Approaches to the evaluation of rapidly evolving radiotherapy technologies

Lucinda Billingham,[1] Neil Burnet,[2] Jenni Macdougall,[3] Emma Hall,[4] Chris M. Nutting,[5] Fergus Macbeth,[6] David Sebag-Montefiore,[7] Jane Blazeby,[8] Andrew Jackson,[9] Corinne Faivre-Finn,[10] Gillies McKenna,[11] Søren M.Bentzen,[12] Mahesh Parmar,[13] Tim Maughan[11] on behalf of the NCRI Clinical and Translational Radiotherapy Research Working Group

Version: 17 July 2012

1. MRC Midland Hub for Trials Methodology Research and Cancer Research UK Clinical Trials Unit, University of Birmingham, Birmingham, UK
2. University of Cambridge Department of Oncology and Oncology Centre, Addenbrooke's Hospital, Cambridge, UK
3. National Cancer Research Institute, London, UK
4. Clinical Trials and Statistics Unit, The Institute of Cancer Research, London, UK
5. The Institute of Cancer Research and Royal Marsden NHS Foundation Trust, London, UK
6. National Institute for Health and Clinical Excellence, UK
7. University of Leeds, St James's Institute of Oncology, Leeds, UK
8. School of Social and Community Medicine, University of Bristol, UK
9. Department of Clinical Oncology, Southampton University Hospitals Trust, Southampton, UK
10. Department of Clinical Oncology, The Christie NHS Foundation Trust, Manchester, UK
11. Gray Institute for Radiation Oncology and Biology, University of Oxford, Oxford, UK
12. University of Wisconsin School of Medicine and Public Health, Madison, USA
13. Medical Research Council Clinical Trials Unit, London, UK

ACKNOWLEDGEMENTS

This report is based on the CTRad Methodology Workshop on 'Approaches to the Evaluation of Rapidly Evolving Radiotherapy Technologies', funded by the Medical Research Council, and held on May 27th 2010 in London.

We are grateful for contributions to the day's discussions from Dr Peter Davidson who represented the NIHR Health Technology Assessment (HTA) programme, and Ms Julie Hearn from the Clinical Trials Awards and Advisory Committee (CTAAC) at Cancer Research UK.

Sally Stenning and Matthew Nankivall kindly provided notes from the discussions.

Professor Lucinda Billingham is supported by the Medical Research Council [grant number G0800808].

Professor Jane Blazeby is supported by the Medical Research Council [grant number G0800800].

Dr Neil Burnet is supported by the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre.

INTRODUCTION

Radiotherapy is a highly potent and cost-effective treatment for cancer [Bentzen, Heeren et al. 2005], and forms a central part of modern cancer treatment. Of those patients cured of cancer, 40% have radiotherapy as the principle component of their treatment [Tubiana 1992], and it has been estimated that the addition of radiotherapy to other modalities increases five-year survival by 16% [Barton et al. 1995].

Modern radiotherapy is both a technology-dependent and multi-faceted intervention, with variables including total dose, dose per fraction, overall treatment time, and number of fractions per day. The traditional randomised controlled trial (RCT) methodology is clearly the gold standard where the research question is amenable to this, for example comparing the use of radiotherapy versus no radiotherapy. However, the complexity of assessing how technologies are used means there are situations where randomisation is not practical, not likely to be informative, and potentially not ethical. As such, there is scope for exploring alternative ways to evaluate radiotherapy in cancer treatment.

The National Cancer Research Institute (NCRI) Clinical and Translational Radiotherapy Research Working Group (CTRad) was launched in 2009, and part of their remit is to explore when RCTs are needed for evaluation of technological developments, and when more pragmatic approaches could be taken [NCRI 2008]. In 2010, a workshop was held to engage the radiotherapy community in considering some of the methodological challenges and opportunities in radiotherapy research. The workshop was funded by the Medical Research Council (MRC) and organised by Workstream 3 of CTRad. This paper provides a summary of the presentations given in the workshop and expands on the key issues raised to outline the agenda for further research.

The workshop was introduced by Prof Max Parmar from the MRC Clinical Trials Unit, who set out the rationale for considering methodology issues in relation to clinical radiotherapy research, especially the evaluation of new techniques and technologies. Evidence is needed that such developments are doing more good than harm. It is accepted that RCTs are the conventional best method to evaluate standard interventions, when they can be used. The methods to evaluate a new drug are well established, but how do such approaches map onto the area of radiotherapy technology? Prof Parmar noted that the audience and speakers were made up of researchers and funders, and threw down the gauntlet to the participants to identify areas for further research, identify methodology issues and provide solutions.

AN INTRODUCTION TO RADIOTHERAPY TRIALS

An introduction to radiotherapy trials was presented by Dr Chris Nutting. RT trials can be split into two broad groups: those where the radiotherapy is the main focus, and those where the radiotherapy is integral to the treatment strategy but is not the main focus of the trial. A subtly different approach may be required for each.

For trials where radiotherapy is the main focus, the endpoint is often a reduction in toxicity, and large differences are sought. This means that only a small sample size is required, but consequently the trials are underpowered to look for equivalence in tumour control. Survival or disease control are, and should remain, gold standard endpoints for radiotherapy trials. This highlights the problem of the need for two separate endpoints (toxicity and tumour control), arguably both primary endpoints, both of which are required to establish the therapeutic ratio.

Trials where the radiotherapy is integral but not the main focus, i.e. trials of radiotherapy +/- pharmaceutical agents, ask the question of whether this additional or altered systemic therapy in combination with radiotherapy improves the therapeutic ratio – that is increases benefit more than toxicity. Modern radiotherapy treatment techniques are required for these studies to have clinical relevance.

THE COMPLEXITY OF MODERN RADIOTHERAPY

Radiotherapy, particularly the most modern image-guided, intensity-modulated radiotherapy, involves a substantial number of components, many of which interact with each other. Quantification of the uncertainties associated with each component is required. Each component is dependent on specific behaviour, and multiple groups of professional staff are involved in the preparation and the delivery of treatment.

For some aspects, such as the calculation of the margins required to achieve accuracy in treatment delivery (Planning Target Volume (PTV) margins), complex feedback loops are required. For example, PTV margins depend on the quality of imaging, then the interpretation of that imaging, reproducibility in delineation (intra- and inter-operator variation), patient immobilization (reproducibility) systems, intra-fractional motion of patient and/or target, and on-treatment imaging protocols [BIR 2003]. The PTV margin also depends on the patient's ability and motivation to cooperate. Thus, margins depend on the particular technique, the particular anatomical location and tumour type, and on the particular department. Although most of these elements vary from one patient to another, the population variation needs to feedback into both the PTV margin used in plan preparation (upstream) and the departmental image guidance strategy (downstream).

Interactions in the radiotherapy pathway divide into those that are *physical* and those that are *behavioural*. Physical interactions include, for example, the relationship between the installation of planning CT equipment and radiotherapy treatment machines, whose localisation systems need to be precisely linked in three-dimensional space. These relationships are certainly complicated, but can be identified relatively easily by following the physical pathway of the patient proceeding through the planning and treatment pathway. These physical interactions are often referred to as the 'Radiotherapy Chain' (Table 1); some steps are performed once and can contribute to systematic uncertainty or inaccuracy, and others are performed with each dose, providing additional random uncertainty or inaccuracy.

Behavioural interactions are substantially more complicated and obscure (Table 2). These interactions commence with the first treatment decision, and include availability of technology, staff experience, waiting lists and targets, and the clinician's view of risk-benefit.

