Public Paper no. 1

CHI+MED: Multidisciplinary Computer–Human Interaction research for the design and safe use of interactive medical devices

Ann Blandford, Paul Curzon, Harold Thimbleby, and others

Publicly released version of the original proposal accepted by EPSRC

PP release date: 10 March 2010

file: WP001.pdf
Objectives

List the main objectives of the proposed research in order of priority [up to 4000 chars]

This programme has two overall aims:

1) to improve the Computer-Human Interaction (CHI) science and engineering of widely used safety-critical systems, taking interactive medical devices as exemplars, and

2) to build a dialogue between medical stakeholders (device designers, investigators, health policy makers, health service managers, patients, carers) and the CHI (also known as HCI) community.

This dialogue will result in stakeholders having a better understanding of HCI as pertinent to the design, selection and use of medical devices, and HCI researchers developing a deep understanding of stakeholder cultures and concerns. The focus will be on interactive devices such as infusion pumps and monitoring devices that are widely used by both clinicians and (increasingly) patients. CHI+MED complements ongoing initiatives in Design for Patient Safety, addressing HCI design with much greater depth and rigour.

The specific objectives are to:

* improve understanding of how medical devices are used in practice (by different user populations and in different situations), focusing on how situations can provoke errors (e.g., through distraction) or mitigate errors (e.g., through independent checks of actions, or by well designed reminders), and how interaction design can improve usability and minimise errors.

* investigate how systematic descriptions of medical devices (focusing on their interactive behaviour) can support reasoning about device properties of safety and usability.

* deliver an improved understanding of human error, particularly cognitive slips, and the role of design in provoking or mitigating slips, particularly by being grounded in accurate device and situated models.

* provide evidence-based theories, tools and techniques that support developers and evaluators in reasoning about medical device design from an HCI and safety perspective.

* develop and test device redesigns that improve user interaction (particularly reducing human error).

* develop a dialogue with stakeholders, including patients and the general public, on issues of human error and device design, cognition, computing and HCI, using medical devices as a shared focus.

* move interaction design up the stakeholder agenda as a recognised but manageable source of error, and hence of medical incidents. This will include engaging in dialogue with regulators, policy makers, investigators, purchasers, etc., to support them in reasoning about the quality of interaction design, empowering them to make informed decisions, integrate HCI understanding in investigations, and inform future policy, standards and regulations regarding device design.

* develop a transferable framework for linking scientific developments with stakeholder engagement (such that the dialogue with stakeholders and the scientific investigations mutually inform each other).

Summary

Describe the proposed research in simple terms in a way that could be publicised to a general audience [up to 4000 chars]. Note that this summary will be automatically published on EPSRC’s website in the event that a grant is awarded.

Patient safety is a major concern. Reliance on interactive medical devices is growing, both in clinical settings and, increasingly, for patients without direct clinical supervision. The usability and reliability of such devices is critical. For example, decimal points are a well-known source of error (e.g., .5mg misread as 5mg), yet few devices detect decimal keying errors. Considering the broader context of use, a nurse familiar with one kind of infusion pump may absent-mindedly use the same set-up procedure on a similar one, leading to incorrect dosage. These and other user programming errors cause patient deaths. Even when devices are programmed correctly, interaction difficulties raise workload and stress, increasing overall system vulnerability. For example, it is common for clinicians to turn pumps off and on to reset the state, but patient data may be lost in the process and need to be re-entered. Better interaction design, the focus of CHI+MED, will improve safety by a scientific approach to understanding and designing out latent errors. It will have very broad impact because interactive systems are encroaching everywhere in healthcare. CHI+MED complements other initiatives on the ergonomics and engineering of safer devices.

CHI+MED is multidisciplinary. A combination of empirical approaches, drawn from the disciplines of HCI, computer science, psychology and the social sciences, will be applied. Empirical findings will build on, be subjected to, and be challenged by precise computational representation and reasoning. This will support integration across perspectives and generalisation of findings, as well as enforcing a distinctive rigour in all aspects of analysis. The CHI+MED approach will advance, and transform, the science of user-device interactions and the design of interactive medical devices.
We will investigate which device properties are most important from safety and usability perspectives (relating to whether adverse incidents can occur, whether there are unsafe inconsistencies between device behaviours under similar conditions, etc.), and which system descriptions best support reasoning about them. We will study the causes of errors when interacting with devices and how a better science of error can improve design. This will include factors affecting slips such as stress, attention, task structure and the design of cues and feedback. We will also study the broader situational factors that provoke or mitigate interaction errors (interruptions may provoke errors; colleagues checking may mitigate them, etc.). CHI+MED will deliver a significantly better understanding of interaction design and associated cognitive and situational factors, new tools to support design and evaluation, underlying better informed standards and incident investigations, and more resilient practices.

As well as addressing new scientific questions that generalise beyond the medical domain, CHI+MED will develop a dialogue with stakeholders, to raise awareness of the role of interaction design in medical errors and to develop evidence-based techniques for designing medical devices that mitigate potential errors while accommodating the cultures and processes of stakeholder groups. It will draw on research in Public Engagement and Knowledge Transfer, and explore the appropriateness and effectiveness of different approaches for different contexts. We will investigate how informed principles and tools can help developers to design more reliable and usable devices, identify factors that influence design and procurement decisions, and explore how stakeholders can be better informed about interaction design.

CHI+MED involves and integrates three leading groups on a single, focused problem. It will contribute to the UK's status as an international leader in research in multidisciplinary, rigorous HCI, and make significant and informed contributions to the international agenda on design for patient safety.

Beneficiaries
Describe who will benefit from the research [up to 4000 chars].

The aims of this research are both to inform practice (and thereby improve patient safety) and to extend the science of HCI.

CHI+MED will inform practice: this research will save lives and improve the standard of living of those with long-term illnesses that critically depend on living with medical devices. Patients will be less vulnerable to human errors provoked by interaction design, and less stressed by difficulties of interacting with devices that they are expected to use unsupervised (e.g., ambulatory or in the home). Medical practitioners will be better able to focus on their primary task of caring for patients, with fewer distractions due to poor device design. Healthcare (in the UK, primarily the NHS) will benefit through improved training and performance and reduced threats to patient safety which will reduce the frequency and duration of hospitalisation. Regulatory and policy organisations such as the Medical and Healthcare Products Regulatory Agency (MHRA) and the National Patient Safety Agency (NPSA) will gain new ways of assessing devices in terms of interaction design. Those investigating clinical incidents will be better equipped with information about interaction design where medical devices are involved.

UK (and international) manufacturers will acquire new techniques they can apply for improving interaction design in product development, which will give competitive advantage in the marketplace. This will grow in importance as the critical role of device design becomes more widely recognised, leading to increasing pressure on manufacturers to take account of designing for safety (including, but not limited to, interaction design): they will need the understanding and tools, such as this project will deliver, to ensure better design. This will create a sustained and achievable change to research and industry.

The core scientific developments will benefit the HCI and related research communities: the representation and analysis/testing of device designs will progress new interaction programming approaches as well as developing and testing novel design solutions, while the novel methods being developed and tested will have value beyond the medical domain. The study of human error will contribute to research in cognitive science and experimental psychology. The computational reasoning work will extend the applications of formal and mathematical methods. The situated studies will extend the understanding of the use of medical devices across all the settings in which they are used. The multidisciplinary approach being taken can serve as an exemplar for other areas of HCI (few projects have brought together the range of expertise we propose here or integrated the approaches so tightly).

The framework for science-stakeholder engagement will advance the understanding of public engagement in science and engineering, which is of concern to science funders (e.g. EPSRC) as well as many scientists, engineers and policy makers.

CHI+MED will raise the profile of UK HCI research: the UK already has a strong track record in rigorous, engineering approaches to HCI, but this project will have the critical mass to make a significant impact in the area of HCI for safety-critical systems interaction design, making the UK an international leader in this area. As a distributed, multisite programme, the impact will be distributed across the UK rather than being restricted to a single university group. CHI+MED will complement and augment recent initiatives such as the Digital Economy programme as well as Design for Patient Safety.
CHI+MED: Track record

The CHI+MED proposal brings together investigators with international reputations in interactive medical systems and public engagement, with an internationally leading team in Computer-Human Interaction who have extensive experience of working successfully together (e.g., Thimbleby et al., 2002). CHI+MED will be led by Blandford at UCL Interaction Centre (UCLIC); Thimbleby and Curzon will lead the work at Swansea and QMUL, respectively.

UCLIC and the Future Interaction Technology Laboratory (FIT Lab) at Swansea have recently been jointly awarded a Platform Grant (EP/G004560 and EP/G003971), Healthy Interactive Systems. Blandford and Curzon have recently completed a joint project (GR/S67494 and GR/S67500) Human Error Modelling, which delivered over 40 publications on empirical, formal and integrated approaches to understanding human error. Blandford and Thimbleby have been co-investigators on two earlier successful EPSRC grants: GR/S73723 and GR/M81748. Blandford, Thimbleby and Curzon have all led doctoral consortia in HCI.

Prof Ann Blandford is Director of UCL Interaction Centre, a leading HCI research centre spanning Computer Science and Psychology, with 20 members. She has extensive experience of leading and managing research: she has been PI on 8 EPSRC/ESRC projects, including an ongoing Platform grant, and Co-I on one EPSRC and one NHS SDO grant. All projects that have received a final assessment were graded as internationally leading or tending towards internationally leading. She has over 200 publications, of which 130 are full papers in international refereed journals and conferences. Blandford has been a member of over 40 international conference/workshop committees; she was technical co-chair for HIM-HCI 2001 and HCI 2006, Associate Programme Chair for CHI 2007, and will be technical co-chair for NordiCHI 2010. She is a member of the UK Computing Research Committee and the EPSRC College, and is on the editorial board of Health Informatics Journal. Blandford is an expert in both qualitative and quantitative research methods for studying human error (Back et al., 2008a), in the development and testing of methods for evaluating interactive systems (Blandford et al., 2008a), and in working with professionals who use technology (Adams et al., 2005) and system developers (Furniss et al., 2008).

Prof Harold Thimbleby is a Royal Society-Leverhulme Trust Senior Research Fellow (2008–9), an Emeritus Gresham Professor of Geometry, and was a Royal Society-Wolfson Research Merit Award Holder (2001–5). He won the BCS Wilkes Medal. His work has been recognised by, for instance, 35 invited conference keynotes, over 300 seminars in 20 different countries, numerous conference committees, editorial board memberships and exhibiting at the Royal Society Summer Science Exhibition. He has 400 publications (over 150 of which were invited); he has lectured at the Royal Institution and the House of Lords, and at many science festivals, most recently TECHFEST08 in India. He has written for the New Scientist and Encyclopedia Britannica. He is a member of the EPSRC College, on the Board of Informing Health Care (the national IT programme to transform health services for Wales), and was elected to the UK Computing Research Committee in 2007. He leads EP/G003971, Healthy interactive systems for healthcare, with Blandford, and EP/F059116, A state of the art information-driven health care demonstrator: Combining advanced technologies in the gastro-intestinal field, with Ford and Williams. He also holds EP/F020031, Formally-based tools for user interface analysis and design; he previously held/jointly held 18 EPSRC grants. Thimbleby’s work on modelling a syringe pump (Thimbleby, 2007a) led to an invited keynote at the Usability and HCI for Medicine and Health Care conference, Graz. His recent book, Press On (Thimbleby, 2007b) won the American Publishers Association best book in computer science 2007 award. He is an expert in formal methods for HCI (Thimbleby, 2004).

