-analysis, there might be outcomes that will only have adequate power in the context of a meta-analysis. Among the issues to consider are whether the trial is justified if the existing meta-analysis is significant, whether it is possible to conduct a trial involving fewer patients and thus reach a decision on the most appropriate treatment earlier, and whether statistical heterogeneity might require the trial to be larger than the estimate if it was powered in isolation.

We present a sources of data on the use of systematic reviews in the design and conduct of randomised trials: a cohort of HTA-funded studies in which we explore how the trial design was informed by the existing evidence base. Our findings will provide important information for trialists and trial funders.

Acknowledgements: Thanks go to HTA for providing details on successful funded applications between 2006 and 2008.

A55
Central statistical monitoring in clinical trials
Amy A Kirkwood, Allan Hackshaw
CRUK and UCL Cancer Trials Centre, London, W1T 4J, UK
Trials 2011, 12(Supp 1):A55

Background: On-site monitoring is a common but time-consuming and expensive activity, with little evidence that it is worthwhile. Centralised statistical monitoring (CSM) is a much cheaper alternative, where data checks are performed by the co-ordinating centre, reducing the need to visit every site. Although some publications have outlined possible methods, few have applied them to data from real clinical trials.

Methods: R programs were developed to check data at either the patient or site level, for fraud or data errors. These included finding anomalous data patterns, digit preference, rounding, incorrect dates (eg weekends/holidays),
values of variables too close or too far from the means, odd correlation structures and extreme values or variances. We applied these to 3 trials: (i) where data had already been checked, (ii) an ongoing trial where our findings could be checked in real-time, and (iii) where data errors and fake patients were created.

Findings: The programs were designed to be run automatically and produce simple tables or figures. Few errors were detected in the trial where data had already been checked (as expected). Most data errors were found in the two other trials. The programs were able to detect data errors, as well as fabricated patients that we generated to have values that were too close to the multivariate mean (fig. 1). They also detected centers that had too few or too many serious adverse events (fig. 2). It might be difficult to reliably apply some of the programs to centers with few patients. Several patients that were fabricated were not detected because the data did not follow the assumptions used by the R-programs, or the number of fabricated patients within a center was too small. Examples of the different output produced, including easy-to-read diagrams and how they are interpreted, could be shown and discussed, along with their strengths and limitations.

Conclusions: CSM appears to be a cost-effective and worthwhile alternative to on-site monitoring. It can identify incorrect patient data, or center where the data considered together is too different to all other sites and therefore should be reviewed. However, more research is needed to identify which situations CSM does not work well in.

Meta-analysis of individual participant data (IPD) is widely accepted as the most reliable approach for systematic reviews. Advantages include standardising outcome definition across studies, increased potential to investigate subgroups, reducing bias by analysing on an intention to treat basis, minimising the possibility of within study selective reporting, thorough analyses of time to event outcomes, opportunities to identify unpublished studies through collaboration with the original researchers, and incorporating additional follow-up.

IPD provides a rich source of information that allows clinical and methodological developments to extend beyond exploring the main effects that are traditionally of interest in a single trial or systematic review. These opportunities, coupled with the resources required for the IPD approach which are often prohibitive for reviewers, make it essential that as much use as possible is made of IPD that have been collected.

We propose that a secure central repository be established to store previously collected IPD. Restricted access to the central repository would only be granted following an approval process that would involve the original reviewers and a nominated committee. The central repository would facilitate exploring additional clinical and methodological questions across a range of studies and reviews.

To assess the feasibility of developing and managing a central repository, we have undertaken an on-line survey of 70 IPD reviewers registered with the Cochrane IPD Meta-analysis Methods Group. We asked about their willingness to provide anonymised IPD from their review and asked about practical issues that this may raise. Non-respondents have been reminded about the survey up to three times. Analyses are ongoing and will be presented, along with future plans at the conference.

Trials 2011, Volume 12 Suppl 1
http://www.trialsjournal.com/supplements/12/S1

A56
Feasibility of establishing a central repository for the individual participant data from research studies
Catrin Tudur Smith, Kerry Dwan, Mike Clarke, Richard Riley, Douglas Altman, Paula Williamson
1 North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, UK; 2 All-Ireland Hub for Trials Methodology Research, Queen’s University Belfast, UK; 3 Public Health, Epidemiology and Biostatistics, University of Birmingham, UK; 4 Centre for Statistics in Medicine, University of Oxford, UK
Trials 2011, 12(Suppl 1):A56

A57
Individual participant data meta-analyses compared with meta-analyses based on aggregate data
Catrin Tudur Smith, James Oyee, Maura Marcucci, Marooska Rovers, Alfonso Ioris, Richard Riley, Paula Williamson, Mike Clarke
1 North West Hub for Trials Methodology Research, University of Liverpool, UK; 2 Department of Medicine, University of Perugia, Perugia, Italy; 3 Departments of Epidemiology, Biostatistics & HTA and Operating Rooms, UMC St Radboud, The Netherlands; 4 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada; 5 Public Health, Epidemiology and Biostatistics, University of Birmingham, UK; 6 All-Ireland Hub for Trials Methodology Research, Queen’s University Belfast, UK
Trials 2011, 12(Suppl 1):A57

Background: Meta-analysis based on individual participant data (IPD) is widely accepted as the most reliable approach and has been described as the ‘gold standard’ for systematic reviews. An IPD approach often allows more powerful, consistent and thorough analyses but does require additional resources compared with meta-analyses based on aggregate data (AD). Several empirical comparisons of IPD with AD meta-analyses have been published, some of which show that IPD meta-analyses can differ in important ways from meta-analyses based on AD. For example, the importance of including as much follow-up as possible on all randomised participants and data from all relevant trials was shown in separate empirical comparisons [1-3] whilst Duchateau [4] found substantial differences between IPD and literature-based AD meta-analyses, mainly due to different approaches to analysis. An unpublished review [5] summarised results from across 25 studies, showing that for two thirds of the comparisons AD estimated effect sizes with less precision and tended to overestimate the IPD effect but differences were small in most comparisons.

Objectives: We have undertaken a Cochrane systematic review of empirical studies that compared IPD and AD meta-analysis to explore key reasons for the differences.

Methods: The Cochrane Methodology Register, CENTRAL, MEDLINE and EMBASE were searched using a predefined set of search terms. Studies that report an empirical comparison of IPD meta-analysis against AD meta-analysis of randomised trials were assessed for inclusion, by two reviewers independently. Data were extracted by two reviewers independently and stored in a central database.

Results: Over forty empirical studies have satisfied the inclusion criteria for this review. Estimates of effect size and precision obtained from IPD and AD
will be compared and differences will be discussed. Results will help inform the ongoing debate about whether, and when IPD may be most valuable.

References


A58
What are the characteristics of an application (or team) which are associated with success or failure of an application to the NIHR HTA programme for research funding?

Sheila Turner1, Peter Davidson1, Louise Dent1, Elke Streit1, Victoria Bowness1
1NIHR Evaluation Trials and Studies Coordinating Centre at the University of Southampton S016 7NS, UK, 2University of Southampton Clinical Trials Unit, Southampton, SO16 6YD, UK.
Trials 2011, 12(Suppl 1):A58

Background: Previous studies have shown that gender or seniority of applicants can be linked to success in obtaining research funding, but we found little evidence about other features of successful applications for trials funding. The aim of this study was to investigate the association between items or features in a funding application and success when considered by a research funding board. This knowledge will be useful to potential applicants to the HTA programme and to other programmes, and to funding boards of this and other research programmes. The work was not intended to set ideal standards for applications, rather to extend knowledge of current practice and to explore what may be learnt from this.

Methods: We undertook a retrospective cohort study using applications to the NIHR HTA programme. The sample comprised 296 outline proposals for primary research, submitted to the commissioning board of the HTA programme between January 2006 and December 2009. Proposals to the commissioned work stream were selected as these proposals addressed issues which the HTA programme already deemed to be important, hence the priority of the research question was not considered as one of the selection criteria for success or failure. The dependent variables were characteristics of the applications (or teams) selected a priori by a panel of experts drawn from the board and informed by existing literature. They included characteristics relating to the principal investigator, the team and the bid. The outcomes were success or failure at short-listing and in obtaining research funding.

Data were analysed using uni-variate and multivariate analysis and conditional logistic regression modelling.

Results: Results of the analysis indicated that the characteristics most strongly associated with success were: whether the proposal met the specifications of the commissioning brief; multi-disciplinarity of the team; experience of writing and critiquing research bids; and whether or not a pilot or feasibility study had been conducted. The gender of the chief investigator and proposed trial costs were not significantly associated with success or failure.

Conclusions: The association between meeting the commissioning brief and success is in line with expectations and is reassuring regarding the decision making processes of the board. It suggests to applicants that deviating from the brief is risky. The relative success of multi-disciplinary teams and those including a statistician, a health economist or another methodologist were also anticipated by the experts. This offers evidence to applicants of the preferred make-up of teams. Further research in other funding streams or organisations would be useful.

A59
Strategies for handling missing data in randomised trials

Iain R White
MRC Biostatistics Unit, Cambridge CB2 0SR, UK
E-mail: ian.white@mrc-bsu.cam.ac.uk
Trials 2011, 12(Suppl 1):A59

Missing outcome data in randomised trials are a major potential source of bias in trial results, and their correct handling can be a major source of difficulty for investigators [1]. Missing outcome data matter for three main reasons:

1. They lead to a loss of power. This cannot be reversed, and so all efforts should be made to maximise completeness of follow-up. 2. Any analysis of incomplete data makes untestable assumptions and is likely to be biased if those assumptions do not hold. Because the correct assumption is usually unknown, it is important to justify the assumption used from subject-matter knowledge, and to perform sensitivity analyses.
3. Some popular analysis methods give inefficient estimates and/or biased standard errors, even if their assumptions are correct.

This talk will outline some of the statistical methods available for handling missing data in randomised trials, and what their underlying assumptions are. Some methods are simple to implement but harder to support: for example, “Last Observation Carried Forward” is based on an assumption that is rarely justified, while analysis of complete cases can be statistically inefficient. More complex methods such as mixed models and multiple imputation are usually based around a missing at random assumption, which is popular because it is perceived to become more plausible as models become more complex.

The intention-to-treat principle is sometimes seen as requiring missing values to be imputed. By focussing instead on assumptions, I will show that it is sufficient for all observed data to be included in the main analysis, but that sensitivity analyses must take account of all randomised individuals. This approach will be summarised in an intention-to-treat analysis strategy [2].

I will also briefly describe how to perform a sensitivity analysis and how to handle missing baseline variables.

References


A60
Using linear increment models for the imputation of missing composite outcomes in randomized trials

Aidan G O Keeffe1, Daniel M Farewell2, Brian DM Tom1, Vernon T Farewell3
1MRC Biostatistics Unit, IHP Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK, 2School of Medicine, Cardiff University, Neuadd Meirionnydd, Heath Park, Cardiff, CF14 4XS, UK.
Trials 2011, 12(Suppl 1):A60

In randomized trials, it is typical that a number of outcome variables are collected at each follow-up time. Sometimes, a composite outcome may be of scientific interest. A composite outcome is composed as a function of several patient-specific outcomes and could measure, for example, improvement or deterioration in the condition of a patient. Multiple imputation is one method for handling patient drop-out in randomized trials, and usually involves a maximum likelihood-based model fitted to complete cases, which is then used to draw imputations for the missing data. In principle, imputation could be used to impute composite outcomes. However, composite outcomes may be complicated combinations of many outcome variables and, as a result, it can be difficult to impute composite outcomes, since they may not possess a distribution amenable to statistical modelling.

Using trial data on early rheumatoid arthritis patients, we examine the use of linear increment models, introduced by Diggle, Farewell and Henderson [1] as a basis for the multiple imputation of missing outcome information.\n
\[ A60 \]
These imputations are used to create random draws of the composite outcome ACR20, a binary indicator of disease improvement defined by the American College of Rheumatology. We compare the multiple imputation of ACR20 using linear increments methodology to that using more established maximum likelihood methods. We observe some evidence to suggest that the use of linear increment models may result in a more accurate imputation of missing ACR20 values. The methodology presented has broad applicability in randomized trials.

Reference

A61
CONSORT: missing missing data guidelines, the effects on HTA monograph reporting
Yvonne Sylvestre
Affiliation tbc
Trials 2011, 12(Suppl 1):A61

Objectives: The reporting of randomised controlled trials (RCTs) has improved since the publication of the first CONSORT statement. However, word limitations on journal articles still preclude full description of methodology. The extended format of HTA monographs enable trialists to fully articulate methods. We assess the impact of the lack of reference to missing data in CONSORT on the reporting of RCTs in HTA monographs and compare with randomisation, blinding, and allocation concealment which are included.

Methods: The methodology for the comparison for the randomisation, blinding and allocation concealment followed the structure used by Hopewell et al. when assessing the impact of the updates of CONSORT. Detail of these on methodological CONSORT items were extracted for each monograph identified using a 1996 and 2001 CONSORT item check list for the periods 1997-2001, 2002-2005 and 2006-2010 respectively. Screening and data extraction was carried out independently by two authors determining whether each item was reported. Any discrepancies were referred to the other authors and resolved through discussion and consensus. We extended this methodology for whether missing data occurred and was described, and whether imputation and a sensitivity analysis were used. Additionally the methods of treatment of missing data were recorded.

Results: We identified 517 monographs which yielded 93 parallel group, crossover, cluster, factorial and cost-effectiveness RCTs for assessment. Across the three periods randomisation reporting increased from 70% to 92%, allocation concealment from 70% to 90% while blinding varied between 70% and 83%. Missing data issues were rarely discussed in a missing data section. Some details could be identified by careful reading of methods section, however the details of imputation methods used were only rose from 40% to 71%. The details of a sensitivity analysis are only reported in 30% of studies.

Conclusions: Most topics examined were more frequently reported in recent years, however missing values information is still not present in one quarter of RCT reports, and is more poorly reported than items specifically listed in CONSORT. We often experienced difficulty in establishing which methodology was used and whether the statistical and health economics analyses used the same methods. We recommend that missing data be included within the CONSORT checklist to improve the reporting of an important element of trial analysis.

OUTCOMES

A63
Patient reported outcomes: past, present, and future
Jane A Scott
Global Strategic Marketing & Market Access, Janssen Global Services, High Wycombe, Buckinghamshire, HP12 4DP, UK
Trials 2011, 12(Suppl 1):A63

For decades, clinical trials have relied on information from patients about their symptoms, functioning, and treatment experiences to evaluate treatment safety and efficacy. Today, information reported directly by patients (patient reported outcomes or “PROs”) using various technologies serves as the primary source of data about symptoms, functioning, health events, and the impact diseases and treatments have on the lives of patients and their families. The methods used to record and process that information has evolved over time so that today regulators require that trials use rigorous methods for all subjective assessments employed to evaluate treatments. This presentation is a brief review of the past, current, and future direction of PRO research and its implication for clinical trial measurement from the perspective of a PRO researcher working in medical product development trials.

A64
The effect of diabetes complications on health-related quality of life: estimating the bias due to unobserved heterogeneity using the UKPDS
Maria L Alva1*, Alastair Gray2, Borislava Mihaylova1, Philip Clarke2
1Department of Public Health, University of Oxford, Oxford, UK; 2Sydney School of Public Heath, University of Sydney, Australia
E-mail: maria.alvachiola@dph.ox.ac.uk
Trials 2011, 12(Suppl 1):A64
The impact of six diabetes related complications: myocardial infarction, ischemic heart disease, stroke, heart failure and amputation on quality of life is studied based on EQ-5D data on 3,380 patients collected in the UKPDS Post Trial Monitoring Study over the period 1997-2007. Analysts tend to resort to cross-section data to estimate the determinants of quality of life, as these are easier and less expensive to collect. The use of cross-section data may come at a cost in terms of biased estimators due to unobserved heterogeneity that can be addressed by the use of panel data.

The importance of non response (e.g. sample selection) is tested, and cross-sectional and longitudinal approaches to study effects on Qol, are compared. The variation in utility across patients in our data is more than twice that observed within a patient over-time and, this heterogeneity is often correlated with the likelihood of having complications, making the results of pooled estimations biased. Cross-sectional analysis in our sample over-estimates the effects of complications on utility. We address this problem by using fixed effects (FE). The bulk of the variation in the unexplained utility appears to be largely determined by time-invariant differences across patients' characteristics: gender, race, socio economic status and habits. We therefore link panel data methods with cross sectional analysis in a two-stage analysis, using the fixed effects as the dependent variable in the second stage regression to examine the extent to which certain time-invariant patient characteristics and variables recorded only at baseline in the trial, are related to patients' utility score. The results highlight the importance of studying quality of life changes over time to distinguish between time invariant correlates of wellbeing and the effect of diabetes complications.

A65 Patient Reported Outcomes: misinference from ordinal scales?

Mike Horton1, Alan Tennant
Psychometric Laboratory for Health Sciences, Department of Rehabilitation Medicine, The University of Leeds, UK
Trials 2011, 12(Suppl 1):A65

Patient Reported Outcomes (PROs) are widely used in medical outcome studies, and usually take the form of administered or self-completed questionnaires. The data that these questionnaires produce is of the type known as ordinal scaling, where magnitudes of the attribute may be ascertained. At the same time, most outcome studies rely on the calculation of means, standard deviations, change scores, and concepts such as Minimally Important Difference (MID) or effect sizes. Yet, ordinal scales do not support the mathematical operations needed to calculate these type of statistic [3]. Indeed when several items are measured on an ordinal scale it is far from certain that the sum of scores has even ordinal properties [1]. Despite these constraints, these limitations are largely ignored, and thus statistics such as means and MID are widely reported for PROs. This runs the risk of drawing an incorrect inference from data based upon PROs [5].

This risk can be illustrated by considering the concepts of the ‘plateau’ and the calculation of the MID. Both are investigated by contrast of the ordinal raw score against the cardinal metric derived from fit of data to the Rasch measurement model [4]. It can be shown that as the raw score from a scale moves towards the margins, then a smaller and smaller raw score change is associated with a standard metric unit of change. Thus patients may seem to be ‘slowing down’ in their improvement, or even ‘plateauing’, yet they are still moving the same metric distance. Likewise, when considering a magnitude of improvement such as an MID, the raw score distance associated with the MID can be shown to vary across the scale, depending upon the starting point. Thus for one patient the same MID may involve a change in the metric distance four times greater than that of another patient.

PROs provide ordinal estimates of the magnitude of a patient on the trait being measured. Appropriate non-parametric statistics should be used. Else, where possible, the data should be converted to the cardinal metric through use of the Rasch model, which is consistent with the requirements of the theory of Additive Conjoint Measurement [2][6].

References

A66 Making continuous outcomes meaningful to clinicians

Janet L Peacock1, Odile Sauzet2, Sally M Kerry3
1Division of Health and Social Care Research, King’s College London, SE1 3QD, UK; 2Public Health Sciences and Medical Statistics, University of Southampton, Southampton, SO16 6YD, UK; 3Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, E1 2AB, UK
Trials 2011, 12(Suppl 1):A66

Objectives: Dichotomizing continuous data prior to analysis is rarely a good idea as power is lost. However expressing results as a difference of means is not always meaningful for clinicians; small differences in means may seem clinically unimportant while the same study results expressed as a difference/ratio of proportions or number needed to treat might reveal an effect that is considered worthwhile. We therefore sought an estimate based on comparing proportions that gives the same p value as the corresponding comparison of means.

Method: We considered the proportion above a chosen cut-point of a Normal distribution as a function of the mean and standard deviation. This was used to derive the standard error and confidence interval (CI) for a difference in proportions that maintains the power of the corresponding comparison of means. This therefore provides the difference in proportions with a 95% CI that corresponds to a shift of 0.5 standard deviations.

To illustrate this method we use published data from a study of teenage motherhood and birth outcomes where 9.7% of 164 young adolescents had a low birthweight LBW, <2500g) baby compared with 3.5% of 423 adults; difference 6.2% 95% CI 1.3 to 11.1%. Using these data we estimated the mean birthweights that would produce these proportions assuming a Normal distribution with standard deviation 500g. Further, we simulated the effects of the method assuming the birthweights were Normally distributed and computed two different 95% CI : i) for the difference in proportion derived from a difference in means as described above, and ii) directly from the difference in proportions. The widths of the two CIs were compared, and the percentage of derived CIs containing the “true” difference of 0.062 (6.2%) was calculated.

Results: Mean birthweight is estimated as 257g higher in adult mothers. The derived 95% CI for the difference in percentage LBW is 3.5% to 8.9%. Simulation of 10,000 pairs of samples gave mean (SD) ratio of the width of the CI for the difference in proportions, direct/derived, of 2.09 (0.12). In approximately 6% of cases, the 95% CIs for proportions derived from the difference in means did not contain the expected value 0.062 (6.2%).

Conclusions: A difference in proportions with 95% CIs can be calculated for continuous outcomes without the loss of power usually associated with dichotomising given certain reasonable assumptions. This methodology provides estimates that are meaningful to clinicians, but retains statistical power.

A67 Efficiency gains resulting from the ordinal analysis of a functional outcome scale: a case study of a major phase III stroke trial

Gordon D Murray1, Elie Charlotte Sanders*, Philip MW Bath, Eivind Berge2
1Edinburgh MRC Hub for Trials Methodology Research, University of Edinburgh, Edinburgh, EH8 9AG, UK; 2Department of Internal Medicine, Oslo University Hospital Ullevål, Oslo, NO-0407, Norway; 3Stroke Trials Unit, University of Nottingham, Nottingham, NGS 1PB, UK
Trials 2011, 12(Suppl 1):A67

Background: Phase III clinical trials in areas including acute stroke and traumatic brain injury commonly use ordinal functional outcome scales as
their primary outcome measure. Conventionally these scales are analysed by dichotomising the ordinal scale into a binary scale: ‘dead or dependent’ versus ‘independent’. This can potentially discard much relevant information, reducing both the clinical relevance of the results and the statistical efficiency of the analysis.

**Methods:** Methodological work in stroke (by the OAST Group) and in traumatic brain injury (by the IMPACT Investigators) has demonstrated that using more appropriate approaches to the analysis of ordinal outcome scales, such as proportional odds regression or the ‘sliding dichotomy’, can potentially lead to substantial efficiency gains relative to the conventional dichotomous analysis. However, to date relatively few trials have prospectively adopted ordinal techniques for their primary analysis. We report here how in SCAST [1], a major Phase III trial of blood pressure reduction in acute stroke, ordinal methods were adopted for the primary analysis of the modified Rankin Scale (mRS), an ordinal functional outcome scale.

**Results:** Since ordinal methodology was evolving in parallel with the conduct of the trial, the Statistical Analysis Plan was not finalised until close to database lock. It was decided to use proportional odds regression for the primary analysis of the mRS with the sliding dichotomy as a sensitivity analysis. Relative to a conventional dichotomous analysis both of these approaches did indeed lead to substantial efficiency gains, equivalent to more than doubling the sample size.

**Conclusions:** SCAST shows that the potential efficiency gains demonstrated in basic methodological research can be realised in practice. This has major implications for the design and analysis of future trials based on ordinal outcome scales.

**Acknowledgements:** This work is presented on behalf of the SCAST Study Group.

---

**A69**

**The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities**

Duoiao Wang1, Con Arni, Tim Collier, Stuart Pocock

Medical Statistics Department, London School of Hygiene and Tropical Medicine, London, UK

Trials 2011, 12(Suppl 1):A69

**Background:** The conventional reporting of composite endpoints in clinical trials is unsatisfactory in that it emphasises each patient’s first event which is often of lesser clinical importance. The objectives of this talk are to introduce the concept of the win ratio for reporting composite endpoints to address this problem and to assess the new method by analysing real trials.

**Methods and results:** Patients in the new and control treatment are randomised into matched pairs based on their risk profiles. Consider a primary composite endpoint, e.g. cardiovascular (CV) death and heart failure hospitalisation (HF hosp) in heart failure trials. For each matched pair the new treatment patient is labelled a “winner” or a “loser” if it is known who had a CV death first. If that is not known, they are labelled a “winner” or “loser” if it is known who had a HF hosp first. Otherwise they are considered tied. The win ratio is the total number of winners divided by the total numbers of losers. If formation of matched pairs is impracticable then an alternative win ratio can be obtained by comparing all possible unmatched pairs. This method is insightfully illustrated by re-analyses of the EMPHASIS-HF, PARTNER B and CHARM trials.

**Conclusions:** The win ratio is a powerful and easy to use method for reporting composite endpoints, and gives appropriate priority to the more clinically important event, e.g. mortality. We encourage its use in future trial design and analysis.

---

**A70**

**The COMET (Core Outcome Measures in Effectiveness Trials) Initiative**

Paula R Williamson1, Doug G Altman2, Jane M Blazeby3, Mike Clarke4, Elizabeth Gargan1

1Department of Biostatistics, University of Liverpool, Liverpool, UK; 2Centre for Statistics in Medicine, University of Oxford, Oxford, UK; 3School of Community and Social Medicine, University of Bristol, Bristol, UK; 4School of Medicine, Dentistry and Biomedical Sciences, Queen’s University Belfast, Belfast, UK

E-mail: pnv@liv.ac.uk

Trials 2011, 12(Suppl 1):A70

**Background:** Systematic reviews are hampered by inconsistencies in outcomes assessed and reported in otherwise eligible studies. Many meta-analyses have to exclude key studies because relevant outcomes were not reported. Much could be gained if an agreed minimum set of appropriate and important outcomes was measured and reported in all clinical trials in a particular area.

**Why standardise outcomes?:** The design of new trials would be simplified, the risk of measuring inappropriate outcomes would be reduced, and selective reporting of outcomes less likely. It would be easier to compare, contrast and combine studies in systematic reviews. Core outcome sets would help review authors to present their findings clearly and succinctly, for example within Summary of Findings tables.

**Aims of the COMET Initiative:** The COMET Initiative brings together researchers interested in core outcome sets, with well attended international meetings held in 2010 and 2011. COMET aims to foster and facilitate research by providing guidance on developing a core outcome set, methods to include user involvement in this process, and preparing reporting standards for such projects. Further information about COMET can be found at [http://www.comet-initiative.org]. Work is ongoing to identify, collate and maintain relevant resources in an on-line searchable database [http://www.comet-initiative.org/studies/search]. More than 50 completed projects in
various areas of health/health care have been identified. Several examples of planned and ongoing work have also been recorded. The COMET database will be demonstrated and progress to date presented.