There is an additional set of interactions which relate to the delineation of the target volume (TVD). This process also interacts with the availability of relevant technology, experience, available imaging, assessment of image co-registration accuracy, data on set up accuracy to derive the PTV margin and the planning organ at Risk Volume (PRV) margin, and the balance of target edge dose versus PRV dose. Contained within the TVD step is the interpretation of imaging and the uncertainties inherent in any and all imaging. For example, it is well understood that the relationship between imaging and pathological specimen is not exact [Daisne 2004]; it is also very difficult to investigate. Individual variation in delineation, both inter- and intra-operator variation, also needs to be addressed in this stage. The process of plan preparation involves interactions between the availability of IGRT technology, planning system software, patient anatomy, analysis of errors and uncertainties within the systems for planning, delivery and quality assurance, and clinician specification of dose limits. The availability of IGRT also bears directly on the PTV and PRV margin considerations. Treatment delivery also has behavioural interactions which at this stage clearly involve the patient who must be willing and able to comply with the requirements for accurate treatment. This includes the element of intra-fraction stability. Here, both individual patient anatomy and planning (i.e. PTV and PRV margins) considerations are relevant, and for the latter the specific IGRT strategy employed links into a feedback loop on PTV and PRV margins. At this point in the treatment process training and experience of the treating staff, departmental waiting lists, the number of patients in the treatment queue at any one moment and the reliability of the equipment interact, and link back, to all the elements of the accuracy of set up and treatment delivery.

In addition to the factors above, the known variation which occurs in patient positioning can lead to variation in dose delivered to the target and to critical normal structures, in a way that is difficult to quantify for the individual patient. For this reason population estimates must be used, which adds a level of complexity to the correlation of both tumour control and toxicity with outcome. Moreover, there is substantial variability in outcome, even considering a uniform treatment. A large component of this variation is thought to relate to underlying genetic differences between individuals, which convolves an additional complexity with the other uncertainties.

An additional issue of complexity relates to the fact that the biological effectiveness of radiotherapy depends not only on total dose, but also on the dose per fraction and the overall treatment time. There are also questions about relative biological effectiveness (RBE) which can be relevant in treatments using protons or other light ions. The variation in outcome with total dose, dose per fraction and overall treatment time, including differences between tumour and normal tissue, provides a need for specific trial evaluation methodology.

RADIOTHERAPY AS A COMPLEX INTERVENTION

The MRC has issued guidance on developing and evaluating complex interventions, first in 2000 and more recently in 2009 [MRC 2009]. A complex intervention by their definition is one with several interacting components. There are several dimensions to this complexity, which relate to the range of possible outcomes, or their variability in the target population, rather than the number of elements in the intervention itself [MRC 2009]. This terminology has been developed in relation to public health interventions, not radical treatments, and so its use in the context of radiotherapy is not clear.

At the workshop, a vote indicated that 88% of attendees believed that radiotherapy did meet the definition of a 'complex intervention,' given that there are several interacting components to the treatment. However, some noted that it was unclear, at that stage, what the benefit was of being so classed. The most important part of the answer is that recognition of this complexity means alternative trial methodologies may need to be considered, for the evaluation of new (radiotherapy) technologies.

The implications of the definition of 'complex'

There are important implications for any treatment intervention which is considered complex. Firstly, for an intervention fitting the MRC definition, there may be important and appropriate alternative methodologies for evaluation other than the RCT [MRC 2009]. This issue of the need for alternative methodologies has also been highlighted elsewhere [Rawlins 2008, Sullivan et al. 2011]. In fact, the guidance encourages methodological experimentation, while requiring uncertainties to be quantified. Next, the presence of interactions between components requires careful consideration of the effects of changing any single one (Tables 1 and 2). Finally, it underlines the need for some element of quality assurance in clinical trials, the relevance of which has been demonstrated clinically [Peters et al. 2010]. This has been successfully addressed for radiotherapy in England by the central funding of the Radiotherapy Trials Quality Assurance (RTTQA) Group by the National Institute for Health Research and comparable bodies in the other UK countries, allowing a critical mass of expertise and support to be available for QA delivery.

THE EVIDENCE REQUIRED TO CHANGE CLINICAL PRACTICE

The evidence required to change clinical practice from the viewpoint of the National Institute for Health and Clinical Excellence (NICE) was presented by Dr Fergus Macbeth. Evaluation by NICE is considered in three areas:

- an interventional evaluation where safety and efficacy is assessed;
- a technical appraisal of clinical and cost effectiveness;
- a clinical guideline, which also looks at clinical and cost effectiveness.

Thus far, the only area where radiotherapy has been addressed is the third.

Typically NICE don't ask whether A is better than B, but whether A represents better *value for money* for the NHS than B. They therefore need information on relative effectiveness, additional benefit across all outcomes (including Quality Adjusted Life Years [QALYs]), relative costs, the reliability and relevance of evidence and estimates, and a judgement of the value society puts on an intervention.

There are several problems associated with evaluating radiotherapy techniques and technologies:

- a new intervention is not always identified as a distinct product;
- advances are often made in small incremental changes;
- numerous manufacturers are involved in creating technology;
- multiple indications (including rare tumours) are often involved;
- there are often large capital costs and other additional costs
- there are issues surrounding the assessment of learning curves once a new method has been introduced (e.g. speeding up time taken to plan)
- small modifications may be made over time;
- outcomes are often delayed.

Although evidence from RCTs is not always needed, the best available evidence is required [Rawlins 2008]. This distinction may be relevant where randomised trials may not be possible, ethical or generalisable to routine practice, or when observational evidence may be compelling. The specific question of when randomised trials might be unnecessary has been addressed by Glasziou and colleagues [Glasziou et al. 2007]. They suggested that a trial is not needed if the probability of two sets of data coming from the same group is less than 0.01, or the presence of a large treatment effect, for example when the event rate ratio is greater than 10:1.

There can be difficulties in judging how much additional cost should be covered by the NHS where a new treatment is more expensive than an existing one, and NICE has to trade outcomes against cost when assessing a treatment.

In discussion, costs formed an important theme. Improving cure rates was considered likely to prove cost effective, whereas showing cost effectiveness based on a reduction in toxicity is much more difficult. It was also noted that although the CHART trial of continuous, hyperfractionated, accelerated radiotherapy in lung cancer demonstrated an improvement in 2-year survival from 21% to 30% [Saunders et al. 1999], with a cost per QALY of approximately £2000, it is not widely used. Although it is available in a few centres, there are practical obstacles to implementation, such as radiographers and institutional reluctance regarding working late evenings and weekends.

The cost to improve survival would reduce if it was known which patients would benefit, and biomarkers may help more accurately identify who will benefit from certain treatments. For example, there is good evidence from the CHART head and neck cancer trial that the level of tumour expression of epidermal growth factor receptor (EGFR) predicts benefit from dose acceleration [Bentzen, Atasoy et al. 2005]. There is scope to develop this area, and work is

being undertaken to investigate possible biomarkers that predict benefit from different dose/fractionation schedules in prostate cancer.

It may be possible to develop evidence of benefit from a dose change, such as a reduction in toxicity or increase in tumour control for some of the more common tumours, but can this information be generalised? Indeed, can the results of such a trial at one tumour site be extrapolated to others? It was felt that although results from a trial cannot directly be applied to other tumour sites, randomised trials encourage the installation and use of new equipment, and develop training and experience. Clinical trial evidence for a technique may also encourage users to seek ways to apply it to other sites.