Dr Paul Curzon has over 15 years’ experience using proof in the verification of hardware, software and human-computer systems (Curzon, 1995; Curzon et al., 2007; Ruksenas et al., 2008b; Xiong et al., 2007). Both he and McOwan have wide experience in outreach activities, including a 5-year PPE grant (EPSRC EP/F032641) supporting Computer Science for Fun (www.cs4fn.org), which was singled out for praise in the EPSRC International Review of Computer Science (2007). Curzon was technical co-chair of HCI 2006 and of two International Workshops on Formal Methods for Interactive Systems (Macau, 2006; Lancaster, 2007), and PI on EPSRC project GR/M45221, which received a final rating of Outstanding, as well as on the recently completed Human Error Modelling project with Blandford. He is Co-I on the Extreme Reasoning Platform Grant (EPSRC EP/F02309X). He has partnered schools on two Royal Society Partnership Grants (2007, 2008) and was PI on a 2007 RCUK Science Week award. He won the EPSRC Non-professional Computer Science Writer award in 2007.

UCL CHI+MED team and fit with UCL strategy

Along with Blandford, the UCL team comprises Dr Duncan Brumby, Dr Anna Cox and Dr Astrid Mayer. Brumby has focused on demonstrating how cognitive strategies are adapted to the demands of the task and constraints on human performance. He has shown how eye-movement strategies are adapted to efficiently locate information content on a computer screen (Brumby and Howes, 2008) and how, in safety-critical multitasking situations, attention is allocated to trade off performance between tasks. He recently submitted a First Grant application. Cox is an expert in quantitative research methods for studying human error and visual search and in computational modelling of cognition (Cairns and Cox, 2008). She was PI on EPSRC grant GR/T28225 (recently completed) in which she used eye-tracking to study how people attend to information on the screen. Mayer is a medical oncologist with expertise in the development of therapeutics, early phase clinical trials and translational medicine. She has an interest in informatics applications for cancer care such as decision support systems and has presented her work at national and international conferences.

CHI+MED contributes to UCL’s research strategy: in particular, to the institutional Grand Challenge on Well-being and cross-disciplinary theme on Design. High profile dissemination activities around these are planned, in which CHI+MED’s outcomes would play a significant role. UCL is a major partner in several national health initiatives
including UCL Partners (with UCL Hospitals, the Royal Free Hospital, Moorfield and Great Ormond Street); these initiatives will support engagement with both professional and patient stakeholders in CHI+MED. As well as its extensive involvement in health and design, UCL is a Public Engagement Beacon site, having recently received funding to develop innovative approaches to engagement with the public and professional stakeholders. CHI+MED will work closely with these initiatives, contributing to major London-centred developments in both the delivery of clinical care and public engagement.

**QMUL CHI+MED team and fit with QMUL strategy**

Prof Peter McOwan will work with Curzon at Queen Mary, University of London. McOwan has an extensive record of developing and empirically testing mathematical models for human visual processing (Anderson and McOwan, 2006). As well as the 5-year PPE grant (EPSRC EP/F032641) supporting Computer Science for Fun, his outreach experience includes two EPSRC PPE grants (GR/R78909, EP/C528921) and his work has been included in Royal Society Summer Exhibitions (2005, 2007). He is a member of Sciencewise-ERC, funded by the Department for Innovation, Universities and Skills (DIUS), which helps policy makers commission and use public dialogue to inform policy decisions in emerging areas of science and technology. McOwan is a member of the EPSRC College, is co-investigator on EP/F033133, *Building a New Community: Modelling, Visualisation and Verification of Large Scale Systems*, and was General Chair of the First International Conference on Evolvability and Interaction Symposium.

CHI+MED contributes directly to QMUL’s research strategy to exploit the opportunities presented in health sciences and their interface to other disciplines, and in increasing QMUL’s effectiveness in transferring research outcomes to business and the wider community. Furthermore mathematical modelling of IT systems, as in CHI+MED, is a priority area of the science and engineering sector of the college. The School of Electronic Engineering and Computer Science holds both a complementary Platform Grant on Antennas for Healthcare and Imaging and a digital economy related grant, DIADEM: data analysis for clinical decision making. Public and Stakeholder Engagement is a major part of the college’s strategy: it is already playing a leading role in national public engagement projects in Mathematics, Physics and Computer Science. QMUL is also the London Excellence Hub for Gifted and Talented students and leads the London East Thames Gateway ‘Aimhigher’ Partnership.

**Swansea CHI+MED team and fit with Swansea University strategy**

With Thimbleby, the team at Swansea comprises Dr George Buchanan, Dr Parisa Eslambochilar, David Ford, Dr Matt Jones and Prof John Williams. Ford and Williams are Co-Is with Thimbleby on EPSRC grant EP/F059116. Ford is Director of the Health Information Research Unit for Wales (HIRU), with approximately £1.5M current funding from a variety of research and public sector sources, and is a Director of MediWales. Williams is Professor of Health Services Research, and was Director of the Wales Office of R&D from 2002 to 2007. He is a member of the HEFCE RAE 2008 Health Services Research Panel, and was a Special Adviser to the House of Commons Select Committee on Electronic Patient Records. Williams has been Director of the Health Informatics Unit, Royal College of Physicians since 2000; he has been awarded nearly £6M grant income, covering a wide range of topics relevant to this proposal.

Jones is an internationally-recognised researcher in mobile interactive devices (he co-authored the standard text in the area (Jones and Marsden, 2006)), and is on numerous editorial boards and programme committees; he currently holds a prestigious Nokia Research Fellowship; he is providing research leadership in the area of computer science and social impact including community “wellness”; he has had several grants in these areas: e.g., EP/F035071, EP/E042171, EP/E006396 and EP/E006418. In 2001, Jones organised a research tour of all New Zealand universities as a major science engagement activity.

Buchanan and Eslambochilar are recently appointed lecturers who have very promising research careers. Buchanan is an expert in the design and evaluation of mobile devices and digital libraries (Buchanan et al., 2002). He has been awarded the ACM Ted Nelson Award and numerous best paper awards, and holds EPSRC grant EP/F041217. Eslambochilar is an expert in control theory and the design of mobile devices (Eslambolchilar and Murray-Smith, 2008); she recently submitted an EPSRC First Grant. Ford is an FRSA; Buchanan and Thimbleby are Honorary FRSA.

CHI+MED exactly fits Swansea University’s research strategy, and is strongly supported by the University. Swansea’s mission includes developing interdisciplinary research, and fostering links with national bodies such as NHS Wales and MediWales, as well as international links. CHI+MED contributes to these strategic directions, bringing together Computer Science and the Institute of Life Sciences within the University, contributing to national priorities (health and wellbeing) and stimulating the Welsh economy and developing internationally leading collaborations.

**Other information**

CVs are included for named researchers Dr Jonathan Back, Dr Dominic Furniss, Dr Rimvydas Rukšėnas and Patrick Oladimeji; the first three named have worked together successfully on previous projects including the *Human Error Modelling* Project. CVs are also included for Prof Gregory Abowd, an international expert in software engineering, HCI and healthcare systems, who will be a member of the CHI+MED steering committee, and for Laura Meagher, who will act as a consultant on evaluating the stakeholder engagement activity and also participate in the steering committee. The steering committee will be chaired by Prof David Cousins, Head of Safe Medication Practice and Medical Specialities at the National Patient Safety Agency.

Following discussions with EPSRC staff, and to facilitate cross-reference (since many publications are referenced from multiple documents within this proposal), all references are gathered together at the end of the Technical Annexes.
CHI+MED: Case for support

Vision

Our vision is that interactive medical devices will be made safer through better interaction design, based on an explicit and rigorous foundation. Explicitness and rigour will be achieved through the complementary application of empirical and computational reasoning techniques. Impact on practice will be achieved through productive dialogue with stakeholder groups and the delivery of analytic, theoretically founded and empirically tested methods and tools to support the interaction design of medical devices.

User interaction is central to medical device safety, but is under-explored. CHI+MED complements developments in related areas, e.g., on the ergonomics and engineering of such devices, by taking an in-depth approach, developing a multidisciplinary understanding of the design of user–device interactions. There are three important concepts that we define explicitly here, and use throughout this proposal with these defined meanings:

1. **Interactive medical device:** There are many kinds of interactive devices in use in health care, small and large, used by professionals or patients, for diagnosis, therapy, monitoring and mobility. We are focusing our attention on medical devices that incorporate interactive software and a user interface, and that are widely used by nurses and patients as well as more specialist clinical staff. Such devices include infusion pumps, syringe drivers, ambulatory pumps, blood glucose monitors and vital signs monitoring devices. We are focusing on such devices for several reasons:
   - There are many existing designs of each of these products and many known design issues, giving a space of alternatives to study, and many potential redesign possibilities to investigate.
   - There is movement towards patients taking an increasingly active role in their own care. Over the next few years, it is likely that many more devices will come onto the market intended for patients and their carers to use unsupervised. If these devices are to be used safely, it is essential that the user–device interactions (as well as other features such as ergonomics) are well designed. While devices for patient use are unlikely to have direct safety consequences (e.g., dosage errors), difficulties of use may have indirect safety consequences such as misreading of patient data.
   - Focusing on devices that are widely used, including by people with relatively little training in their use, greatly simplifies the design of empirical studies because there is a much larger pool of potential study participants than there would be for more specialised devices. This minimises the risk of being unable to gather the necessary empirical data.
   - Simulations of devices that are widely familiar will be a very useful resource for dissemination and training, in a way that simulations of more specialist technologies would not be.
   - Focusing on widely used devices also increases the potential impact of the research.

2. **Stakeholder:** We will engage with various stakeholder groups. These include device manufacturers, organisations and individuals developing policy around the design, procurement and use of medical devices, incident investigators, device users (professional, patients and carers), and others with an interest in such devices, including medical students and future computer scientists. In this proposal, we use the phrase “stakeholder engagement” to encompass approaches to developing an ongoing dialogue with all these groups, including techniques usually referred to as “public engagement” and “knowledge transfer.”

3. **Interaction Design:** Our focus is on transforming the design of the interaction between users and devices. This interaction is achieved through a programmed interface, and takes place in a situation that includes time (when in the day, when relative to broader clinical processes) and location (where the device is physically situated, who else is interacting, e.g., in the clinical team). All these factors influence the experience of the interaction (e.g., distractions in the environment may cause the user to lose track of where they are in a procedure), and will be taken into account while retaining our focus on rigorous interaction design.

We propose a research agenda on developing techniques that are useful and used for designing user interactions with medical devices. This is not simply reusing and reapplying known findings from Human–Computer Interaction (HCI), but extending HCI theory and methods while contextualising them to the needs of interactive medical device design.

For a technique to be useful, it must support reasoning about key factors that make such user interactions less-than-perfect. The first research challenge is thus to deliver design and evaluation techniques that support developers in designing human–device interaction, focusing on factors that affect efficiency and safety in medical settings. No one technique can cover all factors in detail. We will therefore develop and test techniques that focus on the device, on the user, and on the situation of use. For the device, we have identified the ability of developers to be able to reason rigorously about interaction programming as an important focus, as the design of the interaction determines both usability and likelihood of user errors. Focusing on users, theories of how people exploit their knowledge in interactions are relatively well developed, whereas the causes of cognitive slips (in which users know how to perform a procedure but inadvertently get it wrong) are comparatively poorly understood. We will conduct research into factors underlying cognitive slips, the role of interaction design in promoting system resilience (i.e., minimising the probability of errors, and supporting people in recovering from failures) and the interrelationships between interaction programming and human factors. Of course, device use is situated, and it is essential to reason about use in context; we will investigate approaches to such reasoning, taking prior work on Distributed Cognition (Hollan et al., 2000) as a suitable starting point to develop and test a method for reasoning about the design and effective deployment of medical devices.