**Implications:** If successful, COMET will help trialists to choose outcomes, and will therefore increase the likelihood that these outcomes will be measured, thereby decreasing the likelihood of important studies being excluded from systematic reviews. By improving the evidence base, COMET will make it easier for people to make well-informed decisions about health care.

**Acknowledgments:** The COMET Initiative has been funded by the MRC North West Hub for Trials Methodology Research, the MRC ConDuCT Hub and the MRC Hub for Trials Methodology Research Network.

---

**A71**

A mixed methods approach to assess the reliability of a test for the success of blinding in a surgical randomised controlled trial (RCT)

Caroline E Boulind, Kerry Avery, Chris Metcalfe, Nader K Francis, Jane M Blazey*

University of Bristol, School of Social & Community Medicine, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK

Trials 2011, 12(Suppl 1):A71

**Background:** Blinding participants in randomised controlled trials (RCTs) with patient reported outcomes is important to reduce bias and widely used. The success of blinding is rarely tested, but may be assessed with tools such as the Bang Blinding Index (BBI). There is controversy, however, about the value and interpretation of test results.

**Objectives:** To investigate the reliability of the BBI in a feasibility RCT comparing two types of pain relief following colorectal surgery.

**Methods:** Participants were asked to guess which trial arm they were in within a day of surgery and at discharge. Responses could include arm A, arm B or ‘don’t know’ (DK). Participants responding DK were also asked to give a ‘forced guess’ and reasons for responses were documented. Data were analysed using the Bang Blinding Index (BBI). Proporions of correct guesses for the two arms of the trial were compared using Fisher’s exact test. Two analyses were performed, first using the original responses (including DK), then replacing DK with the forced guesses. Audio-recorded semi-structured interviews with participants were undertaken after discharge, exploring the reasons behind participants’ guesses and beliefs about treatment allocation. The interviews were analysed using a constant comparison method. Results from the qualitative analysis were triangulated with the BBI data.

**Results:** Twenty six participants were included. In arm A (n = 13), 62% correctly guessed treatment allocation, compared to 70% in arm B (n = 13, p = 0.41). The BBI result for people in arm A was 0.15: 15% more correct guesses than expected by chance. In arm B the result was 0.31, suggesting that 31% more than expected by chance correctly guessed treatment allocation. When DK responses were replaced with the forced guesses, the proportions of correct guesses changed to 69% for arm A and 63% for arm B (p = 0.043). The corresponding BBI results were 0.38 and 0.08 respectively.

Qualitative interviews suggested that there is variable understanding of the reason for blinding among trial participants, but that they were accepting and appreciative of need for blinding and that their guesses reflected true beliefs about their treatment allocation.

**Conclusions:** The BBI can be used to reliably estimate the rate of unblinding among trial participants and is likely to accurately reflect participant beliefs about treatment allocation; however DK responses may not represent successful blinding, and the use of information from forced guesses is important for the correct interpretation of the BBI.

---

**A72**

Patient reported clinical outcomes: the challenges and implications for randomised controlled trials

Suzanne Breeman, Alison McDonald, Gladys McPherson, Graeme MacLennan, Marion Campbell, Kath Starr, Seonaidh Cotton

Health Services Research Unit, University of Aberdeen, Aberdeen, AB25 2ZD, UK

Trials 2011, 12(Suppl 1):A72

**Objectives:** Clinical outcomes are an important component of randomised controlled trials (RCTs) and are often used to complement patient reported outcomes measures such as health-related quality of life. Although clinical outcomes were traditionally collected through clinical examination or laboratory results, routine data sources and self-reporting by patients are now being increasingly used. We describe patient reported clinical outcomes and the challenges associated with collection of data in this way.

**Methods:** Four RCTs that collected patient reported clinical outcomes through postal questionnaires were examined. In each RCT the patient reported clinical outcome was verified using either medical records, routine data sources or by contacting the patient’s general practitioner or consultant to ascertain the accuracy of reporting by the patient.

**Results:** The accuracy of patient reporting of clinical outcomes is dependent on a number of factors, including the nature and timing of the clinical outcome and the phrasing of the clinical questions. For example, it may be easier for a patient to report a knee-related hospital re-admissions than self-report a urinary tract infection. Nevertheless, approximately 15% of patient reported knee-related hospital re-admission (collected through annual postal questionnaires) could not be verified through routine data sources and/or medical records. Such inconsistencies were shown to be a combination of misunderstanding by the patient and inaccuracies of the routine data sets.

**Conclusions:** Obtaining clinical information from the perspective of the patient remains important, especially if the outcome of interest is subjective or symptomatic. However with the potential inaccuracies associated with patient reporting of clinical outcomes, it may be necessary to consider verifying such outcomes with medical professionals and/or routine data sources. Such a strategy has implications in terms of staff time and cost and therefore has to be considered during the design stage of the RCT. We will discuss some inconsistencies between self-reporting and medically confirmed clinical outcomes. We will also highlight the process involved in verifying patient reported clinical outcomes and how adopting such a verification strategy may impact on the overall trial results.

---

**A73**

Treatment effect bias in randomised controlled trials using surrogate outcomes: a preliminary cohort study analysis

Oriana Cianí*, Rod S Taylor

Peninsula Technology Assessment Group (PenTAG), Institute of Health Services Research, Peninsula College of Medicine & Dentistry, University of Exeter, Exeter, Devon, EX1 2EU, UK

E-mail: oriana.ciani@pg.soton.ac.uk

Trials 2011, 12(Suppl 1):A73

**Background:** Ideally, decisions on the value of health technologies should be based on evidence from well-conducted clinical trials that assess clinically important final patient-relevant outcomes, such as mortality or impaired quality of life. Pressure to reduce the delay in the availability of technologies to patients has led to an increased reliance on the use of surrogate outcomes [1]. A key tenant of surrogate outcomes is unbiased quantification of the predictive treatment effect. To estimate the final patient-relevant clinical outcomes, this study compares the treatment outcome of trials that use a surrogate versus a final patient-relevant primary outcome.

**Materials and methods:** We searched for all randomized controlled trials (RCTs) published in JAMA, NEJM, Lancet, BMJ, PLoS Medicine and Annals of Internal Medicine in 2005 and 2006 [2]. We distinguished between trials that used a surrogate or a patient-relevant primary outcome. An outcome was defined as a surrogate if it did not directly measure “how a patient feels, functions, or survives” [3] or was judged to be a substitute and predictive of a final outcome [1]. We excluded non-RCTs, composite (of both surrogate and final outcomes), economic evaluations and non-interventional technologies. Surrogate and final outcome trials were matched on patient population, intervention, journal and year of publication. In this preliminary analysis we compare the two groups of trials based on the statistical significance and direction of their outcome results.

**Results:** Of the 639 citations identified by our searches, we included 138 trials that used a primary surrogate outcome (‘surrogate trials’) and 132 trials that used a final patient-relevant outcomes (‘final trials’). Table 1 summarises the trial characteristics used for matching. Other trial characteristics also appeared to be well balanced except for the length of follow-up (i.e. more studies with follow up <30 days and >1 year for final trials).
Conclusions: This preliminary analysis supports our initial hypothesis that the use of surrogate outcomes is more likely to lead statistically significant treatment effects than patient-relevant primary outcomes. We are currently undertaking additional analysis using actual effect sizes in meta-analytic/meta-regression framework. These results have important implications for payers faced with making coverage decisions on the effectiveness and cost-effectiveness of new treatments based on surrogate rather than final clinical trials data.

Acknowledgements: We thank Toby Pavey for the assistance in data checking and professor Peter Gøtzsche and co-authors who kindly provided access to the data files of their study. Ciani Q. is currently in receipt of a Peninsula College of Medicine and Dentistry PhD studentship.

References

Table 1 (abstract A73) Characteristics of included surrogate and final trials

<table>
<thead>
<tr>
<th>Surrogate Trials (N=138) N(%)</th>
<th>Final Trials (N=132) N(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention clinical area</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>31(22)</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>11(8)</td>
</tr>
<tr>
<td>Gastrology/hepatology</td>
<td>11(8)</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>28(20)</td>
</tr>
<tr>
<td>Nephrology/urology</td>
<td>4(3)</td>
</tr>
<tr>
<td>Neurology</td>
<td>2(1)</td>
</tr>
<tr>
<td>Obstetrics</td>
<td>5(4)</td>
</tr>
<tr>
<td>Oncology</td>
<td>4(3)</td>
</tr>
<tr>
<td>Other</td>
<td>36(26)</td>
</tr>
<tr>
<td>Pulmonology</td>
<td>6(4)</td>
</tr>
<tr>
<td>Population clinical area</td>
<td></td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>34(25)</td>
</tr>
<tr>
<td>Endocrinology</td>
<td>14(10)</td>
</tr>
<tr>
<td>Gastrology/hepatology</td>
<td>9(7)</td>
</tr>
<tr>
<td>Infectious disease</td>
<td>21(15)</td>
</tr>
<tr>
<td>Nephrology/urology</td>
<td>2(1)</td>
</tr>
<tr>
<td>Neurology</td>
<td>2(1)</td>
</tr>
<tr>
<td>Obstetrics</td>
<td>7(5)</td>
</tr>
<tr>
<td>Oncology</td>
<td>4(3)</td>
</tr>
<tr>
<td>Other</td>
<td>36(26)</td>
</tr>
<tr>
<td>Pulmonology</td>
<td>9(7)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Journal</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annals</td>
<td>14(10)</td>
</tr>
<tr>
<td>BMJ</td>
<td>11(8)</td>
</tr>
<tr>
<td>JAMA</td>
<td>31(22)</td>
</tr>
<tr>
<td>Lancet</td>
<td>28(20)</td>
</tr>
<tr>
<td>NEJM</td>
<td>51(37)</td>
</tr>
<tr>
<td>PLoS</td>
<td>3(2)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Year</th>
<th>Surrogate Trials (N=138) N (%)</th>
<th>Final Trials (N=132) N (%)</th>
<th>P-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>2005</td>
<td>64(46)</td>
<td>63(48)</td>
<td>0.006c</td>
</tr>
<tr>
<td>2006</td>
<td>74(54)</td>
<td>69(52)</td>
<td></td>
</tr>
</tbody>
</table>

Surrogate trials were less likely to have adequate evidence of randomisation sequence generation and adopt the ITT principle. We also found clear evidence that final trials were more likely to observe a non-statistically significant (neutral) treatment effect than surrogate trials (49% vs 23%) (Table 2).

Table 2 (abstract A73) Comparison of outcome results

<table>
<thead>
<tr>
<th>Study outcomeab</th>
<th>Surrogate Trials (N=138) N (%)</th>
<th>Final Trials (N=132) N (%)</th>
<th>P-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>49 (36)</td>
<td>41 (31)</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>5 (4)</td>
<td>4 (3)</td>
<td></td>
</tr>
<tr>
<td>Neutral</td>
<td>32 (23)</td>
<td>65 (49)</td>
<td></td>
</tr>
</tbody>
</table>

| Risk of bias   |                               |                           |          |
| Statement of ITT| 95 (69)                       | 106 (80)                  | 0.031    |
| Automated sequence generation | 82 (59)                   | 93 (70)                   | 0.058    |
| Allocation concealment | 96 (70)                     | 97 (63)                   | 0.476    |
| Blinding/Placebo | 72 (52)                       | 60 (45)                   | 0.27     |

aChi-square test, unless otherwise specified.
bpositive: treatment group superior to control (P<0.05); ‘negative’, control group superior to treatment (P<0.05); ‘neutral’, treatment and control indifferent, P>0.05.
cFisher’s exact test. Multiple-interventions trials are excluded from this comparison.
To optimise reporting, the PERSEPHONE team have developed Patient Diary Sheets which list the common expected toxicities. For each trastuzumab cycle, patients are requested to record how much they are troubled by each of the possible side-effects using patient-friendly severity ratings: 0=Not at all, 1=A little, 2=Moderate, 3=Quite a bit, 4=Very much).

**Results:** To date 1260 patients have been recruited into the PERSEPHONE trial with adverse event information summarised over 10,000 trastuzumab doses. At the 3-monthly clinic visits, the site research teams discuss the detailed patient diary information with the patient and accurately interpret the information into incidences of toxicities and relevant Common Terminology Criteria for Adverse Events (CTCAE) grades for entry onto the trial CRFs. The patient diaries are not entered onto the database for analysis as they are used solely as memory aids.

**Conclusions:** Feedback from the clinical research teams at sites indicates that the method adopted in the PERSEPHONE trial appears to work well. The Patient Diary Sheet is indeed well received and utilised by the patients and reported as a very useful memory aid at clinic visits. Whilst it may seem counterintuitive to ask for additional data which will never be scrutinised by the PERSEPHONE trial team, the strength of the final data collected has been optimised by the use of these data triggers.

### A75 Applying Rasch analysis to the SF-36 physical function scale: effect of dependent items

**Mike Horton, Alan Tennant**

Psychometric Laboratory for Health Sciences, Department of Rehabilitation Medicine, The University of Leeds, UK

Trials 2011, 12(Suppl 1):A75

Medical outcome studies often use Patient Reported Outcomes (PROs), and these often take the form of administered or self-completed questionnaires. Often, the responses from these scales are simply added up as a total, and this total score is utilised for outcome purposes. However, for this approach to be valid, there are a number of underlying assumptions that are being made. One of these assumptions is that of Local Independence, which comprises two aspects; the items making up the scale should all be unidimensional (trait dependency) and, the response to an item should not directly influence the response to another item within the set (response dependency) [1].

These assumptions apply to the Rasch model (and other IRT models). Thus the process of Rasch analysis provides a means to test these assumptions, along with other key properties such as invariance. Where data satisfy these assumptions, and fit to the model, an interval scale transformation becomes available.

A secondary analysis was performed on data from the Physical Function scale (PF-10) of the SF-36 [2] to demonstrate the application of Rasch Analysis and to investigate the influence of response dependency within the dataset.

Initial results showed evidence of a lack of unidimensionality (t-tests = 8.39%; lower bound CI = 5.9%), along with apparent response dependencies between the three walking items of the PF-10 (walking 100 yards, walking half a mile & walking more than a mile), the two climbing stairs items of the PF-10 (climbing one flight of stairs & climbing several flights of stairs) and a dependency between the ‘moderate activities’ item and the ‘lifting or carrying groceries’ item. When adjustments had been made for the response dependency within the item set, it then appeared to be unidimensional (t-tests = 2.1%).

Also to be noted is the artificial inflation of the reliability index due to the dependency within the item set. After dependency adjustments had been made, the reliability index (PSI) dropped from 0.851 (initially) to 0.789 (Cronbach’s alpha dropped from 0.91 to 0.80).

In conclusion, response dependency can inflate reliability, and lead to spurious (dependency) factors. It has also been shown to influence both item and person parameters in IRT analysis. Rasch analysis provides a framework to assess, as well as adjust for, response dependency within an item set.

**References**


### A76 Incorporating patient reported outcomes (PROs) in gastro-intestinal (GI) cancer randomised controlled trials (RCTs): the need for adequate rationale and integrated reporting

Rhiannon C Macerfield1, Angus GK McNair, Natalie S Blencowe, Sara T Brookes, Jane M Blazeby

School of Social and Community Medicine, University of Bristol, Bristol, BS8 2PS, UK

E-mail: r.macerfield@bristol.ac.uk

Trials 2011, 12(Suppl 1):A76

**Objectives:** Despite the widespread use of PROs in RCTs evidence suggests that results may not influence practice. Absence of an a priori rationale for measuring PROs and poor integration with trial clinical outcomes may contribute to this problem. This hypothesis was addressed by examining current reporting of PROs in GI cancer RCTs.

**Methods:** A systematic review in MEDLINE, EMBASE and Cochrane databases searched for RCTs of radical treatments for GI cancer using validated PROMs, published between 2000 and 2009. Trials with a potential high risk of bias (Cochrane Collaboration tool) were excluded. Independent data extraction (3 reviewers) recorded the rationale for PRO measurement classifying this as 1) no rationale, 2) general rationale e.g. to examine QOL, 3) partial rationale e.g. hypothesis for a specific PRO domain or an expected direction of change or 4) complete rationale specifying a PRO domain and direction of change. Integrated reporting of clinical and PROs was investigated by examining whether trials reported PROs with or separately to clinical data, if PRO results were included in abstracts, and publication dates and journal impact factors where PRO and clinical results were published separately.

**Results:** 43 papers reporting PROs from 40 trials were included. Interventions were mostly chemotherapy (52.5%) and in colorectal cancer (77.5%). 16 (37.3%) papers did not report a PRO rationale, 14 (32.6%) gave general reasons e.g. to examine QOL, 3) partial rationale e.g. hypothesis for a specific PRO domain or an expected direction of change or 4) complete rationale specifying a PRO domain and direction of change. Clinical and PROs were reported together in 30 papers (70.0%), in which PROs were typically a secondary trial endpoint (27/30, 90.0%). Of these, 11 had significant PRO results, with 10 (90.1%) reporting this in the abstract. 13 papers (30.3%) were separate reports of PRO data, supplementary to clinical findings. Median time between clinical and PRO publications was 21 months (range 5-51). The median journal impact factor for clinical and PRO papers was 15.6 and 6.3 respectively (p=0.03). Eight (61.5%) of the corresponding clinical papers did not report any PRO data and four (30.7%) made no indication that PROs had been measured despite the subsequent PRO publication.

**Conclusions:** Few GI cancer RCTs provided a detailed rationale for measuring PROs and integration of clinical and PRO results was often poor. Standards for reporting PROs alongside clinical outcomes are required to improve clinical understanding and facilitate use of all trial data in decision-making.

### A77 The case for a HRQL core outcome set: outcome reporting bias in oesophageal cancer studies

Rhiannon C Macerfield1, Marc Jacobs2, Natalie S Blencowe1, Ida J Korfage1, Joanna Nicklin1, Sara T Brookes1, Mirjam A Sprangers2, Jane M Blazeby1

1School of Social and Community Medicine, University of Bristol, Bristol, BS8 2PS, UK; 2Department of Medical Psychology, Academic Medical Center/University of Amsterdam, 1105 AZ Amsterdam, The Netherlands; 3Department of Public Health, Erasmus MC, 3000 CA Rotterdam, The Netherlands

E-mail: r.macerfield@bristol.ac.uk

Trials 2011, 12(Suppl 1):A77

**Objectives:** The large number of scales and items within generic and disease specific HRQL measures may lead to outcome reporting bias of HRQL data. We explored this hypothesis in publications of HRQL results from studies of curative treatment for oesophageal cancer.
Methods: Systematic literature searches (MEDLINE, Embase, PsychINFO and CINAHL) identified articles reporting HRQL data after curative treatments for oesophageal cancer from validated patient-completed questionnaires. The most frequently used instruments were examined (EORTC QLQ-C30, QLQ-ES52, SF-36 & FACT-E). Outcome reporting bias was explored by examining which of the instrument’s scales and single items were documented in the methods and the results of papers, with each article checked by two independent reviewers.

Results: Of 52 included papers, three specified to focus on a restricted set of the measured scales in the study methods, with one providing a rationale for doing so. 31 papers (60%) reported EORTC QLQ-C30 results, of which 20 (65%) reported all 15 of the instrument’s scales and single items. Global health status was consistently reported (30/31 papers, 97%), followed by physical function (29/31, 94%) and role function (28/31, 90%). Insomnia and constipation were least reported (22/31, 71%). 25 papers additionally included the QLQ-ES52 or QLQ-ES18 oesophageal-specific module (consisting of 11 and 10 scales and single items respectively). All of the modules, scales and items were reported in 14/25 papers (56%). Dysphagia, eating and reflux scales were most often reported (in 21, 23, and 22 papers respectively). 12 papers reported SF-36 results, of which 9 (75%) reported all 8 scales. SF-36 physical functioning, role-physical, role-emotional, social functioning, vitality and general health scales were reported most often (11/12 papers), followed by bodily pain and mental health (10/12 papers). Three studies used the FACT-E, and all reported the 5 subscales (physical well-being, functional well-being, social/family well-being, emotional well-being and the oesophageal cancer subscale).

Conclusions: Selective outcome reporting was evident in some studies of curative treatment for oesophageal cancer, increasingly when a higher number of scales and items were available. The development of a core outcome set of HRQL domains for RCTs of curative treatments in oesophageal cancer may reduce the risk of outcome reporting bias and ensure important domains are consistently reported.

A78
Optimising trial monitoring on the AZURE trial
Geraldine A Matthews1,2, Claire Davies1, Vicky Hiley1, Helen C Marshall3, Robert E Coleman4
1Clinical Trials Research Unit, University of Leeds, Leeds, UK; 2Cancer Research Centre, Academic Unit of Clinical Oncology, Weston Park Hospital, Sheffield, UK; 3EORTC Breast Cancer Group, Brussels, Belgium; 4University of Bristol, Bristol, BS10 5NB, UK
E-mail: g.a.matthews@leeds.ac.uk
Trials 2011, 12(Suppl 1):A78

Objectives: To assess telephone monitoring as a method of central monitoring for clinical trials.

Methods: Determining how academic clinical trials units can optimise data quality via central monitoring methods is an important factor in trial management. In the AZURE trial, it became apparent from site monitoring that the endpoint dates as defined in EORTC literature [1] were mis-interpreted at site. Although a training programme was initiated, some sites were seen to report date of confirmation rather than date of onset/suspicion as required. Therefore the focus of site monitoring was changed to review this critical endpoint data. Due to limited trial monitoring resources, a review of telephone monitoring was implemented.

A pilot phase was implemented to review the practicalities and possible results which could be obtained using telephone monitoring. During the pilot phase 17 trial endpoints were reviewed which had been verified by both site and telephone monitoring.

Results: Agreement between site and telephone monitoring was achieved on 12/17 events during the pilot phase. Telephone monitoring was implemented and a total of 105 events were reviewed, some cases with multiple queries. Agreement between telephone monitoring and the Case Report Forms was found on 81/105 events. Discrepancies between the Case Report Form and telephone monitoring lead to a median amendment of the endpoint date of 1.25 days (IQR 0.75 - 2.00 days). In total, 22/105 queries were also identified for telephone monitoring which required clarification of diagnoses and addition of missing dates of suspicion with censoring of patients’ data if unresolved.

Conclusions: Our findings indicate that although site monitoring remained the gold standard method of source data verification, telephone monitoring is a useful method for validating endpoint data when site monitoring resources are not available. The success of telephone monitoring was strongly related to the experience of the person at site involved with the review. When talking to less experienced staff, it was difficult to remotely navigate through the notes and pick up earlier scans/reurrences. Also, if a participant had more than one recurrence it was found to be difficult to piece together all the available information over the phone.

Acknowledgments: Thank you to the AZURE patients, the AZURE investigators and research staff at recruiting centres, the AZURE trial teams at the 7 clinical trials units (CTRU, University of Leeds; ICR CTU, London; ISD CCITT, Edinburgh; CRU, University of Birmingham; VCG, Ireland; ICORG, Ireland; SOLT, Spain), the independent TSC members, the National Cancer Research Network and Novartis Pharmaceuticals for their academic grant support.


A79 Pain assessment: the relationship between pain thresholds and pain severity in osteoarthritis
Vikki Wyld1,2, Shea Palmer1, Ian D Learmonth3, Paul Dieppe3
1University of Bristol, Bristol, BS10 5NB, UK; 2University of the West of England, Bristol, BS16 1QY, UK; 3Peninsula Medical School, Universities of Exeter and Plymouth, PL6 8BU, UK
Trials 2011, 12(Suppl 1):A79

Objectives: When pain is assessed in research, it is most commonly assessed using patient-reported outcome measures (PROMs). These measures can provide information about pain severity, distress and quality, but they are unable to provide information about the underlying mechanisms contributing to pain perception. Quantitative Sensory Testing (QST) comprises non-invasive tests which can identify abnormalities in pain processing on a localised and widespread level. The aim of this study was to explore the relationship between the measurement of pain thresholds by two different QST methods and the measurement of pain severity by a validated PROM.

Methods: Patients on the waiting list for a total knee replacement because of osteoarthritis were approached about this study, and 107 patients attended a QST session. A digital algometer (Somedic) with a 1cm probe was used to assess pressure pain thresholds (in kPa) three times at each site. A MSA Thermotest (Somedic) was used to measure hot pain thresholds (in Celsius) four times at each site. Pain thresholds were tested at the painful knee and the pain-free forearm. Self-reported pain severity was assessed using the WOMAC questionnaire. Spearman-Rank Correlation Coefficients (CC) were calculated to determine the strength of correlation between pain thresholds and pain severity.

Results: Pain severity in the knee significantly correlated with pressure pain thresholds at the knee (CC=0.275, p=0.005) and the forearm (CC=0.208, p=0.03). No significant correlation was found between pain severity in the knee and heat pain thresholds at the knee (CC=0.118, p=0.236) or the forearm (CC=0.134, p=0.173).

Conclusions: This study found a small but significant positive correlation between self-reported knee pain severity and pressure pain thresholds, but not heat pain thresholds. This suggests that as pain worsens, patients become more sensitive to pressure pain at the osteoarthritic knee and the pain-free forearm. Increased pain sensitivity at a pain-free distant body site, such as the forearm, suggests involvement of the central nervous system and widespread pain sensitisation. The addition of Algometry as an outcome measure in clinical trials and pain studies could improve understanding of the mechanisms contributing to self-reported pain severity, by identifying the presence and extent of pain sensitisation. We are currently using QST in a large randomised controlled trial of pain control in arthroplasty (the APEX study), to explore the relationship between pain sensitivity and pain severity in osteoarthritis patients, and the influence of pain sensitisation on outcomes after joint replacement.

Acknowledgements: The authors would like to thank the Bristol Orthopaedic Trust for funding the purchase of the QST equipment, Professor Sarah Hewlett for input into study design, and Dr Rachael Gooberman-Hill for her helpful comments on the abstract.
PATIENT INVOLVEMENT

A80
What have we learned from a decade of patient involvement in OMERACT and its effect on trial outcome assessments?
John R Kirwan1*, Maarten de Wit2
1School of Clinical Sciences, University of Bristol, Bristol UK and Academic Rheumatology Unit, Bristol Royal Infirmary, Bristol, UK; 2Department of Medical Humanities, VU Medical Centre, Amsterdam, The Netherlands
E-mail: John.Kirwan@bristol.ac.uk
Trials 2011, 12(Suppl 1):A80

Background: Since 2002 people with rheumatic diseases have participated in the bi-annual 5-day international conference and workshop on Outcome Measures in Rheumatology (OMERACT), which uses evidence-based consensus to determine core outcome measures for clinical trials of rheumatological conditions. Patients contribute their perspective and collaborate as equal partners in all conference sessions. In all, 46 patients with 7 different conditions from 12 countries have participated in OMERACT. With the help of professionals they have organized themselves and developed various support mechanisms.