CASE STUDY OF AN EMBEDDED IMRT QUESTION: THE ARISTOTLE TRIAL

A discussion of embedding a radiotherapy question into another trial was presented by Prof David Sebag-Montefiore, using the ARISTOTLE trial as a case study. ARISTOTLE is a phase III trial in rectal cancer. The primary objective of the trial is to investigate the benefit of adding a second targeted drug to chemo-radiotherapy, so is not a pure radiotherapy question. In development, it was considered what radiotherapy questions could be answered within the trial. Questions could potentially have been asked about radiotherapy dose (e.g. escalate to all or just high risk subsites) and volume (IMRT vs Conformal RT) in a 2x2 design. These were considered but rejected in ARISTOTLE, partly as the result of a perceived lack of clinical support and momentum regarding radiotherapy research when the trial was being designed and the complexity of incorporating separate radiotherapy questions into the study design. Many important questions remain, for example relating to complex imaging, target volume delineation, technique, delivery, and QA, but the timing was not right as ARISTOTLE was being set up.

It was felt that the IMRT versus conformal 3D question would not be helpful since benefits from IMRT were most likely to be identified if the standard RT volume was large, which was not the case here. Additionally there was little data on IMRT with chemoradiotherapy. If a specific radiotherapy question is to be asked within a trial, it may be better done at the phase II stage, and not as a bolt-on to a phase III study.

There are questions remaining over target volume definition, especially with regards to microscopic areas at risk and the fact that anatomical knowledge can vary between operators. This was addressed in a detailed trial protocol and a quality assurance programme coordinated by the RTTQA group.

It is vital for us as a community to realise that the UK can deliver a high quality trials structure, with dedicated study groups, international collaboration, and a high quality of pathological assessment.

METHODOLOGICAL ISSUES AROUND EMBEDDING A RADIOTHERAPY QUESTION WITHIN A CLINICAL TRIAL

A number of methodological issues arise when embedding a radiotherapy question within a clinical trial, which were discussed by Dr Emma Hall. Many trials allow radiotherapy to be given in accordance with a centre's standard practice if it is 'background' treatment. It is often possible to exploit this centre heterogeneity to ask specific radiotherapy questions within a clinical trial, even if the radiotherapy is not the main focus. If there is a good question that needs asking and relevant endpoints, it is worth considering whether it is possible to address it within the trial, and consider whether a randomisation can be used.

Careful consideration is needed as to whether the main trial endpoint is likely to have an impact on the radiotherapy question endpoint, and conversely whether the addition of a radiotherapy question is going to compromise the main trial question. An example might be that the additional complexity affects patient or centre participation rates.

There are two ways in which a radiotherapy question can be embedded into a trial:

- Add a radiotherapy randomisation, using either a second randomisation, a factorial design, or a split plot factorial design (explored further below)
- Add the radiotherapy question without an additional randomisation.

The following issues need to be considered at the design stage:

- Should all patients be included? Should all centres or only certain centres be included in the embedded study?
- What proportion of patients will enter?
- Is the radiotherapy randomisation stratified by the main randomisation?

If a second randomisation is used, its timing is important: will it be done at the same time as the main randomisation (an advantage is the need for a single consent process), or just prior to the radiotherapy if given later? Can the second randomisation be adequately powered? Non-inferiority designs are often the relevant approach, but typically need to be large, while sample sizes are generally driven by the main trial question. If the target recruitment number is more than sufficient, do all centres need to take part in the second randomisation? If a fully powered radiotherapy question is not possible, is an underpowered feasibility question of value? Would this provide useful pilot data, could the data still contribute to a meta-analysis, or could more data be acquired by asking the same question across a number of trials and disease sites? If a radiotherapy question is to be added after the trial has started, there are a number of practical issues that will need addressing. For example, a major protocol amendment will be needed, ethics approval will need to be sought, additional resource may be required, and it will need to be implemented before the main trial finishes.

A factorial design can be very efficient and can look at interactions between treatments, but the study may be underpowered if significant interactions exist. Patients also have to be appropriate for all possible arms. An alternative is to allow sites/patients to opt out of one of the randomisations leading to an unbalanced factorial design. This design can increase accrual, but

be more complicated, with respect to the patient information sheets and the statistics required for analysis, and the trial may not be powered to detect interactions. An alternative option is a split-plot design which mixes cluster and individual randomisation, and in which you randomise centres, and then randomise patients within each centre. This may have use in radiotherapy trials where the cluster is the centre.

An important question is whether radiotherapy heterogeneity matters. Can different schedules be allowed? If so, this will maximise recruitment as centres are not forced to perform radiotherapy that may be different from their standard practice. It does not pose a big problem with respect to analysis, because the trial can be stratified by centre. However, it may affect the power to detect radiotherapy effects. If heterogeneity of radiotherapy is to be allowed, subgroup analyses must be defined *ab initio*, and sufficient radiotherapy data need to be collected for heterogeneity to be explored.

If a second randomisation for a radiotherapy question proves too difficult, options include:

- patient preference designs, with groups of randomised patients and groups on non-randomised patients
- observational studies
- non-randomised subgroup comparisons
- use of other sources of data.

Creating a standard minimum radiotherapy dataset, as is now happening in the UK [NatCanSAT 2011], may help answer radiotherapy questions, particularly if data collection can be expanded to include physics data.

In discussion, the issue of the need to fund preparatory work, which may need to be very detailed, was raised. Such work might include evaluation of target volume delineation and redefining anatomical limits for target volumes. Most centres do not have funding available for this preparatory work, so it can be very difficult to undertake and may be possible only at centres (currently only 1) with core funding. A process is needed to address this issue.

Collecting data on technologies and protocols at different centres, or about patients who refuse to be randomised, can provide observational data that adds a richness to what is already available from a trial and helps to generate hypotheses.

Thus embedding a radiotherapy question in a trial may be a way to obtain high quality data on a radiotherapy question but it needs to be considered early in the trial design process, preferable starting at the phase 2 stage, with detailed statistical and methodological input from the start.

LEARNING FROM SURGERY

Prof Jane Blazeby presented the challenges and experience of evaluating surgical innovation. She outlined the enormous developments that have taken place in major surgical procedures, many of which are now carried out as day cases and using minimally invasive surgery. She

noted the lack of regulation over the introduction of new techniques, which often evolve from a single surgeon trying something new.

The question of how to evaluate new techniques and whether the drug trial phase I–IV paradigm could be applied to surgery was considered several years ago. A meeting of methodologists and surgical experts was convened at Balliol College in Oxford for four 3-day meetings – the Balliol Collaboration – to discuss the evaluation of surgical innovation. Ultimately three papers appeared together in *The Lancet* in late 2009 covering current practice, challenges and a framework for progress, respectively [McCulloch et al. 2009, Barkun et al. 2009, Ergina et al. 2009]. These three topics formed the basis for the presentation.