For a technique to be used, it must address a recognised need and be usable. In the medical context, the “elephant in the room” is that interaction problems have become culturally entrenched, effectively invisible to most observers, with
clinicians encouraged to work around them, often whilst several other tasks are demanding attention at the same time. However, these invisible problems have very serious consequences: poor design induces adverse incidents, patient injury and fatalities, and inappropriate allocation of blame. This raises the second research challenge: to develop a dialogue with stakeholders about interaction problems. Some factors (e.g., consistency of interaction) are already well known within the HCI community but need to be presented clearly for the medical domain, recognising that stakeholders also bring valuable understanding and that there are opposing forces—e.g., standardisation and innovation are often in opposition to each other. Other factors (e.g., slip errors) are less well understood, so need to be researched and be packaged in a form that supports clear understanding. We will therefore investigate techniques for engaging with stakeholder groups, drawing on our work in Public Engagement and exploring the appropriateness and effectiveness of different approaches for different contexts, to engage in an effective dialogue around needs and possibilities.

Background

Within the healthcare domain, work on Human–Computer Interaction (HCI) has focused on the design of next-generation computing solutions, i.e., systems that cope with the increasing complexity of devices and decision support technologies (e.g., Hajdukiewicz, 2001). The design of seemingly simple and mundane medical devices such as infusion pumps and monitoring devices has received less attention. Many such devices have been designed without an understanding of how their interactive behaviour can increase the likelihood of medical errors during situated use.

Human error has caused and been implicated in many patient deaths, with the National Institutes of Health report (Kohn et al., 2000) drawing it to widespread attention: fatalities per year from human error in US hospitals exceeds car accidents, AIDS and breast cancer combined. CHI+MED is more specifically concerned with human error in the use of interactive devices; e.g., in paediatrics, a high proportion (Kaulah et al., 2001) of adverse drug incidents are caused by calculation or operator errors. In the UK, the Medicines and Healthcare Products Regulatory Agency (MHRA) receives many reports of incidents involving infusion pumps that result in patient harm or death, primarily from over-infusion (1,495 reported between 1990 and 2000 (MHRA, 2004)). The MHRA asserts that in most fatal incidents no fault is found with the infusion device, which suggests to them that “user error” is the most significant contributing factor (MHRA, 2004) — latent error in the design is a largely unexplored causal factor but will be a key area for CHI+MED. The EU Medical Device Directive (2007) and the IEC 60601:1-1-6 (2006) are recent publications that highlight the need for manufacturers to test and evaluate the usability of devices. These publications build on recommendations made by the US Food and Drug Administration (e.g., FDA, 1996) who argue that interactive medical devices should conform to established user interface design principles (e.g., Norman, 1988) and that designs should be evaluated using healthcare practitioners. This is conventional, if not rather basic, HCI wisdom (itself largely developed in non-safety critical domains), which we argue can be substantially refined for healthcare. In particular, since medical devices are now used by patients and carers (e.g., ambulatory pumps for drug administration) as well as operators in clinical environments, research must be progressed in these situations.

The recognition of the role of good design in improved patient safety has resulted in a number of studies investigating the usability of individual medical devices, most notably infusion pumps (Liljegren et al., 2000; Lin et al., 1998). Although valuable, these studies are limited by relying on the traditional retrospective self-report paradigm when identifying usability problems. This is also an approach advocated by the National Patient Safety Agency (NPSA, 2009). Evidence suggests that simple routine tasks are most likely to be performed incorrectly under conditions where there are competing demands and limited resources (Back et al., 2007), making retrospective analysis difficult. Such an approach is ideally complemented by other techniques such as controlled experiments or observation, to build up a more reliable understanding of the causes of adverse incidents.

In the UK, the MATCH programme is currently working on case study projects (e.g., Martin et al., 2008) in collaboration with partners from industry, healthcare and academia, to investigate the development of medical devices. This approach recognises the need for facilitating communication between different stakeholders in an attempt to understand usability requirements. Our proposed work is complementary but shifts the focus away from how to elicit known problems, of which there are many, towards analytical methods. These methods aim to reveal the consequences that an interaction design has on patient safety at different levels of granularity ranging from the patient at home to specialist clinical teams. Our results will thus provide useful additional input to MATCH. Instead of trying to identify why things went wrong (e.g., Gaba et al., 1987), we will deliver tools that ensure medical devices have properties that enable errors to be avoided or mitigated. This will enable designers to reason about conditions that they have not encountered before.

Designers and those involved in the procurement of medical devices recognise that a wide variety of operational conditions can lead to problems that cause medical device errors (Howard, 1993). Designing safer medical devices requires a more sophisticated understanding of human cognition (e.g., Gray, 2000), distributed cognition (e.g., Klien et al., 1993), and technical work analysis (Luff et al., 1994). Although the roles that all these factors play is recognised by many safety practitioners within the healthcare domain, the interplay between these factors during human-device interaction is often considered to be too complex (Cook et al., 1994). This has led to the exemption of medical devices from adhering to the IEC 61508 functional safety guidelines (which govern devices in other safety critical domains). There is nevertheless growing recognition of the role of human factors when improving safety (e.g., NPSA, 2009).

Other industries, notably transport (particularly aviation) and nuclear power generation, have well-developed safety cultures. System components are designed to facilitate safe operation within the system as a whole. The traditional approach to safety involves designing barriers to prevent dangerous scenarios from occurring (e.g., Reason, 1990). This
The core science takes three distinct but interlinked perspectives on the design and use of medical devices such as infusion pumps and monitoring devices: device design (WP2); individual use of devices (focusing on cognition and error: WP3); and situated use of devices (focusing on how device use is interleaved with other activities and interactions: WP4). Each scientific perspective will be studied in a work package that integrates empirical studies with computational representation and reasoning, based on real devices. Further work packages address integration (WP1); the development and testing of structured techniques to support stakeholders in reasoning about interaction design (WP5); and stakeholder engagement (WP6). All are inter-institutional. Flexibility will be built into all work packages, to make it possible to respond to developments in both research and practice as the programme proceeds, as described in the Management Annex.

**WP1: Integration:** This is a multidisciplinary programme of work involving three research groups and also external stakeholders. Resources will be dedicated to integration across themes. This will include integration between the scientific strands and between the science and the stakeholder engagement work. Means for integration will include joint meetings, working on shared exemplars, developing shared documents and researcher exchanges.

**WP2: Device design:** This work package develops basic science on the design of existing and future medical devices. It will start by reverse engineering existing device designs. This has four important outcomes: identification of both strengths and vulnerabilities of existing designs (linked to WP3 and WP4); underpinning methodological investigations (in WP5); delivering system simulations that can be used for testing in WP3 (and that can be adapted to test experimental hypotheses in ways that commercial products cannot); and delivering simulations for demonstration in engagement activities (WP6). This work package will also investigate computational modelling of device interactions as a way of reasoning about system properties. As the programme develops, we will design, implement and test alternative device prototypes that are informed by findings from all strands of the programme.
**WP3: Individual use:** This work package investigates individual use of devices. Laboratory-based studies (e.g., controlled experiments) will investigate use, with a focus on cognitive slips and the factors that provoke or mitigate slips. Empirical studies will be mutually informed by computational representations and reasoning. Findings will guide future device design (WP2); they will inform and be challenged by findings from WP4 (where we will observe situated behaviour); and will be communicated through WP5 and WP6.

**WP4: Situated use:** The third of the scientific work packages will investigate use in context. Observational studies in natural medical settings, medical research settings and a training setting (purpose-built medical room simulators at Swansea) will establish how practitioners use devices: how they interleave their tasks, deal with distractions, interact with colleagues, recover from errors, etc. Studies will also be conducted on patients’ uses of devices – in wards, at home, and out-and-about. Observations will be complemented by interviews to establish mental models of device operation and perceptions of use, and to gather essential data about critical incidents. As in other work packages, empirical research will be codified and tested through computational reasoning. Findings from situated use research will be integrated with those from analysis of device designs (WP2) and controlled laboratory studies (WP3). These findings will form the basis for method development (WP5), and will be communicated through stakeholder engagement (WP6).

**WP5: Method development and testing:** CHI+MED will develop and test theoretically grounded methods and tools for medical device developers to reason rigorously about interaction design. Methods based on the science of work packages 2, 3 and 4 will be developed and tested, working with WP6 (stakeholder engagement) to ensure that the methods are usable and useful.

**WP6: Stakeholder engagement:** We will develop and test novel approaches to engaging with stakeholders to explore how a richer understanding of interaction design and human error can guide both design and procurement practices. It will be necessary to better understand existing perceptions, cultures, values and practices relating to design and procurement to design effective techniques and materials. Both direct and mediated approaches to engagement will be researched; direct approaches will include workshops, participation in science festivals and informal interactions; mediated approaches will include written and audio-visual forms (e.g., device simulations and short videos made available via the internet). The CHI+MED Advisory Group will include representatives of stakeholder groups. Outputs from other strands will create essential material to support this work.

**Programme progression**

All work packages in CHI+MED include three key elements. First, particularly early in the programme but continuing throughout, one focus will be on understanding ‘what is’. An important tool for this will be linking empirical findings with computational reasoning. Second, as understanding develops, the focus will shift towards developing and testing methods and tools – for both professional stakeholders and broader public engagement – that can transform understanding and practice. Finally, throughout the programme but particularly in the later years, attention will be paid to evaluating the efficacy of the approaches being taken. Here is an overview of these three elements of the programme:

1. **Understanding by linking empirical studies with computational reasoning**

   Computational modelling techniques are transforming the sciences by providing new ways to tackle complex problems. We integrate empirical work, based on real systems, with formal modelling and state-space exploration techniques. This opens new ways to study the complex interplay between multiple causal factors that underpin interactive phenomena such as difficulty in reaching particular device states; human error in interactions; or multi-agent interactions around a device. The value of computational reasoning about critical computer systems is widely accepted (e.g., Rushby, 2001). Our foundational work has shown that abstract non-deterministic models provide powerful new ways to model and rigorously analyse interactive behaviour. By drawing from and feeding into empirical enquiry, computational reasoning becomes a powerful tool to better understand interactive phenomena and device design. This is achieved through tightly coupled engagement between researchers conducting the empirical work and those developing and working with the corresponding computational representations. This form of linking will be adopted in all three scientific work packages.

2. **Developing methods and tools from understanding**

   Tools are specialised to support particular goals and activities. In CHI+MED, we will develop and test mechanisms, based on our core scientific work, to facilitate rigorous software development, inform procurement decisions and more broadly support understanding by stakeholders. Some of these will be recognisably design or evaluation methods for devices, in a fairly classical HCI tradition, aligned to WP3 2-4 but developed in WP5. Others will be designed for broader stakeholder engagement (e.g., digital stories, short videos for distribution via the Web, science writing and tutorials); this activity will be led by researchers in WP6, supported by researchers in the scientific work packages.

3. **Evaluating tools and activities**

   Evaluation will be done throughout CHI+MED, building in intensity as the programme proceeds. It will be possible to draw on standard academic evaluations for some parts, e.g., reviews of scientific papers. We will develop a specific plan for evaluating the stakeholder engagement work. Evaluation will be both formative (e.g., how usable do developers find this evaluation tool? How can it be improved?) and summative (e.g., measuring the outcomes through metrics).

**Programme Grant assessment criteria**

**Quality of research**

CHI+MED will deliver internationally leading research on HCI for medical devices. The focus will be on applying rigorous scientific methods, drawn from various disciplines, to a well-defined class of problems. The programme will
deliver both core scientific understanding and tools and techniques for sharing that understanding with stakeholders, engaging with their concerns as well as with the science.