Methods: The authors have called on their personal experiences of OMERACT, a review of OMERACT publications and discussions with long-term OMERACT participants in setting out results and drawing conclusions.

Results: The effect of patient involvement has been threefold:
First, the OMERACT research agenda has been enriched by the identification of new domains that are relevant from a patient perspective. The most important example has been the increased efforts to gain more knowledge about the nature and impact of fatigue in rheumatoid arthritis. Fatigue has been added to the rheumatoid arthritis core-set and is now widely measured as an important outcome in clinical trials.
Second, involvement of patients in the structure of the meetings has given insights into barriers to patient participation and to ways of facilitating the incorporation of the patient perspective in outcome research. Before the start of the conference patient participants need appropriate information about the objective and program of the conference and the expected contributions. They need training, support and encouragement to be able to fully participate in all parts of the program. And challenges like the patient-doctor relationship, ethics, confidentiality and communication need to be addressed.
Third, OMERACT has set a standard for patient involvement in healthoutcome research that has led to new local, national and international initiatives.

Conclusions: Building the involvement of patients directly into the OMERACT meeting program has significantly contributed to the success of this biannual conference. It has broadened the research agenda of OMERACT, it has had a major spin off on patient involvement in health outcome research. However, there remains the challenge of fully developing patient participation in the research working groups that meet between the conferences.

A81
Consumer involvement at the MRC Clinical Trials Unit: results of a survey
Claire L Vale1, Lindsay C Thompson, Silvia Forcat, Claire Murphy, Bec Hanley
MRC Clinical Trials Unit, London, WC2B 6NH, UK
Trials 2011, 12(Suppl 1):A81

Background: We aimed to establish levels of consumer involvement in randomised controlled trials (RCTs), meta-analyses and other studies carried out by the MRC Clinical Trials Unit across the range of research programs, predominantly in cancer and HIV.

Methods: Staff responsible for studies included in a Unit Progress Report (MRC CTU, April 2009) were asked to complete a questionnaire survey regarding consumer involvement. This was defined as active involvement of consumers as partners in the research process and not as subjects of that research. The electronic questionnaires combined open and closed questions, intended to capture quantitative and qualitative information on whether studies had involved consumers; types of activities undertaken; recruitment and support; advantages and disadvantages of involvement and its perceived impact on aspects of the research.

Results: Between October 2009 and April 2010, 138 completed questionnaires (86%) were returned. Studies had been conducted over a 20 year period from 1998, and around half were in cancer; 30% in HIV and 20% were in other disease areas including arthritis, tuberculosis and blood transfusion medicine. Forty-three studies (31%) had some consumer involvement, most commonly as members of trial management groups (TMG) (88%). A number of positive impacts on both the research and the researcher were identified. Researchers generally felt involvement was worthwhile and some felt that consumer involvement had improved the credibility of the research. Benefits in design and quality, trial recruitment, dissemination and decision making were also perceived. Researchers felt they learned from consumer involvement, albeit that there were some barriers. Additional results will be presented.

Conclusions: Whilst most researchers identified benefits of involving consumers, most of studies included in the survey had no involvement. Information from this survey will inform the development of a unit policy on consumer involvement, to guide future research conducted within the MRC Clinical Trials Unit.

A82
Public involvement in the design and conduct of clinical trials: a narrative review of case examples
Jonathan Boote1*, Wendy Baird2, Anthea Sutton2
1NIHR Research Design Service for Yorkshire and the Humber, South Yorkshire, S1 4DA, UK; 2School of Health and Related Research, University of Sheffield, South Yorkshire, S1 4DA, UK
E-mail: jboote@sheffield.ac.uk
Trials 2011, 12(Suppl 1):A82

Background: Public involvement is health research policy in the UK and internationally. There is a need to establish a robust evidence base on the impact of public involvement on research processes and outcomes.

Aim: To review case examples of public involvement in the design and conduct of clinical trials, to synthesise the contributions of the public, as well as the identified tensions and facilitating strategies.

Method: Systematic literature search and narrative review.

Findings: Nine papers were identified, covering the following topics: breast-feeding, antiretroviral and nutrition interventions [1]; paediatric resuscitation [2]; exercise and cognitive behavioural therapy [3]; hormone replacement therapy and breast cancer [4]; stroke [5,6]; chronic suppurative otitis media [7]; Paget’s disease [8] and shared decision-making in patient consultations [9]. Six papers reported on public involvement at the trial design stage, while three reported on public involvement at the design and conduct stages of clinical trials. It was found that the public contributed at the consultation, collaboration and publicly-led levels of involvement. Four main public contributions to trial design were identified: review of consent procedures and patient information sheets; suggestion of additional trial outcomes; review of trial data collection procedures; and recommendations on the timing and location of trial follow-up data collection. Two main contributions that the public made to the conduct of trials were identified: scrutiny of the conduct of the trial through membership of the Trial Steering Committee; and delivering the trial protocol after completing relevant training. Four main tensions were identified with regard to involving the public in trial design and conduct: tensions between stakeholder groups when designing trials; public understanding of trial methodology; the added time, complexity and cost of public involvement; and the representativeness of the public involved. Four main facilitating strategies were identified with regard to involving the public in trial design and conduct: cultural sensitivity; clear explanation of trial methodology; independent facilitation of trial design planning meetings; and adequate funding for public involvement.

Conclusions: Papers on public involvement in the design and conduct of clinical trials may have been overlooked due to the difficulty of searching for, and identifying, papers in this area. Only publications published in English were searched for and the review focused on evidence contained in peer-reviewed journal articles only.

Conclusions: The issues raised in this review should assist researchers in developing and conducting clinical trials with the involvement of the public.
References
6. Courzos S, Lea T, Murray K, Colbung M. We are not just participants—we are in charge: the NACCHO ear trial and the process for Aboriginal community-controlled health research. Ethnicity & Health 2005, 10:91-111.

A83
Including service users in trials and rigorous studies in health and social care: developing a standard operating procedure for researchers
Bridie Angela Evans1, Emma Bedell2, Philip Bell3, Hayley Hutchings1, Lesley Lowes1, David Rael1, Anne Seagrove1, Stefan Siebert1, Graham Smith1, Helen Snook1, Marie Thomas5, Kym Thorne1, Ian Russell1
1College of Medicine, Swansea University, Swansea, SA2 8PP, UK; 2North Wales Organisation for Randomised Trials in Health (NWORTH), Bangor University, Bangor, LL57 2DG, UK; 3Service User, Cardiff University, CF14 4XN, UK; 4School of Nursing and Midwifery Studies, Cardiff University, CF14 4XN, UK; 5College of Human and Health Sciences, Swansea University, Swansea, SA2 8PP, UK; 6School of Education, University of Wales Newport, Newport, NP18 3QT, UK
Trials 2011, 12(Suppl 1):A83

Background: Involving service users in research is encouraged as a way to improve research quality, relevance and accountability and is a prerequisite for many funding bodies. Existing guidance for researchers on how to do this mainly discusses general principles. Some researchers may question the value, feasibility and impact of including service users and limit the scope of involvement. We defined service users as patients, carers, people eligible for a service or anyone relevant to the trial inclusion criteria.

Objective: To develop a standard operating procedure (SOP) to give researchers detailed guidance on how to include service users in trials and other rigorous studies.

Methods: Researchers with experience of service user inclusion and service users currently involved in trials adopted by the West Wales Organisation for Rigorous Trials in Health (NWORTH), the Clinical Trials Unit based in Swansea University, collaborated to develop the SOP for service user inclusion.

Results: We articulated core principles of equality underpinning the SOP and guidance on how to achieve these. Processes to recruit and engage service users were set out. We developed a framework for inclusion in research which defined minimum core levels of collaboration plus additional consultation and collaboration opportunities. A flow diagram identified when and how to include service users at each of the Medical Research Council’s five stages of developing and evaluating complex interventions in health [1]. We listed people across the research team responsible for including service users in studies and promoting an inclusion culture and highlighted importance of training for researchers and service users. We stated that service users should be included as early as possible in the research process with a minimum of two on all formal trial groups and committees. We proposed a minimum 1% of total research budget should be set aside to include service users and sufficient additional time built in to allow for full inclusion.

Conclusion: Supporting good practice when including service users in research could benefit the relevance, accountability and quality of health and social care research. This SOP provides guidance to researchers to successfully involve service users in developing proposals, undertaking rigorous research and creating a culture of routine service user inclusion in research at all stages. The UK Clinical Research Collaboration should require trials to demonstrate service user inclusion and research funders must set aside sufficient funds and time for this in research proposals.

Acknowledgements: We are grateful to the service users and researchers who contributed to this Standard Operating Procedure. This study is presented on behalf of the West Wales Organisation for Rigorous Trials in Health (NWORTH).

Reference

A84
Self-help therapy and recovery in psychosis: methodological considerations and service user involvement in a partially randomised preference trial
Samantha Hartley1*, Gillian Haddock2
1Psychology Services, Greater Manchester West Mental Health NHS Foundation Trust, Prestwich, UK; 2Division of Clinical Psychology, University of Manchester, Manchester, UK
Trials 2011, 12(Suppl 1):A84

A new Recovery Guide for people with experience of psychosis has been developed (Barrowclough et al, 2009), drawing on existing cognitive behavioural interventions and founded on collaborations between service users, psychologists, researchers and healthcare professionals. This poster showcases the involvement of ‘experts by experience’ in the design, delivery and evaluation of this Recovery Guide, and the therapeutic structure that supports its use. The project has input from service users by various methods, including service user researchers, a service user reference group and service user consultants. Previous work has highlighted the strong and rational preferences that people experiencing psychosis hold about their potential treatment options (Gerrard et al, 2009); service users were presented with detailed information about the Recovery Guide and about different levels of therapeutic intervention to support it (low support= telephone cognitive behavioural therapy; high support= as low support, plus group sessions facilitated by a cognitive behavioural therapist and service user researcher). Findings revealed that most would be willing to take part in a trial of psychological treatments, whether as participants receiving extra therapeutic support or receiving treatment as usual only and assisting with the progress of the research. Participants also held strong feelings about which option they would prefer to receive. Participant preferences of this kind have not been taken into account in previous psychological treatment trials for psychosis. The poster describes the trial currently being implemented to evaluate the Recovery Guide; STAR-T (Self-help Therapy and Recovery-Trial), is an NIHR-funded preference trial in which participants can elect to join their treatment option of choice, or be randomly allocated to a condition. The trial emphasises the value of choice for people experiencing psychosis, and the poster will also discuss the methodological considerations necessary to support such a preference trial. The trial also benefits from the involvement of two service user researchers, who have personal experience of psychosis, and a panel of service user consultants. The ways in which this input has helped shape the design, implementation and evaluation of the trial will be outlined.

Acknowledgments: The authors would like to acknowledge the input of the project team (past and present): Christine Barrowclough (Co-Principal Investigator); Katherine Gerrard, Rachel Watts, Sarah Woodward, Kimberly Drummond (Research Assistants); Liz Pitt, Jason Price (Service user Researchers); James Kelly, Sandra Neil, John Mulligan, Chris Taylor, Zoe Rivers, Katherine Berry, Mary Welford (Therapists),
the Recovery Programme research team including programme lead Tony Morrison, service user reference group and service user consultants. We are also grateful for the involvement of all participants, clinical teams, and the support of the North West Mental Health Research Network.

A85
The development of a DVD to aid patients’ understanding of surgical breast reconstruction clinical trials: QUEST Trials A & B
Judith Mills1, Maria Emson1, Judith M Bliss1, Zoe E Winters1*
1Clinical Trials & Statistic Unit (CTSU-ICR), The Institute of Cancer Research, Sutton, Surrey, UK; 2University of Bristol, Department of Clinical Science South Bristol, Bristol, UK
Trials 2011, 12(Suppl 1):A85

Introduction: There is a pressing need for further clinical evidence to better inform both patients and clinicians when recommending the optimal type and timing of breast reconstruction. The QUEST trials are unique; the first surgical trials comparing different types (A) and timing (B) of Latissimus dorsi (LD) reconstruction with a primary outcome of quality of life (QoL). Surgical trials are challenging, therefore necessitating a feasibility study to assess the acceptability of randomisation from both the perspectives of the patients and healthcare professionals. It was decided to develop a DVD to complement the patient information sheet, in order to help patients understand the concepts of the clinical trials, surgical techniques, randomisation and clinical equipoise.

Methods: All the women filmed in the DVD had a prior breast reconstruction and were invited to participate in a Q&A session with the Chief Investigator (CI) of the QUEST trials. The CI and 2 breast care nurses were also interviewed about the surgical techniques, complications and side-effects and the time taken to return to normal everyday activities after a breast reconstruction. All women will be given a short questionnaire to complete, assessing their level of understanding of randomisation and clinical equivalence irrespective of their entry into the QUEST trial.

Results: Filming took place over 1 day in Bristol and editing over a further 2 months. Drawings were produced to pictorially explain clinical equivalance, randomisation and the different types of LD reconstruction. The DVDs and the patient information sheets will be given to all women who are eligible and interested in the QUEST trials. We will be evaluating the impact of this DVD on patient recruitment by both quantitative (questionnaires) and qualitative (interviews) methodology.

Conclusions: It proved feasible to develop a patient targeted DVD based on the experience of patients and healthcare professionals to enable potential QUEST participants to make fully informed decisions.

A86
Research to meet the needs of the NHS: a review of published NIHR HTA clinical trials
Amanda Young1*, Ruairidh Milne2, James Raftery1, Louise Dent3, David Turner1, Peter Davidson1, Andrew Cook1
1University of Southampton’s Wessex Institute, Southampton, SO16 7NS, UK; 2NIHR, Evaluation, Trials and Studies Coordinating Centre (NETSCC), University of Southampton, Southampton, SO16 7NS, UK; 3University of Southampton Clinical Trials Unit, Southampton General Hospital, Southampton, SO16 6YD, UK
E-mail: A.J.Young@southampton.ac.uk
Trials 2011, 12(Suppl 1):A86

Background: The NIHR HTA programme was established in 1993 with the aim of being ‘needs-led’. This was achieved through commissioning trials to address questions specified by the programme rather than by applicants. At the time this was an unusual funding model and the programme has developed experience that is now globally significant. There is dissatisfaction worldwide with the focus of much existing clinical research (the arrival in the USA of Comparative Effectiveness Research is one attempt to address this) and many trials focus on outcomes that are of little relevance to patients [1]. However, apart from early work reviewing the sources of topics suggested to the HTA programme [2], there has been no systematic attempt to review the programme’s success in being needs-led.

Table 1 (abstract A86) Summary data

<table>
<thead>
<tr>
<th>Description</th>
<th>No (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sources used to identify research suggestions:</td>
<td></td>
</tr>
<tr>
<td>Widespread consultation</td>
<td>77 (68.8)</td>
</tr>
<tr>
<td>Systematic review / DARE review</td>
<td>25 (22.3)</td>
</tr>
<tr>
<td>Horizon Scanning Centre</td>
<td>3 (2.7)</td>
</tr>
<tr>
<td>Reconsidered / recycled topics</td>
<td>5 (4.5)</td>
</tr>
<tr>
<td>No data available</td>
<td>2 (1.8)</td>
</tr>
</tbody>
</table>

Priority given by the programme to the research:

<table>
<thead>
<tr>
<th>(up to and including publication date 1999)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recommended for commissioning – must commission</td>
</tr>
<tr>
<td>Recommended for commissioning</td>
</tr>
<tr>
<td>Category unknown (FT and NSCAG)</td>
</tr>
<tr>
<td>(Post 1999)</td>
</tr>
<tr>
<td>Commissioning requested by PSG</td>
</tr>
<tr>
<td>Direct commissioning in priority area</td>
</tr>
</tbody>
</table>

Actual type of primary outcome reported in the monograph:

| Patient important (including those with additional outcomes) | 84 (75.0) |
| Surrogate | 7 (6.3) |
| Physiological / laboratory | 1 (0.9) |
| Other | 18 (16.1) |

No information available / unable to judge primary outcome | 2 (1.8) |

Objectives: The establishment of a database of metadata for HTA trials allowed us to examine the experience from a long series of published HTA trials. We therefore set out to review how far trials funded by the NIHR HTA programme were indeed needs-led by looking at:

□ The source of the original suggestion
□ The priority given by the programme to the research
□ The patient-relevance of the primary outcome

Methods: The study used a variety of methods to assess the extent to which HTA trials published up to March 2011 were needs-led (n=112). The sources of topic identification and the prioritisation methods were examined over a 10 year period. The type of primary outcome measure reported during the commissioning stage, research protocol and HTA monograph were assessed using Gandhi et al. (2008) [1] classification list.

Results: The source of the original suggestion: 77/112 trials (69%) addressed questions that came out of the widespread consultation, a mix of postal and online questionnaires of managers, clinicians and patient and professional groups. Majority of the remaining trials were from systematic reviews (22%, 25/112). The priority given by the programme to the research: 58/84 had been recommended as ‘must commission’ priorities by the HTA programme, meaning that they would be advertised and if that failed to result in funding, further work would be done to ensure that they were taken forward.

The patient-relevance of the primary outcome: Three quarters of trials addressed patient relevant outcomes (75%, 84/112). This compares with 46% and 45% in previous studies [1, 3].

Conclusions: These analyses suggest that the first 112 published HTA trials can indeed claim to be meeting the information needs of the NHS. Further work is required to compare these results with elsewhere and to develop more robust measures for the future.

Acknowledgements: The authors would like to thank members of the advisory group for their valuable comments during the development of the metadata database. We also wish to thank the Horizon Scanning Centre for helping with the most recent data collection.

References


PHASE II TRIALS

A87
Identifying appropriate phase II trial designs
Sarah R Brown
Clinical Trials Research Unit, University of Leeds, Leeds, UK
Trials 2011, 12(Suppl 1):A87

The phase II to III transition in the drug development process is associated with the highest risk compared to transition rates between other phases [1]. With increasing pressure to improve efficiency in this process it is essential that phase II trials are designed based on informed decisions, and to provide reliable results.

With over 120 different phase II trial designs available [2], identifying which designs are appropriate can be difficult and is based on a number of elements. Randomisation, endpoint selection, and statistical design all contribute to the decision as to which trial design to use. Additionally, in an environment of ever-changing treatments and newly developed biomarkers, it is vital that the way in which treatments may work is incorporated into the decision making process. As a solution to identifying designs, a structured thought process, guidance manual and library of phase II trial designs has been developed [2]. This considers key elements associated with identifying a phase II trial design, and is intended to facilitate interaction between the clinician and statistician, as well as providing a structured and systematic approach to identifying appropriate trial designs.

Challenges remain, however, in choosing between a number of designs identified that fit trial-specific design criteria. Researchers may consider practical elements of conducting a trial, or previous experience, to determine which design to use. However further consideration to the performance of different designs may be necessary. Simulation provides an ideal opportunity to evaluate this under differing trial scenarios, and is often used in the design of phase I trials.

An overview of the role of phase II trials in the drug development pathway will be presented, highlighting current issues and solutions to identifying appropriate trial designs, including a worked example. Further discussion will include the challenges in choosing between designs, with an example of the use of simulation to evaluate trial design presented.

Acknowledgements: This research was carried out in collaboration with Julia Brown, Walter Gregory, Chris Twelves, Marc Buyse, Mahesh Parmar and Matt Seymour.

References

A88
Phase II investigation of a PARP inhibitor (olaparib) in castration resistant prostate cancer (CRPC) which incorporates the possibility that treatment effect may be restricted to biomarker defined subgroups
Roger AHern*, Johannes DeBono*, Shahnene Sandhur*, Elethena Kalatzaki*, Martine Ustin*, Emma E Hall*
1Clinical Trials and Statistics Unit, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK; 2Department of Experimental Cancer Medicine, Institute of Cancer Research, Sutton, Surrey, SM2 5NG, UK
Trials 2011, 12(Suppl 1):A88

Objective: To undertake a phase II trial to identify the best CRPC patient group(s) to be studied for sensitivity to olaparib in a randomised, controlled phase III trial.

Methods: The TO-PARP trial has a multistage phase II design consisting of a non-randomised component with response as the primary endpoint followed by a randomised component with overall survival as the primary endpoint.

Non-randomised component: This part of the trial design allows rapid progression to a randomised comparison if there is evidence of a high response rate (50% or more) in unselected patients. If the evidence for a high response rate in unselected patients is weak, biomarker defined groups will be investigated for their sensitivity to olaparib.

The first stage will include 30 CRPC patients. If 15 (50%) or more respond then no more patients will be entered and the randomised component will be undertaken in unselected patients. Should 2 (7%) or fewer respond then olaparib will be rejected. If between 3-14 (10%-47%) patients respond then a further 15 patients will be entered. Should 23 (51%) or more respond overall then the randomised component will be undertaken in unselected patients and if 5 (11%) or fewer respond then olaparib will be rejected. Otherwise with 6-22 responders (13%-49%), biomarker analysis of tissue collected from all 45 patients will be undertaken. This will aim to identify a sensitive subgroup, with a response rate which is compatible with a 50% response rate, warranting further assessment in the randomised component. If such a subgroup is found, a confirmatory single stage 44 patient trial will be undertaken in this subgroup; this will also serve to pilot prospective biomarker testing in a multi-centre clinical trial setting.

Randomised component: This is a phase II controlled, definitive endpoint assessment of the results generated in the non-randomised part, offering a more secure foundation for efficacy before proceeding to phase III. 180 patients will be randomised 2:1 to olaparib or an appropriate standard of care (i.e. 1-sided 10%, power 80%).

Conclusions: Randomised phase II trials are the gold standard to establish a basis for treatment efficacy but only a subset of patients may be sensitive to a new therapy. This possibility was incorporated into the design of the TO-PARP trial by introducing a non-randomised initial component with the potential to identify a sensitive subgroup.

A89
Design choices for small-scale phase II trials with non-inferiority (NI) intention
Hong Sun*, Shu-Fang Hsu Schmitz
Syngenta Group for Clinical Cancer Research (SACK)
E-mail: Hong.sun@sakk.ch
Trials 2011, 12(Suppl 1):A89

Background: So far, NI trials are almost exclusively performed in phase III setting. However, the NI question might already be relevant in phase II setting in several scenarios, like: treatment optimization in diseases for which standard treatments already exist, or investigation of cytostatic agents, or rare diseases for which a phase III trial will not be feasible, or try to establish a reduced dose for frail patients. Before pursuing a phase III NI trial, a phase II trial is warranted. If resources are limited, then the scale of the phase II trials might be constrained. Feasible design choices for small-scale phase II trial with NI intention are desired.

Methods: Both frequentist hypothesis testing and confidence interval (CI) approaches are considered for the intended NI phase II trials. Both single-arm and two-arm trials with a binary endpoint were performed, sample sizes obtained from the two approaches are compared under different parameter settings with the NI margin in the hypothesis testing approach equal to the interval width in the CI approach.

Results: If the success rate of standard treatment is 0.5, with the NI margin 0.1 for the new treatment, the required sample size for one-sided hypothesis testing approach is at least 283 patients in two-arm trial with power 80% and 20% type I error (alpha) and 158 in single-arm trial with same power and 5% alpha. Using one-sided CI. approach with 95% confidence level and allowed width of 0.1, only 76 patients are needed. The smaller the margin or the width is, the larger sample size is needed for either approach. The sample size from the CI approach is always lower than that from the other approach under the same settings.

Discussions: The approach of the hypothesis test here is basically the same as for phase III trials, but could have a larger NI margin or alpha [1,2]. However, the determination of the NI margin is often problematic in a phase III setting and even more difficult in phase II setting as reliable
previous data about the efficacy may not be available at this early stage. The C.I. approach does not require the NI margin or impose a rigid go/no-go decision rule only based on the primary endpoint. This allows the investigators to consider other aspects of the treatment for the final decision, hence might be preferred.

References

A90
A Bayesian dose-finding procedure applied to a seamless phase I/II trial in rheumatoid arthritis
Anne Whitehead1, Helene Thysen2, Vladimir Dragalin1, John Whitehead1
1Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, Lancashire, LA1 4YE, UK; 2Aptiv Solutions, Morrisville, North Carolina, 27560, USA
E-mail: p.a.whitehead@lancaster.ac.uk
Trials 2011, 12(Suppl 1):A90

There is a growing interest amongst clinical investigators in the conduct of single trials combining the safety evaluation of phase I with the initial investigations of efficacy usually made during phase II. This is being made increasingly possible through the use of biomarkers that show early signs of physiological changes that are associated with a therapeutic effect. Such a combined study calls for complex statistical models, able to capture the joint distribution of the safety and efficacy outcomes. Bayesian models are particularly attractive in such early phase studies because in interpreting small data sets, judicious use of investigators’ opinions becomes worthwhile. We will describe a dose escalation procedure for a combined phase I/II clinical trial, based on a Bayesian model for the joint distribution of toxicity and efficacy (both considered binary variables) making no assumptions other than monotonicity: that is the risk of toxicity and the chance of benefit are both assumed to be non-decreasing as functions of dose level. The procedure will be discussed in the context of a placebo-controlled, sequential trial in rheumatoid arthritis, in which patients, in each stage, are randomized across all dose levels that appear safe and non-futile at the time of recruitment. The primary efficacy outcome is a binary response at 16 weeks related to an assessment known as the ACR20, but an earlier efficacy assessment based on the ACR20 assessment and reduction in C-reactive protein at 4 weeks is used during the dose escalation phase for making decisions on doses for the next cohort. The measure of safety is the occurrence of a dose limiting toxicity within 4 weeks of treatment. Based on data from a pilot study, we constructed five different scenarios for the dose-response relationships for which we simulated the trial and assessed the performance of the procedure. The new method appears to have satisfactory operational characteristics, and is flexible in that it can be adapted to the logistics of a particular trial and incorporate a placebo arm.