Current practice

How do new surgical techniques arise? New operations may evolve from emergency situations where innovative solutions are needed or from ideas within surgical team. When new procedures are undertaken there is often planning and discussion with colleagues and careful selection of appropriate (slim, fit) patients. Animal work occasionally may precede the introduction of procedures in practice. Operations are evaluated on a case by case basis by the surgeon, who uses the experience and immediate outcomes to judge what did or did not work and why. Techniques are gradually refined based on iterations in selected patients. The outcomes that influence innovation are immediate harm and technical outcomes rather than longer term patient reported outcomes. Case reports are published, promising results are discussed with others who may decide to take up the new procedure and this leads to reports of case series, and later retrospective studies. Finally, approaches that appear 'to work' move gradually into practice. Laparoscopic cholecystectomy evolved in this way and is now practiced almost universally. However, there have been few well design and conducted randomised trials comparing laparoscopic with open cholecystectomy, so those carrying out open operations can claim justification in doing so. If problems arise with a new technique, alternative approaches may be tried, but if unsuccessful they are simply abandoned and generally not reported.

Challenges

Several challenges exist:

i. Like radiotherapy, surgery is a complex intervention as defined by the MRC (with several independent and interdependent variables that influence outcomes). This may include variables of the operation or pre or post operative care Operative variables include the components of a procedure, surgeon expertise, availability of equipment and team experience.

ii. Selection of outcome measures to evaluate surgical interventions is difficult. For example, 'anastomotic leak' has 56 definitions in the literature and few studies make an attempt to blind the outcome assessor – which is most often the surgeon themselves. Comprehensive evaluation of many surgical interventions with patient reported outcomes are lacking.

iii. The surgical culture is not evidence-based. This may occur because surgeons are often unfamiliar with how to recruit patients, and because of challenges in selecting the best time to evaluate a new procedure. If a trial is performed while the procedure is still evolving the results will be criticised as being reflective of the learning curve, and trials carried out at a later date may be difficult because surgeons no longer want to perform the standard comparison arm of the trial.

Framework

The Balliol Collaboration proposed a research framework, developed at the Balliol Collaboration, for surgical innovation is 'IDEAL' – Idea, Development, Exploration, Assessment and Long-term study**.** For example:

- Idea – first patient, highly selected, technical outcomes assessed
- Develop – more patients, selection processes documented, short term clinical outcomes published as case series
- Explore – larger series, broader inclusion criteria, more surgeons – carefully document and learn from mistakes (maybe small single centre efficacy trials)
- Assess – compare outcomes with the standard approach in a prospective multi-centre randomised trial
- Long-term study – audit and registries.

An example of attempted regulation of new surgical procedures is that of laparoscopic surgery for colorectal cancer. In 2002, NICE recommended that laparoscopic surgery should be performed within the context of RCTs. The UK and international trials were performed and after RCT results became available, the NICE guidance was updated in 2006 to note that it was an acceptable approach, and now it is recommended. Traditionally, trial QA is done through review of the pathology specimens. The MRC complex intervention guidance emphasises qualitative research. With respect to surgery, this may mean using electronic means, such as DVD recordings in theatre, to allow independent analysis of techniques and to identify problems and for data to be analysed in depth to provide surgeons participating in trials with feedback.

How should the process of surgical innovation be regulated? Should the consent process be taken away from the surgeon? As surgeons are notoriously bad at writing up papers, releasing DVDs of operations on the internet may be an alternative approach for allowing peer review. As for radiotherapy, there is a lack of funding for developmental research but schemes for feasibility trials do exist. The competitive spirit typical of surgeons should be harnessed with respect to trial recruitment. While the Balliol meetings led to an understanding for the methodologists of the surgeon's personal involvement in patient outcomes, there was perhaps less acceptance that roadblocks to trials are just excuses – RCTs are needed and are possible. To quote Prof Blazeby: "Any alternative is less than IDEAL".

THE DIFFICULTIES OF USING TRIALS TO ADDRESS QUESTIONS IN SPECIFIC SITUATIONS OF TECHNOLOGICAL DEVELOPMENT

Several situations had been identified as presenting special problems for evaluation using conventional trial approaches. One common theme with technology based interventions is to define the appropriate level of trial-specific and patient level QA necessary to demonstrate adequate implementation and use of the technology. QA is increasingly seen as a barrier for trial participation by many departments, and it is a research priority to define an adequate and evidence-based system of trial QA [Bekelman et al. 2012].

Image-guided radiotherapy – Dr Andrew Jackson

IGRT uses 3D imaging in the treatment room to maintain accuracy and/or refine treatment delivery. IGRT can be used in a number of ways, such as correcting patient or target position based on the position of bony anatomy or soft tissue structures, or by changing the shape of the volume to be treated in response to imaging during a course of treatment, so called adaptive radiotherapy. Studies show that IGRT is more accurate – but is it 'better'? From first principles, by allowing greater accuracy of treatment, IGRT could increase local control if it avoids geographical miss of tumour. However, we already treat a margin around tumour including a planning target volume (PTV) margin to account for variability in tumour position.

By giving greater accuracy, IGRT can allow the use of smaller margins, which in turn should reduce side effects. Often though, smaller margins are used to allow escalation of the radiotherapy dose, to increase local control while maintaining side effects at an acceptable (but not necessarily reduced) level.

Perhaps IGRT should be considered as a form of quality assurance, or even a natural evolution of technology, which therefore would not need evaluation in RCTs. However, potential disadvantages of IGRT include the expense of the equipment [Van de Werf et al. 2012], greater need for intervention by staff who require training, and additional X-ray exposure from in-room IGRT CT imaging. There are also concerns that reduction of PTV margins, even with IGRT may compromise tumour coverage, suggesting the need for some form of evaluation. With respect to trials, IGRT is probably best regarded as a tool and not a single concept, and it was argued that there is little value in a trial of IGRT versus no IGRT. If used specifically to reduce margins then an embedded RCT with outcome measures of local control and toxicity would be important.

NRAG recommended in 2007 that all new linear accelerators should be capable of IGRT [NRAG 2007], while an NCRI rapid review of the technology posited that if RCTs are seen as the only way to formally assess new technology they may well be bypassed, with technology creeping into practice without evaluation [NCRI 2008].

Stereotactic body radiotherapy (SBRT) for lung cancer – Dr Corinne Faivre-Finn

Though outcomes after surgery for stage I NSCLC are good, with ~60-70% survival at 5 years [Ost et al. 2008], few patients are suitable. For many, radiotherapy is a more appropriate alternative, but local control with standard conformal radiotherapy is poor.

SBRT combines a high biologically effective dose with extreme hypofractionation, providing acceleration, with precise delivery. Is there any evidence to support its use? Publications began appearing in 1994, but even now the 'evidence' comes from phase I/II studies, often single institution, and retrospective reviews. There is clear similarity to the surgical paradigm. SBRT has become the standard of care in highly specialised centres in Europe, Asia and the USA, but there are only five centres in the UK delivering this treatment.

Can we randomise SBRT vs standard radiotherapy? Those against argue that the strikingly better local control rate and reduced acute toxicity of SBRT, coupled with a lower number of patient visits, mean that equipoise does not exist. However, results with respect to survival are far from clear and there is little long term toxicity data. The improvement in local control may not translate in improved overall survival in this population with a high burden of comorbidities. Interestingly animal studies have suggested that extreme hypofractionation may increase the risk of distant metastasis. A possible hypothesis to explain this phenomenon is that local angiogenesis could be stimulated through an increase in TGFB and VEGF levels after local irradiation [Camphausen et al. 2001].

A possible study approach would be to establish a randomised trial comparing two different SBRT schedules with equivalent EQD2, keeping the overall treatment time below 3 weeks. Such a trial might, for example, test 60 Gy in 5 fractions versus 84 Gy in 15 fractions. The hypothesis of the study is that hypofractionated RT (84 Gy in 15 fractions over 3 weeks) will lead to reduced risk of metastatic spread with similar local control compared to extreme hypofractionated RT (60 Gy in 5 fractions over 2 weeks).