**Added value**

The strands of our work are inherently interdependent, as illustrated in the example presented in Annex 1. The core scientific approaches validate and guide each other, and hence are mutually dependent.

Without the core scientific work, the stakeholder engagement activity would have little substance; conversely, the scientific work will be of less value if not communicated and explored with stakeholders; the directions this work takes must be informed by clinical knowledge, and both developer and user practices. As noted, there is limited stakeholder awareness of how a richer understanding of interaction design could improve practice; raising awareness will involve identifying compelling and persuasive examples; conversely, a better understanding of professional concerns will motivate scientific work, ensuring that the science addresses problems that have a substantial impact on practice in the longer term. The stakeholder engagement activities mutually inform each other, focusing on a deeper understanding of the rich context within which design and purchasing decisions are made (necessary for understanding where changes are possible, and how the culture can adapt) and on approaches to engage with stakeholders. The critical mass of CHI+MED will facilitate meaningful engagement with parallel activities such as the “Designing for Patient Safety” programme, MATCH (Multidisciplinary Assessment of Technology Centre for Healthcare) and DOME (Designing Out Medical Errors, EPSRC EP/F064802/1), which have broader agendas around the design of technologies in healthcare. It will also contribute to capacity building, delivering people and relationships to lead future activities in this area and train future experts in interaction design for medical devices.

**Vision and ambition**

What we propose is transformational research. Interaction design is only one facet of the design of medical devices, but it is an essential one, directly affecting users and patients, which has been overlooked for too long; this programme of work will bring rigorous interaction design into focus as an important contributor to patient safety, and provide usable and useful techniques for reasoning about interaction design in development and procurement.

In parallel, CHI+MED will contribute to the UK’s standing as a centre of excellence in rigorous and scientifically grounded approaches to HCI – both in the medical domain and more generally in the design of safety-critical systems.

**Alignment to EPSRC’s seven criteria for success**

1. **Stimulating creativity and adventure in research and research processes:** The proposed work is highly creative and adventurous in the exploration of novel ways of understanding, reasoning about and representing interaction design and human error in medicine. No other research groups have taken the multidisciplinary approach proposed here, where findings from user studies are directly represented in computational models and applied in design practice, and where clinicians and HCI specialists work together to investigate the use of devices in clinical practice. We are not aware of any previous work that has linked ongoing scientific developments with public engagement techniques to build a dialogue with professional stakeholder groups as we propose in CHI+MED. There are fundamental challenges, such as identifying representations of devices, people and situations that are both powerful and tractable, and exploring ways of communicating effectively with diverse stakeholder groups about a subject that has wide-reaching effects and yet is so often effectively invisible.

2. **Attracting, nurturing and supporting the most talented people at every stage of their career for the benefit of the UK:** CHI+MED will have the critical mass to have a significant international profile, and to attract talented researchers. The programme will provide a structure for developing staff at all levels. As well as the individuals formally aligned to the programme, we will also involve undergraduate, Masters and other doctoral students in projects aligned to CHI+MED. Through the programme, a wide range of people will be trained in HCI and medical devices, resulting in a long-term sustained impact beyond the life of CHI+MED. Everyone involved will receive both cross-training in other programme angles (facilitated by the requirement to deliver stakeholder engagement materials and activities) and specialist training in their own areas and in public engagement. Some such cross-training will take place within programme meetings; researchers will also learn from each other through their various interactions (see Annex 1).

The programme will recruit nine PhD students across the three sites. All will be trained in the research methods that are pertinent to their individual projects and more broadly in research methods and practices. They will receive the training that is provided for all research students at their respective institutions and follow recognised development paths (e.g., with formal registration and upgrade vivas), including participation in doctoral consortia. Their involvement in the programme will provide additional training benefits, including significant contact with users of the research (i.e., stakeholders), as well as with researchers from contrasting sub-disciplines of HCI.

CHI+MED will involve twelve post-doctoral researchers. It is to be expected that most of the PDRAs will initially have relatively focused backgrounds; the multidisciplinary nature of this programme provides an excellent context in which to cross-train PDRAs in other sub-disciplines of HCI, as well as in the application context (medical device design and use) and in stakeholder engagement. Where appropriate, researchers will be given autonomy to run well-defined sub-projects on their own, and also to supervise Masters students on dissertations related to the theme of the programme grant (suitable MSc programmes run at UCL and Swansea).

The multidisciplinary nature of CHI+MED will provide many opportunities for investigators to extend their research and project management skills in new directions, including stakeholder engagement.

3. **Building collaborations that achieve a two-way flow of knowledge between the research base and industry:** Dialogue between industry and research is central to CHI+MED. One of the key aims of the programme is to develop
The proposed programme is fundamentally multi-disciplinary, involving design, cognitive science, social sciences, computational modelling and medical practice, and it is also cross-institutional.

5. Developing a shared vision of tomorrow’s major challenges and opportunities with stakeholders: society, industry, universities and other partners: One of the central aims of CHI+MED is to develop a shared vision with agencies and industry of both what is desirable and what is possible in terms of designing safer interactive medical devices. We have already embarked on this path, initiating discussions with many of the key stakeholders.

6. Building a better understanding of where we should focus our effort to benefit both UK society and the UK economy and increase its global competitiveness: The broad agenda of improving patient safety is widely recognised; there are many possible areas of investment for achieving this. CHI+MED focuses on interaction design, and will be evaluating the costs and benefits of research on this aspect of medical devices, which will contribute to the broader debate on where to invest research effort.

7. Creating and sustaining research scientists and engineers in the UK so that they are recognised worldwide as leaders in their field: Most of the research team are already recognised worldwide leaders in their research, and the UK is a world leader in computational and other rigorous approaches within HCI. The problems addressed in CHI+MED are worldwide problems, affecting all health services: the programme will enhance the UK standing, particularly in rigorous HCI, medical device design and public engagement. It will create an environment in which a critical mass of young researchers will be trained in a stimulating combination of research and engagement approaches that will make a long term contribution to the UK profile in this area of research and practice.

Leadership quality
The leadership qualities and experience of Blandford (as PI), Thimbleby and Curzon (as principal co-investigators) and all other co-investigators are presented in the Track Record.

Advocacy for Engineering and the Physical Sciences
The basic premise of the proposed programme is that an engineering approach is essential for improving the safety of interaction designs; this involves not just engineering the technology, but also representing users and situations in ways that support design reasoning (sometimes called “cognitive engineering”). Medical devices and human error are compelling topics that provoke wide interest. Thus, CHI+MED has the capacity to raise awareness among stakeholder groups of the value and importance of engineering, particularly a rigorous, scientifically-based engineering approach to HCI. Extensive stakeholder engagement is central to the proposal; as well as transforming practice, we anticipate that the research in CHI+MED will be an exemplar of how important engineering is to safe medical practice.

Economic impact
The focus of CHI+MED is on patient safety, and the contribution of high quality interaction design to safety. This will make at least three contributions to economic impact:

- Devices that are easier to use, and less error prone, will increase clinical efficiency.
- Fewer medical errors will result in fewer resources being expended on dealing with the aftermath (including enquiries and litigation) of incidents. Improved patient outcomes will improve recovery times, resulting in fewer demands on health service resources and fewer lost working days.
- As the value of interaction design becomes more widely recognised, it will become a more important factor in purchasing decisions, and developers who engage with this programme will be well placed to develop and market products with better interaction designs, increasing their market share.

If CHI+MED achieves even a fraction of what we anticipate, the benefits to society (social, financial, etc) in terms of improved patient safety, better recovery and more efficient and resilient clinical practices will be enormous. This research will open up new avenues for further investigation and long term dialogue between stakeholders. Relevance to beneficiaries is discussed in more detail on the accompanying form.

Summary
As the reliance on medical devices grows, there is a very urgent need to develop approaches to minimising error in their use, and to develop an informed culture where the role of HCI in reducing errors is recognised and exploited. In this proposal we have presented a vision: that the interaction design of medical devices (as defined above) should be recognised as an important contributor to patient safety, and that a suite of scientifically grounded and tested methods should be available to support development and procurement decisions. We have outlined a programme of work, moving from understanding to development and testing of methods and design solutions, with parallel work on stakeholder engagement, to transform understanding and practice in this area. As described in the Management Annex, all projects will be reviewed regularly to ensure flexible response to developments both within and external to CHI+MED.
CHI+MED: Work Plan

The programme of work comprises six work packages, each consisting of several projects. Broadly, early projects focus on understanding and representing existing situations, progressing towards later projects that investigate tools, methods and strategies for effecting change, with an ongoing but increasing emphasis on evaluating the effectiveness of tools, methods and interventions.

The outline project plan is shown below (Figure 1). Projects in italics are PhD projects (with provisional themes proposed). As discussed in the Management annex, detailed project plans will be reviewed at annual Steering Committee meetings. Therefore, at the outset of the programme commitment is being made to the early phases (up to 3 years) of projects that start within the first year, and to studentships, and decisions about both commitment to new projects and the directions of ongoing projects will be made at Steering Committee meetings (or as necessary). Initial commitments are shown as solid bars, while those proposed are hashed (diagonal). We are committed to PhD projects starting at the specified times, but will review the details at the immediately preceding Steering Committee meeting (and of course for PhD projects the details are likely to further change as students start to explore their individual topics). PhD projects are indicated by horizontal bars, while vertical bars indicate activities that will be ongoing throughout the programme. The researchers and investigators listed against each project will be the main participants in that project, though others will be involved to a lesser extent. Conversely, not every researcher involved in a project will be working on it full-time for the entire project duration.

Each work package is described in more detail in the accompanying technical annexes. Interdependencies between work packages are presented in Annex 1.

Management and other meetings

As noted in the Management annex, Steering Committee meetings will take place annually and Advisory Group meetings every 6 months. These meetings will generally be aligned to other project meetings to ensure exchange between CHI+MED members and external stakeholders. Other project meetings will be as follows:

- Whole project meetings at the beginning and every 6 months (months 3 and 9 each year). To support monitoring, every researcher will prepare a report and presentation for each meeting, summarising recent activities, achievements, ongoing issues and plans (this is a technique we have found very effective in the past).
- External themed workshops (sometimes with tutorials) every 6 months. These workshops represent one approach to stakeholder engagement. Future themes will be agreed at Steering Committee meetings.
- Researcher exchanges averaging a week a year.
- The management team will meet or hold a teleconference 6-weekly to monitor progress, ensure coordination between the various investigations, and plan directions.

The annual calendar of major meetings will be as follows:

<table>
<thead>
<tr>
<th>Month</th>
<th>Meeting Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Themed workshop (except first year, which will be a whole project meeting)</td>
</tr>
<tr>
<td>3</td>
<td>Whole project meeting with Steering Committee, Advisory Group and management meetings</td>
</tr>
<tr>
<td>8</td>
<td>Themed workshop with Advisory Group and management meetings</td>
</tr>
<tr>
<td>9</td>
<td>Whole project meeting with management meeting (except final year, when this will be delayed to month 11 and include a final Steering Committee meeting)</td>
</tr>
</tbody>
</table>

Figure 1: Diagrammatic Workplan
**Milestones and deliverables**

Steering Committee meetings at which milestones will be reviewed will take place at month 3 in each year of the project, to review progress so far and future directions. A final milestone review will take place at the end of the project.

**Milestone 1 (@ 3 months):**

**Work package 1:** An inaugural project meeting will have been held and the first Steering Committee meeting organised.

**Work package 2:** “R-S2” (provisionally named researcher Oladimeji) will have been recruited and be ready to start (building on foundational work that precedes CHI+MED) on WP2.1. Recruitment of “R-S1” (also for WP2.1) will be underway and a PhD student at Swansea (for WP2.3) will have been recruited.