A91
Screened selection design for randomised phase II oncology trials: an example in chronic lymphocytic leukaemia
Christina Yap1, Lucinda Billingham1-3, Andrew Pettitt1
1MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham, B15 2TT, UK; 2Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, B15 2TT, UK; 3Molecular and Clinical Cancer Medicine, University of Liverpool, Liverpool, L69 3GA, UK
E-mail: cyap@bham.ac.uk
Trials 2011, 12(Suppl 1):A91

Background: There is a growing number of potential new treatment regimens for chronic lymphocytic leukaemia (CLL). As there is limited numbers of patients, it is important that statistical methodologies in Phase II trials efficiently select promising regimens for subsequent evaluation in a confirmatory, larger-scale Phase III trial.

Methods: In this study, we propose the screened selection design, which combines two conventional Phase II trial designs to provide a practical multi-stage approach for evaluating the efficacy of two experimental arms in high-risk CLL patients. Our aim is to select the most promising regimen in terms of effectiveness to be recommended for further testing. The proposed Phase II randomised design is divided into two different segments. In the first segment, patients are randomised equally into two experimental arms. By applying Simon’s two-stage design [1] in each of the two parallel experimental arms, this allows for initial determination of efficacy and early stopping for futility in any of the arms. If there are an insufficient number of responses in the first stage, recruitment will not continue for that particular arm. Otherwise, the study proceeds to stage 2 to randomize further patients to each arm. The second segment of the study involves the play-the-winner selection strategy as proposed by Simon, Wittes and Ellenberg [2], which only applies if results from both arms are found to be positive. Our proposed design allows the treatment arm with the highest response rates to be recommended only if the efficacy rate is greater than a specified clinically-relevant value. The number of subjects required for each treatment arm in the first segment is selected to be as close as possible to the number required for the selection strategy in the second segment according to pre-specified error rates. The operating characteristics of the trial design are explored via a simulation study.

Results: The proposed approach has the advantage of substantial reduction in the probability of incorrectly selecting an ineffective arm whose rates are not clinically significant or when no true difference exists between the arms. The only compromise is a slight reduction in the probability of correctly selecting an effective treatment arm if one exists. This approach is comparable to the Bayesian Selection Strategy proposed by Estey and Thall [3].

Conclusions: The proposed approach provides an easy to implement Phase II design to select a most effective and most promising treatment regimen for further testing in Phase III.

References

A92
Experiences in the design and implementation of phase II trials in CLL
Dena R Cohen1, Peter Hillmen2, Julia M Brown1, Walter M Gregory3
1Clinical Trials Research Unit, University of Leeds, Leeds, UK; 2Leeds Teaching Hospitals NHS Trust, St James’s University Hospital, Leeds, UK
Trials 2011, 12(Suppl 1):A92

Since 2004, we have developed five phase II trials in Chronic Lymphocytic Leukaemia (CLL), utilising six different statistical methods. Two of the trials have closed to recruitment, two are currently open and a further one is in development. The rationale behind the different designs chosen for each trial will be explained. Difficulties and learning experiences with the implementation, wider understanding and interpretation of the trials will be discussed.

CLL201 used Gehan’s two-stage approach to assess response, and randomised to a control arm which was not included for formal comparison, but to give validity of the study results. Challenges included the timing of the stage I analysis without halting recruitment, and the temptation to formally compare the two arms even though there was not power to do so. The inclusion of the control arm proved to be valuable since the response rates were not as expected.

CLL207 is a single arm trial designed using Bryant and Day’s two-stage design, incorporating toxicity considerations as well as efficacy. The two-stage aspect worked well in this trial due to the short treatment duration and assessment time. However, implementation was difficult due to the definitions of unacceptable toxicity and unacceptability bounds, and the overlap with the role of the Data Monitoring Committee.

ARCTIC and ADMIRE are two large, randomised phase III trials, both formally powered to compare responses against a common control arm. One of the trials assesses non-inferiority. Difficulties were experienced in convincing reviewers that these were not underpowered phase III trials. This design was necessary for the non-inferiority question, as it provides an acceptable certainty of finding the treatment inferior in terms of response before proceeding to a much larger trial to assess longer-term endpoints.

Trials 2011, Volume 12 Suppl 1
http://www.trialsjournal.com/supplements/12/S1
COSMIC is a randomised selection design with two experimental arms. The A’Hern one-stage design is used to determine which of the treatments are eligible to be taken forward for further investigation. In the case where both are acceptable, Sargent & Goldberg’s selection criteria will be applied to determine whether to take forward the treatment with better response rate, or to use alternative selection criteria. The sample size was inflated to ensure acceptable power for selecting the best treatment.

### PLENARY

**A93**

Statistical validation of surrogate outcome measures

Marc Buyse
IDIII, Louvain-la-Neuve, and Hasselt University, Diepenbeek, Belgium
E-mail: marc.buyse@idiid.com
Trials 2011, 12(Suppl 1):A93

With the large number of promising new molecules that are currently available for clinical testing, clinical trials need to detect a drug’s benefit (or harm) as quickly as possible. In parallel with the need for speed in clinical development, advances in molecular biology, high throughput technologies and imaging techniques provide investigators with an ever growing number of biomarkers which can be used for a variety of purposes: to inform go / no go decisions in early clinical development, to stratify patients, to target subsets, to adjust therapy, or to replace clinical endpoints in the comparison of the new drug with standard treatments. This talk will focus on the latter goal, and will discuss the type of statistical evidence required for a biomarker (or a clinical endpoint) to be an acceptable outcome measure for use in clinical trials [1].

Historically, the first formal definition of surrogacy is due to Prentice [2]. While this definition has had a huge role in focusing attention on the need for formal statistical criteria to validate a potential surrogate, it may also have led to excessively pessimistic views about the potential for any outcome measure (whether it be a clinical endpoint or a biomarker) to ever qualify as a good (let alone “perfect”) surrogate. A large amount of research has been devoted to operational criteria to implement Prentice’s definition in practice.

An even larger amount of research has been devoted to a different approach based on statistical associations between the endpoints (surrogate and true), and between treatment effects on these endpoints. It has been proposed that a good surrogate must be tightly correlated with the true endpoint (the so-called “individual-level” association), and that the treatment effect on the surrogate must be tightly correlated with the treatment effect on the true endpoint (the so-called “trial-level” association) [3].

Showing that both criteria are met usually requires a meta-analysis of randomized trials, or one large trial that can be broken down in smaller units (such as participating countries). When such data are available, the predictive value of potential surrogate biomarkers can be investigated, and the “surrogate threshold effect” can be estimated as the minimum effect on the surrogate biomarker that predicts a statistically significant effect on the clinical endpoint [4].

A very different line of research has evolved from concepts of causal inference, in particular the concept of “principal stratification”, in which treatment effects on the true endpoint are estimated within strata defined by different surrogate values [5]. The conceptual elegance of this approach has not yet led to convincing applications, in large part because it has proven challenging to find good estimation methods for the counterfactual probabilities that are required to validate a surrogate [6]. It is likely, however, that causal inference will play a key role in future attempts to validate surrogate endpoints.

**References**


**A94**

Investigator led trials: challenges and opportunities

Peter Sandercock
University of Edinburgh, Department of Clinical Neurosciences, Western General Hospital, Edinburgh EH4 2XU, UK
Trials 2011, 12(Suppl 1):A94

Investigator led trials, whether of novel pharmaceutical agents, surgery or complex non-drug therapies, face a particular set of challenges. This is a selective review, based on personal observations and experiences of trials in my own field of stroke and clinical neuroscience over the past 32 years. The guiding principles of clinical trials – all too often lost in the process of designing and managing clinical trials - are: trials should address a ‘burning question’; to provide sufficiently reliable evidence to answer that burning question (i.e. sufficient to change practice), trials need to be large; to achieve large sample sizes, trials need to be affordable and efficient in design and conduct; and, investigators need rigorously to apply the principles of the ‘business model’ of clinical trials to help overcome the many obstacles that inhibit trials.

From the 1980’s and into the early 2000’s, a series of very successful large-scale investigator-led trials in cardiovascular medicine and neuroscience had a substantial impact on clinical practice worldwide: the ISIS 1,2,3 & 4 trials in acute myocardial infarction, IST and CAST in acute stroke, CRASH in traumatic brain injury, ISAT in the management of ruptured intracranial aneurysms. All of these trials were conducted on remarkably modest budgets by today’s standards. After 2004, the regulatory environment changed significantly – especially for trials in the UK – with the implementation of the EU directive on clinical trials, the changes to UK ethical approvals system, and the introduction of the research governance framework. These changes have undoubtedly substantially raised the costs of, and delayed implementation of, clinical trials (and I would argue those costs and delays have disproportionately increased for investigator-led trials).

There is little disagreement that trials like those listed above simply could not be undertaken in the current regulatory environment. Clinical trials in the 21st century must therefore adapt to the new environment (or die). However, for investigators planning their own trials now, there are grounds for a degree of optimism, since in the UK at least, several changes are underway that reduce the risk that investigator trials might become extinct. To name a few of these changes: the regulatory environment will ease a little; the number of registered clinical trials units in the UK has substantially increased for investigator-led trials.

### A95

Seven myths of randomisation in clinical trials

Stephen Senn
CRP Santé, Strassen, L-1445, Luxembourg
E-mail: stephen.senn@crp-sante.lu
Trials 2011, 12(Suppl 1):A95

Randomisation in clinical trials continues to be controversial and to be attacked on ethical, practical and philosophical grounds. Whatever the practical, moral or philosophical limitations of randomisation may be, many of the attacks reveal that the critics do not understand why randomisation is carried out and what its purpose is. In some cases one may even claim that the defenders of randomisation are equally confused. Seven common myths are examined in this paper, namely, that 1. patients are treated simultaneously in clinical trials 2. balance of prognostic factors in necessary for valid inference 3. observed covariates may be ignored because one has randomized 4. blinding can be carried out effectively without randomisation 5. randomisation is inefficient 6. randomisation precludes balancing covariates and 7. large trials are more balanced than small ones.
Many of these myths are related to the problem of distinguishing between conditional and unconditional inference and the relevance of this distinction is explained with a simple game of chance involving two fair dice, and played in three variants, and a statistician who has to correctly call the odds of obtaining a total score of ten. Others simply arise, because many who have written on clinical trials have not designed them.

In fact, what some critics have overlooked is that when it comes to allocating treatments to patients, ‘the devil is in the detail’. A useful discipline that can be recommended to any would-be critic of randomisation in clinical trials is to attempt to write a detailed protocol, capable of being followed by another. This ‘thought experiment’ is useful in revealing potential problems with any scheme.

Finally a technical problem to do with the estimation of nuisance parameters in analysis of covariance is covered.

It is concluded that the debate on the role of randomised clinical trials in evidence based medicine would be improved if those debating it paid careful attention to what randomisation can and cannot do to strengthen the validity of inferences regarding the effects of treatment.

**PRAGMATIC TRIALS**

**A96**

Alternative approaches to tuberculosis treatment evaluation: the role of pragmatic trials

Daniel J Bratton, Andrew J Nunn

MRC Clinical Trials Unit, 125 Kingsway, London, WC2B 6NH, UK

E-mail: djb@ctu.mrc.ac.uk

Trials 2011, 12(Suppl 1):A96

Clinical trials are sometimes classified as following either an explanatory or pragmatic design. Explanatory trials seek to investigate the efficacy of new treatments by imposing strict limitations on the design of the trial, including, for example, recruiting only those patients who are most likely to respond, considering only the most adherent patients in analyses or standardising trial treatment. In contrast, the objective of pragmatic trials is to assess whether treatments work in conditions more appropriate to routine practice and are therefore often much less restrictive in their design.

This dichotomy is, however, over simplistic; there is a continuum which exists due to there being many areas of trial design that can vary between the extremes of the explanatory and pragmatic approaches. The PRECIS (Pragmatic-Explanatory Continuum Indicator Summary) wheel is a tool that proposes a way of enabling researchers to assess the extent to which the trials they are designing could be considered explanatory or pragmatic. It is based on ten aspects of trial design, including inclusion criteria, flexibility of delivery of the intervention and intensity of follow-up. We consider how the PRECIS tool can be applied in the design stage of trials of tuberculosis treatment, to help teams design trials in line with their purpose. In view of the well recognised delay in getting results from well conducted clinical trials into practice, we would suggest that if more pragmatic trials are conducted, physicians would better understand the implications of the results for their own practice and be more ready to adopt new treatments [1].

**Reference**


**A97**

Treatment success in pragmatic randomised controlled trials: a review of trials funded by the UK Health Technology Assessment programme

Louise Dent*, James Raftery

1University of Southampton Clinical Trials Unit, MP131, Southampton General Hospital, Southampton, SO16 6YD, UK; NHSR Evaluation, Trials and Studies Coordinating Centre, University of Southampton, Southampton, SO16 7NS, UK

E-mail: ldent@southampton.ac.uk

Trials 2011, 12(Suppl 1):A97

**Background:** Equipoise implies that given a random unbiased sample of trials, no significant difference would be expected in the proportion favouring the new treatment to the proportion favouring the standard treatment [1-3]. Previous research reviewed treatment success and whether the collective uncertainty principle is met in RCTs in the US National Cancer Institute portfolio [4-6]. This paper classifies clinical trials funded by the UK HTA programme by results using the methodology applied to the US Cancer Institute trials, and compares the two portfolios [7].

**Methods:** Data on all completed randomised controlled trials funded by the HTA programme 1993-2008 were extracted. Each trial’s primary results was classified into six categories; 1) statistically significant in favour of the new treatment, 2) statistically significant in favour of the control treatment 3) true negative, 4) truly inconclusive, 5) inconclusive in favour of new treatment or 6) inconclusive in favour of control treatment. Trials were classified by comparing the 95% confidence interval for the difference in primary outcome to the difference specified in the sample size calculation. The results were compared with Djulbegovic’s analysis of NCI trials.

**Results:** Data from 51 superiority trials were included, involving over 48,000 participants and a range of diseases and interventions. 85 primary comparisons were available because some trials had more than two randomised arms or had several primary outcomes. The new treatment had superior results (whether significant or not) in 61% of the comparisons (32/85 95% CI 49.9% to 71.6%). The results were conclusive in 46% of the comparisons (19% statistically significant in favour of the new treatment, 5% statistically significant in favour of the control and 22% true negative). The results were classified as truly inconclusive (i.e. failed to answer the question asked) for 24% of comparisons (20/85). HTA trials included fewer truly inconclusive and statistically significant results and more results rated as true negative than NCI trials.

**Conclusions:** The pattern of results in HTA trials is similar to that of the National Cancer Institute portfolio. Differences that existed were plausible given the differences in the designs of the trials - HTA trials are more pragmatic. The results indicate HTA trials are compatible with equipoise. This classification usefully summarises the results from clinical trials and enables comparisons of different portfolios of trials.

**References**


trial procedures, including randomisation, maintaining the counterfactual, recruitment; and relevance of findings for and translation into policy and practice.

Objectives: The current NPRI funded trial of SFP 10-14 UK is presented as a case study to discuss these issues, solutions and remaining barriers. The SFP 10-14 UK programme aims to strengthen areas of family life that protect against substance misuse, for example, parenting, communication, and young people’s resilience skills. The SFP 10-14 UK is being delivered by statutory and voluntary agencies in six local authority areas across Wales, and is offered to mixed groups comprised of families from the general population, and families who may experience/present challenges within a group setting.

Methods: The trial aims to recruit 748 families, 374 of whom will be randomised to receive the usual services available to families within their local area. 374 families will receive the SFP 10-14 UK in addition to usual care. Families are identified by staff employed within the statutory services and voluntary sectors and referred to embedded research staff for recruitment.

Results: Challenges encountered related to a lack of awareness of the randomised trial as a research paradigm among staff and key referring agencies, related concerns about the ethics of randomisation and the maintenance of the counterfactual among the usual care group, and challenges regarding the maintenance of recruitment and intervention fidelity. Whilst a challenge in itself, partnership working with delivery agencies, programme trainers, and the Welsh Assembly Government at all stages of the development, funding and conduct of the trial has proved an important strategy to overcome these issues.

Conclusions: This trial seeks to generate evidence on the effectiveness and cost effectiveness of the SFP10-14 UK which is of direct relevance to general practitioners, commissioners and practitioners. The trial highlights that strategic partnership working, the winning of ‘hearts and minds’ regarding the ethics and operationalisation of randomisation, and maintaining the balance between internal and external validity are key areas of focus for the successful conduct of pragmatic trials in non-NHS settings. The lessons learnt from its implementation will be important for future multi-sector agency projects and for role out of the intervention if found to be efficacious.

A99

Generalisability of trials of home blood pressure monitoring: a comparison of two UK primary care trials

Sally M Kerry1,2, Hugh S Markus3, Teck K Khong4, Pippa Oakeshott4

1Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, E1 2AB, UK; 2Stroke and Dementia Research Group, St George’s, University of London, London, SW17 ORF, UK; 3Basic Medical Sciences, St George’s, University of London, London, SW17 ORF, UK; 4Population Health Sciences and Education, St George’s, University of London, London, SW17 ORF, UK
Trials 2011, 12(Suppl 1):A99

Objective: Blood pressure (BP) monitors are widely available and easy for patients to use. Systematic reviews [1-3] show that home monitoring of BP improves BP control but there is significant heterogeneity between studies, and meta regression has only been able to explain part of the heterogeneity, with concomitant interventions being a possible factor. Most evidence comes from trials of patients who have poorly controlled blood pressure at baseline, although not usually explicitly stated in trial or review title. However many hypertensive patients using home monitors may have BP below the recommended target.

Two large recent UK RCTs [4,5] have been carried out with very different inclusion criteria and interventions. The objective is to compare the main findings of these trials and assess how far difference between the trial populations might explain the apparent difference in the efficacy of the intervention between the two trials.

Methods: We identified two hypertension stroke patients were recruited from stroke services in London, and randomly allocated, at home, to either home BP monitoring (n=187), with ongoing telephone support from a nurse, or usual care (n=194). Patients were included without any restriction on baseline BP [4].

The TAMINH trial [5] randomised 527 primary care patients with baseline BP between 140/90 and 200/100, to a self management program (n=263) or control (n=264). The program involved two training sessions for patients, so that they could increase their medications when necessary, without consulting the GP.

The primary endpoint for both trials was change in mean systolic BP between baseline and 12 months follow up after adjusting for baseline BP. A post hoc subgroup analysis of those patients in the stroke study with baseline BP between 140/90 and 200/100 was carried out for comparison.

Results: Follow up rates for survivors were over 90% in both trials. Intervention effect in the TASMINH trial was 5.4(2.4 to 8.3); n=480, and was 0.3(-3.6 to 4.8) in the stroke trial; n=337. Of 1650 patients assessed for eligibility in the TASMINH trial, 916 (55%) had BP less than 140/90 and 48 had BP above 200/100. These patients were excluded prior to randomisation.

133 (40%) patients in the stroke trial had baseline BP between 140/90 and 200/100; the intervention effect in this group was 5.9(-0.3 to 12.0); p=0.02 for interaction.

Conclusions: Whether or not patients with controlled hypertension, who comprise more than half the hypertensive population, are included, may explain the difference in effect size between these 2 trials.

Acknowledgements: The trial of home blood pressure monitoring in stroke patients was funded by The Stroke Association TSA 2006/05 and recruitment to the study was supported by the English National Institute of Health Research Clinical Stroke Research Network.

References
incorporates objectives measures and central blinded assessments of these measures to reduce potential bias.

Timing of randomisation can also be problematic due to the need for theatre planning. Preferably surgery should take place as soon as possible after randomisation however in ROLARR, up to 28 days after surgery has had to be permitted. Monitoring timings will take place to allow prompt action on any possible problems that may increase bias.

Conclusions: Surgical trials are complex to design and implement. Careful consideration needs to be given to the additional issues that arise to ensure an accurate and unbiased interpretation of the results.

A101

Holding onto power: why confidence intervals are not (usually) the best basis for sample size calculations

Chris Metcalfe
School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, UK
Trials 2011, 12(Suppl 1):A101

Objectives: It has recently been suggested in a high profile paper that statistical power is no longer a useful basis for sample size calculations (Bland, BMJ 2009). It is proposed instead to calculate the sample size to achieve a narrow confidence interval width for the treatment effect estimate. My objective is to critically appraise this proposal.

Methods: I compare the proposed approach to sample size calculations with the traditional statistical power based approach, and to the sample size calculations employed for equivalence studies which are also based on confidence interval width.

Results: With a little simplification, the sample size calculations for the traditional power-based approach, for equivalence studies, and following the new proposal can be shown to be much the same. The single fundamental difference is that the new proposal does not include a multiplier to increase the statistical power beyond 50% (i.e. only a 50:50 chance of detecting a true treatment effect of clinically important magnitude). The attempt to having to define a minimum clinically important difference on a predefined primary outcome is wholly unsuccessful. The calculation of confidence interval width must be on a particular outcome measure, still requires the size of an unimportant difference to be defined if the confidence interval is to exclude it, and additionally requires a likely true effect of treatment to be defined about which the confidence interval will be centred.

Conclusions: The proposal to base all sample size calculations on confidence interval width does not avoid the need to pre-define the minimum clinically important difference on particular important outcome measures, and in fact additionally requires that the likely effect of the intervention is specified. Most importantly, the approach does not replace statistical power. Statistical power is simply an inflation of the sample size to allow a good chance that a true treatment effect of clinically important magnitude will be detected, even if by chance it is underestimated in the trial data (as it will be, even if only slightly, with 50% probability). I conclude that statistical power is not the source of dissatisfaction with sample size calculations, and there is no real need to replace it as the basis for sample size calculations.

A102

Making the most of animal data – improving the prospect of success in pragmatic trials in the neurosciences

Kureen J Eqani, Emily S Sena, Hanna M Vesterinen, Malcolm R MacLeod
Department of Clinical Neurosciences, Centre for Clinical Brain Sciences, University of Edinburgh, Royal Infirmary of Edinburgh, UK
E-mail: malcolm.macleod@ed.ac.uk
Trials 2011, 12(Suppl 1):A102

Introduction: Developing disease modifying treatments for neurodegenerative disease and stroke has proved remarkably difficult. It may be that current structures in the pharmaceutical industry encourage early clinical trials designed on the basis of exciting but incomplete preclinical data in order to protect competitive advantage. An overly generous reading of animal data may lead to underpowered clinical trials testing treatments at inappropriate time points and at ineffective (but side-effect free) doses. Trials based on a systematic analysis of animal data may have a better prospect of success. Here, we report such an analysis of interventions tested in transgenic models of Alzheimer’s disease (AD).

Materials and methods: (1) Electronic searching of three online databases to identify publications reporting the use of interventions in transgenic models of AD where outcome was reported as behavioural (probe phase of the Morris water maze (MWM)) or histological (changes in immuno-histochemically stained plaque burden) end-points. (2) DerSimonian and Laird random effects meta-analysis and stratified meta-analysis.

Results: We identified 428 publications testing 353 interventions across 55 different transgenic models of AD. Study quality was low; 16% of papers reported random allocation to group, 22% reported a blinded assessment of outcome and no publications reported a sample size calculation. Blinded assessment of outcome was associated with lower effect sizes for results from the probe phase of the MWM. Longer durations between treatment onset and outcome assessment for plaque burden, and younger ages at assessment of behavioural outcome were each associated with lower effect sizes.

Conclusions: Both study quality and study design characteristics appear to impact observed effect sizes. Study quality was low, and this was associated with larger estimates of treatment effects when a behavioural outcome was measured. Improvement in behavioural outcome (which may be influenced by effects on underlying pathophysiology or by effects on performance in the face of a fixed deficit) is usually larger in older animals, but improvement in histological outcomes was smaller with longer intervals between treatment onset and outcome assessment.

These findings highlight the importance of a detailed and systematic analysis of animal data before embarking on clinical trial. Specifically, if data from animal studies are to be invoked in support of a trial protocol investigators should be able to demonstrate that such data are largely free from bias, and that the treatment has shown efficacy under the circumstances (for instance the stage of disease) in which it is proposed that it be tested.

A103

The CORONIS Trial: international study of caesarean section surgical techniques

Edmund Juszczak*, Barbara Farrell
NPEU Clinical Trials Unit, University of Oxford, Oxford, OX3 7LF, UK
E-mail: ed.juszczak@npeu.ox.ac.uk
Trials 2011, 12(Suppl 1):A103

Objectives: The aim of CORONIS is to examine five specific aspects of caesarean section technique to determine which methods lead to optimum outcomes for women and their babies.

Background: A variety of surgical techniques for all elements of the caesarean section operation are in use. Many have not yet been rigorously evaluated in RCTs, and it is not known whether any are associated with better outcomes for women and babies.

Design: CORONIS is a pragmatic multicentre fractional factorial randomised controlled trial and is being conducted in sites in Argentina, Chile, Ghana, India, Kenya, Pakistan and Sudan [1]. Women are eligible if they are undergoing their first or second caesarean section through a transverse abdominal incision. Five comparisons will be carried out using a 2^3 balanced incomplete block factorial design. Each woman is allocated THREE of the five pairs of interventions using a bespoke secure web-based randomisation system (with 24/7 automated back-up telephone system) hosted by the NPEU Clinical Trials Unit. The 5 pairs of interventions are: i. Blunt versus sharp abdominal entry ii. Exteriorisation of the uterus for repair versus intra-abdominal repair iii. Single versus double layer closure of the uterus iv. Closure versus non-closure of the peritoneum (pelvic and parietal) v. Chronic catgut versus Polyglactin-910 for uterine repair Primary outcome: death or maternal infectious morbidity (one or more of the following: antibiotic use for maternal febrile morbidity during postnatal hospital stay, antibiotic use for endometritis, wound infection or peritonitis) or further operative procedures or blood transfusion. Sample size required: 15,000 women in total; minimum 9,000 women per comparison pair.

Experience: It is possible to conduct a pragmatic trial in a developing world setting. The six week follow-up rate achieved was 98% despite
natural disasters in Pakistan (flooding displacing millions of people) and Chile (earthquake), political unrest in Kenya and Pakistan, and the fact that many women do not have a formal address. Continuous central monitoring of recruitment and trial material usage within sites meant that recruitment targets were met, overall and by pair of interventions. This involved adding new sites in Chile, and switching the allocation used in some sites to compensate for difficulties described above. Adherence to the allocation by surgeons has been exceptional. Final analysis of clinical outcomes is underway.

Conclusions: Good communications (both written and verbal), clear and concise documentation, thorough planning, a strong multi-national collaboration and close working relationships are the building blocks of success.

Acknowledgements: Presented on behalf of the CORONIS Trial Collaborative Group. CORONIS (ISRCTN31089967) is funded by the UK Medical Research Council.