Proton therapy – Prof Gillies McKenna

Proton radiotherapy provides the perfect example of the conflict between those who mandate RCTs to evaluate a new technology and those enthusiasts who consider that withholding the new therapy would be unethical, at least in certain situations. This argument parallels the surgical paradigm.

An example of implementation of technology preceding evidence could be the progression of use from CT to MRI, and to a lesser extent to PET. In the radiotherapy field in particular, the use of port films, simulators, CT-based treatment planning, and even the introduction of linear accelerators to replace cobalt units, are examples of progress made without randomisation, in many cases with good reason. On the other hand, the use of IMRT has been subject to RCT, in the PARSPORT trial [Nutting et al 2011].

However randomised trials are not always a panacea. For example, the PROTECT prostate cancer trial of watchful waiting versus surgery versus radiotherapy [Donovan et al 2003], which

addresses an important question, may not answer it particularly with respect to radiotherapy because of the extensive heterogeneity of radiotherapy practices permitted in the study.

One issue in considering how to evaluate proton therapy is that of equipoise. This can apply at the level of the patient, the physician and society. An example might be a patient with rhabdomysarcoma of the skull base in whom conventional radiotherapy would inevitably blind the unaffected eye, which could be spared with a proton treatment.

Trials of equivalent doses using protons compared to X-rays are unlikely to show differences in tumour control; indeed local control would be expected to be the same. Differences might be seen with respect to normal tissue toxicity. However, randomised trials could address things that are not fully understood, such as the effect of hypofractionation, or the effect of dose-modifying agents. So, selected questions, but not all, may be appropriate for testing in RCTs.

RADIOTHERAPY FRACTIONATION TRIALS

Dr Neil Burnet introduced the concept of the Three Arm Randomised Trial with Two Experimental Arms (TART with TEA) as a means of evaluating different fractionation schemes. A good example is the START-A trial [START Trialists' Group 2008]. In this design, the trial is set up with one control arm using the standard schedule, and two experimental arms which both use the same experimental fractionation scheme but with a difference in dose. Thus the two experimental arms can be considered to be linked. The difference in dose can be achieved by giving one less fraction (e.g. the CHHiP trial [Khoo & Dearnaley 2008]) or a slightly different dose per fraction, with the same number of fractions (the START-A Trial) [START Trialists' Group 2008]. Typically, either strategy leads to a difference in dose of between 5 and 15%. Although this dose difference is small, provided the study is correctly designed, it is sufficient to provide the necessary information.

The three-arm study design has some particular additional advantages. It allows the extraction of biological information from clinical results, in this case defining the alpha/beta ratio for both tumour control and normal tissue. This allows assessment of the therapeutic ratio, though results for both endpoints (local control and toxicity) are required. Equally importantly, it allows projection of fractionation schedules to optimise the therapeutic ratio. Secondly, this method also allows internal validation of the sensitivity of the endpoints of the study, by testing whether they differ between the two experimental arms. Finally, it allows greater security in choice of experimental schedule, which typically requires a degree of guesswork, because the equivalence to the standard arm, for both endpoints, can be interpolated. This is particularly important where the tumour control endpoint may take many years to be reached, to provide a result, which is typical of radiotherapy endpoints.

The TART with TEA design should be considered a standard radiotherapy methodology for phase III trials, in common tumours where there is an interest in changing dose. Although specific to radiotherapy, the methodology may also have relevance for testing the combination

of new pharmaceutical agents in combination with radiotherapy, especially those that might alter the fractionation response by causing DNA damage or interfering with the damage response pathways. This is an important area for further development [Harrington et al. 2011].

GENERAL DISCUSSION

A vigorous and interesting discussion took place that spanned several of the areas presented.

Caution is certainly needed with new technology such as proton therapy, based on an article in the Journal of Medical Ethics [Hofmann 2009]. There can easily be fallacies in the arguments made in favour of a new technology, partly as the result of over optimistic attitudes towards technology and progress. This point was not to argue against proton therapy per se, but rather to point out that flawed arguments can actually hamper their implementation instead of promoting it. Patients deserve the best technology, not only on the basis of the best available evidence, but also on the basis of the best arguments. There are certainly difficulties in business planning for proton therapy centres, which may affect decisions on which patients to treat; the expense of the equipment for proton therapy also requires a large patient base to make a reasonable business case.

A question was raised on the possibility of using the TART with TEA design to test proton therapy against two standard radiotherapy schedules. The reverse, with two proton arms and an X-ray arm might also be possible. It was thought that this could be possible in theory, but would in practice depend on being able to recruit sufficient patient numbers. An alternative design could be to test the question of dose response by specifically dose escalating patients using protons, which could be an ethical approach.

With respect to IGRT, it was suggested that a method for acquiring evidence could be to use this routinely, but collect and review information retrospectively.

The potential risks of selection bias were also discussed. For example, with PET, there has been at least one study in staging lung cancer prior to surgery which showed no benefit to PET (in reducing thoracotomy rates), although there was a degree of selection in the patients enrolled [Viney et al 2004].

CHALLENGES FOR EVALUATING NEW RADIOTHERAPY TECHNOLOGIES

Prof Søren Bentzen noted the unquestionable advances that have been made since the earliest use of radiotherapy in the 1930s, based largely on an engineering paradigm. For example, the greater penetration and skin sparing achieved with megavoltage equipment, compared to orthovoltage X-rays, was clearly seen to reduce or avoid skin toxicity; the results were so clear that no RCT was needed.

Many questions (and much of the expensive new technology) concern dose distribution – how do we know that one dose distribution is better than another with respect to clinical benefit? Unlike drugs, for which evidence of benefit from phase III trials are required for FDA approval, medical devices can gain FDA approval with so-called 510(k) clearance – they are not required to show an improvement over current technology, only safety of the equipment when used as intended. For example, in trials of IMRT, what constitutes the appropriate level of evidence? Prof Bentzen argued that it is not sufficient to simply show better dose distribution, but rather it is essential to prove clinical benefit.

For example, does increased precision inevitably lead to better outcomes? Not necessarily – as seen in a paper by Engels and colleagues that explored the outcomes of patients with prostate cancer (T1-T3N0M0 tumours) [Engels et al. 2009]. In 213 patients, daily X-rays were used for positioning based on the location of bony structures, and in 25 patients positioning was based on direct prostate visualization with X-rays of implanted markers. Unexpectedly, on multivariate analysis, the use of implanted markers was associated with a worse rate of freedom from biochemical failure (FFBF). In retrospect, the group considered the margins around the clinical target volume to be inadequate in the cases in which markers were used due to over-reliance on the expected improved geometrical precision. This case demonstrates that technology developments can bring their own problems, and emphasises the need for clear protocols and vigilant quality assurance.

Certainly the UK community can deliver radiotherapy RCTs: CHHiP and START are excellent examples, and both finished patient recruitment ahead of schedule. Fractionation studies are very suitable for randomisation because you plan first, and then randomise. The issues are different with technology-based interventions, for example, IMRT versus standard radiotherapy, because the intervention is not necessarily well defined, and there are many possible alterations and potential biases in treatment planning that could depend on the technology itself.