**Work package 3:** “R-U1” and “R-Q1” (provisionally named researchers Back and Rukšenas) will have been recruited and work will have started on controlled studies (WP3.1) as described in Annex 3.

**Work package 4:** “R-U2” (provisionally named researcher Furniss) will have been recruited and pilot studies on situated use (WP4.1) as described in Annex 4 will be underway. Two PhD students (at UCL and QMUL) will have been recruited and recruitment will be underway for “R-Q2” (for WP4.1).

**Work package 5:** No activity is planned.

**Work package 6:** PDRAs “R-S4” and “R-U3” will have been recruited and will have started working on WP6.1 and WP6.3 (see Annex 6). Initial plans for evaluation of the SE work will be under development. The themes for the first two themed workshops will be agreed. An initial project website will have been developed.

**Milestone 2 (@ 15 months):**

**Work package 1:** Two further project meetings and a Steering Committee meeting will have been organised.

**Work package 2:** Publications will be being produced from WP2.1. The focus and extent of WP2.1 will be reviewed at this Steering Committee meeting. The aims of WP2.2 will be reviewed in the light of findings to date. PhD project WP2.3 will be progressing.

**Work package 3:** Publications will be being produced from WP3.1. The focus and extent of WP3.1 will be reviewed at this Steering Committee meeting. Two PhD students will have been recruited (at UCL and QMUL).

**Work package 4:** Publications will be being produced from WP4.1. The focus and extent of WP4.1 will be reviewed at this Steering Committee meeting. PhD students on WP4.2 and WP4.3 will have completed first year vivas.

**Work package 5:** No activity (the post-doc working on WP5.3 will have already been recruited to work on WP4.1).

**Work package 6:** Publications will be being produced from WP6.1 and WP6.3; the focus and extent of these projects will be reviewed at this Steering Committee meeting. The aims of WP6.2 and WP6.4 will be reviewed in the light of findings to date. Recruitment for WP6.4 (“R-Q4”) will be underway. The initial framework of core questions and evaluation plan will have been developed. Two themed workshops will have been organised and run. The themes for the next two will be agreed. The website will be being populated with material as it is produced.

**Milestone 3 (@ 27 months):**

Publications will be being produced from all work packages. The final three PhD students will have been recruited and recruitment will be underway for the remaining PDRAs (there may, of course, be further recruitment as PDRAs move to develop their careers – e.g. into lecturships). The first three PhD students will have been through MPhil-PhD upgrade at their respective institutions. Regular meetings and workshops will have been held and the website updated. Strong collaborations will have been developed with outside organisations (other research groups specialising in medical devices, both in the UK and overseas; device manufacturers; other stakeholder groups).

**Milestones 4-6 (@ 39, 51, 63 months):**

The basic programme infrastructure is likely to have been retained, but with adaptations in the light of experience and advice; all projects will be progressing (with regular review). The basic review strategy is outlined in the Management plan.

**Milestone 7 (@ 72 months):**

Towards the end of the project, there will be an increasing focus on evaluation. There will also be a major review of future directions. Final deliverables will include publications, tutorials, stakeholder engagement materials, an evaluation framework, a final evaluation report from the external evaluator and a final website as described in the various technical annexes.
CHI+MED: Management
The contributions of the individual projects to the overall vision of CHI+MED are discussed in the Case for Support. As described in Annex 1, the various projects and work packages within CHI+MED are interdependent, but none of the interdependencies is time-critical. Thus, it will be important for the success of the overall programme that communications between the various projects are effective, but no individual project will be threatened by unexpected delays in another. CHI+MED will have an internal management team, a small Steering Committee and a larger Advisory Group.

Management structure
The management structure is illustrated in Figure 1.

![Figure 1: core management structure](image)

Programme leader: Ann Blandford
Management team: Paul Curzon, Harold Thimbleby, Project manager
Project leaders: Peter McOwan, Anna Cox, Duncan Brumby, Astrid Mayer, George Buchanan, Parisa Eslambolchilar, Matt Jones, David Ford, John Williams
Assistant Technician

The programme will be led by Blandford, and managed by a professional project manager (0.5FTE), supported by an assistant (0.4FTE) and a technician (0.3FTE) with complementary skills. Their responsibilities are listed in the Justification of Resources. Blandford, Curzon, Thimbleby and the project manager will form the management team. Blandford, Curzon and Thimbleby have worked closely together for over 10 years, including collaborating on EPSRC grants. Each will lead the work at their respective sites (NB many individual projects are cross-site; in these cases, one site is identified as the lead).

The detailed terms of reference for the Steering Committee and Advisory Group will be agreed at the inaugural meetings of each. Broadly, the role of the Steering Committee will be to advise on direction, review progress and agree on the future of individual projects. Therefore, annual Steering Committee meetings will include formal review of progress against targets. These will include formal decisions about the continuation and initiation of individual projects. The Steering Committee will be chaired by a patient safety expert (Cousins) and other members will be an international expert (Abowd), and a stakeholder engagement expert (Meagher). We have included letters of support from members of the Steering Committee and a small but representative sample of other stakeholders. The exception is David Cousins, who has confirmed his very enthusiastic support and contributions verbally, but was unable to provide a letter because he was away (for unanticipated personal reasons) in the period leading up to the submission of this proposal.

The role of the Advisory Group will be to represent stakeholder communities in advising the programme, to alert programme team members to new opportunities and ideas and to disseminate CHI+MED findings back to their own communities. It therefore comprises key UK stakeholders (both research users and technology users) and also specialists with skills that complement our own. The Advisory Group will meet every 6 months. The following have accepted invitations to participate in the Advisory Group:

<table>
<thead>
<tr>
<th>Name</th>
<th>Organisation/Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Cousins</td>
<td>National Patient Safety Agency</td>
</tr>
<tr>
<td>Nathan Moore</td>
<td>Head of Business Development, Centre for Evidence-Based Procurement (MHRA)</td>
</tr>
<tr>
<td>Oliver Wells</td>
<td>Association of British Healthcare Industries</td>
</tr>
<tr>
<td>Nick Komaromy</td>
<td>Royal Free Hospital</td>
</tr>
<tr>
<td>Paul Lee</td>
<td>Singleton Hospital, Swansea</td>
</tr>
<tr>
<td>Duncan McNeil</td>
<td>Royal Free Hospital</td>
</tr>
<tr>
<td>Laura Meagher</td>
<td>Independent consultant</td>
</tr>
<tr>
<td>Richard Lilford</td>
<td>MATCH Programme</td>
</tr>
<tr>
<td>Ed Matthews</td>
<td>Helen Hamlyn Centre, RCA</td>
</tr>
<tr>
<td>Chris Johnson</td>
<td>University of Glasgow</td>
</tr>
<tr>
<td>Michael Harrison</td>
<td>University of Newcastle</td>
</tr>
<tr>
<td>Gregory Abowd</td>
<td>Georgia Institute of Technology</td>
</tr>
<tr>
<td>Nathan Moore</td>
<td>Head of Business Development, Centre for Evidence-Based Procurement (MHRA)</td>
</tr>
<tr>
<td>Oliver Wells</td>
<td>Association of British Healthcare Industries</td>
</tr>
<tr>
<td>Nick Komaromy</td>
<td>Royal Free Hospital</td>
</tr>
<tr>
<td>Paul Lee</td>
<td>Singleton Hospital, Swansea</td>
</tr>
<tr>
<td>Duncan McNeil</td>
<td>Royal Free Hospital</td>
</tr>
<tr>
<td>Laura Meagher</td>
<td>Independent consultant</td>
</tr>
<tr>
<td>Richard Lilford</td>
<td>MATCH Programme</td>
</tr>
<tr>
<td>Ed Matthews</td>
<td>Helen Hamlyn Centre, RCA</td>
</tr>
<tr>
<td>Chris Johnson</td>
<td>University of Glasgow</td>
</tr>
<tr>
<td>Michael Harrison</td>
<td>University of Newcastle</td>
</tr>
<tr>
<td>Gregory Abowd</td>
<td>Georgia Institute of Technology</td>
</tr>
<tr>
<td>Nathan Moore</td>
<td>Head of Business Development, Centre for Evidence-Based Procurement (MHRA)</td>
</tr>
<tr>
<td>Oliver Wells</td>
<td>Association of British Healthcare Industries</td>
</tr>
</tbody>
</table>

Membership of the Advisory Group will evolve over time, as new research challenges and engagement possibilities emerge.

Individual projects will be managed by a designated (academic) project leader, employing best-practice project management (see Work Plan and Justification of Resources for details).

Flexibility in planning
As noted above, progress on individual projects will be reviewed regularly. The initial direction of early projects is well defined, and preliminary goals of all projects, and methods for addressing those goals, have been identified. However,
as the research progresses, we anticipate that changes will be made in response to early findings, to research findings by others, and to changes in external policy and practices. While it is anticipated that UCL, QMUL and Swansea will be approximately equal partners in CHI+MED, resources will be managed from UCL to give flexibility in resource allocation for individual projects as the programme develops. Where appropriate, post-doctoral researchers will be given relative autonomy to manage projects and develop their own project management skills. At the outset, resources will be committed for the first two-to-three years (see workplan), with rolling review and commitment at Steering Committee meetings (or other management meetings if resourcing issues need to be resolved quickly – e.g. in response to an unanticipated opportunity).

**Risk management**

We have identified the following risks and mitigations in the programme:

**Interdependencies between projects:** Although the strands of work are interdependent, each has pre-existing resources (the device simulation of a syringe pump; a preliminary computational model of cognition; early empirical results) that will form initial input into other projects. In the unlikely event of any investigation failing to deliver results to another strand of work in a timely way, there are external resources (e.g., other simulations; empirical findings from other groups) that can be exploited to minimise knock-on effects. While PhD projects are planned to deliver useful results to the rest of the project, they will not be on a critical path such that delay would have a deleterious effect. As described in Annex 1, we have built frequent communications between individual projects into the overall programme to ensure effective interchanges and many individual projects are planned to involve participants from multiple sites.

**Access to equipment users and observational study sites:** We already have extensive connections into clinical settings and with patients (two members of the project team are clinicians), and prior experience of data gathering in clinical settings, so we do not anticipate there being access difficulties. There are multiple clinical study sites in both Swansea and London, so that if there are access problems at any time in one location, it will be possible to plan studies in another. We include letters of support from the Royal Free Hospital in London and Abertawe Bro Morgannwg University NHS Trust (Swansea) confirming their participation in CHI+MED.

**Access to professional stakeholder groups:** we already have very good links with many stakeholders, including the NPSA, the Association of British Healthcare Industries (ABHI – see letter of support), MediLink and MediWales (of which Ford is a Director and past Chairman). We have very good informal relations with manufacturers, and we have attached an example letter of support (from Zühlke), but other manufacturers were not able to provide letters of formal support, mainly citing legal reasons. However, we have every expectation of working constructively with manufacturers as we know individuals within them who are very excited about CHI+MED, and even if a few manufacturers have issues in engaging formally as collaborators (e.g., because of overseas head office practices, legal reasons, or because of mutual confidentiality issues) there are many who will participate fully with the planned open stakeholder activities (WP6). In fact, the investigation of ways to engage with such stakeholder groups has been identified as a research challenge, and is a focus of planned studies. It is anticipated that engagement with patient groups and policy makers will result in a culture change that will, in turn, influence manufacturers. Evaluation of the effectiveness of different forms of engagement will take place throughout the project.

**Staffing problems:** while many of the individual projects are designed in sets such that a single post-doctoral researcher can move seamlessly between projects, there will also be a need to recruit additional post-doctoral researchers at various points in the programme. We do not anticipate recruitment difficulties, but there is flexibility in the start times of most individual projects to accommodate any recruitment delays.