Reference


A104

Randomised Evaluations of Accepted Choices in Treatment (REACT) trials: large-scale pragmatic trials within databases of routinely collected electronic healthcare records

Tjeerd-Pieter van Staa1,2*, Ber Goldacre3, Martin Guifford, Jackie Cassell4, Munir Pirmohamed5, Adel Taweel2, Brendan Delaney3, Liam Smeeth6

1 General Practice Research Database, Medicines and Healthcare products Regulatory Agency, London, UK; 2 Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands; 3 London School of Hygiene & Tropical Medicine, London, UK; 4 Primary Care and Public Health Sciences, King’s College, London, UK; 5 Division of Primary Care and Public Health, Brighton and Sussex Medical School, University of Brighton, Brighton, UK; 6 The Wolfson Centre for Personalised Medicine, Institute of Translational Medicine, University of Liverpool, Liverpool, UK

E-mail: tjeerd.varsta@gpdr.com

Trials 2011, 12(Suppl 1):A104

Pragmatic trials determine the effects of an intervention under the usual conditions in which it will be applied. Databases of electronic health records (EHRs), such as the General Practice Research Database (GPRD), provide a unique opportunity to conduct large scale pragmatic trials. This paper describes the infrastructure as being implemented for two feasibility pragmatic trials within GPRD. Clinicians willing to participate first need to complete a web-based training (covering both the study protocol and key aspects of good clinical practice). The Erfr database will be searched for potentially eligible patients and a list including codes of potentially eligible patients and those of clinicians who completed the training will be compiled. This list will then be sent, using secure file transfer, to the desktop computer of the clinician. When a potentially eligible patient then visits the clinician for a consultation, a flag at the clinician’s computer screen may then appear, notifying of the possibility to recruit. This flag provides a link to the study website, including the consent form and patient information sheet. After obtaining consent, the patient can then be randomised (using the website) and a prescription for the randomly allocated treatment can then be issued by the clinician. Daily downloads of the anonymous EHR records from the clinician’s office to the central research site will allow the reporting of suspected side-effects. The trial database can be compared periodically to the full GPRD for purposes of fraud detection and comparison of the trial patients to the real-life population of users. GPRD is linked periodically to other NHS datasets (including hospital episode statistics, disease registries and death certificates), using an anonymous process through a trusted third party. This linkage allows an anonymous long-term data collection of major clinical outcomes with no intrusion. It is possible to contact the clinician in order to confirm the data as provided by the various sources. One trial concerns a comparison of two statins by randomising patients with primary hypercholesterolaemia and high CVR risk between simvastatin and atorvastatin (with 300 patients; ISRCTN33113202). The other trial concerns a comparison of antibiotic to no treatment in patients with a COPD exacerbation and non-purulent sputum (with 150 patients; ISRCTN72035428).

A105

QUEST Perspective Study (QPS) to measure the understanding by patients and healthcare professionals of surgical breast reconstruction clinical trials: QUEST Trials A & B

Zoe E Winters1,2, Judith Mills2, Marie Emson3, Judith M Bliss3, Rob Horne3

1 University of Bristol, School of Clinical Sciences, Bristol, UK; 2 Clinical Trials & Statistic Unit (CTSU-ICR), The Institute of Cancer Research, Sutton, Surrey, UK; 3 School of Pharmacy, London University, UK

Trials 2011, 12(Suppl 1):A105

Introduction: The QUEST trials are the first surgical trials comparing different types (A) and timing (B) of Latissimus dorsi (LD) reconstruction with a primary outcome of quality of life (QoL) in an attempt to improve clinical evidence. Surgical trials are challenging, necessitating a feasibility study, and the QUEST Perspectives Study (QPS), to assess the acceptability of randomisation from the perspectives of patients and healthcare professionals (HCPs).

Methods: To maximise information from both patients and HCPs who either accept or decline to participate in the trial, it was agreed that a mixed methodologies should be used in QPS. The questionnaires ‘Patient views on QUEST (PVQ) would be given to patients after they have or have not consented randomised to QUEST. Two acceptability will be purposively selected for telephone interview 1 month post-operatively. The HCP questionnaire called PEERS (PERCEPTIONS OF EQUIPOISE EVIDENCE and RANDOMISATION SURVEY) will be requested from all PIs approached to participate in QUEST.

Results: QPS for patients consists of an ‘acceptors’, and a ‘decliners’ questionnaire regardless of trial participation. Both questionnaires have a free writing task’ where patients can record any views they might have about QUEST. By comparison, the clinicians have formulated questions to respond to a number of statements about randomisation, their surgeon and their surgical preferences. PEERS comprises a number of statements on which the HCPs are asked to rate their agreement on their views regarding clinical equipoise and the randomisation design in QUEST.

Conclusion: By integrating the QPS, we will gain a far better insight into why patients wish to enter surgical randomised clinical trials and so facilitate the design of future trials in breast reconstruction. Furthermore, capturing HCPs views are essential in this trial because they need to relay the meaning of clinical equipoise to patients based on the current evidence.

Acknowledgements: We are grateful to CRUK, BUPA Foundation and Above & Beyond Charities for funding this study.

A106

Strategic use of new generation antidepressants for depression, SUN (> ^ : study design and rationale

Nachiro Yonemoto1,2*, Tatuo Akeshi3, Shinji Shimodera2, Mitsuhiro Yamada1, Kazuhira Miki3, Norio Watanabe4, Masatoshi Inagaki6, Toshi A Furukawa4

1 Department of Epidemiology and Biostatistics, Translational Medical Center, National Center of Mental Health, National Centre of Neurology and Psychiatry, Kodaira, Tokyo 187-8502, Japan; 2 Department of Psychiatry and Cognitive-Behavioral Medicine, Nagoa City University Graduate School of Medical Sciences, Nagoa, Nagoya, 467-8601, Japan; 3 Department of Neuropsychiatry, Kochi Medical School, Nankokushi,Kochi 783-8505, Japan; 4 Department of Neuropsychopharmacology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Tokyo 187-8553, Japan; 5 Miki Clinic, Yokohama, Kanagawa, 220-0023, Japan; 6 Center for Suicide Prevention, National Institute of Mental Health, National Centre of Neurology and Psychiatry, Kodaira, Tokyo 187-8553, Japan; 7 Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/ School of Public Health, Kyoto, Kyoto 606-8501, Japan

Trials 2011, 12(Suppl 1):A106

Background: After more than half a century of modern psychopharmacology, with billions of dollars spent on antidepressants annually worldwide, we lack good evidence to guide our everyday decisions in conducting antidepressant treatment of patients with major depression. First we did not know which antidepressant to use as first line treatment. Second we do not know which dosage we should be aiming at with that antidepressant. Because more than half of the patients with major depression starting

Page 39 of 57

Trials 2011, Volume 12 Suppl 1

http://www.trialsjournal.com/supplements/12/s1
treatment do not remit after adequate trial with the first agent, they will need a second line treatment. Dose escalation, augmentation and switching are the three often recommended second line strategies but we do not know which is better than the others. Moreover, we do not know when to start considering this second line treatment. The recently published multiple-treatments meta-analysis of 12 new generation antidepressants has provided some partial answers to the first question [1]. Starting with these findings, this proposed trial aims to establish the optimum 1st line and 2nd line antidepressant treatment strategy among adult patients with a non-psychotic unipolar major depressive episode.

Methods/design: SUN(^_^)D, the Strategic Use of New generation antidepressants for Depression, is an assessor-blinded, parallel-group, multi-centre, pragmatic randomised controlled trial. The trial composites three-steps [2]. Step I is a cluster-randomised trial comparing titration up to the minimum vs maximum of the recommended dose range among patients starting with sertraline. The primary outcome is the change in the Patient Health Questionnaire (PHQ)-9 scores administered by a blinded rater via telephone at week 1 through 3. Step II is an individually randomised trial comparing staying on sertraline, augmentation of sertraline with mirtazapine, and switching to mirtazapine among patients who have not remitted on the first line treatment by week 3. The primary outcome is the change in the PHQ-9 scores at week 4 through 9. Step III represents a continuation phase to Steps I and II and aims to establish longer-term effectiveness and acceptability of the above-examined treatment strategies up to week 25.

Discussion: The trial is first pragmatic mega trial of psychiatry in Japan. We are now going a pilot phase in the trial. The pilot phase is supported by the Grant-in-Aid by the Ministry of Health, Labour and Welfare, Japan. Trial registration ClinicalTrials.gov identifier: NCT01109693.

References

RARE CONDITIONS

A107
Trials in rare diseases: the need to think differently
Lucinda Billingham1,2, Kinga Malottki2, Neil Steven2
1MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham B15 2TT, UK; 2Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham B15 2TT, UK
Trials 2011, 12(Suppl 1):A107

Background: In general, large-scale clinical trials are needed to make definitive conclusions regarding changes in treatment practice. This is because treatment effects are often modest and large numbers of patients are needed to detect such differences with sufficient power for robust conclusions. Such trials are difficult and often impossible in rare diseases and their worth in this setting could be questioned and this drives the need to consider a different approach.

Methods: A review is undertaken of methods that may be appropriate for making conclusions from clinical trials of treatment in rare diseases. In particular, the practicalities of applying a Bayesian strategy [1] are assessed by application to a trial in Merkel cell carcinoma.

Results: It may be possible to run a traditional trial in rare diseases by adapting logistical aspects in order to maximise recruitment but alternative statistical approaches are normally needed. In rare diseases, if the focus is on estimation rather than hypothesis testing then important information can be gleaned from the results of small trials, with interval estimates providing a measure of the uncertainty inherent in small studies. Using a Bayesian approach to trial design in the rare disease setting has been promoted as this enables information gathered from relevant previous studies to contribute to the estimation process. Proposed methodology suggests incorporating all levels of evidence, including case studies, into the prior distribution with appropriate weighting to reflect quality [1]. Practical application of this approach highlights that the substantial effort in incorporating lower levels of evidence may not be worth the gain in information, especially when such information is prone to bias.

Conclusions: Treatment decisions in rare diseases should be based on evidence but the traditional approach to trials is difficult and therefore there is a need to think differently. Small trials focused on estimation rather than hypothesis testing are worthwhile as they will reduce the uncertainty about treatment effects. Incorporating prior knowledge together with trial data using a Bayesian approach can further reduce the uncertainty but the acceptability of this approach is subject to the believability of the prior information.

Reference

A108
Comparison of anticipated and actual control group outcome in randomised trials in paediatric oncology provides evidence that non-randomised studies are biased in favour of the novel treatment
Veronica Moroz1, Jayne Wilson, Keith Wheatley
Cancer Research UK Clinical Trials Unit, University of Birmingham, UK
Trials 2011, 12(Suppl 1):A108

Introduction: Historically controlled studies (HCSs) compare data from two or more separately conducted studies – the new treatment arm, usually prospectively collected, is compared to a control arm of retrospective data. HCSs are frequently undertaken in paediatric oncology (PO), where there is a widespread belief that RCTs cannot be performed in rare diseases. This is despite the potential biases in HCSs being well known – e.g. other factors change over time.

Aim: To compare the outcome of the control group of RCTs in PO with that anticipated in the sample size calculation. The assumption being that, had an HCS been carried out instead, the control arm data in the historical control study would have likely been that utilized in the RCTs sample size calculation.

Methods: A search for published PO RCTs was conducted using the Cochrane Central database from March 2011 to database inception. Search terms were “randomised” plus the disease name in all fields. Only papers reporting sample size parameters and observed estimates were included. Data were extracted and compared on anticipated and observed outcomes in the control arms (and experimental arms).

Preliminary results: To date 45 RCTs have been included from 16 tumour sites: 36 superiority trials, 9 equivalence/non-inferiority; 12 with dichotomous primary endpoints, 33 time-to-event; mean recruited number of patients was 231 (range: 50 to 2619). In 33 trials the control group did better than anticipated (in 13 cases >10% absolute difference), in 11 trials the control group did worse (by >10% in four); outcome was the same as anticipated in one trial. The median absolute difference between control groups’ observed outcome and anticipated outcome was 5% (range: -21% to +35%); median difference was also 5%. In superiority trials, the median difference was 6%; it was 4% in equivalence/non-inferiority trials. In trials with a dichotomous endpoint, the median difference was 7%; it was 5% in trials with a time-to-event endpoint (8 out of 9 equivalence/non-inferiority trials had time-to-event endpoint). The median observed difference between the experimental and control groups was 0% (range: -16% to 21%); the median difference between the observed experimental arm outcome and anticipated control group outcome was 6% (range: -20% to 45%).

Interpretation: Since the control group (i.e. standard treatment arm) in RCTs did better than anticipated, non-randomised HCSs that use similar assumptions for outcome with standard treatment will overestimate the benefit of new treatments, potentially leading to children with cancer being given inappropriate therapy.
Objective: To highlight a successful method for improving recruitment into a trial of a rare skin condition.

Background: Pyoderma gangrenosum (PG) is a rare, ulcerative condition that is often associated with underlying autoimmune disease. Most dermatologists in the UK only see 1-2 PG patients per annum, making this a very difficult condition to evaluate empirically.

Methods: The STOP GAP trial is a randomised controlled trial (RCT) of oral prednisolone compared to oral ciclosporin for the treatment of PG. The study includes a parallel observation study (case series) for patients requiring topical therapy.

Patients who are enrolled into the observational study continue to be followed up and contribute data relating to the efficacy and acceptability of topical treatments. These data provide an important contribution to the available literature on PG, which to-date is based largely on case reports and retrospective case series. An additional benefit of this approach is that patients remain in contact with the research team, and should systemic therapy be indicated, participants are considered for inclusion into the RCT of systemic treatments. This design is efficient and means that most patients who are willing and able to give informed consent are able to contribute to the research activity; this contributes to a broader evidence-base for the treatment of PG patients.

Results: The STOP GAP trial is ongoing and has currently enrolled 73/140 (52%) participants into the RCT, and 35 participants into the observational study. Five (14%) of the patients given topical therapy (observational study) have subsequently gone on to take part in the RCT, having required systemic therapy.

Conclusions: Whilst it is more work to conduct an observational study alongside an ongoing RCT, and the added value of these data is limited by the lack of a randomised comparator, the benefits of this approach outweigh the disadvantages when the condition of interest is rare, and where the existing evidence base is poor.

Trial registration: ISRCTN 35898459

RECRUITMENT STRATEGIES

A110

Recruitment to trials - why is it hard and how might we make it less so?

Shaun Treweek

Population Health Science, University of Dundee, Kirsty Semple Way, Dundee, DD2 4BF, UK

Trials 2011, 12(Suppl 1):A110

Background: Recruiting participants to trials can be extremely difficult. Estimates of just how difficult vary but around half of trials fail to recruit their original target sample size. Poor recruitment can lead to an adverse effect on the trial as a product that needs to be marketed to participants, rather than recruitment planning. Methodological work around thinking of a real trial as a product that needs to be marketed to participants, rather than assuming the value of the trial is obvious to all, is worthy of further investigation.

Conclusions: Trialists and methodologists can help themselves and future recipients of healthcare innovations by making better use of existing research evidence on recruitment and by explicitly aiming to plug gaps in this evidence where they exist.

A111

Investigating strategies to improve attendance at screening visits in a randomized trial

Fang Chen, Kh Rahimi, R Haynes, K Naessens, M Taylor-Clarke, C Murray, Jane Armitage*

Clinical Trial Service Unit & Epidemiological Studies Unit, University of Oxford, OX3 7LF, UK

Trials 2011, 12(Suppl 1):A111

Background: A common method of recruiting for randomized trials is to send letters to potentially eligible patients inviting them to a screening appointment. In 3 consecutive UK studies the proportion attending from those invited fell from 49% [1] in 1995-1997 and 42%[2] in 1998-2001 to 13% at the beginning of an ongoing study in 2007. Procedures were similar in the 3 trials except that in 2007 the Patient Information Leaflet (PIL) was enclosed with the invitation letter. In order to understand whether the contents and/or style of invitation would explain the declining trend, 2 separate randomized comparisons were undertaken during the recruitment for the ongoing study.

Methods: Five (14%) of the patients given topical therapy (observational study) have subsequently gone on to take part in the RCT, having required systemic therapy.

Results: Between July and October 2008, 20,759 personalized invitation letters were randomized to have the PIL or brief summary enclosed. There were no significant differences in either the proportions attending:

- PIL: 1222/10,566 (11.6%) versus not 1181/10,590 (11.2%) [OR 1.06; 95%CI 0.97-1.16]; or in the proportion entering the pre-randomization run-in: 720/1181 (6.8%) of those invited versus 690/1122 (6.5%) [OR 1.05; (0.94-1.17)].

From November 2009 to January 2010, 12,164 patient invitations were randomized to enclose either the modified or the original PIL. A 17% higher attendance was detected for the modified PIL: 580/6104 (9.5%) versus 499/6060 (8.2%) for the original PIL [OR=1.17; (1.03-1.33); p=0.01]. However there was no significant difference in the proportion entering the pre-randomization run-in: 373 (6.1% of those invited) vs 339 (5.6%) for modified versus original (OR 1.10; 0.94-1.28).

Conclusion: Whether the full PIL or brief summary was enclosed with the invitation did not affect the likelihood of attending or entering the run-in.

Enclosing a more patient friendly PIL modestly improved the chance of attending, but not whether patients agreed to enter the study.

References:
**A112**

Rates of recruitment from systematic and opportunistic methods: preliminary results from the DDELPHI study

Melwyn Hillsdon, University of Exeter, Sport and Health Sciences, St Luke’s Campus, Heavitree Road, Exeter, Devon, EX16 7NW, UK

**Trials 2011, 12(Suppl 1):A112**

**Background:** Problems with recruitment at both practice and patient level frequently lead to low response rates in primary care trials that can confound the study results. As part of a trial feasibility study the effectiveness of brief advice from general practitioners to encourage daily walking, we examined the association between practice characteristics and the recruitment of practices and patients. We also examined differences in recruitment rates between two methods of patient recruitment.

**Methods:** Practices in Coventry, Bristol and Devon were invited to take part in the study (n=114) via letter and asked to indicate an expression of interest (EOI) for participation. Of the 37 interested practices, 24 were randomly assigned to cluster randomisation (n=6) or individual patient randomisation (n=6) and then randomised to recruit patients either opportunistically via practice waiting rooms or systematically by post from patient lists. One hundred and forty four patients were recruited. Practice list sizes were categorised as low (<3,500), medium (3,500-8,000) and large (>8,000). Using the home postcodes of patients and the English Index of Multiple Deprivation, practice deprivation levels were categorised as low (lowest quartile of IMD in England) middling (inter-quartile range) and high (highest quartile of IMD).

**Results:** It took an average of 22.6 days (SD 14.7) for practices to give an EOI in Devon, 31.5 days (SD 13.3) in Bristol and 44.3 days (SD 20.4) in Coventry. Time to EOI was 40.7 (SD 22.2), 26.9 (17.2) and 29.6 (13.6) days for small, medium and large practices respectively and 25.0 (SD 2.6), 33.6 (SD 20.0) and 31.2 days for low, middling and high deprivation practices. Devon practices averaged 163.6 (SD 33.5) days to recruit the first patient after EOI compared to 123.5 (SD 35.2) days in Bristol and 134.6 (SD 32.7) days in Coventry. Time from EOI to recruitment of first patient was unrelated to practice size but low deprivation practices averaged 19.4 days less than midlevel practices and 17.2 days less than high deprivation practices to recruit the first patient. Opportunistic recruitment took an average of 132.3 (SD 35.0) days to recruit compared to 151.3 (SD 37.0) days for patients recruited by letter.

**Conclusions:** There are important differences in time taken to recruit both practices and/or patients according to geographical location of the practice, practice list size and deprivation score as well as the method of patient recruitment.

**A113**

FARSITE: evaluation of an automated trial feasibility assessment and recruitment tool

Sarah Thew1, Gary Leeming1, John Ainsworth2, Martin Gibson1, Iain Buchan2

1North-west e-Health, Manchester, UK; 2Community Based Medicine, University of Manchester, Manchester, UK; 3Greater Manchester Comprehensive Research Network, Manchester, UK

**Trials 2011, 12(Suppl 1):A113**

**Objectives:** In a review of UK-supported clinical trials more than half of the investigators asked the funding agency for an extension and a third did not hit their recruitment targets [1]. Study feasibility is often assessed on an ad hoc basis by asking clinical staff to estimate how many patients with particular characteristics they might expect to see in a given time period, and over-estimation is common. We have developed FARSITE (Feasibility Assessment and Recruitment System for Improving Trial Efficiency), a system to support the evaluation of trial feasibility by providing accurate assessments of numbers of patients eligible for a particular trial. Furthermore, FARSITE automates patient recruitment whilst preserving consent for consent.

**Methods:** Previously we investigated the requirements of clinical and protocol development experts for a system to support trial design and recruitment. We developed a software architecture which addresses their primary concerns: preservation of consent for consent; improving the efficiency of the trial design process and automation of as much as possible of the recruitment workflow [2]. The FARSITE software is based on this architecture and has been deployed in collaboration in the Greater Manchester Clinical Research Network and the NHS in Salford. We present the results of an analysis of the benefits of using the FARSITE application when compared to previous modes of working.

**Results:** We have worked with local clinical research organisations to evaluate FARSITE, running recruitment criteria for on-going clinical trials through FARSITE and comparing our estimates of numbers of patients eligible for the trial with the trials’ actual recruitment rates. A strong correlation was observed between protocols with a low FARSITE recruitment treatment and trials struggling to recruit participants.

**Conclusions:** We have shown that FARSITE can improve the speed and efficiency of clinical trials feasibility, allowing researchers to quickly assess the feasibility of trials in advance.

**References**


**A114**

Exploring meaning of participation in a clinical trial in a developing country setting: implications for recruitment

Joanna Reynolds1,2, Peter Mangesho1, Lasse S Vestergaard1, Clare Chandler1

1London School of Hygiene & Tropical Medicine, UK; 2National Institute for Medical Research, Amrani Centre, Tanzania; 3University of Copenhagen, Denmark

**Trials 2011, 12(Suppl 1):A114**

**Objectives:** This study aimed to explore the experiences of people participating in a clinical trial in Tanzania. We sought to understand the meaning attached to participation and how experiences of being in the trial related to participants’ original motivations for consenting, in order to explore appropriate strategies for recruitment in a developing country setting.

**Methods:** We designed a qualitative study alongside a clinical observational trial of the efficacy and safety of artesinisin-based combination therapy (ACT) for malaria in patients concomitantly receiving antiretroviral therapy (ART) for HIV in Muheza, Tanzania. Focus-group discussions have been held with HIV-positive and HIV-negative people who have participated in the trial, and with HIV-positive people who were screened but who did not participate. This data has been triangulated with in-depth interviews with staff conducting the trial and delivering HIV care at the hospital where the trial was conducted. Data is being analysed using an iterative, line-by-line approach based on the principles of grounded theory, to identify units of meaning and develop themes and constructs from the data.

**Results:** Analysis of data of eight FGDS and IDIs indicates a disconnect between the information given to trial participants in the recruitment and consent process and their understanding of the trial and its aims, with a few participants stating they did not realise they were part of a research study. This reflects, and may be attributable to trial participants’ frequent conflation of the clinical encounter - testing for malaria - with the research encounter - the recruitment process - and an inability to distinguish between these as separate events. When describing recruitment, many participants framed their narrative around clinical events such as the malaria test and seeking treatment, the latter being considered the most important reason for joining the study. The clinical context in which participants were screened and recruited appeared to influence their ability to interpret information about the trial and expectations for what may happen to them, raising questions about the nature of ‘informed consent’ in the recruitment process. Participants reported overwhelmingly positive experiences of participating in the trial, largely based around their access to numerous tests and free treatment, as well as reimbursement for transport and telephone costs. Being part of the trial was frequently conceptualised as receiving a ‘service’, valued chiefly for enabling participants to be observed and to know their health status. The perceived value of this ‘service’ was reflected in many participants’ reports of encouraging friends and family to attend for malaria testing, in order to access the service associated with the trial. Although this indicates again some confusion between clinical and
research activities, it appeared that the ‘enactment’ of the trial – giving and receiving the service – offered the social space in which participants were likely to raise any concerns about aspects of the trial. Such concerns included questioning the ‘true’ aim of the study, and fears over blood taking. This suggests that it is within a relationship of interaction with trial staff and activities that comprehension about the meaning and value of participation can begin to emerge, thus highlighting the potential shortcomings of the standard recruitment process.

Conclusions: The findings from this study raise questions about how information is presented and used in the recruitment and consenting process for a trial, particularly one located within an existing clinical context. Recruitment strategies should take into consideration where, when and how information will be conveyed to participants, and explore likely expectations of these contexts which may shape how and if information is interpreted and utilised by potential participants. The findings also indicate that the meaning participants place on participation, its risks and benefits, develops largely through enactment of the relationship with trial staff and activities, rather than through the consent process. To ensure better informed consent, researchers could seek to activate this relationship earlier on in the trial process, for example through meaningful community engagement activities prior to recruitment.

This study has also underscored the value participants in this low-resource setting places on having a ‘service’ of health care provision, wherein they perceive their health status and needs to be the focus of trial activities. This service is an influential motivator for continuing to engage with a trial, beyond more material compensations such as money for transport. These findings also suggest the potential for word-of-mouth strategies to aid informed recruitment in contexts where the perceived benefits of participation are highly valued amongst the local community.

Recruitment and continued participation in trials involve logistical and ethical challenges. Understanding values and concerns of participants can pose practical solutions.

A115 Improving recruitment to clinical trials with a register of a million patients

Objectives: The aim of the study was to explore the views of parents and clinicians of their recruitment experiences to the Scottish Health Research Register (SHARE) in order to consider the feasibility and potential impact of extending this resource to other Scottish clinical research. The study sought to understand and address the factors contributing to the recruitment shortfall in the wider population and to develop a suitable recruitment strategy for Scottish clinical research.

Background: The SHARE register was developed to address the recruitment shortfall in clinical research in Scotland. The register was set up in 2011 and aims to recruit a million patients from the general population. The register is open to all researchers and has been used in a number of clinical research studies. The register is intended to be a valuable resource for researchers in Scotland and is supported by the Scottish Health Research Register (SHARE) and the Scottish Health Research Register (SHARE) 2011-2016.

Methods: The recruitment strategies used in this study were based on the principles of community engagement and patient-centred care. The strategies were designed to ensure that patients were informed about the potential benefits of the trial and were given the opportunity to participate.