Although radiotherapy QA appears to be of less concern where radiotherapy is the *background* (i.e. non-randomised) treatment, and represents standard practice, this is now known not to be a safe approach [Peters et al 2010]. Poor QA can lead to poor outcomes, potentially overwhelming the effect of the treatment being tested. When a radiotherapy technique or technology forms a major part of the intervention, careful protocol definitions and QA are vital, since a lack of precision in the radiotherapy planning and delivery allows everyone to interpret the results their own way. In addition, improved technology brings issues of its own. For example, patient motion presented less of a problem, clinically or within trials, when fields were bigger and less conformal.

Prof Bentzen reflected that the comparison of protons versus photons is challenging. Trials are difficult in settings such as this because people have developed fixed ideas about which patients are most likely to benefit from protons. Equipoise requires that only patients for whom there is uncertainty over benefit to be randomised, but sometimes this eliminates the majority of patients for whom benefit is likely from the trial population. This raises the question of whether the RCT will provide any useful answer. Prof Bentzen argued it would be unlikely to change practice whatever the outcome – positive results reinforcing the views of those in favour of

protons and negative results being dismissed as relating to an irrelevant population. There may also be specific issues to do with defining the intervention: physicians concerned about motion and range uncertainty in proton therapy may tend to be more liberal with target volume coverage; or for a photon plan, some physicians may compromise the tumour dose distribution to spare neighbouring organs at risk. In a trial patient, the two plans could be developed before randomization – but it does not really remove the problem: the planner would obviously not be blinded to the radiation modality when optimizing the two plans.

RCTs rank highly in level of evidence hierarchies but the quality of RCTs is rarely addressed. A key issue is to ensure trials are adequately powered for both efficacy and toxicity endpoints to define the therapeutic ratio. For example, it is easy to reduce toxicity by 10% – simply reduce the dose by 10%! The real question is whether reduced toxicity from an improved dose distribution, for example using IMRT or proton therapy, can be achieved without compromising efficacy.

Concerning efficacy endpoints, survival is of obvious importance. However, local control or time to local progression may be a more sensitive endpoint for the effect of modified radiotherapy, and therefore requires a smaller sample size of an RCT to test a difference. In addition there are competing causes of mortality. Patient reported outcomes (PROMS) are important; for instance, with parotid-sparing radiotherapy and xerostomia, if PROMS were not improved it would be questionable whether there would be any clinical benefit.

In summary, Prof Bentzen suggested that RCTs may be less problematic for questions linking technology to clinical or biological questions (for example, narrow versus wide margins, selective volume boost) or for studies of inter-patient dose escalation based on individual estimates of normal tissue complication probability. Depending on the answer to the biological question the required technology will be defined. In terms of modified dose distribution, pattern of failure or pattern of toxicity studies, provide useful alternative methodologies for evaluation. Likewise, when alternative technologies can achieve similar reduction in for example CTV to PTV margins or dose to a critical organ at risk, cost-utility analysis may be informative. Finally, it was noted that while we claim to want the unbiased answers that only an RCT can bring, many radiotherapy RCTs have failed to change practice.

EVALUATING NEW RADIOTHERAPY TECHNOLOGIES

To round off the day's formal presentations, Prof Lucinda Billingham explored the need for formal evaluation of radiotherapy technologies in clinical trials and considered the alternatives. Although some interventions are 'self-evidently' better due to their theoretical basis, there is a case to argue that formal evaluation is needed to investigate how the theoretical background translates into clinical reality, establishing the actual size of the clinical benefit and the cost implications. Sometimes a clinical trial can demonstrate findings that are unexpected, such as the trials of neutrons versus photons. Treatment of tumours with neutron radiotherapy was assumed by many (but not all) to make good radiobiological sense but in fact a trial in pelvic

tumours carried out at Clatterbridge Hospital showed that patients treated with neutrons had a higher rate of mortality due to metastatic disease within a year of treatment than those treated with photons [Errington et al 1991]. The results were controversial but such trials highlight that real world application does not always conform to expectations. This trial also highlighted one of the key conundrums with evaluating radiotherapy technologies, which health care providers need evidence of benefit before making major investment in a new technology, but investment in that technology is required to be able to assess its benefits. For example, the Clatterbridge Cancer Centre had to invest millions of pounds on a cyclotron to then find out that such technology did not appear to be providing benefit to certain groups of patients.

Having argued that formal evaluation of new radiotherapy technologies is important, it is acknowledged that the gold standard RCT is not always necessary or possible. Indeed, Sir Austin Bradford Hill, the architect of the RCT, said that, “Any belief that the controlled trial is the only way would mean not that the pendulum had swung too far but that it had come right off its hook” [Hill 1966]. Three different alternative approaches to the RCT were considered:

- Prospective non-randomised clinical trial of a new treatment with either historical or concurrent controls. Methods to improve the acceptability of results from non-randomised intervention studies have been reviewed [Deeks et al 2003], and using a prospective design with a clear statistical plan for case-mix adjustment ensures the regulatory rigor of a trial and deals with the problem of selection bias. A recommended approach for case mix adjustment is to use ‘propensity scoring’ but its role needs further evaluation [Deeks et al 2003].

- Stepped wedge design [Brown & Lilford 2006, MRC 2009]. In this, the order in which sites have new technology rolled out is randomly assigned. Adjustment is needed for the time effect, and each time period needs to be analysed carefully. Its potential role in application to new radiotherapy technologies needs further evaluation.

- Decision modelling. A model is constructed identifying possible patient pathways for different treatment options, each involving different health states and costs. Existing evidence in the literature is used to determine (i) the different probabilities of patients moving from one health state to another, (ii) the utilities attached to each health state reflecting the differing quality of life and (iii) the costs associated with each pathway. Hypothetical cohorts of patients are put through this model to compare clinical effectiveness and cost effectiveness of treatment. For example, a decision model was used to determine that a single fraction of RT was the most cost-effective palliative treatment for patients with hormone-refractory prostate cancer and bone metastases [Konski 2004], and decision modelling was used to conclude that particle therapy should not be adopted as standard therapy in non-small cell lung cancer and further evidence is required [Grutters et al 2010]. This approach has the advantage of being able to compare interventions that may be impossible to compare in a clinical trial but the legitimacy of the results are reliant on the validity of the model and the quality of the data attached to it.

In summary, the session concluded that the RCT remains the gold standard for evaluating new radiotherapy technologies but properly designed prospective non-randomised assessments of

radiotherapy technologies can provide worthy evidence where RCTs are not feasible and within this setting the stepped wedge design is an option that requires further research. Decision modelling provides an alternative to prospective evaluation and has been shown to be a useful tool in the field of radiotherapy but for definitive conclusions that properly inform future clinical practice it requires existing robust evidence to be incorporated into a valid complex clinical model.

INTERACTIVE PANEL DISCUSSION

A set of questions were posed to the participants concerning evaluation of new technologies, as part of the discussion, and views from the audience were collected by means of a short questionnaire (n=33 evaluable responses). The panel commented on each question thereafter to generate discussion. Overall, there was a good degree of openness to exploring alternative methodological options for radiotherapy and it was felt that some evidence is nearly always better than none.

Q1: Is radiotherapy a complex intervention, based on this workshop, and using the MRC definition?

Panel and audience comments: Given the number of interacting processes and uncertainties in the processes, radiotherapy would seem to fit the MRC definition of 'complex' intervention (Tables 1 and 2). As noted, this may help our thinking in developing studies, including developmental phase studies, which are currently not valued and therefore difficult to fund.