More generally, progress on all individual projects will be carefully monitored and action will be taken to minimise the chances of any problems escalating.

All individual projects will be regularly reviewed, e.g., considering when to finish each and what new investigations to initiate, as well as ensuring that projects are mutually informing each other. The planned review strategy is:

- with the science: investing most in the projects that are proving most fruitful, both in terms of basic science and in terms of ability to inform stakeholder engagement.
- with the stakeholder engagement: investing most in the projects that are likely to have the greatest impact.
- with the integration work: investing most in the approaches that have proven to be most effective to date and (for method development work) in those that appear to be most promising in terms of stakeholder impact.

All projects will follow appropriate ethics procedures.
Annex 1 (Work Package 1): Project integration

The first work package focuses specifically on integration across the core scientific and stakeholder engagement work packages. This will involve four different forms of integration:

- WP1.1: Relationships between the empirical work in the three scientific work packages (WP2, WP3 and WP4);
- WP1.2: Interrelating the computational reasoning of devices (WP2), people (WP3) and situations (WP4);
- WP1.3: Relationships between the core science (WP2, WP3, WP4) and stakeholder engagement (WP6); and
- WP1.4: Activities to synthesise across all projects and sites in the programme.

These different forms of integration are a focus for research in their own right: we will be investigating and reflecting on the success of the various approaches to integration as we apply them. Our initial plans for integration build on our prior experience (e.g., in AMODEUS, an ESPRIT Basic Research Action, BRA7040, 1992-1995, and HUM, EPSRC GR/S67494 and GR/S67500, 2004-2008). We will investigate further approaches to integration as the programme evolves.

WP1: Illustrative example

To illustrate likely forms of CHI+MED integration, consider an example based on a real incident report (there are many cases on MAUDE and elsewhere). This starts with a case in which a nurse who was familiar with syringe pump A (with which pressing the units button enough times causes it to cycle into the tens) was using the similar syringe pump B from the same manufacturer, and expected pressing the units button to have the same effect; it did not (rather, it cycled around the unit values), resulting in an incorrect dosage being delivered to the patient. There are many contributory factors to this incident: e.g., the two pumps are visually similar, with similar buttons for data entry, and neither alerts the user to differing effects when the units button is pressed more than nine times: key beeps are provided identically. It may be that the nurse had received insufficient training on the differences between the models, or that they were distracted and/or performed data entry as if for working with pump A.

This case study illustrates typical project activities in all work packages. Starting points will arise from any of the CHI+MED work packages.

- The device reconstruction project (WP2.1) will deliver simulations of both pumps, A & B, investigating which design features are identical and which differ, to deliver a specification of the interaction design features of the “family” of pumps as well as specific features of each model, as well as a realistic simulation. This process forms the basis for an evaluation method, the Interaction Walkthrough (Thimbleby, 2007a), which will be further developed in WP5.1 (device-focused method development). That method development will take account of the needs of device designers (WP6.1: understanding development), and testing of the method will include testing with device designers (WP6.2: stakeholder engagement with developers).

- The formal analysis of the two pump designs (also within WP2.1) will include automated identification of inconsistencies between the designs. These are likely to include both inconsistencies of which we are already aware and many others of which we were not initially. Those inconsistencies will require empirical investigation to evaluate their potential impact in practice (WP4.1: situated interaction). Many inconsistencies are likely to be formalisable, and hence leverage future work.

- Studies of individual cognition (WP3.1) will include a systematic series of experiments on capture errors (errors in which a familiar sequence of actions is performed in an inappropriate situation) and their avoidance, studying the effects of interruptions, multitasking and various possible mitigations such as visual indicators and auditory alerts on performance. The computational model of individual cognition will be extended to model capture errors and may highlight further questions for study, as described in Annex 3. Both the realistic simulations from WP2.1 (device reconstruction) and carefully designed adaptations of them will be used as experimental instruments to investigate capture errors (Norman, 1981), while the corresponding formal specifications will be used for computational reasoning (WP3.1). Findings will inform device redesign (WP2.3). Narratives from studies will be used in stakeholder engagement activities (WP6.4) and findings will also be essential for cognition-focused method development and testing (WP5.2), informed by WP6.1 and WP6.2 as outlined above.

- The situated effects of practitioners using different devices from the same ‘family’ will be investigated, generalising beyond pumps A and B to other medical devices that are within the scope of CHI+MED (WP4.1). These studies will include analyses of incident reports and critical incident interviews (Flanagan, 1954) as well as naturalistic observations in hospital settings and observations with interventions in training settings. Situated studies may highlight further problems in the use of families of devices that demand new investigations in WP3 (individual cognition) or that identify additional system properties that should be a focus for WP2 (devices). Conversely, situated studies may highlight behavioural workarounds that extend the understanding of cognition or highlight new design possibilities. Situated studies will be integrated with WP6 (stakeholder engagement) and WP5.3 (method development and testing) in similar ways to the empirical studies.

- Dialogues with the developers of pumps (WP6.1) and with other stakeholders such as the Centre for Evidence-based Practice specialists in pumps (WP6.3) will clarify their concerns, values and practices. This will, in turn, guide the design and use of materials to support engagement with them (including methods for developers as well as general information materials and joint events).

A common theme across CHI+MED activities is that there is a strong shared focus for activity: the example above shows how concepts become “boundary objects” (Star & Griesemer, 1989) for communication between disciplines and
teams, and indeed for other activities such as the development, delivery and evaluation of a tutorial for practitioners and other stakeholders.

Clearly, the different activities will not take place simultaneously, but will involve cycles of activity through different projects within the programme. This is just one example that starts from a situated finding: other devices and issues will also form focuses for study. Starting from WP2.1 (device reconstruction): a simulation of a Gnasby 3400 has already been developed, and is therefore ready for use as an experimental instrument and for computational representation in WP3.1, while early situated studies (WP4.1) will consider pump use more generally. Similarly, starting from WP3.1 (individual cognition): earlier work has highlighted the role of enforced reflection in mitigating errors, which will be tested in redesigns in WP2.3, while also being a focus for situated studies (WP4.1).

WP1: Background

Few studies have explicitly explored techniques for interrelating findings and methods from different research traditions. One approach that is discussed involves ‘triangulation’ across techniques and their findings (e.g., Wilson, 2006; Mackay and Fayard, 1997). Implicitly, triangulation involves applying different approaches to the same situation or object of study (e.g. a system design) and testing the findings of one approach against those of another in order to improve confidence in the findings or develop a richer understanding of the situation. Garmer et al. (2004) present such an approach to the redesign and test of an infusion pump, highlighting the complementary roles that different human factors and empirical approaches can play in identifying user requirements by creating a table of methods against the requirements identified by using them. Blandford et al. (2008a) employed a similar approach in comparing the findings of eight different methods for evaluating the usability of a robotic arm for use by people with limited movement.

Other approaches, still taking shared exemplars as boundary objects with which to explore integration, have included using Questions, Options and Criteria (QOC), a design rationale representation, as a unifying framework for interrelating findings from user- and device-focused analysis techniques (Bellotti et al., 1996) and “collational co-modelling,” which involved developing understanding across techniques by iterative explanation and questioning (Buckingham Shum et al., 1996).

While there is no agreed approach to triangulation across methods, all reported approaches share two features in common. The first is that they are applied to shared examples in order to relate findings; the second is that a shared representation framework is adopted to support interrelationships. That framework might just be a list of features (e.g. requirements or usability issues) or a more structured representation (e.g., QOC). Sometimes, it is simply a list of questions or propositions that are passed from one approach to another to seek clarification or validation.

The process of going from empirical or analytical studies to user-centred requirements, specification, design and implementation is one of successive translations between representations that are more or less formal. None of these steps is mechanistic: all involve interpretation and translation into an alternative representation with a different underlying ontology. One approach that will be explicitly explored in this programme is the use of computational representations that share the same ontology as a means of integrating findings from the core scientific work packages (i.e., devices, users and situations), as discussed below (WP1.2).

WP1.1: Integration across the scientific work packages

The illustrative example presented above highlights many of the forms of integration across the scientific work packages. For example, the device simulations developed in WP2 will be used as experimental instruments in WP3 (individual cognition) and will be tested in terms of the errors they actually provoke, which will be compared with the results of theoretical analysis of the behaviours they support. Conversely, findings from WP3 will be used to define user requirements that will guide the redesign work of WP2 (devices), and later studies in WP3 will use redesigns from WP2.

A similar relationship will exist between WP2 (devices) and WP4 (situated interaction): the device modelling highlights predictions about user behaviour that need to be validated in the situation. So WP4 will also be checking predictions from WP2. Conversely, findings from WP4 will inform redesign work in WP2.

The relationship between WP3 and WP4 will be different: whereas devices do not behave differently in laboratory and real-world settings, people often do, so it is essential to relate the findings of laboratory studies (WP3) with situated studies of device users (WP4). Whereas laboratory studies facilitate the controlled study of cognitive phenomena that underpin slips, situated studies facilitate the study of behaviours that emerge in response to the situation. To this end, we will test hypotheses in the individual (laboratory controlled) context, but then investigate whether the same behaviours manifest in the natural setting, or whether people have developed mitigating strategies that do not emerge in the lab. Conversely, we expect to observe behaviours in the natural setting that merit further investigation in the lab (e.g., we anticipate that people’s responses to time-outs will merit controlled study).

As discussed above, shared examples – both existing medical devices and new designs – will be central to the integration. Different approaches to integration will be explored through the programme, some more structured than others. At the unstructured end of the spectrum, we will be creating opportunities for ‘round the photocopier’ interactions between programme participants – through regular programme meetings, jointly organised (themed) workshops and researcher exchanges between sites. At the structured end of the spectrum, computational representations and reasoning techniques will be applied in all three work packages, and will in turn be related to each other to investigate how the ontologies of the different approaches interrelate. Between these, techniques such as those
outlined above (comparing issues or challenging each other with questions) will be investigated. These forms of integration across the scientific approaches will be used throughout the programme.

WP1.2: Integrating the approaches to computational reasoning

As described in the Case for Support (and in more detail in Annexes 2-4), the core science will be developed by applying both empirical and computational reasoning methods to the study of interaction design to ensure validity and rigour. One vehicle for integration across the three disciplinary perspectives will therefore be integration across the computational representations. This will be a separate project involving all the researchers that are directly involved in computational reasoning. In general terms, the interrelationships between the three approaches to computational reasoning are as follows; a more detailed investigation and development of the interrelationships will be pursued within the programme.

The computational work in WP2.1 (devices) will generate complex device specifications that are realistic; it will compute properties of the device (e.g., concerning consistency of interaction patterns and network properties such as reachability of states) and ask whether or not they matter from a user perspective. The computational work in WP3.1 (individual cognition) will be based on abstracted device specifications of features that are important for reasoning about user behaviour (including features of the interface); it will answer questions about what matters, and may identify important properties that will require a change to the representation developed in WP2.1. The relationships between the device representations developed in WP2.1 and WP3.1 will be explored in depth throughout the programme.

While WP2.1 develops formal models of devices and WP3.1 develops formal models of users, WP4.1 (situated interaction) puts them together, along with a situation/environment representation, to test properties of the system. In WP4.1, the aim will be to study the effects of communicative strategies and information flows; for this, it will be necessary to develop more abstract models in order to investigate emergent and situated behaviours. It will then be possible to identify situational properties, such as whether a system is resilient to errors by individuals within that system. Reasoning about abstract situational properties will serve to identify elements of representation that may be incorporated within the representations of devices (WP2.1) and people (WP3.1).