Results: The results of the study showed that the recruitment strategies used in this study were effective in increasing the number of patients who agreed to participate in the trial. The strategies included the use of social media, the development of a patient information website, and the provision of patient support groups.

Conclusions: The results of this study indicate that the recruitment strategies used in this study were effective in increasing the number of patients who agreed to participate in the trial. The strategies should be adopted by other researchers in Scotland as they are likely to be effective in increasing the number of patients who agree to participate in clinical research.
Conclusions: The concerns of some practitioners, that parents would be overburdened, stood in sharp contrast to the perspectives of parents. Contrary to what practitioners often expected, parents were positive about being approached to enter their child into a clinical trial. Helping practitioners to understand how families perceive clinical trials and providing them with ‘moral’ support in approaching families may enhance recruitment to children’s clinical trials. This strategy may also potentially benefit recruitment in trials with other vulnerable patient groups [1].


RECRUITMENT STRATEGIES IN TRIALS

A117
Recruiting patients cost-effectively by mail

Introduction: Large randomised trials have been successfully conducted using mailed drug supply and follow-up [1-4]. ASCEND (A Study of Cardiovascular Events In Diabetes) is a randomised “2x2 factorial design” study of aspirin versus placebo and of omega-3 fatty acid versus placebo, for the primary prevention of cardiovascular events in people with diabetes. In order to be able to study 15,000 people with diabetes for about 7 years at low cost, ASCEND is streamlined and run mainly by mail with back-up from a 24-hour Freefone service.

Methods: In collaboration with consultants around the UK, potentially eligible people with diabetes have been identified from various sources including: centrally-held registers (e.g.: Retinopathy screening registers); GP-held local registers and self-referral via a website. For patients identified from centrally-held registers, permission was gained from the National Information Governance Board to allow centrally generated letters in the name of the holder of the register, to be sent to patients listed on registers. In addition, with the collaboration of the Diabetes Research Network and the Primary Care Research Network, 730 general practices agreed to send pre-assembled invitation packs to people on their locally held registers. To facilitate recruitment, the design is straightforward with simple inclusion and exclusion criteria and treatment packages for easy mailing. Double sided A3 forms are used for screening, randomisation and follow-up. On the Screening form, patients confirm their eligibility and consent that they are happy to take part by ticking a series of boxes. If they have any questions then they can telephone study staff via a Freephone number. Screening forms of those entering the study are checked centrally by clinical staff. Patients enter a 2-month pre-randomisation run-in phase and are randomised if they complete a randomisation form and remain willing and eligible.

Results: In total, 423,286 people with diabetes were invited to take part and 15,481 randomised (see table).

Conclusion: If sufficient numbers of potentially eligible patients can be identified centrally and trial treatments require little in the way of monitoring, the recruitment and follow-up of patients in clinical trials by mail is feasible and cost-effective. Wider use of these methods could allow more large randomised trials to be undertaken successfully and cost-effectively.


Table 1 (abstract A117)

<table>
<thead>
<tr>
<th>Centrally-held register</th>
<th>GP practices (Local register)</th>
<th>Others*</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invitations sent</td>
<td>300,188</td>
<td>120,875</td>
<td>2223</td>
</tr>
<tr>
<td>Patients enter-run-in</td>
<td>16104</td>
<td>9,741</td>
<td>635</td>
</tr>
<tr>
<td>Patients randomised</td>
<td>9013 (3.0%)</td>
<td>6038 (5.0%)</td>
<td>430 (19.3%)</td>
</tr>
</tbody>
</table>

* HPS follow up/Self/friends/Hospital referral.

A118
Analysis and validation of a Parkinson’s disease register as a recruitment tool for clinical studies
Camille B Carroll1, Amy Palmer2, Christine Cosby2, John P Zajicek3
1Peninsula College of Medicine and Dentistry, Dept Clinical Neurobiology, University of Plymouth, John Bull Building, Tamar Science Park, Plymouth, Devon, PL6 8BU, UK; 2South West Dementias and Neurodegenerative Diseases Research Network, ITTC Building, Tamar Science Park, Plymouth, Devon, PL6 BX8, UK; 3Peninsula College of Medicine and Dentistry, University of Plymouth, Dept Clinical Neuroscience, ITTC Building, Tamar Science Park, Plymouth, Devon, PL6 BX8, UK. E-mail: camille.carroll@pms.ac.uk
Trials 2011, 12(Suppl 1):A118

Background: Many patients with Parkinson’s disease (PD) are not afforded the opportunity to participate in clinical studies. A register of research-interested patients could improve involvement. We have established a register of research-interested PD patients within the South West of England, with pragmatic inclusion criteria and multiple routes of recruitment.

Purpose: To determine whether a register of PD patients interested in research could be established in a resource-efficient manner, and whether in comparison with traditional recruitment methods, the register would provide more representative patient cohort and facilitate rapid and inclusive recruitment to clinical studies.

Methods: We undertook a comprehensive analysis of the register three years after it was initiated, including documentation of the pitfalls and benefits of its establishment, investigation of its utility as a recruitment tool and a survey of recruiters.

Results: There were 529 active participants (M:F = 1.6:1) (589 recruits, 60 withdrawn): mean age 71.4 yrs; mean disease duration 8.8 yrs from symptom onset, 7.2 yrs from diagnosis. 30% of register participants were self-referred; 70% were recruited by a healthcare practitioner. Local factors such as the availability of research support staff influenced recruitment. Response rate to annual questionnaires was 86.5%. There was a self-reported PD diagnosis rate of 92% at baseline, 88% at month 12 and 83% at month 24. Total staff time required for pack preparation, recruitment and data entry was 15 minutes for each new recruit, and 5 minutes for each follow-up questionnaire. 85% of recruiters felt the register was a useful means of facilitating research and providing data for planning of service provision. In their feasibility study, a single mailing to participants resulted in a final recruitment rate that was double that achieved by traditional face-to-face recruitment.

Limitations: Despite 30% of participants self-refering to the register, all patients on the register were being seen in secondary care, either by a neurologist (40%) or a geriatrician with a specialist interest in movement disorders (60%). We have therefore yet to demonstrate access to a population that could not be accessed by traditional secondary-care-based recruitment methods, and will be targeting register recruitment specifically in primary care to address this.

Conclusions: We have established a register of research-interested PD patients in a resource-efficient and pragmatic manner, which has the potential to maximise inclusivity and clinical research opportunities.

Acknowledgements: This work was supported by the Dementias and Neurodegenerative Diseases Network.

A119
Recruiting patients with advanced malignant and non-malignant disease: lessons learned from a palliative care RCT
Morag C Farquhar1,2, Barbara Braffman-Kennedy1, Irene J Higginson3, Sara Booth1
1Department of Public Health & Primary Care, University of Cambridge, Cambridge CB2 0SR, UK; 2Palliative Care, Cambridge University Hospitals NHS Foundation Trust, Cambridge, CB2 0QQ, UK; 3Department of Palliative Care, Policy & Rehabilitation, King’s College, London, London, SES 9PJ, UK
Trials 2011, 12(Suppl 1):A119

Objectives: Recruiting patients to palliative care randomised controlled trials (RCTs) is particularly challenging. This paper describes and analyses the differing recruitment trajectories for patients with advanced malignant and non-malignant disease to a palliative care RCT, outlining activities undertaken to achieve targets. It will outline the lessons learned in order to inform design and conduct of future studies.

Methods: Analysis of descriptive recruitment statistics (patient identification, recruitment and completion rates) to a Phase II pilot pragmatic single-blind fast track RCT, and subsequent Phase III RCT, of a breathlessness intervention service for advanced disease. Phase II piloted with chronic obstructive pulmonary disease (COPD) patients only, whereas the Phase III RCT incorporated two sub-protocols: one for patients with malignant and one for non-malignant disease. Documentary analysis of: recruitment activity log, Trial Management and Advisory Group minutes and field notes.

Results: Recruitment targets for patients with non-malignant disease were achieved. The pathway to recruitment was through referral to the service therefore referral rates impacted on recruitment alongside response rates. Recruitment of cancer patients was considerably slower despite concerted efforts to increase referrals by raising the service profile. Referrals only improved for the latter when a researcher attended clinics, supporting clinical staff in patient identification: recruitment tripled from 0.8 to 2.4 patients per month. Three possible reasons for the effectiveness of this are (1) dedicated time, (2) reciprocity & (3) established relationships. Predictably, response rates remained lower for patients with malignant disease than for those with non-malignant disease.

Conclusions: Recruitment was partly referral-driven, therefore gate-keeping did not explain the differences. Clinical inter-professional relationships consolidated in Phases 0-II drove early referrals of non-malignant disease patients. Local palliative care services were available for patients with cancer. Consideration of the natural history and context of a service is therefore important when predicting recruitment. Pilot trials are informative, but should include qualitative elements and all disease groups. Placing researchers in relevant clinical settings is helpful.

A120
Recruitment to the pilot phase of Start2quit – lessons learned
Leanne M Gardner, Hazel Gilbert, Irwin Nazareth
Research Department of Primary Care and Population Health, University College London, London, London, UK.
Trials 2011, 12(Suppl 1):A120

Background: The recruitment of smokers to trials investigating smoking cessation is often difficult and participation rates are low. Programs using reactive recruitment methods reply upon smokers contacting the program in response to an advertised service or treatment. Generally only smokers motivated to quit will respond and thus participation rates for reactive recruitment are less than 5% of the target population. Conversely proactive recruitment, where a service is offered to potential participants directly, can recruit larger proportions and a more diverse range of population. Start2quit, a randomised controlled trial aimed at increasing the use of NHS Stop Smoking Services, will utilize proactive recruitment. We explored response rates and the effects of sending reminder questionnaires in the pilot phase of this study. Furthermore, we compared recruitment in a London-based PCT versus an outside-London-based PCT.

Methods: In the pilot phase of Start2quit, GP practices were identified through the local Primary Care Research Network in Camden and Oxfordshire. Current smokers aged 16 and over were identified from their medical records and sent an invitation to participate in the study and a baseline questionnaire. Replies were processed three weeks later. Those not responding were sent a reminder invitation letter and questionnaire. Of those responding, eligible smokers were randomized to one of the intervention groups. Replies were processed a further two times at three and six weeks following mail out of the reminder. The main outcome measure was the recruitment rate, i.e. the proportion of eligible participants.

Results: For the pilot phase, three GP practices from Camden and four GP practices from Oxfordshire were selected for the recruitment of smokers. There was variability in the response rate between GP practices (2.7% to 9.4%), however rates increased in all practices between the first invitation letter and the reminder invitation. The average response rate in Camden (3.2%; London-based PCT) was lower than in Oxfordshire (6.8%; outside-London PCT). An overall a response rate of 5.5% was obtained over the two PCTs.

Conclusions: We found in the pilot phase of Start2quit that practice location and sending reminder invitation letters and questionnaires can influence recruitment rates. Response rates in London were found to be lower in general and thus other strategies may be required to increase participation in this area. The proactive method used in Start2quit to recruit smokers, identified from their computerised medical records in GP practices, is effective at inviting participants to smoking cessation trials.

A121
Analysis of patient information leaflets (PILs), used in clinical trials using the Informed Consent Evaluation instrument (ICEI)
Katie Gilleece1,2, Wan Huang1, Zoe Skea1, Seonaidh Cotton1
1Health Services Research Unit, University of Aberdeen, Aberdeen, AB25 2ZB, UK; 2Academic Urology Unit, University of Aberdeen, Aberdeen, AB25 2ZB, UK
Trials 2011, 12(Suppl 1):A121

Background: Informed consent (IC) is regarded as a cornerstone of ethical healthcare research and is a requirement for most clinical research studies [1]. Guidelines suggest that prospective clinical trial participants should understand a basic amount of information about trials in order to provide valid IC. However, poor participant understanding of the research processes, a lack of knowledge about the expectations and demands of trials and insufficient support when faced with the decision has been demonstrated across a range of clinical areas [2,3]. As such, the existing approach to obtaining IC for clinical trials is not optimal. Therefore, it is important to investigate the effectiveness of the current consent process to examine how well information PILs conform to empirically derived standards for promoting high quality decisions. We propose that the process could be improved by drawing on existing research in the fields of decision making and decision support interventions.

Methods: Websites of Clinical Trial Units registered with the UK Clinical Research Collaboration were screened to identify open access PILs from ongoing or recently completed randomised controlled trials (RCTs). A total of 60 PILs were identified and 20 were purposively sampled, for analysis using the Informed Consent Evaluation instrument (ICEI). The ICEI was developed by combining informed consent guidelines and the International Patient Decision Aid Standards (IPDAS), which describe detailed recommendations about the content and delivery of information to facilitate high quality decisions for treatment or screening. Two independent raters scored PILs according to the items presented in the ICEI; any disagreement was resolved by a third rater.

Results: Variation existed amongst the PILs in terms of overall scores and scores for specific items. Some aspects were consistently poor across all PILs analysed, namely: presenting probabilities; clarifying and expressing values; and structured guidance in deliberation and communication. Better informed decisions about participation may result in patients being retained throughout the duration of the trial, as their decisions will be
linked to more realistic expectations and be more in line with their personal values and preferences.

References

A122
The complexities of maximising recruitment to complex intervention trials in child and adolescent mental health services
David Cottrell1, Liz Graham1, Amanda Farrin2
1School of Medicine, University of Leeds, Leeds, LS2 9NL, UK; 2Clinical Trials Research Unit, University of Leeds, Leeds, LS2 9JT, UK

Trials 2011, 12(Suppl 1):A122

Objectives: The HTA-funded SHIFT Trial (Self-Harm Intervention: Family Therapy) aims to recruit 832 young people following self-harm, along with their primary care givers. Recruitment is scheduled to take place over 3 years, across three ‘Hubs’ in the UK – Greater Manchester, London and Yorkshire – each involving approximately 15 Child & Adolescent Mental Health Services (CAMHS). Due to the complexities of set-up and the involvement of many clinicians in largely research-naïve services, recruitment has been slower than anticipated. The TMG thus agreed an intensive strategy to enhance the existing recruitment plan.

Methods: The key approach to immediate improvement in recruitment rates was intensive work with existing services. The common approach of involving additional sites was not a straightforward option, given that trial-specific Family Therapists (FTs) work across services and form a team for rotational delivery of FT in each service. Inclusion of new sites would involve either an increase in FT workload (which may or may not be possible should existing services increase recruitment), or development of an entirely new group of services involving lengthy and costly employment of additional Therapists, plus the governance approvals required for each Trust.

The main elements of this approach are a) passing on guidance regarding strategies employed in services already recruiting well to other CAMHS, b) a systematic, site-by-site approach with those identified as poor recruiters – this involves phone calls from the Hub Leads to each PI, increased Researcher contact, identification of SHIFT ‘Champions’ in each service location to ensure SHIFT is flagged at all relevant team meetings, c) MHRN C50 attendance at referral meetings to specifically flag eligible cases.

Results: This approach was initiated at the end of February 2011. Sites have responded positively and have been keen to identify champions and look at strategies to ensure SHIFT is flagged as part of routine practice. To date (August 2011) we have seen a promising increase in recruitment.

Conclusions: Strategies to ensure optimal recruitment in CAMHS require good local knowledge of the varying ways in which services operate to ensure a tailored approach to recruitment. Identification of Champions is proving successful. Increased Researcher and MHRN contact with sites, attendance at relevant team meetings, and maintenance of trial profile within each service is key.

Acknowledgements: The authors are presenting on behalf of the SHIFT trial team. The SHIFT Trial is supported by HTA-470772.

A123
Can we improve recruitment to trials and informed consent by improving participant information sheets? - A nested RCT
Peter Knapp1, Natasha Mitchell2, David K Raynor3, Jonathan Silcock4, Brian Parkinson3, Janet Holt4, Yvonne Birks5, Simon Gilbody1
1Department of Health Sciences, University of York, York, UK; 2School of Healthcare, University of Leeds, Leeds, UK; 3School of Pharmacy, University of Bradford, Bradford, UK; 4Making Sense Design, Sheffield, UK

Trials 2011, 12(Suppl 1):A123

Objectives: - Develop a more readable participant information sheet (PIS) for an ongoing trial (CASPER) through re-writing, graphic design and user testing; and - Assess the impact of the enhanced sheet on trial recruitment and informed consent.

Methods: CASPER is a UK NIHR-funded trial of Collaborative Care to prevent depression in adults aged 75+. During the CASPER set-up phase we produced an information sheet compliant with National Research Ethics Service (NRES) guidance (the ‘standard sheet’). We decided to try to enhance recruitment by improving the PIS (the ‘enhanced sheet’). The nested study comprises: first, development and testing of an ‘enhanced’ PIS; second, comparison of the ‘standard’ and ‘enhanced’ versions in a nested RCT.

The comprehensibility of the NRES-approved standard sheet was assessed by asking cohorts of 10 older adults to take part in individual user testing interviews. After reading the sheet, participants were asked to find answers in the sheet to 18 factual questions and show their understanding of found information. The sheet was then re-written, re-structured and re-designed, drawing on graphics and writing expertise and the user testing data. The ‘enhanced sheet’ was user-tested on further cohorts of 10 people and amended according to the data obtained.

In a nested RCT, potential CASPER participants will be posted either the ‘standard’ or ‘enhanced’ sheet, to assess effects on interest in participation and recruitment. Responders will be asked to complete an abbreviated version of the Joffe Quality of Informed Consent measure [1]. We will also conduct 2 focus groups with participants to explore the role of the sheet in their decision to participate (or not).

Results: Despite it being approved by NRES, testing of the standard CASPER trial sheet revealed limitations in document organisation and writing, resulting in difficulty understanding such issues as trial benefits, sources of patient data and trial withdrawal. Revision included the use of lay language and short sentences; new font, layout and sub-headings; document re-organisation including adding a contents list and summary section. Testing of the ‘enhanced sheet’ on 30 people showed significant improvements in finding and understanding of information. During 2011-12 we will assess the impact of the two versions of the sheet on recruitment and consent in our trial of older people with low severity depression.

Conclusions: User testing, expert re-writing and graphic design produced a more readable trial information sheet. To our knowledge this will be the first randomised evaluation of an enhanced trial sheet remaining NRES-compliant. These results will interest others concerned with the improvement of information for trial participants.

Reference

A124
Addressing patient treatment preferences at trial recruitment
Nicola Mills1, Julia Wade1, Athene J Lane1, Freddie C Hamdy2, David E Neal3, Jenny L Donovan4
1School of Social and Community Medicine, University of Bristol, Bristol, UK; 2Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK; 3Department of Oncology, University of Cambridge, Cambridge, UK

Trials 2011, 12(Suppl 1):A124

Background: Patient recruitment is one of the main challenges in conducting randomised controlled trials (RCTs). Patients’ treatment preferences are viewed as a barrier to RCT recruitment yet there is little research to understand them. This study explored the expression of preferences by potential trial recruits, the response to them by recruiters, and their influence on trial participation decisions during trial recruitment appointments.

Methods: We undertook an analysis of audio recordings of consecutive recruitment appointments to a UK multi-centre RCT of three different treatments for prostate cancer (the ProtecT - Prostate cancer testing and Treatment - study) over a three month period. 93/108 appointments with
men aged 51-70 years were recorded successfully and analysed using techniques of content and thematic analysis.

Results: Most potential participants expressed a desire for a particular treatment early in appointments, with their desires ranging on a continuum from hesitant to well-formed opinions. Recruiters explored these initial treatment views in the context of evidence-based treatment and study information which resulted in many men becoming uncertain about their initial views and open to RCT recruitment, often accepting a different treatment from their original ‘preference’. Only a quarter of men who initially expressed a wish for a particular treatment sustained or developed a clear treatment preference as the consultation proceeded and ultimately received this treatment. In most of these cases the recruiters established the rationale for the preference then provided specific information that counter-balanced their reasoning by emphasising the position of clinical equipoise, uncertainty of the prognosis and the pros and cons of their desired and non desired treatment. This counter-balancing of information continued until they were sufficiently satisfied that the man was making a fully informed decision.

Conclusions: Many potential trial recruits will present initially with a preference and study information which resulted in many men becoming uncertain in the baseline characteristics of patients recruited via DR. The majority of participants (72%) were recruited via DR. The response from attendees at the meeting was very positive. Recruitment to randomised controlled trials (RCTs) is a known problem, with many failing to reach recruitment targets [1]. RCTs involving participants with mental health problems often struggle to recruit. This is a particular problem in primary care [2]. The multi-centre REEACT (Randomised Evaluation of the Effectiveness and Acceptability of Computerised Therapy) trial led by the University of York [SIRC199147481; http://www.reeact.org.uk] recently completed recruitment ahead of target. It recruited primary care patients with depression using two strategies - database screening (DS) to identify potential eligible recruits to target, and traditional GP direct referral (DR) from face-to-face consultations. These strategies were used in combination in the hope of expediting recruitment, which was achieved. With recruitment complete, we examined this recruitment strategy in more detail.

Materials and methods: We tabulated the overall contributions of each method of recruitment to the overall number of participants and examined the trend in recruitment over the course of the trial. We checked for statistically significant differences in the baseline characteristics of participants recruited via each method. In order to see if there was regional variation in use of recruitment methods, we compared the number of participants recruited from each method by study site. The conversion rate to trial participants for those patients identified through each method was also compared.

Results: The majority of participants (72%) were recruited via DR. The participants recruited through DS were older on average, and had a higher probability of having had a previous episode of depression. The proportion of participants entering the trial via each method was consistent with the overall recruitment figures across all sites except York, where the contribution from DS was slightly higher. The proportion of participants entering the trial through each referral method remained consistent from about a year before the end of recruitment. A higher proportion of DRs assessed for inclusion converted into participants and a lower proportion were ineligible as compared to those identified via DS.

Conclusions: The pragmatic design of the REEACT trial resulted in target recruitment ahead of schedule. A detailed examination of the re-cruitment trend suggests that DR was a more effective method of recruitment, although the use of DS has been a favoured tool in primary care trials. The findings from the REEACT suggest that DRs may be a better strategy when recruiting patients with depression in the primary care setting.

Acknowledgments: This abstract has been prepared on behalf of the REEACT trial researchers. The REEACT trial is funded by the NIHR Health Technology Assessment programme (project number 06/43/05). The views and opinions expressed herein are those of the authors and do not necessarily reflect those of the HTA programme, NIHR, NHS or the Department of Health.

References
A127
Recruiting and retaining medical patients in trials of complex interventions: practical experience of challenges and solutions in three multi-centre trials
Jane Walker1*, Michael Sharpe2
1Psychological Medicine Research, Department of Psychiatry, University of Edinburgh, Edinburgh, EH10 5HF, UK; 2Psychological Medicine Research, Department of Psychiatry, University of Oxford, OX3 7IX, Oxford, UK
E-mail: jane.walker@ed.ac.uk
Trials 2011, 12(Suppl 1):A127

Objectives: 1. To consider the challenges of recruiting and retaining medical patients in trials of complex interventions, that include ‘psychological’ components. 2. To discuss practical solutions to these challenges that ensure trials recruit to target and maximise outcome data collection.

Examples are taken from three trials of complex interventions: SMART Oncology-2, SMART Oncology-3 and SMART Neurology-1.

Methods: Challenges faced in the recruitment to three trials included: identifying eligible patients, achieving ‘buy-in’ from local clinicians, addressing the stigma associated with ‘psychological’ interventions and describing novel interventions to patients. Solutions included: identifying potential participants through systematic screening, minimising burden on clinicians, embedding the trial team in services and novel approaches to delivering trial information to patients. Challenges of retaining patients to ensure outcome data collection included: trial intervention withdrawals, patients’ physical deterioration and the collection of data at multiple time-points. Solutions included: ensuring a positive relationship between patients and the trial team, building relationships with local services and using financial incentives for patients.

Practical examples of each of these solutions are taken from the following trials: SMART Oncology-2 is a 500 patient cost-effectiveness trial comparing a complex intervention for depression in cancer patients to usual care, with the primary outcome at six months. SMART Oncology-3 is a 150 patient efficacy trial evaluating a similar intervention for depression in patients with lung cancer, who have limited life expectancy, with data collected over eight months to give an average depression score for the primary outcome. SMART Neurology-1 is a 130 patient efficacy trial of a guided self-help intervention for patients with medically unexplained symptoms that ensure trials recruit to target and maximise outcome data collection.

Conclusion: Recruiting and retaining medical patients in trials of complex ‘psychological’ interventions poses special challenges. Innovative practical solutions can ensure recruitment to target and collection of adequate outcome data.

RETENTION AND ADHERENCE

A128
Strategies to reduce attrition in randomised trials
Valerie Brueton1*, Jayne Tierney2, Sally Stenning3, Irwin Nazareth4, Sarah Meredith5, Seeromanie Harding4, Greta Rait4, Jane Glenton6
1MRC General Practice Research Framework, London, WC2B 6NH, UK; 2UCL Institute of Vaccine Preventable Diseases, London, NW3 2PF, UK; 3MRC Clinical Trials Unit, London, WC2B 6NH, UK; 4MRC Social & Public Health Sciences Unit, Glasgow, G12 8ZD, UK; 5Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, L69 3GS, UK; 6MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK
E-mail: s.r.dodd@liv.ac.uk
Trials 2011, 12(Suppl 1):A128

Background: Attrition from randomised trials can introduce bias and reduce study power affecting the generalisability, validity, and reliability of results [1]. Many strategies are used by trialists to reduce attrition, including motivating and engaging participants and sites to optimise data return or compliance to follow-up procedures [2].

Objective: To quantify the effect of strategies to reduce attrition from randomised trials in any healthcare setting.

Methods: Included studies were randomised evaluations of strategies to reduce attrition embedded within randomised trials from all disease areas and settings. The following sources were searched for eligible studies [3]: MEDLINE (1950 to present), EMBASE (1980 to present), PsycINFO (1806 to present), DARE (most recent issue), CENTRAL (most recent issue), MR000032.

Results: From 19,281 abstracts 31 unique RCTs were identified from the following sources: MEDLINE, CENTRAL, MR000032 n=9; SCT abstracts 1980-2010 n=4; reference lists of relevant reviews n=7; and of included trials n=8 (7 duplicates); word of mouth n=4; and CTUs survey n=6. Six types of strategies to reduce attrition were identified: a) communication e.g. email, letters signed by different study personnel, type of post, and delivery method; b) questionnaire length i.e. short versus long; c) incentives i.e. monetary incentives, offers of monetary incentives or vouchers, and gifts; d) case management i.e. trial assistants assigned to manage participant follow-up; e) behavioural e.g. workshops giving participants information about goal setting; and f) methodological interventions e.g. blinded versus open trials. Final results of the review will be presented.