For the HTA, 'complex' therapies are typically 'black box' therapies, e.g. talking therapies, where it is not clear what the 'active' ingredient is. Radiotherapy, though more complex than a drug, is perhaps different in that the complexities can be identified and measured in many cases.

Participant questionnaire: This statement was overwhelmingly supported, with 87.9% agreeing or strongly agreeing that RT is a complex intervention.

Strongly agree	51.5%
Agree	36.4%
No strong opinion	3.0%
Disagree	3.0%
Strongly disagree	6.1%

Q2: Should cost effectiveness be a major influence on the decision to introduce a new radiotherapy technique into clinical practice?

Panel and audience comments: There are some very simple measures of cost that can be useful, e.g. speed of delivery of treatment. Quantification of cost needs to take account of the fact that initial costs almost always come down, and this cost reduction would need to be modelled within a cost effectiveness analysis.

Participant questionnaire: Two thirds (67.7%) agreed or strongly agreed that cost effectiveness should be a major influence on the introduction of new technology.

Strongly agree	20.6%
Agree	47.1%
No strong opinion	8.8%
Disagree	20.6%
Strongly disagree	2.9%

Q3. In general, the problems with including a second radiotherapy randomisation embedded within a phase III clinical trial outweigh the benefits.

Participant questionnaire: Two thirds (67.8%) disagreed or strongly disagreed that problems outweighed potential benefits of embedding a radiotherapy question in a Phase III trial.

Strongly agree	0%
Agree	12.9%
No strong opinion	19.4%
Disagree	48.4%
Strongly disagree	19.4%

Q4: Which of the following RT techniques do you believe can and should be evaluated in RCTs vs standard RT?

- **SBRT for lung cancer**
- **IMRT to reduce proctitis in prostate cancer**
- **IGRT in pancreatic cancer**
- **Proton therapy for paediatric tumours**
- **Alternative fractionation for H&N tumours**.

Panel and audience comments: In discussion, the question was raised whether prospective, well designed studies that are not RCTs could get funded. It was noted that the HTA will consider non-randomised trials provided they are very well designed and address selection bias; experience shows that the problem of selection bias is often not addressed well enough. The HTA also consider that evidence synthesis and decision analysis modelling is an area of interest. Similarly, the Clinical Trials Awards and Advisory Committee (CTAAC) at Cancer Research UK has a remit to support both RCTs and other well designed studies, provided the design is justified.

In considering alternative methodologies, it should be considered that positive results of some RCTs have proved difficult to implement. It is perhaps even more likely that effective treatments evaluated in *non*-RCTs will not be implemented.

Participant questionnaire: In general, participants did not strongly favour RCTs to evaluate these new approaches, with only 13–29% supporting an RCT. The greatest support (29%) for an RCT was in the evaluation of SBRT for lung cancer and fractionation in head and neck cancer.

Setting	**Percentage saying 'yes' to RCT**
SBRT for lung cancer	28.9%
IMRT to reduce proctitis in prostate cancer	14.4%
IGRT in pancreatic cancer	13.3%
Proton therapy for paediatric tumours	14.4%
Alternative fractionation for H&N tumours	28.9%

Q5: A funder is interested in building a new £multi-million proton therapy facility – what would you advise?

- **Don't build and funnel resource to other RT initiatives, sending patients to existing facilities abroad if appropriate**
- **Delay building, wait for more data from existing facilities**
- **Build facility and use to evaluate proton therapy in prospective studies**
- **Build facility and assume beneficial to patients without conducting prospective study**
- **Other**

Participant questionnaire: In response to this question, three-quarters (75.7%) felt that a facility should be built, and the vast majority of those in favour of building felt that the facility should be used to evaluate proton therapy in prospective studies.

Don't build and funnel resource to other RT initiatives, sending patients to existing facilities abroad if appropriate	9.1%
Delay building, wait for more data from existing facilities	9.1%
Build facility and use to evaluate proton therapy in prospective studies	72.7%
Build facility and assume beneficial to patients without conducting prospective study	3.0%
Other	6.1%

Q6: After this workshop, how do you feel about more creative approaches for evaluating RT interventions, such as those discussed today?

- **Keen and intend to incorporate in future projects**
- **Interested but would need further information/ guidance**
- **Open-minded but concerned about practicality of delivery**
- **Sceptical and see barriers as being too great**

- **The traditional RCT approach will always be best.**

Participant questionnaire: Participants were polarised in answer to this question, with a quarter citing RCTs as their preferred method of evaluating new technology, but with 75% supportive of the concept of alternative methodologies. While 16% of the group felt empowered to incorporate such methods in future research, an additional 52% felt that they would be interested in alternatives though would need guidance to do so. This strongly suggests the need for further work both to refine the details of alternative methodologies to apply specifically to radiotherapy, as well as to promote and explain these within the community.

Keen and intend to incorporate in future projects	15.9%
Interested but would need further information/ guidance	52.3%
Open-minded but concerned about practicality of delivery	6.8%
Sceptical and see barriers as being too great	0%
The traditional RCT approach will always be best	25.0%

CONCLUSIONS

As a community, radiotherapy researchers must think harder about trial design and engage methodologists in the planning process. It was encouraging to note that at the conclusion of the meeting, three-quarters of participants were open to alternative methodologies beyond the RCT. The challenge remains to design studies to provide the strongest possible evidence base for our developing technologies.

REFERENCES

Barkun JS, Aronson JK, Feldman LS, et al. Evaluation and stages of surgical innovations. *Lancet* 2009; 374:1089–96.

Barton MB, Gebski V, Manderson C, Langlands AO. Radiation therapy: are we getting value for money? *Clin Oncol (R Coll Radiol)* 1995;7(5):287–92.

Bekelman JE, Deye JA, Vikram B, Bentzen SM, Bruner D, Curran WJ Jr, Dignam J, Efstathiou JA, Fitzgerald TJ, Hurkmans C, Ibbott GS, Lee JJ, Merchant TE, Michalski J, Palta JR, Simon R, Ten Haken RK, Timmerman R, Tunis S, Coleman CN, Purdy J. Redesigning radiotherapy quality assurance: opportunities to develop an efficient, evidence-based system to support clinical trials-report of the national cancer institute work group on radiotherapy quality assurance*. Int J Radiat Oncol Biol Phys* 2012; 83(3):782-90.

Bentzen SM, Atasoy BM, Daley FM, Dische S, Richman PI, Saunders MI, Trott KR, Wilson GD. Epidermal growth factor receptor expression in pretreatment biopsies from head and neck squamous cell carcinoma as a predictive factor for a benefit from accelerated radiation therapy in a randomized controlled trial. *J Clin Oncol*. 2005; 23(24): 5560-7.

Bentzen SM, Heeren G, Cottier B, et al. Towards evidence-based guidelines for radiotherapy infrastructure and staffing needs in Europe: the ESTRO QUARTS project. *Radiother Oncol* 2005;75:355–65.

British Institute of Radiology: Geometric Uncertainties in Radiotherapy – Defining the Planning Target Volume. British Institute of Radiology, London, 2003.

Brown CA, Lilford RJ. The stepped wedge design: a systematic review. *BMC Medical Research Methodology* 2006, 6:54.

Camphausen K, Moses MA, Beecken WD, Khan MK, Folkman J, O'Reilly MS. Radiation therapy to a primary tumor accelerates metastatic growth in mice. *Cancer Res* 2001;61*:* 2207-2211.