WP1.3: Linking the science with the stakeholder engagement

This programme is charting new territory in science/stakeholder engagement (SE) integration. As described in Annex 6, we will be exploring a variety of ways of developing effective dialogue with a broad range of stakeholder groups. It is likely that different aspects of the science, as well as different approaches to engagement, will prove to be most engaging for different groups. Our explorations of these topics will be developed into a framework for science/SE integration, which will be worked up into a form that supports ready application by other research groups.

Most of the communication techniques will directly link scientists with the researchers specialising in engagement and will involve, for example, scientists aligned to WP2.4 participating in science fairs or co-authoring materials with the stakeholder engagement researchers. The stakeholder engagement activities are described in Annex 6, while the development of design and evaluation methods for use by professional stakeholders are described in Annex 5.

WP1.4: Overall integrational activities

The integrational activities outlined above will be achieved through various mechanisms, including video conferencing, teleconferencing, document exchanges, and meetings. The formal meeting structure is outlined in the Management Annex. There will be two contrasting kinds of meetings: regular programme meetings (generally involving all programme participants) and themed workshop meetings that will bring together a small number of CHI+MED participants to work with stakeholders in a variety of ways (e.g., practitioner workshops, tutorials, a doctoral consortium, policy briefings). The latter are described in Annex 6.

Programme meetings will take place every 6 months (plus a start-up workshop). They will generally be 2-3 day meetings (some residential, some at a partner site) to work jointly on shared ideas. They will include research presentations, break-out groups, shared sessions with the CHI+MED Advisory Group, and team-building activities.

There will also be regular researcher exchanges between the three institutions, involving both post-doctoral researchers and PhD students (equivalent to a week a year per researcher). In these visits, the researcher will work closely with CHI+MED participants at the local site on problems of mutual interest; many of these visits will result directly in research papers or other deliverables.

WP1: Resources and deliverables

18 post-doctoral researcher (PDRA) months are allocated explicitly to the comparison of computational representations work. Additional effort from all researchers will be assigned to this work package throughout the programme (equivalent to approximately 20% of the total researcher effort).

Deliverables from this work package will naturally include scientific papers in international journals, conferences and workshops reporting findings of the integration work.
Annex 2 (Work Package 2): Device design

The first of the scientific work packages is concerned with interaction design of devices. Fatalities have been caused by details “as trivial” as decimal points: e.g., “.5” may be misread as “5” (ISMP, 2001, Zhang et al., 2003). We also know (Gray & Boehm-Davis, 2000) that barely perceptible details influence user behaviour. We know that neither manufacturers (Thimbleby, 2000; 2007b; 2007c) nor users (Payne, 1991) of devices know exactly how they work. This is ironic, because in principle interactive behaviour is fully specified by the program code of the devices. In fact, even with the benefits of current formal methods, there is very little work on the interactive behaviour and properties of programs, and what work there is (e.g., in the DSVIS conferences) is almost entirely based on simplified device models, typically analysing the behaviour of interface features or idealised features in isolation. WP2 will investigate real device models, rigorous approaches to reasoning about interaction design, and novel design solutions that improve interaction design based on a full modelling approach. WP2 comprises four projects: WP2.1 and WP2.2 start near the beginning of CHI+MED and focus on understanding current designs, while WP2.3 and WP2.4 are projected to start two years in and focus on redesign. WP2.2 and WP2.4 are PhD projects while WP2.1 and WP2.3 will be conducted by postdoctoral researchers.

WP2.1 & WP2.2: Understanding current designs: Background

“Latent failures” are defined by Reason (2001) as errors that lie in wait for the operator, but he primarily views latent errors as organisational failures. We include design errors that later impact user behaviour (e.g., features implemented in interactive device programs or in ergonomics decisions) in latent errors, but further, we define “contributory latent errors” as lying earlier in the causal chain. Thus bad training may be a latent error, but a contributory latent error may be that the devices being designed are inappropriate to the user’s task and situation such that any training would be insufficient. For example, in a radiation overdose (Johnston, 2006), the treatment centre was running two versions of software, old and upgraded, both of which were being used for a single treatment. The versions differed in “normalisation” and a radiographer made a mistake in normalisation. Why were the functionalities of the versions so different that the treatment centre needed to use both simultaneously; why were the designs of software different in a critical way; and why did the software not help the radiographer do a routine, and evidently error-prone, calculation they had to do on paper? The incident report noted: “[At] no point in the investigation was it deemed necessary to discuss the incident with the suppliers of this equipment since there was no suggestion that these products contributed to the error.” The key recommendation from the report, then, was to improve management processes (i.e., to cope with the systems), not to improve design to better match the requirements of use. Latent error in design was almost certainly a contributory factor (certainly it is not clear why radiographers were required to do hand calculations on paper which could have been automated), despite the report’s findings. Fixing the latent error of training at that hospital is worthwhile given poor device design, but it is not as worthwhile in the long run as resolving the contributory latent errors that lead to bad device design that in turn is a causal factor in adverse incidents. To do that requires all the research, understanding of use and situation this CHI+MED research covers, as well as closely researching the devices (and potential devices) themselves, that is the key focus of WP2.

Clearly latent error, with its contributory factors reaching back into inadequate design, is an “elephant in the room”: most commentators take the design of interactive systems as a given. On the contrary, Thimbleby (2007a) argues that device redesign can improve usability, and Lin et al. (2001) actually show that device redesign eliminates errors as well as speeding up use. Our reading of Lin et al.’s promising work is that more needs to be done, with more devices, and with more faithful device models, and with more thorough empirical work (see other Annexes): CHI+MED will leverage much similar research. In contrast, Lin et al.’s work is process-based (detect defects; improve; iterate) and not based on analytical approaches that promise to more effectively avoid problems in the first place (Thimbleby, 2007c). Not surprisingly, even leading work such as that of the National Patient Safety Agency (NPSA, 2009) is, in terms of the requirements of programmers, very vague and hard to know how to satisfy.

Recent papers such as Appert & Beaudouin-Lafon (2007) review the complexity of interactive system programming and propose their own approaches; they focus on powerful and expressive interaction programming rather than the design principles and properties of interaction programs. In contrast, we believe that only by knowing the appropriate programming properties (and having appropriate tools and techniques to use them) can user interfaces be improved. In turn, this begs deep questions, including: what are the relevant interaction properties; of the broad user interface design principles, how can they be converted reliably into properties that can be used clearly in development, particularly for real devices with unavoidable domain complexity? Manufacturers create extremely complex devices and the state of the art in software engineering and HCI is not adequate to design assuredly safe interactive devices. Guidance of the National Patient Safety Agency, the Institute for Safe Medical Practices and others on writing doses (e.g., to write “.5” as “0.5”) are not (but could be) enforced by the design of interactive devices. The operator manual for the Graseby 3400 makes the apparently helpful claim that “its input works like a calculator”—when scientific examination reveals that in fact it does not (Thimbleby, 2007a). Arguably this will contribute to errors; certainly it would be trivial to build a copy of the 3400 that differed only in its input working exactly like a specific calculator, and then comparing them experimentally.

Thimbleby (2008) shows that conventional calculators are inappropriate for clinical use: since they are designed to do any calculation, they are error-prone to use for particular calculations (e.g., a nurse fumbling ‘+’ instead of ‘/’ will get no error warning). Few articles in the clinical literature examine the role calculators play in medical error. However, the
available evidence is alarming: Kaushal et al. (2001) estimates 55% of adverse drug events (ADEs) have calculation errors as causal factors. Smart pumps and dose error reduction software (DERS) have been proposed (often without specific details) but these ideas have to be integrated with user interface design and situated use: for example, even if a “smart” pump rejects programming a dose greater than 50mL per day (e.g., using an RFID to identify the drug) it may still allow an operator to confuse .2, 2 and 20mL; furthermore, how it should best “reject” is a matter for WPs 3 and 4.

For many practicing users, interaction is too complex to be examined, and unfamiliar as an object of study. Thus nurse training, incident investigations and procurement all often ignore usability except at the most general level. For such users there is no feasible way to check how a device handles, say, numbers once it has been built. These limitations also apply in part to interaction experts: the only feasible way to conduct an analysis is formally, on the specification (or with code analysis on the source code)—and that presupposes formal models, such as CHI+MED will build.

**WP2.1 & WP2.2: The proposed work**

The foundational work (WP2.1) is to develop a resource library of faithful, well-defined, executable design implementations that can be used in both theoretical and experimental work in all parts of CHI+MED, by using rigorous reverse engineering and formal modelling. Reverse engineering will be informed by health trust trainers, trained users, technical manuals, user manuals, actual devices and manufacturers; these sources of information are usually inconsistent (indicative of design problems!), but the device itself will be the arbiter. Initially, these will be discrete models implemented in Java (because it is well-known and portable, thus supporting wide collaboration), but as WP2 progresses we will generalise them to continuous models. The derived models will look and work exactly like real devices, and will be useful as simulations on-screen, or as physical simulations (see WP2.3), running programs on PCs but realised in a realistic physical form by replacing the electronics of commercial devices. We will thus develop a precise, well-defined library of uncompromised devices that we can recruit trained users (nurses, anaesthetists, etc) to use exactly like real devices, whether in labs or situated in our simulated wards in the “NHS in a lab” in Swansea University’s HIRU (directed by Ford)—we will know exactly what is going on, and be able to record exact interaction behaviour (WP4.1).

Our approach is programming language independent: it does not rely on inventing or developing new interactive program development languages or environments. After developing a formal framework (e.g., with the help of Alloy), we will develop application program interfaces (APIs) in several languages (e.g., C, Java, etc) that extract abstract models of interactive programs in a portable exchange format (e.g., SCXML) that can then feed further CHI+MED activities.

Although our methods will be refined as CHI+MED proceeds, the initial approach will be reverse engineering into Java, then formal analysis with the benefit of an executable program (the Java simulation) to triangulate with the formal analyses. This process will stimulate interaction with the reverse engineering as “defects” in the simulation appear: the defects may be due to real device design defects or due (particularly initially) to errors in reverse engineering. Note that the device programs can be dynamically analysed to build exact formal models: e.g., an abstract state space search (e.g., pressing buttons to do a breadth first search) will generate a transition system over an equivalence class of states. We have already used this approach successfully in modelling many devices, including a Alaris GP pump, and in getting a Finite State Machine (FSM) from the Graseby 3400 simulation discussed in Thimbleby (2007a).

The formal models in turn can be analysed for many properties, e.g., path lengths, overrun errors, completion times, etc; see Thimbleby (2007c; submitted) for many examples of network properties. We will also use model checking and theorem proving techniques to analyse logical properties (Curzon et al., 1995; 2007b)—see WP3 for further discussion. There is no shortage of formal properties (we have started generating them automatically (Thimbleby, Gow & Cairns, 2006)), but we will also formalise informal properties from sources such as (NPSA, 2009). Annex 3 discusses how we will explore such properties empirically. The intention is that we will establish which formal properties or classes of properties have a real impact on safety and usability. One would then envisage designing systems from requirements based on those properties, or (until that can be achieved by industry) using the properties as an aid to procurement, or (with more immediate application) using the properties in incident analysis.

A useful outcome of this work will be categorising and structuring guidelines such as NPSA (2009) to provide clearer guidance to developers. We anticipate our fundamentally analytical approach will bring many insights into current guidance (e.g., NPSA’s notions of “software personality,” short-cuts, interface consistency, and error recovery).

It is proposed that a self-contained PhD project linked to the formal analysis will explore the cross-fertilisation of automated software testing and user interface testing (WP2.2).