References

A129
Departure from treatment protocol in published randomised controlled trials: a review
Susanna Dodd1, Ian White2, Paula Williamson1
1Department of Biostatistics, Institute of Translational Medicine, University of Liverpool, L69 3GS, UK; 2MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 0SR, UK
E-mail: s.r.dodd@liv.ac.uk
Trials 2011, 12(Suppl 1):A129

Objectives: This review aimed to ascertain the extent to which the issue of departure from treatment protocol (DTP) is reported and addressed in published analyses of RCTs.

Methods: One hundred publications of RCTs were randomly selected from those published in the BMJ, NEJM, JAMA and Lancet during 2008. Each trial report was reviewed to determine the extent and nature of reporting on DTP and whether statistical methods were used to deal with DTP for both benefit and harms analyses.

Results: Even the most basic adherence information was not presented in some trials. Forty-two publications did not state how many patients actually received the primary endpoint of attrition. Ninety-eight publications reported at least one form of DTP, including non-receipt of allocated treatment (39 trials), incomplete treatment in those who initiated allocated treatment (78), switching trial treatments (12), treatment disallowed/non-trial treatment (42), starting open label treatment out of trial (7), contamination across groups (3) and other nonadherence to treatment dose or schedule (23). Treatment providers were reported to be nonadherent when delivering treatment in 12 trials. More than half (50 (51%)) of the publications that reported DTP used some method to deal with it, but none were based on randomisation-preserving techniques. The most common method was based on per protocol (PP).
analysis (46) (including one instance of using inverse probability of censoring weighting) often labelled as intention to treat (ITT) (18) or modified ITT (5), but missing data techniques (2) and as treated analyses (3) were also implemented. Less than 40% (26) of the 69 trials which presented harms analyses specifically defined harms analysis populations, and the majority of these definitions were based on actual treatment received (18). The majority (31) of the 43 trials that did not explicitly specify harms analysis population appeared to analysed harms outcomes using ITT. Twelve reports explicitly commented on the fact that DTP was likely to have influenced the observed treatment effect.

Conclusions: DTP data presented in RCT publications, particularly related to treatment initiation and discontinuation, may be ambiguous or scant. Trials often attempt to deal with DTP using variations of PP analysis, although they may be labelled as ITT; randomisation-preserving methods are not typically used. There appears to be confusion among trialsists over the appropriate analysis population for harms outcomes in the presence of DTP.

A130 The use of the Medication Event Monitoring System (MEMS) for assessing medication adherence for chronic conditions: use and results from a 12 month trial of patients in remission with ulcerative colitis (UC) David Gillespie1,2, Kerenza Hood1, Andrew Williams1, Rachel Stenson1, Christopher Probert1, Antony Hawthorne3
1South East Wales Trials Unit, School of Medicine, Cardiff University, UK; 2Gastroenterology, University Hospital of Wales, UK; 3School of Clinical Sciences, University of Bristol, UK E-mail: gillespie1@cardiff.ac.uk
Trials 2011, 12(Suppl 1):A130

Background: The Colitis Once Daily Asacol® (CODA) study assessed the efficacy and safety of once daily (OD) dosing with Mesalazine (Asacol®), 2.4g as 3 x 800mg tablets) versus one 800mg tablet three times daily (TDS) over a 12 month period for patients in remission with ulcerative colitis (UC). Emerging evidence suggests OD Mesalazine is as effective as divided doses [1]. While this has been attributed to better adherence, detailed measures of adherence have been lacking in previous studies. A sub study was run alongside in order to provide a more intense monitoring of adherence using the Medication Event Monitoring System (MEMS).

Objectives: To (i.) describe the use of the MEMS for a detailed assessment of medication adherence and (ii) present results from the CODA sub study, comparing adherence data collected using the MEMS with the methods used in the main trial.

Methods: The main CODA trial collected tablet counts at patient visits and asked questions on perceived adherence. The CODA sub study used the MEMS cap data, which electronically recorded the time and date of each cap opening. It was assumed that each cap opening represented a patient taking the correct medication from the bottle [2].

Results: A total of 58 patients had usable adherence data (49 with complete data (12 months or until relapse), 9 with partial data, 3 patients were withdrawn). The frequency of cap openings split by trial arm will be presented. The percentage of days adherent was significantly different between the two trial arms, with OD patients considerably more adherent than TDS patients. The impact of controlling for adherence on relapse rates will be presented. A comparison will be made between the MEMS adherence data and the adherence data obtained in the main trial. More detailed analysis of patient adherence; including (i.) weekday versus weekend adherence; (ii.) adherence around visit dates versus regular adherence and (iii) patterns of adherence over time will be considered.

Conclusions: Collecting adherence data electronically using products such as the MEMS provides an adequate representation of the complexities of patient adherence that may not be possible to obtain through other means (e.g. tablet counts and patient perception).

Acknowledgements: The trial was supported by an unrestricted health Research (NISCHR).

References

A131 Adherence in a randomised controlled trial comparing liberal and restrictive red blood cell (RBC) transfusion protocols after cardiac surgery (TITRe2) Katie Pike1, Rachel Brieler2, Chris A Rogers, Gavin J Murphy, Barney C Reeves
1Department of Gastroenterology, University Hospital of Wales, UK; 2South East Wales Trials Unit, University of Wales, UK; 3School of Clinical Sciences, University of Bristol, Bristol, BS2 8HW, UK
Trials 2011, 12(Suppl 1):A131

Objectives: The TITRe2 trial is comparing two haemoglobin (Hb) thresholds for RBC transfusion after cardiac surgery, Hb<9.0g/dl (liberal) vs. Hb<7.5g/dl (restrictive). Based on historic data, and with complete adherence, transfusion rates should be 100% in the liberal and 30% in the restrictive group. Convergence of these rates due to non-adherence severely threatens the power of the trial; there is also concern about differential non-adherence, with transfusion being delayed or withheld in the liberal group when Hb remains close to the 9.0g/dl threshold.

Methods: In order to capture non-adherence, research staff collect data describing:
- The lowest daily Hb;
- Date and time of each RBC transfusion and the preceding Hb measurement;
- Number of breaches of the allocated threshold before a transfusion is prescribed.

These data allow non-adherence to the randomisation and transfusion protocols to be detected:
- Failure to randomise when the 9.0g/dl threshold is breached;
- Randomised >24 hours after first breaching Hb 9.0g/dl threshold;
- Randomised without or before breaching Hb 9.0g/dl threshold;
- After randomisation, transfusion given when allocated threshold not breached (‘extra’), or transfusion withheld when allocated threshold breached (‘withheld’);

Instances of extra and withheld transfusions are classified as mild, moderate or severe depending on their likely influence on overall transfusion rates.

Results: 56% of participants are being randomised; about 8% of the remaining 44% consented participants breach the 9.0g/dl threshold but are not randomised. 3% of randomised participants are randomised >24 hours after first breaching, but none has been randomised without or before breaching the 9.0g/dl threshold. 32% of participants have had ≥1 instance of non-adherence to the transfusion protocol; in 6%, non-adherence was judged severe (extra – transfused and patient did not breach at any point post-randomisation; withheld – not transfused and patient had no post-randomisation transfusions). Site-specific rates of non-adherence are being fed back to try to improve adherence. Rates of transfusion in the liberal and restrictive groups are confidential to the Data Monitoring and Ethics Committee (DMEC).

Conclusions: We believe that this is the first attempt to measure withheld transfusions in trials of this kind. Data collection to do this is burdensome but satisfactory. The current rates of transfusion in the liberal and restrictive groups are, so far, judged by the DMEC to be consistent with the sample size justification.

A132 Complexities of retention in primary care randomised trials: A thematic analysis of in-depth interviews Valerie Brueton1, Fiona Stevenson2, Claire Vale3, Greta Rat1,2
1MRC General Practice Research Framework, London, WC2B 6NH, UK; 2UCL Research Department Primary Care and Population Health, London, NW3 2RP, UK; 3MRC Clinical Trials Unit, London, WC2B 6NH, UK E-mail: vb@gpf.mrc.ac.uk
Trials 2011, 12(Suppl 1):A132

Introduction: Loss to follow-up in randomised trials can cause bias, compromise study power, and affect the generalisability and reliability of results [1]. Many strategies are used to try to retain participants, however
little is known about trialists' experiences of implementing such strategies, or their perspectives of effectiveness. The complexity of these experiences and perceptions may influence the type of strategies used in different disease areas and with different population groups. We explored factors that trialists think contribute to loss to follow-up in primary care randomised trials, and whether some strategies to improve retention are perceived to be more successful than others.

Methods: 29 purposively sampled UK trialists including principal investigators n=10, research nurses n=9, and trial managers n=10 were invited for an in depth interview. Trialists were sampled from randomised trials conducted in UK primary care settings and published between 2000-2010. Randomised trials with high (>20%), moderate (5-20%) and low (<5%) rates of attrition were included in the sampling frame [2]. In-depth interviews were digitally recorded, transcribed verbatim, and anonymised. Concurrent thematic analysis was conducted. ATLAS ti 6.1 was used to organise and explore coded transcripts. Themes around each category were verified and confirmed by constant comparison and searching across all interviews for similar themes and categories for analysis.

Results: 29 in-depth interviews were conducted with 10 principal investigators, 10 trial managers, and 9 research nurses from primary care randomised trials in mental health, nutrition, elderly care, and chronic diseases. A major theme emerging across all interviews is the importance of communication between participant and trialist. Factors thought to contribute to retention include: rapport between participant and trialist, participant altruism, and flexibility around appointment schedules. Giving information about what the trial involves at the initial recruitment visit was considered to influence retention. Reducing burdens, both financial and physical, by provision of transport and reimbursement of costs were also considered useful.

Conclusions: The findings provide a deeper understanding of the complexity of retention in randomised trials and may inform trialists' choice of potentially effective strategies in future trials. In combination with an ongoing systematic review of randomised trials of retention strategies, we will highlight strategies or combinations of strategies that should be evaluated prospectively.

References:

A133
Maximising adherence to study protocol within pharmaco-rehabilitation clinical trials
Suzanne Hartley1, Sharon Ruddock2, Bipin Bhakta2, John Pearm2, Lorna Barnard1, Alison Ferguson1, Dean Langan1, Amanda Farry1
1Clinical Trials Research Unit, University of Leeds, Leeds, LS2 9JT, UK; 2Academic Department of Rehabilitation Medicine, University of Leeds, LS2 9JT, UK
E-mail: s.hartley@leeds.ac.uk
Trials 2011, 12(Suppl 1):A133

Background: The Dopamine Augmented Rehabilitation in Stroke (DARS) trial is a double-blind placebo controlled trial investigating impact of co-careldopa/placebo in combination with routine NHS occupational/physical therapy on functional outcome in people with acute stroke. The trial involves participants taking Investigational Medicinal Product (IMP)/placebo prior to therapy sessions while in the acute stroke unit and following hospital discharge. Stroke survivors may have significant residual impairments such as weakness, aphasia, visual disturbance, cognitive problems and mood disorders which may affect their ability to comply with DARS medication/therapy schedule.

Objectives: To identify issues in adherence and retention to medication/therapy schedule and implement processes to maximise compliance across hospital and community settings.

Methods: Key aspects of compliance with the medication/therapy schedule included (a) ensuring IMP packaging was clear and useable by people with hemiparesis; (b) IMP was taken at the correct time in relation to therapy intervention.

Packaging & labelling: We involved 19 stroke survivors in small group discussions about different aspects of IMP labelling and packaging. Different examples of IMP packaging (developed using NPSA guidance) were presented and preferences/opinions were obtained through standardised questionnaires.

Compliance with therapy / medication schedule: We developed a DVD for the participant/carers to view in their home environment to provide an audio visual aid to supplement the Patient Information Sheet. The DVD included voice over explaining trial processes in particular the therapy/IMP schedule, safety issues and contact details. IMP/therapy schedule compliance was also discussed with community therapists.

Results: The patient feedback was incorporated into IMP packaging to allow one handed opening and prompts for adherence to IMP schedule. The DVD content is presented in a manner accessible to patients with aphasia or hemisensory neglect and uses graphics to illustrate abstract concepts such as randomisation. The DVD will be given to trial participants as part of the recruitment information pack.

A process has been implemented for the therapist (a) to call the patient 45 minutes before the therapy session to remind the patient to take their IMP and (b) to conduct a compliance check at each therapy session.

Conclusions: The trial opened to recruitment in May 2011. An evaluation of the above approaches has been built into the trial follow up outcome visits. Recruitment, adherence to trial protocol and patient satisfaction with information provided will be used as outcomes to judge impact of above strategies.

Acknowledgements: This abstract is being presented on behalf of the DARS Trial Management Group.

A134
Understanding factors influencing questionnaire response rates to maximise retention in a long term complex intervention trial
Athene J Lane1, Liz Down1, Julia Wade1, David Neal2, Freddie Hamdy2, Jenny Donovan2, John Pearn1
1School of Social and Community Medicine, University of Bristol, Bristol, BS8 2PS, UK; 2University Dept of Oncology, Addenbrooke's Hospital, Cambridge, CB2 0QQ, UK; 3Nuffield Dept of Surgical Sciences, University of Oxford, Oxford, OX3 9DU, UK
Trials 2011, 12(Suppl 1):A134

Objectives: Participant retention is challenging in long term trials and attrition can threaten trial validity, especially for patient-reported outcomes which cannot be ascertained otherwise. The evidence base to maximise retention is limited so we investigated factors influencing questionnaire response rates with a mixed methods research approach.

Methods: ProtecT is an ongoing randomised trial of localised prostate cancer treatments in men currently aged 53-76 years (ISCRTN 20141297) with 8 years follow-up. Symptoms and QoL were collected annually by postal questionnaire with a reminder letter for non-responders at six weeks (phase I). Three additional interventions were sequentially investigated to increase retention. Firstly, clinical centre staff commenced telephoning non-responders at four weeks (phase II). A study pen was included with reminder letters and a shortened questionnaire was sent to non-responders at nine weeks (phase III). Questionnaire response rates over six months were compared by phase. 1279 participants were asked in the questionnaires about online completion. 18 participants gave in-depth interviews about trial follow-up which were transcribed verbatim and analysed thematically.

Results: The questionnaire response rate was 76.4% (phase I: 1045/1367) with a median return of 13 days (IQ range 8-22 d) and the reminder letter increased this to 86.8% (1187). Telephoning increased rates to 90.5% (phase III: 1105/1221). The additional incentive of a study pen was ineffective (phase II: 1026/1142, 89.8%) whilst the short questionnaire had some impact (phase III: 9/84 posted, 10.7%, 1033/1142, 90.5%). One quarter of men wished to complete on-line questionnaires (430, 25.2%). In interviews, most men found questionnaires acceptable and understood their purpose although they were regarded as a less enjoyable aspect. Some men saw questionnaires becoming less relevant over time either because they felt they were cured or they had become repetitive, although they continued to complete them. Participant newsletters were interesting and gave a sense of belonging to a group although some men wanted preliminary trial findings. The study website was infrequently accessed.
A key risk indicator approach to central statistical monitoring in multCentre clinical trials: method development in the context of an ongoing large-scale randomised trial
Valédis-Méarqué Elsa1, Hopewell C Jemima, Landray Martin, Armitage Jane
Clinical Trial Service Unit and Epidemiological Studies Unit University of Oxford, Oxford, Oxford, UK
Trials 2011, 12(Suppl 1):A135

Background: Monitoring in randomized trials is recommended as part of ongoing international Conference on Harmonisation - Good Clinical Practice standards. On-site monitoring in multCentre trials is common but is costly and can be inefficient. Central statistical monitoring can be used to detect unusual data patterns, identify intentional or unintentional trial misconduct, and to prioritise on-site visits and additional training. Motivated by an ongoing international multCentre clinical trial of over 25,000 randomized participants with electronic data capture, we developed key risk indicator (KRI) methods for central statistical monitoring in multCentre trials.

Method and results: Statistical monitoring of KRIs such as the rate of serious adverse event reporting, compliance with study treatment visit duration (based on how long the electronic case report form is open for) and blood results may be informative for identifying poor site performance in randomized trials. We have used these KRIs to illustrate an approach to central statistical monitoring.

For examining the rate of serious adverse event reporting, centres were assigned a dichotomous score depending on whether they showed extreme deviation from comparable sites (arbitrarily defined as half the observed median rate across sites). A similar approach was used to identify sites with short visit duration and these methods were also extended to examine repeated measurements of compliance with study treatment.

Blood results and other similar continuous variables were examined for unusual patterns that may suggest fabricated data or training issues. Residuals were calculated from linear regression models allowing for important covariates. The moments of the distribution (mean, variance, skewness and kurtosis) of the residuals within a centre were compared with those across sites. Subsequently, centres were assigned a dichotomous score depending on whether the observed data showed relevant (clinically or otherwise) deviation from comparable sites. Varying combinations of KRIs may be informative for different questions of interest. For example, KRIs that are useful for identifying potential data fabrication may not be relevant for assessing unintentional protocol violations. Consequently, we used flexible weighting methods to formulate summary scores involving different KRIs of interest.

Conclusions: KRI methods provide a flexible approach to central statistical monitoring in multCentre clinical trials. However, a KRI approach needs to be tailored to each study, and the KRIs selected using knowledge of what is important to individual clinical trials.

SURVIVAL ANALYSIS

A136
Survival analysis coping with non proportional hazards in randomized trials
Patrick Royston*, Mahesh KB Parmar
Hub for Trials Methodology Research, MRC Clinical Trials Unit, London, UK
Trials 2011, 12(Suppl 1):A136

Almost all trials with a censored time-to-event outcome are designed, powered andanalysed with a target hazard ratio for comparing experimental and control treatments in mind. Differences in survival experience between trial arms are tested with a logrank test and usually illustrated using a Kaplan-Meier plot. We describe this as an analysis in the probability domain. The focus is often on estimating survival probabilities at particular times after randomization.

In the absence of censoring, we would almost certainly prefer to work in the time domain. We would analyse times to event directly, summarizing results in terms of means, SDs and confidence intervals for differences. We would illustrate the different survival experiences using comparative scatter plots of times to event, histograms and the like.

The probability domain paradigm works well in many cases, but the hazard ratio, as a meaningful summary measure, requires one crucial assumption: proportional hazards of the treatment effect. In some recent high-profile trial reports, the PH assumption has clearly been invalid. We consider and recommend a general analysis strategy, relevant to the time domain, that does not assume PH. It based on the idea of the restricted mean survival time (RMST). We discuss the motivation, definition and interpretation of RMST and suggest how it can be woven into a strategy for time to event trials in which non-PH is anticipated or discovered.

A137
Crossing survival curves: alternatives to the log-rank test
George Boulouros1, Lucinda Billingham1,2
1MRC Midland Hub for Trials Methodology Research, University of Birmingham, Birmingham, B15 2TT, UK; 2Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, B15 2TT, UK
Trials 2011, 12(Suppl 1):A137

Background: It is not uncommon for clinical trials to present results on survival time as Kaplan-Meier survival curves that cross, indicating non-proportional hazards. A recent example was given in a pivotal trial in advanced non-small cell lung cancer (The IPASS study [1]). Trials such as these present a hazard ratio and log-rank test for treatment comparison as this is their planned primary analysis. However, the validity of such analysis is questionable and has received published criticism. This paper reviews the use of the log-rank test with crossing curves and considers alternatives that have been proposed.

Methods: The review of the alternative approaches includes weighted log-rank tests (Wilcoxon, Tarone-Ware, Peto-Prentice and Fleming-Harrington), supremum versions of the log-rank test (modified Kolmogorov-Smirnov and Renyi-type tests) which are based on the maximum difference between estimates of two survivor functions and modified log-rank tests (Lin and Wang test using squared differences at each time point, and Levene-type test focusing on variance differences). In addition, methods based on splitting the analysis at the crossing point have also been proposed. Methods are compared and evaluated using both real and simulated datasets using Weibull and Weibull-Cox distributions representing realistic situations.

Results: Crossing survival curves is generally a result of the survival times having greater variance in one treatment group than another. The performance of the log-rank test and alternatives depend on the type of crossing (early, mid or late) but in general the probability of a Type II error is increased for log-rank and weighted log-rank tests but performance is improved with the alternatives. The choice of time-point for the split-analysis is problematic. Standard software such as sts test (Stata), proc lifetest (SAS) and survfit (R) and routines-on-demand support some but not all the tests considered.

Conclusions: There is a need in the clinical community to clarify methods that are appropriate when survival curves cross. Statistical analysis plans for clinical trials with survival as primary outcome measure should specify an analysis dependent on the proportionality of hazard rates and explicitly consider non-proportionality issues, powering the analyses based on log-rank alternatives. Modelling the survival data may be more appropriate than simple univariate hypothesis tests when hazards are not proportional. Finally, there are some feasibility issues regarding software for such analysis that remain to be tackled.

Reference
A new measure of predictive ability for survival models
Babak Choobari-Oskoei, Patrick Royston, Mahesh KB Parmar
MRC Clinical Trials Unit, London, WC2B 6NH, UK
E-mail: Babak.choobari@ctu.mrc.ac.uk
Trials 2011, 12(Suppl 1):A138

In clinical research, an understanding of prognostic factors is important in the design and analysis of clinical trials and retrospective reviews of clinical experience. The results of prognostic factor studies are usually summarized in the form of statistics resulting from statistical significance testing, i.e. estimated parameters, confidence intervals, and p-values. These statistics do not inform us whether prognostic factor information will lead to substantial improvement in the prognostic assessment. Predictive ability measures can be used for this purpose since they provide important information about the practical significance of prognostic factors. R^2-type indexes are the most familiar forms of such measures in survival models, but they all have limitations and none is widely used.

Bura and Gastwirth (2001) [1] proposed a new predictive ability measure, named total gain (TG), for a logistic regression model. TG is based on the binary regression quantile plot, otherwise known as the predictiveness curve, which was first proposed by Copas (1992). Gu and Pepe (2009) [2] showed that TG is related to the ROC summary index, but it does not have the reported shortcomings of the ROC index.

In this paper, we extend the proposed TG measure to survival models and explore its properties using simulations and real data. In survival models, the TG statistic is a non-negative, unitless measure of the total cumulative distance between the average survival probability, as expressed by the Kaplan-Meier (KM) estimates of the survival probability, at a fixed time point and the estimated survival probabilities from a given model. Standardised TG ranges from 0 (no explanatory power) to 1 (perfect explanatory power).

In our simulation studies, we investigated the impact of censoring, covariate distribution and influential observations on the measure. The results of our simulations show that unlike most of the other R^2-type predictive ability measures, TG is independent of censoring and follow-up time. TG also increases as the effect of a covariate increases, but it is adversely affected by the categorisation of continuous prognostic factors. Finally, we applied TG to quantify the predictive ability of prognostic models developed in several disease areas. On balance, although TG lacks the intuitive interpretation of the explained variation measures, our results indicate that the estimates of the measure are within the reasonable range of the estimates of explained variation measures and can be recommended as an alternative measure to quantify the predictive ability in survival models.

References

A139
Evaluation of methods to adjust for treatment switching in clinical trials
Richard Fox*, Lucinda Billingham**, Keith Abrams***
1 Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, B15 2TT, UK; 2 MRC Midlands Hub for Trials Methodology Research, University of Birmingham, Birmingham, B15 2TT, UK; 3 Centre for Biostatistics and Genetic Epidemiology, University of Leicester, Leicester LE1 6TP, UK
Trials 2011, 12(Suppl 1):A139

Introduction: Treatment switching in clinical trials dilutes estimates of treatment effect which is problematic in decision making, especially in an economic context. For example 75% of patients crossover from the randomised treatment which emphasises the need to address such bias. We consider methodologies that compensate for this bias.

Methods: The methods assessed are: intention to treat (ITT), per-protocol (PP), adjusted Cox model [1], causal proportional hazards estimator [2], rank-preserving structural failure-time models (RPSFT) [3], iterative parameter estimation (IPE) [4], parametric randomisation based method [5], and the less well known inverse probability of treatment weighting (IPTW) [6].

Data: Survival data having an underlying Weibull distribution was simulated and designed such that probability of switching for patient subgroups could be varied within treatment arms. Other subgroup characteristics could be controlled and 24 scenarios with varied levels of bias were analysed. A review of submissions to the National Institute of Clinical Excellence was performed and used to inform the scenario parameters.

Results: The RPSFT and IPTW methods returned the lowest biases, <8%, in all scenarios. The estimates of the parametric randomisation based method were often far less variable than other methods but were subject to erratic behaviour with extreme biases observed. The other methods performed poorly in general, with biases of up to 50% not uncommon. In particular the IPTW method over-compensates in most scenarios.

Conclusion: Under these conditions the results clearly identified the RPSFT and IPTW methods as most consistent and accurate, with the latter the more consistent of the two. None of the other methods returned consistent results, and as such cannot be recommended.

Further avenues of investigation include exploring the effect of other underlying survival distributions, extending from univariate models to adjust for other covariates, and extending from situations where just control-arm patients switch to scenarios with multidirectional cross-over.

References
difference in competing risks between treatment groups. For the multiple myeloma example, the p-values were 0.0003 (log-rank), 0.005 [1] and 0.051 [2]. When censoring was evenly distributed between groups, the different weighting approaches yielded similar results. However, when there was differential censoring between the groups, the usual cumulative incidence regression produced inflated Type I errors (exceeding 10% when the hazard ratio for censoring was greater than 2) and reduced power when compared to a method where weights were calculated separately within treatment groups.

Conclusions: Competing risk approaches should be included in time-to-event analyses, even when the rate of competing events is expected to be low. When comparing two groups, using weights from separate censoring distributions is recommended as this has a less inflated Type I error and greater power. This may be thought of as analogous to using separate variances when comparing two means rather than using a pooled variance, although the increased efficiency of the pooled approach does not exist here when there is no true difference in censoring risk.

References

A141
A robust parameterisation for the analysis of survival data in the presence of covariates with extreme value observations
Richard Jackson*, Trevor Cox
Liverpool Cancer Trials Unit, Liverpool University, Liverpool L69 3BX, UK
E-mail: R.J.jackson@liverpool.ac.uk
Trials 2011, 12(Suppl 1):A141

We propose a new parameterisation for standard survival models which allow for the effect of a covariate with extreme value observations. We show that the robust model can offer an improved fit to both the parametric exponential model and the Cox proportional hazards model. Furthermore, we demonstrate the application of this approach using data from a pancreatic cancer dataset.