Daisne JF, Duprez T, Weynand B, et al. Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG PET and validation with surgical specimen. *Radiology* 2004;233:93–100.

Deeks JJ. Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C. Song F, Petticrew M, Altman DG. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003: 7(27).

Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, Jewell D, Powell P, Gillatt D, Dedman D, Mills N, Smith M, Noble S, Lane A; ProtecT Study Group. Prostate Testing for Cancer and Treatment (ProtecT) feasibility study. *Health Technol Assess*. 2003;7(14):1-88.

Engels B, Soete G, Verellen D, Storme G. Conformal arc radiotherapy for prostate cancer: increased biochemical failure in patients with distended rectum on the planning computed

tomogram despite image guidance by implanted markers. *Int J Radiat Oncol Biol Phys* 2009;74:388–91.

Ergina PL, Cook JA, Blazeby JM, et al. Challenges in evaluating surgical innovation. *Lancet* 2009;374:1097–104.

Errington RD, Ashby D, Gore SM, Abrams KR, Myint S, Bonnett DE, Blake SW. Saxton TE. High energy neutron treatment for pelvic cancers: study stopped because of increased mortality. *BMJ* 1991;302:1045-1051.

Glasziou P, Chalmers I, Rawlins M, McCulloch P. When are randomised trials unnecessary? Picking signal from noise. *BMJ* 2007;334:349–51.

Grutters JP, Pijls-Johannesma M, Ruysscher DD, Peeters A, Reimoser S, Severens JL, Lambin P, Joore MA. The cost-effectiveness of particle therapy in non-small cell lung cancer: exploring decision uncertainty and areas for future research. *Cancer Treat Rev* 2010;36(6):468-76.

Harrington KJ, Billingham LJ, Brunner TB, et al. Guidelines for preclinical and early phase clinical assessment of novel radiosensitisers. *Br J Cancer* 2011;105:628–39.

Hill AB. Heberden Oration,1965: Reflections on the controlled trial. *Ann Rheum Dis* 1966;25:107-13.

Hofmann B. Fallacies in the arguments for new technology: the case of proton therapy. *J Med Ethics* 2009;35:684–7.

Khoo VS, Dearnaley DP. Question of dose, fractionation and technique: ingredients for testing hypofractionation in prostate cancer--the CHHiP trial. *Clin Oncol (R Coll Radiol)* 2008;20:12–4.

Konski A. Radiotherapy is a cost-effective palliative treatment for patients with bone metastasis from prostate cancer. *Int J Rad Onc Biol Phys* 2004; 60: 1373-1378.

McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374:1105–12.

Medical Research Council. Developing and evaluating complex interventions: new guidance. MRC, London. 2009 [cited 17 December 2011]. Available from: http://www.mrc.ac.uk/complexinterventionsguidance

National Cancer Research Institute (NCRI). Rapid review of radiotherapy and associated radiobiology. 2008 [cited 17 December 2011]. Available from: http://www.ncri.org.uk/includes/Publications/reports/radiotherapyreport08_web.pdf

National Cancer Services Analysis Team (NatCanSAT). Radiotherapy Dataset (RTDS). 2011 [cited 17 December 2011]. Available from: http://www.canceruk.net/rtservices/rtds/

National Radiotherapy Advisory Group (NRAG). Radiotherapy: developing a world class service for England. 2007 [cited 17 December 2011]. Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_074575

Nutting CM, Morden JP, Harrington KJ, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. *Lancet Oncol* 2011;12:127–36.

Ost D, Goldberg J, Rolnitzky L, Rom WN. Survival after Surgery in Stage IA and IB Non-Small Cell Lung Cancer. *Am J Respir Crit Care Med* 2008: 177(5): 516–523.

Peters LJ, O'Sullivan B, Giralt J, et al. Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: results from TROG 02.02. *J Clin Oncol* 2010;28:2996–3001.

Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet* 2008;372:2152–61.

Saunders M, Dische S, Barrett A, et al. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer: mature data from the randomised multicentre trial. CHART Steering committee. *Radiother Oncol* 1999;52:137–48.

START Trialists' Group, Bentzen SM, Agrawal RK, et al. The UK Standardisation of Breast Radiotherapy (START) Trial A of radiotherapy hypofractionation for treatment of early breast cancer: a randomised trial. *Lancet Oncol* 2008;9:331–41.

Sullivan R, Peppercorn J, Sikora K, et al. Delivering affordable cancer care in high-income countries. *Lancet Oncol* 2011;12:933–80.

Tubiana M. The role of local treatment in the cure of cancer. *Eur J Cancer* 1992;28A:2061–9.

Van de Werf E, Verstraete J, Lievens Y. The cost of radiotherapy in a decade of technology evolution. *Radiother Oncol* 2012; 102(1):148-53.

Viney RC, Boyer MJ, King MT, Kenny PM, Pollicino CA, McLean JM, McCaughan BC, Fulham MJ. Randomized controlled trial of the role of positron emission tomography in the management of stage I and II non-small-cell lung cancer. *J Clin Oncol.* 2004; 22(12): 2357-62.

TABLES

Table 1. The Radiotherapy Chain, focusing on physical components. The interactions are shown, and typically indicate the need for feedback loops between the stages.

Component		Interactions
1.	Diagnostic imaging	
2.	Immobilisation	
3.	Imaging for planning (including choice of modality) • CT • MRI, PET, MRS • Ultrasound • Image co-registration	1, 2
4.	Volume delineation – target (TVD) and normal tissue structures • GTV • CTV • PTV • Organs at risk (OARs) and Planning at Risk Volumes (PRVs) • Operator variation in TVD, uncertainties in imaging	1, 2, 3, 5
5.	Preparation of dose plan • Choice of technique – CRT or IMRT, SBRT • Choice of target dose(s) • Choice of normal tissue doses and constraints; prioritisation	2, 3, 4
6.	Quality assurance • Machine QA • QA of individual patient plan (Delivery QA)	2, 4, 5
7.	Pre-radiotherapy verification • Simulator or virtual	2, 4, 5
8.	On-treatment verification • EPI, Film • SBRT frame • In vivo dosimetry – TLD, Diode • IGRT – off line, on line, daily, limited fractions	2, 3, 4, 5, 7

CT = computerised tomography
CRT = conformal radiotherapy
CTV = clinical target volume
EPI = electronic portal imaging
GTV = gross tumour volume
IGRT = image-guided radiotherapy
IMRT = intensity-modulated radiotherapy
MRI = magnetic resonance imaging
MRS = magnetic resonance spectroscopy
PET = positron emission tomography
PTV = planning target volume
SBRT = stereotactic body radiotherapy
TLD = thermoluminescent dosimetry

Table 2. Behavioural interactions in the radiotherapy pathway

Activity	Interacts with
Treatment decision	Technology availability Experience Imaging Waiting list and targets Clinician view of acceptable toxicity risk
Target volume delineation	Technology availability Imaging Assessment of image co-registration accuracy Individual variation in delineation, inter- and intra-operator Set up accuracy data PTV margin PRV margin Balance of target edge dose versus PRV dose
Plan preparation	Clinician specification of dose limits Error analysis Availability of image-guided radiotherapy Planning system software Patient anatomy
Treatment delivery	Patient compliance & intra-fraction stability Patient anatomy Planning margins Availability of image-guided radiotherapy Training and experience Number of patients in queue Machine reliability

PRV = Planning organ at Risk Volume; PTV = Planning Target Volume