Under traditional survival models, covariates enter the model via:
\[ \hat{\lambda} = \exp(\beta x) \]

Whilst this allows for ease of interpretation, we show that this can be a dangerous assumption in the presence of extreme value observations and can lead to misleading survival estimates.

We propose two models which aim to counter this effect. For parametric models, the robust model is of the form:
\[ \lambda = \frac{a + b \exp(\beta x)}{c + \exp(\beta x)} \]

This model is bounded below at \( a/c \) and has an upper asymptote equal to \( b/c \). There are four parameters to be estimated. For the Cox proportional hazards model, the robust model is amended to be:
\[ \lambda = \frac{a + b \exp(\beta x)}{a + b - 1 + \exp(\beta x)} \]

Here there are three parameters to be estimated. The function equals unity when \( |\beta| = 0 \) so that a covariate that is not associated with survival will keep the hazard ratio equal to unity.

These models are fitted to both the parametric exponential and the Cox proportional hazards models in a Frequentist and Bayesian framework using the statistical packages R and WinBUGS. Simulated datasets are used to illustrate the scenarios under which these models are of most use. Initial results show that these models can improve the fit to the data in the presence of extreme value observations when compared using Akaike's Information Criterion (AIC).

Having demonstrated the benefits of robust models, the models are used on a pancreatic dataset. We compare the fit of the post operative CA19-9 value, which has been shown to be a good indicator for overall survival. The robust model here provides a better fit for both the exponential model and the Cox proportional hazards model. We suggest this can lead to improved prognostic models for predicting survival and can also improve comparisons of other effects in a clinical trial by reducing the residual variation observed in the data.

Model residuals are used to assess any departure from the proportionality assumption. These show that extreme value observations are less likely to violate the proportionality assumption when the robust parameterisation is used. An influence function has also been produced to demonstrate the effect of single observations on the model parameters.

TRIALS OF COMPLEX INTERVENTIONS

A142
Randomised controlled trials of complex interventions
John Norrie
Centre for Healthcare Randomised Trials, Health Services Research Unit, Health Sciences Building, University of Aberdeen, UK
E-mail: j.norrie@abdn.ac.uk
Trials 2011, 12(Suppl 1):A142

Although all randomised controlled trials (RCT) share much in common in terms of design, conduct, analysis and reporting, trials of complex interventions (loosely “interventions with several interacting components”) can have special issues which need appropriate consideration to deliver high quality and credible evidence to change or inform practice. For example, blinding the intervention may be impossible (a surgical compared with a medical intervention), and standardising the intervention (as would be obligatory and easy in a drug trial) difficult to define and harder still to achieve (for example, reaching a therapeutic alliance between patient and therapist in a cognitive behavioural therapy RCT). In addition, often complex intervention trials have patient reported primary outcomes (for example, quality of life) that present challenges (e.g. deaths, missing data). Safety reporting in complex intervention trials often presents particular challenges, struggling to adapt methodologies designed and developed for regulatory drug trials.

Designers of complex intervention trials may be forced to make assumptions about critical features of the intervention (e.g. intensity, and duration) and how to measure the treatment effect without the funds or the time to develop these aspects in a similar manner to the Phase I-I V development of drug trials. Frequently complex intervention trials are underpowered to detect important effects due to their expense and perceived difficulties. Often times one is left wondering whether the failure of a complex intervention trial to detect a treatment effect was down to sub optimal design, rather than strong & robust evidence that it does not work. This talk will highlight various methodological challenges in the various stages of the design and running of complex intervention trials, including the development of evidence based interventions and outcomes, through to interpretation of often difficult analyses and convincing skeptical peers and journals of the robustness of findings. The talk will be illustrated by practical examples from many complex intervention trials the speaker has been involved in, including several large, multicentre, pragmatic trials funded by the NIHR/HTA, and more explanatory trials funded by MRC/EME.

It will also draw on first-hand experience of sitting as an independent statistician / methodologist / trialist on various Trial Steering Committees and Data Monitoring Boards, as well as on various funding panels and journals.

A143
Designing trials of complex interventions for efficacy and mechanisms evaluation
Richard A Emsley1, Jonathan Green2, Graham Dunn1
1Health Sciences Research Group, School of Community Based Medicine, The University of Manchester, Manchester, UK, 2Mental Health and Neurodegeneration Research Group, School of Community Based Medicine, The University of Manchester, Manchester, UK
E-mail: richard.emsley@manchester.ac.uk
Trials 2011, 12(Suppl 1):A143

Objectives: There have been recent developments in statistical methods for assessing the mechanisms through which a complex intervention acts, with particular applications in mental health trials. This literature has
focused on causal mediation analysis, which differs from statistical mediation analysis by its explicit acknowledgement and exposition of the necessary assumptions, particularly that of no unmeasured confounding between the putative mediators or mechanisms and outcomes. Instrumental variable estimation procedures can produce valid causal estimates even in the presence of this unmeasured confounding, but the problem then becomes finding valid instruments to identify the statistical models. So far, however, little attention has been given to the issue of designing trials of complex interventions with this analysis in mind.

Methods: We introduce the instrumental variables estimation procedure, and discuss some possible methods for obtaining instruments. We then propose possible alternative designs for trials of complex interventions in mental health which could produce these potential instruments:

1. *Waiting list control*, where the control group receive the intervention after a specified period of time;
2. *Innocuous vaccine*, where all participants are measured on a baseline variable strongly related to the putative mechanism;
3. *Parallel trials*, where a concurrent set of trials are run in parallel with different interventions but common measures;
4. *Mediated moderation*, where a potential instrumental variable is included as a baseline measure.

Results: We demonstrate how these trial designs can help validate the causal mediation analysis and strengthen some of the fundamental assumptions underpinning it. We give examples with illustration of the “parallel trials” design in a randomised trial of cognitive behaviour therapy in psychosis, and a “mediated moderation” design in parent mediated treatment trial in child psychiatry.

Conclusions: Well-designed complex intervention trials should not only consider evaluating the efficacy of the intervention but also the putative mechanisms through which it acts. Adopting a suitable design for the trial both enables and strengthens the assessment of this mechanistic analysis.

Acknowledgements: All authors are investigators on an MRC methodology grant entitled “Designs and analysis for the evaluation and validation of social and psychological markers in randomised trials of complex interventions in mental health” (G000678). RE is supported by an MRC Career Development Award in Biostatistics (G0802418). RE and GD are members of the UK Mental Health Research Network Methodology Research Group.

A144

How do treatments for chronic fatigue syndrome work? Exploration of instrumental variable methods for mediation analysis in PACE – a randomised controlled trial of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care

Kimberly Goldsmith1, Trudie Chalder2, Paul White3, Michael Sharpe3, Andrew Pickles1

1Institute of Psychiatry, King’s College London, DeCrespigny Park, London, SE5 8AF, UK; 2Wolfson Institute of Preventive Medicine, Queen Mary, University of London, EC1M 6BQ, UK; 3Department of Psychiatry, University of Oxford, OX3 7JX, UK

Trials 2011, 12(Suppl 1):A144

Objectives: Chronic fatigue syndrome (CFS) is characterised by chronic disabling fatigue. The PACE trial compared four treatments for CFS and found that for therapies added to specialist medical care (SMC), cognitive behaviour therapy (CBT) and graded exercise therapy (GET) were more effective than adaptive pacing therapy (APT) and SMC alone in improving physical function and fatigue. What are the mechanisms of these treatments? CBT and GET may affect outcomes through thought processes and behaviours (mediators). Traditional Baron, Judd and Kenny (BJK) methods for estimating mediation effects can be subject to bias; instrumental variable methods (IV) can address this problem. The aims were: To explore the potential for causal analysis of mediation in PACE; To compare IV estimates to those obtained using BJK methods, which are unbiased only under restrictive assumptions such as no unmeasured confounding.

Methods: Two treatment arms were compared at a time. BJK methods were applied using three ordinary least squares (OLS) regression models. IV methods were applied by compiling a list of baseline variables that could act as IVs in interaction terms with treatment arm and then assessing these using OLS with the mid-treatment measurement of the putative mediator as the outcome. Instrument strength was assessed using the R² change between models with main effects only and with the interaction term. Two-stage least squares regression (2SLS) was used to estimate effects in the presence of IVs. Collective instrument strength was assessed using an F test and partial R².

Results: The IVs were weak, with a maximum R² change of 0.03. The five strongest IVs were therefore used in the 2SLS in each case. There was modest mediation of CBT and GET effects (approximately 20% of the total effect). The IV-derived estimators were somewhat different in magnitude than the BJK estimators and were less precise. There is scope for modelling a common effect of mediators on outcomes across trial arms.

Conclusions: There was evidence for modest mediation of CBT and GET effects. Potential IVs for the study of PACE treatment mechanisms can be found, however, these were weak. Combining trial arms may allow for more efficient analysis using IVs.

A145

Statistical design of mental health complex intervention trials

Sabine Landau1*, Kings College London, Institute of Psychiatry, Department of Biostatistics, London SE5 8AF, UK

E-mail: sabine.landau@kcl.ac.uk

Trials 2011, 12(Suppl 1):A145

Background: Many mental health interventions constitute complex interventions since they are characterised by multiple components which when given in combination are thought to improve outcome as well as the components changing distal clinical outcome indirectly by inducing change in intermediate process variables (mediation). Ideally the evaluation of a complex intervention should be both pragmatic (assess effectiveness) as well as explanatory (measure the efficacy of the intervention components and separate their direct and indirect effects). Ordinary least squares (OLS) estimates of parameters of interest in explanatory analyses will suffer bias despite random treatment allocation when covariates are endogenous. In mental health trials such endogeneity is likely to arise as the result of non-compliance with randomised treatments, hidden confounding or measurement error in the treatment components received or process variables. In this talk I discuss ways of designing a complex intervention trial so that all causal parameters of interest can be estimated consistently using instrumental variable (IV) methods.

Methods: Pear’s graphical checks of identifiability are applied to establish the most general causal model for mental health complex intervention trials under which parameters of interest remain identified. Having determined this model, experimental actions including the generation of relevant IVs, that can ensure model assumptions hold, are reviewed. Monte Carlo simulations are carried out to compare the performance of naive OLS estimators with that of IV estimators.

Results: Allowing for non-compliance with randomised treatment offers, hidden confounding or measurement errors requires independent randomisation of constituent treatment components and the availability of further instruments for hypothesised process variables. A number of avenues for generating the latter exist. Simulations confirm that the biases suffered by OLS estimators are corrected by IV methods. The variance inflation of IV treatment component effect estimators is acceptable in standard sized mental health trials. However, instruments for process variables may be less strong raising the possibility that the variance inflation of mediation effect estimators renders them impractical for use.

Conclusions: Investigators planning to carry out explanatory analyses of complex interventions should consider the generation of relevant IVs at the design stage. The unbiased estimation of (total) effects of the treatment components actually received (efficacy) is feasible under non-compliance or treatment misclassification. The generation of strong IVs for process variables may be more problematic. In mental health research mediation analyses allowing for hidden confounding and measurement error may only be feasible for large trials or after combining trials.
Acknowledgements: This work was carried out while on sabbatical leave at the MRC Biostatistics Unit Cambridge and the University of Manchester. I would like to thank colleagues Ian White, Graham Dunn and Richard Emsley for helpful discussions.

**A146**

*Modelling multiple outcomes to improve the detection of causal mediation effects in complex intervention trials*

Neil Casey,1 Simon Thompson2, Andrew T Prevost1,*
1Department of Public Health and Primary Care, University of Cambridge, CB2 0SR, UK; 2Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, CB2 0SR, UK. *Department of Primary Care and Public Health Sciences, King’s College London, SE1 3QD, UK

E-mail: toby.prevost@kcl.ac.uk

Trials 2011, 12(Suppl 1):A146

**Aim:** In a trial to increase physical activity in sedentary adults, there was no evidence of effect on the primary outcome, though large significant effects amongst eight related SF36 measures of general health [1]. How could such differences arise? Could they be: a true effect mediated through receiving the intervention, another systematic effect such as self-reporting bias, and/or chance? The aim was to develop reliable methods to investigate whether effects of the intervention on the SF36 outcomes were mediated through truly receiving the intervention delivered in intervention sessions.

**Methods:** We adopted a structural mean modelling approach with a two-stage least-squares estimation algorithm, in order to estimate mediation effects free from confounding bias [2]. It involves predicting the mediator (number of sessions attended), and outcomes, from baseline covariates. Each individual has a personal predicted ‘counterfactual’ treatment-effect difference which is regressed on the predicted mediator using a dose-response model. Although reliably bias-free, typically these methods do not provide sufficiently precise estimates except for the simplest of models. A simulation study was designed to establish the factors driving the lack of precision. We extended the two-stage approach, using linear mixed effects and GEE modelling to enable multiple SF36 outcomes to contribute to estimation of a common mediation effect.

**Results:** In this trial, attendance at sessions was invariably high, adversely affecting the precision of the estimates. From the simulation study, important factors affecting detection were identified to be the size of the trial effect and the degree to which the mediator is predictable from baseline covariates. The effect of analysing four SF36 outcomes to estimate an assumed common mediation effect was to reduce the standard error of the estimated effect by up to 40%, equivalent to offering an increase in power to detect mediation from 50% to 80%. The mediation effect was statistically significant.

**Conclusions:** The extension of bias-free estimation of a mediation effect from one to multiple related outcomes offered an appreciable improvement in the power to detect mediation effects and to estimate them more precisely. The significant effect through sessions indicates that some of the effect may well be genuinely connected with receipt of intervention material. The approach requires assumptions.

**References**

1. Kinmonth AL, Wareham NJ,ハードmen W, Sutton S, Prevost AT, Farshawe T, Williams KM, Ekelund U, Spiegelhalter D, Griffin SJ. Efficacy of a theory-based behavioural intervention to increase physical activity in an at-risk population (eg patients, carers); on the specialty being investigated (eg intensive care, outpatients) and on the rest of the organisation and society and give each a unique number.


**A147**

*MATRICS: A Method for Aggregating The Reporting of Interventions in Complex Studies*

Rumbalay-Thorne1, Gabi S Jerzembek2, Wai-Yee Cheung1, David Cohen1, Hayley A Hutchings3, Frances L Rapport4, Anne C Seagrove5, John G Williams1, Ian T Russell2
1Centre for Health Information Research and Evaluation, Grove Building, College of Medicine, Swansea University, Singleton Park, Swansea, SA2 8PP, UK; 2Institute of Health Service Effectiveness (IHSE), Aston University Business School, Aston Triangle, Birmingham, B4 7ET, UK; 3Faculty of Health Sport and Science, University of Glamorgan, Pontypridd, South Glamorgan CF37 1DL, UK

E-mail: k.thorne@swansea.ac.uk

Trials 2011, 12(Suppl 1):A147

**Background:** There are few rigorous methods for combining qualitative and quantitative findings from studies with complex interventions using multiple research methods and giving appropriate weight to each without introducing bias to the overall conclusions.

*We developed a Method for Aggregating The Reporting of Interventions in Complex Studies (MATRICS) for the ENIGMA study (Evaluating Innovations in Gastroenterology by the NHS Modernisation Agency) – a multi-centre, mixed-methods study to evaluate the impact of the Modernising Endoscopy Services programme [1], funded by the UK National Institute for Health Research (NIHR SDO ref 08/1304/46).*

**Method:** We developed a template that requires researchers to follow the steps outlined below:

1. List the types of effects identified by the study (from the aims objectives and outcome measures), and divide them between effects on: sample population (eg patients, carers); on the specialty being investigated (eg intensive care, outpatients) and on the rest of the organisation and society and give each a unique number. 2. List the methods used to explore each effect listed in step 1 and give each a unique letter (eg GP interviews, patient questionnaires, routine data linkage). 3. Create an alphanumeric code by cross matching the effects identified and the methods used to investigate them (eg patient satisfaction “1” was investigated using a patient questionnaire “A” = 1A).

4. List the explicit findings of the study and label them using the alphanumeric code (eg “patients were dissatisfied with waiting times – A1”). 5. Synthesise all consistent findings and list their alphanumeric codes alongside to characterise mutually confirmatory findings. Synthesis is best done independently by at least two researchers. 6. Reorder all contradictory findings and their alphanumeric codes adjacent to one another to better illustrate all conflicting findings.

**Findings:** The MATRICS tool greatly facilitated the unbiased factual reporting of findings from multiple methods for ENIGMA [1]. Additionally, it was most beneficial for the qualitative synthesis of the findings of ENIGMA, a study unsuited to formal cost-benefit analysis, like most in this field. We have also applied the MATRICS successfully to other complex studies using multiple methods.

**Discussion:** If the experience of the study team regarding the MATRICS approach to synthesising results in complex studies is reflected by others, it could provide a formal structure for reporting the results of complex and/or multiple-method studies. Further application of this methodology will provide evidence of whether this reporting tool will improve a reader’s understanding of a study and its findings.

**Reference**


**A148**

*Understanding the complexity of surgical procedures in RCTs: a pilot study to test the application of the MRC framework for evaluating complex healthcare interventions in the operating theatre*

Mark O Eveleigh1, Natalie S Blencowe2, Nicola Mills3, Jane M Blazeb3
1Surgical Research Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK

Trials 2011, 12(Suppl 1):A148

**Background:** Several methodological challenges make randomised trials (RCTs) in surgery difficult, and whether surgical interventions are themselves complex is an issue that requires further exploration [1]. Surgical interventions have multiple concomitant parts that may independently and inter-dependently influence outcomes such as the operation itself, surgeon expertise and contextual factors such as team.
Qualitative research methods were applied to controlled trials, comparing family intervention with an ongoing RCT.

Materials and methods: Qualitative research methods were applied to two operations within the context of an on-going RCT of pre-operative chemotherapy and surgery for oesophageal cancer (OEOS trial). Non-participant observation in the operating theatre and video-recording of the surgery itself were undertaken. Digital video recordings were collected directly from equipment routinely used in laparoscopic surgery onto an encrypted hard drive, transferred to a secure server and analysed. Clinical and non-clinical interactions in the operating theatre were recorded.

Results: The logistics of video recording surgical procedures were explored and operations successfully recorded and transferred to the secure server. Analyses of procedures proved complex, therefore additional software was identified to allow systematic coding of technical parts of the operation. Non-participant observations were divided into issues relating to the intervention itself, its components (patient, surgeon, assistant, anaesthetist, nurses) as well as contextual factors. Specific comments made by any team members were also documented. This enabled generation of a thematic framework for future analyses and allowed triangulation of findings.

Conclusions: Using qualitative research methods to understand the component parts of surgical interventions is a novel concept but this early work shows it is feasible. Future research will extend to new RCTs in surgery and include case studies of the interventions and how they are delivered (with in-depth interviews with patients, surgeons and other team members as well as video and audio recordings and non-participant observation). This will improve understanding of the complexity of surgical interventions and generate methods to enable analysis of shared outcomes with more validated outcomes and monitoring of treatment fidelity to the protocol, meaning that trial results may be more likely to be believed and accurately implemented in practice.

References:

A149
Designing and assessing pragmatic clinical trials of complex interventions in mental health: challenges and solutions
Amanda J Farrin*, Michelle Collinson
Clinical Trials Research Unit, University of Leeds, Leeds, LS2 9JT, UK
E-mail: a.j.farrin@leeds.ac.uk
Trials 2011, 12(Suppl 1):A149

Objectives: Designing and analysing pragmatic clinical trials involving complex interventions is becoming increasingly complex, particularly in the field of mental health. This paper will describe a variety of challenges encountered during the design of a complex intervention mental health trial and the decisions made on how to overcome these challenges. The objectives are to share knowledge gained through the design of this large multi-centre trial and to facilitate discussion and widen the debate regarding potential issues and possible solutions in the design of such trials.

Methods: The HTA-funded SHIFT Trial (Self-Harm Intervention, Family Therapy) is a pragmatic, randomised, controlled trial, comparing family therapy with treatment as usual in 832 adolescents who have self-harmed at least twice, with a primary outcome of repetition of self-harm leading to hospital attendance. The particular challenges encountered during the trial design and solutions adopted will include the level of randomisation, therapist variation, multiple methods employed to collect the primary outcome, maximising reliable data collection and the planned analytical techniques.

Results: In the SHIFT trial different therapists are involved in delivering interventions; therefore we will explain the rationale for choosing individual over cluster randomisation and methods adopted to minimise therapist variation. In trials of complex interventions dropout from treatment and follow-up is common, so we will discuss the multiple methods employed to collect the primary outcome and how these maximise reliable data collection. Due to the nature of SHIFT’s endpoints, the analysis will use techniques not often encountered in mental health trials. We will discuss the planned interim analysis; the use of Cox’s Proportional Hazards Model to test for differences in repetition rates and the use of multilevel survival frailty models[1] to determine the extent of clustering on outcome due to therapists and the impact on the precision of the treatment effect. Multiple events analysis using Anderson-Gill[2] methodology making use of the timing and number of first and subsequent events will be discussed.

Conclusions: We will show how it is possible to design solutions to the challenges encountered in the design of complex trials in mental health, that are feasible, practical and robust.

Acknowledgements: The authors are presenting on behalf of the SHIFT trial team.

The SHIFT Trial is supported by HTA-470772.

References:

A150
A pilot factorial randomised cohort trial of manual therapy or acupuncture for low back pain
Vivienne C Dascanid*, Yvonne Birks, David Torgerson
York Trials Unit, Department of Health Sciences, ARRC Building, University of York, Heslington, York, YO10 5DD, UK
E-mail: VCF500@york.ac.uk
Trials 2011, 12(Suppl 1):A150

Background: Randomised control clinical trials of acupuncture have been hampered by the challenges of assessing it as a complex intervention. Controlling for and separating placebo effects whilst identifying its efficacy as a treatment can be difficult [1]. The comparison of acupuncture to other complex interventions has been recommended to assess the effectiveness of acupuncture against other interventions [2]. The objective of this pilot trial is to investigate the feasibility of undertaking a novel randomised cohort design study with a nested factorial RCT, investigating acupuncture alone versus manual therapy alone versus a combination of acupuncture and manual therapy versus usual care.

The pilot will investigate recruitment rates to allow for planning a full-scale trial, identify any compliance issues and strategies for reducing these in a full-scale trial and assess participants’ acceptance and therapist delivery of combined therapies for the treatment of their LBP.

Methods: The study will follow a randomised cohort trial design and participants from the cohort will be selected to participate in the pilot trial. The use of this design as a recruitment method for nested trials is relatively new methodology but the cohort design has been suggested as an effective method for the use with chronic conditions [3] and its potential for minimising attrition. Attrition is one of the major threats to the internal validity of any trial. The design of this trial specifically reduces that threat [4]. Using a randomised cohort design will provide a ‘run-in period’ of three months, from collecting baseline data to the first set of outcome data. Only participants who return their three monthly questionnaires will be eligible for randomisation to the pilot trial. As the majority of attrition occurs at the first period of follow-up in an RCT, it is expected subsequent attrition, after randomisation, to be minimal [4]. The factorial pilot RCT will investigate the treatment of low back pain with Acupuncture vs Manual Therapy vs Acupuncture and manual therapy vs Usual GP care. All interventions will be delivered by a chartered physiotherapist.

Results and conclusions: Recruitment and retention rates will be presented. The acceptability and feasibility of the design for use with complex interventions and in a common musculoskeletal condition will be discussed.

Acknowledgements: NIHR Programme grant project. Funding applicant Dr Hugh MacPherson, Professor David Torgerson.
Small project grant award from Acupuncture Association of Chartered Physiotherapists (AACP).

References

**A151**

Practical and methodological challenges in the design and implementation of a cluster-randomised feasibility trial of the management of urinary incontinence after stroke

Chris J Sutton, Lois Thomas, Denise Forshaw, Caroline L Watkins
School of Health, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK
E-mail: cjsutton@uclan.ac.uk
Trials 2011, 12(Suppl 1):A151

**Objective:** To evaluate design and implementation issues in an NIHR-funded feasibility trial of the management of urinary incontinence after stroke in a secondary care setting.

**Methods:** Twelve stroke services were cluster-randomised to 3 intervention groups (systematic voiding programme with/without supported implementation; usual care) in 4 strata based on: having separate/combined acute and rehabilitation units; above/below median performance on the ‘nine key indicators of stroke care’ in the National Sentinel Stroke Audit (NSSA) [1]; number of annual stroke admissions. Target recruitment was 780 patients overall; the recruitment period was 9 or 12 months, depending on a Trust’s annual stroke admissions, to reduce variability in numbers across services. Each service gained an additional 2.8 whole time equivalent health care assistants (HCAs) supporting introduction of the intervention or maintaining parity of staffing in ‘usual care’ services.

**Results:** Recruiting services proved problematic, with excess treatment costs and service support cost being unavailable in some regions locally. The Comprehensive Local Research Network helped negotiate service support and excess treatment costs in England, leading to the recruitment of 8 sites. We were encouraged in recruitment of Welsh services, although differences in funding structure and governance processes caused further delays. The Welsh Assembly’s commitment to stroke services meant that funding from them was accessible, although acquiring Health Board funding was more problematic. R&D approval was slow in all 12 Trusts, in some taking almost a year. Appointment of staff was often delayed by current vacancy control procedures in the NHS.

All English services are now recruiting, with larger ones recruiting satisfactorily. Smaller services typically started later and have had greater difficulty meeting their lower targets. Further recruitment strategies introduced recently include a newsletter about trial progress. Welsh sites are commencing recruitment in late summer/autumn 2011. The unplanned staggered start of sites caused a loss of allocation concealment, although there is no evidence that this impacted on site participation. A 9-month trial extension has been approved, with participant recruitment continuing until July 2012.

**Conclusions:** Designing and implementing a cluster-randomised trial of a complex intervention can be difficult, particularly so in the current financial climate in the NHS. This will impact on the feasibility and planning of a definitive trial to evaluate the effectiveness and cost-effectiveness of the intervention. Lessons learned from this feasibility trial may help others more correctly estimate the timelines and workload involved in setting up and running a multi-centre cluster-randomised controlled trial in the NHS.

**Acknowledgements:** We are grateful for the support of the ICONS Project Team and the ICONS Patient, Public and Carer Involvement Groups. This paper presents independent research commissioned by the NIHR under its Programme Grants for Applied Research scheme (RP-PG-0707-10059). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

**Reference**