**WP2.3 & WP2.4: Future design possibilities: Background**

There is enormous scope for the redesign of medical device user interfaces. Considering the case of calculators, we could build in more robust operational behaviour. For example, we could require (human) operators to first enter estimates, or two operators must get consistent answers before the calculator reveals them. Context awareness would allow devices to compare doses with those previously given in a comparable situation (time, location, patient), warning nurses about differences; voice synthesisers could engage contextual help (e.g., a patient in an adjacent bed will overhear the synthesised “5ml/hr” and might say “didn’t they only have 2ml yesterday?”), and the calculator can either be integrated in the device or communicate directly (e.g. by radio), thus eliminating re-keying problems. Devices within
the same ward could have their alarms designed so that nurses are not overloaded but instead guided in triage. And so on. Few of these ideas have been explored, and none have been explored in the context of rigorous device models supporting all tasks they are designed for.

Another significant factor is that devices are often deployed in sets of mixed types (e.g., an infusion pump with a heart monitor). Ensuring consistent operation across types of devices may contribute to further reductions in error patterns. Device manufacturers may recognise the advantages of identical product design, but surprisingly this often is not carried through to the interaction itself. The device design work of WP2 will identify the conceptual limitations and required supporting toolset (to be developed in WP5) for this approach, which will be evaluated in the work detailed in Annexes 3 and 4.

All current device user interfaces are based on discrete interaction models (e.g., two-state push buttons, multi-click knobs) and neither continuous interaction techniques (e.g., based on accelerometers) nor sensor techniques (e.g., RFIDs, proximity detectors) have been seriously evaluated in this domain. Novel interaction technologies, such as accelerometers and proximity sensors, provide a tight coupling between the user and the system based on a continuous input/output exchange of dynamic information, which happens over a period of time. In continuous, dynamic interactive systems, it is important to adapt to user behaviour, sense different inputs in real-time, coupled activities and provide quick and rich information to the user (Eslambolchilar, 2006; Eslambolchilar and Murray-Smith, 2008). We will examine novel, continuous interaction and its impact on safety and on the user experience more generally, and doing so will generate new research in interaction modelling.

WP2.3 & WP2.4: The proposed work

As we explore the real device designs, we will investigate various modifications to remedy identified problems. We will identify and investigate generic problems to develop effective interaction designs. Our knowledge, methods and principles from areas such as input optimisation and display constraints on mobile devices will be exploited in combination with a model-based approach. One known problem will serve as an illustrative example here: we will design and test an improved number entry protocol and then embed it in all the devices already tested. We can then rigorously explore how number entry details affect errors: error blocking, error detection and error recovery. For example, we can change number syntax and semantics. With iterative design, we will converge on a robust and effective number entry scheme that can be used across a wide range of devices (ultimately leading to standards specifications). Here, ‘number entry’ is simply an example user interface (UI) feature; we will of course explore many throughout the project. Modelling work in WP3.1 will make it possible to conduct repeat automated analysis (using model checking/theorem proving techniques) and hence guide sharp, insightful empirical experiments.

We will not restrict our studies to established interaction paradigms, but will extend those paradigms by introducing continuous interaction techniques (including but not limited to proximity, RFID and other sensors into our physical devices). The enhanced parallelism and multi-user interaction will drive further developments in formal methods in HCI; for example, Norman’s informal notion of ‘natural mapping’ (Norman, 1988) makes direct sense in continuous interaction, and can be explored and made more rigorous.

As the CHI+MED team performs its own development of new device designs, it will scrutinise the contribution made by different methodological approaches (developed in WP5.1). The goal will be to identify the methods for analysing interaction that are both effective functionally and are congruent with the software development process, necessarily including working with practicing developer communities through WP6.

It is proposed that a PhD student will complete a self-contained project (WP2.4) on developing and testing prototype ambulatory devices.

WP2: Resources and deliverables

The work associated with this work package is currently allocated 129 PDRA months (111 at Swansea, 18 at UCL). Two PhD students will be associated with WP2.

Deliverables from WP2 will include:

- Executable, faithful device models, and papers analysing in depth these real devices. All devices and derived material will be made fully available on the web, to allow other researchers to check or build on our work.
- Contributions to the theory of interactive system programming, developing interaction properties perhaps as useful as ‘invariant’ in non-interactive programming (cf, our partial theorem work, Thimbleby et al., 2006), and particularly related to safety-critical properties.
- New designs, and reports on the design and evaluation thereof.
- Two PhD theses.
- Themed workshops on the research and practice of analysing existing devices and designing safe novel devices.

Findings will be communicated to the academic CS community (software engineering, human-computer interaction), the medical device industry, and medical computing field. DSVIS (Design, Specification and Verification of Interactive Systems) is the natural conference outlet for WP2, but clinical and general HCI outlets will also be used for greater impact.
Annex 3 (Work Package 3): Individual cognition

The objective of Work Package 3 is to deliver accounts of cognition that relate particularly to users’ interactions with medical devices. A particular focus will be on understanding, representing and reasoning about cognitive slip errors (Byrne and Bovair, 1997) because this topic is of central importance and has been relatively under-researched. Furthermore, the research that has been done has largely concentrated on the psychological details (e.g., refining the descriptions of cognitive mechanisms that explain slips), rather than the management of slips (e.g., advice for developers). Cognitive slips are errors that occur through inattention rather than lack of knowledge; previous research has shown that these errors cannot be eliminated through simple training or practice, and therefore need to be addressed through good interaction design (Byrne and Bovair, 1997). Such slips include omission errors (e.g., forgetting a clean-up step at the end of a procedure (Li et al., forthcoming)) and capture errors (Norman, 1981), in which people follow a familiar procedure in an inappropriate situation (see example in Annex 1). A primary contribution of WP3 will be a much deeper understanding of cognitive slips, expressed in forms that support design and assessment of systems such as medical devices. The main investigations in WP3 (WP3.1) will be conducted by post-doctoral researchers; WP3.2 and WP3.3 will be supporting PhD projects.

While situated studies such as those of Gaba et al. (1987), as well as those to be conducted in WP4, can identify slips in practice and speculate on their causes, it is not possible to test our understanding of causes in the natural setting. Controlled laboratory studies enable us to test hypotheses about the cognitive causes of error. Designs that mitigate erroneous behaviour by better supporting cognitive capabilities need to be explored and evaluated. Computational reasoning can examine a large number of design alternatives without the need for empirical testing. The best redesign candidates identified during this examination can then be further tested empirically. Some empirical studies will be designed in response to findings from other work packages, as discussed in Annex 1; others are planned from the outset. Planned topics for investigation include: what design factors increase or decrease the likelihood of capture errors when transferring between similar devices; what compensatory strategies people adopt to manage interruptions; how changes in demand or stress affect strategy adaptation; and whether design interventions such as the introduction of salient, just-in-time cues (Ratwani, 2008) or opportunities for reflection (Back et al., 2008a) significantly reduce errors. Our modelling approach provides a computational framework for representing such factors and establishing how they interact with the factors we already know have an impact on error rates.

As in the other scientific work packages, there will be tight coupling between the empirical studies and computational representation and reasoning, with empirical studies providing the data underpinning computational model development, and the modelling challenging our understanding and raising new questions that merit further empirical investigation, as well as enforcing clarity and supporting generalisation beyond the details of the particular studies that are conducted.

WP3: Background

Human error has been classified in various ways; one of the common classifications (Reason, 1990) is into slips and mistakes. Mistakes are a consequence of the person having incomplete or incorrect knowledge of what to do next, while slips are a consequence of inattention. Since the users of medical devices (whether clinicians or patients) are routinely trained in their operation, the main focus of this work package is on slips rather than knowledge-based mistakes.

There is growing interest in cognitive slips, their causes and mitigation. Empirical studies have been conducted on the role of interruptions (Li et al., forthcoming; Altmann and Trafton, 2007), visual cues (Chung and Byrne, 2008), and procedural cues (in which task steps become associatively linked) (Altmann and Trafton, 2002) in provoking or mitigating errors. Various studies (e.g., Hodgetts and Jones, 2006a) have shown that slips are more likely when cues are less salient. If both domain and procedural salience are low, as for device-specific actions that do not move the system state toward a task goal, designs must reduce the risk of users forgetting them. To be resilient, sensory cues must be highly salient and specific, i.e., clear and timely, indicating the action needed (Back et al., 2008a). By analysing a comprehensive range of laboratory experiments (e.g., Chung and Byrne, 2008; Ratwani and Trafton, in press) and conducting our own (e.g., Back et al., 2007; Li et al., forthcoming) we have identified a series of primary factors that influence the strength of cues used to drive human interaction with a device. These include delays, the use of external resources, the inherent difficulty of the task, the load imposed by information that does not contribute directly to the performance of a specific goal, and disruptions to the ways external cues are presented. Perceptual studies have shown that sensory cues are not always noticed under high workload conditions (Lavie et al., 2004). However, further studies are needed to explore the relationships between cue types in various situations (different levels of workload, multiple task demands, device behaviour, etc.).

There are established traditions of computational modelling of cognition, drawing from cognitive science (e.g., Gray, 2000; Byrne and Bovair, 1997) and from formal methods (e.g., Rushby, 2001; Duke et al., 1998). Some approaches (e.g., Rushby, 2001) involve individually crafting a model for each new device, whereas others (e.g., Duke et al., 1998; Cerone et al., 2005) have developed generic models of behaviour based on results from laboratory studies that are then instantiated with the unique details for different scenarios. Our work is in the last of these traditions; it involves constructing a generic model based on empirical findings. Work to date covers non-deterministic choice between salient actions, and reactive, goal-based and task termination behaviour. Erroneous actions emerge from the cognitively plausible behaviour specified. Modelling is based on the higher-order framework SAL (de Moura et al., 2004). Its state-space exploration tools can be used to exhaustively analyse task scenarios (Rukšėnas et al., submitted). Compared to simulation-based approaches, our models are abstract and highly non-deterministic; we thus do not aim to predict likely
behaviours; rather, our models can be used to analyse an exhaustive range of cognitive factors that may have ‘caused’ an error. Such models include consideration of the limitations and capabilities of the user, bringing human error within the scope of reasoning. We have demonstrated that several manifestations of slip errors can be reliably predicted by including various types of cueing mechanisms into a computational user model (Rukšenas et al., 2008). Such models can now be developed and used in parallel with empirical studies. This tightly integrated approach means it is possible for reasoning about models to give deeper insight into the results of experiments as well as vice versa.

**WP3.1: The proposed work**

We will further develop the combined investigative approach integrating empirical and computational work while also extending core human error research. The empirical results will feed the modelling work, but at the same time the latter will add insight to the experiments and drive new hypothesis generation. In this way a deeper understanding of error will be gained. In addition the computational basis will be developed for techniques and tools capable of identifying medical devices that are likely to be prone to such problems. We first present more detail on the modelling approach, then describe illustrative experimental studies to exemplify the approach.

The computational reasoning work will consist of at least two facets. The first, and more direct, one concerns problematic medical device behaviours potentially causing human errors. Such potential will be identified using automatic state-space exploration tools and a tractable device-user model. Formal models of the device-user will be based on generic cognitive rules and an understanding of the behaviour of a specific device. Further computational reasoning will then be applied to examine various design alternatives for medical devices, including a range of designs and devices beyond those that will be experimentally investigated. By determining their error potential, this examination will suggest the best designs that can be further tested empirically. Such computational reasoning can be performed using the current version of our generic user model. However, its predictions will be made more precise by further development of the model, which is the second facet of our computational reasoning work within this package.

In this facet, computational reasoning will be carried out on the same scenarios as experiments. Results will be compared, leading to the generation of new rules of interactive user behaviour, modifications and extensions to the generic user model, and the formulation of new hypotheses to be tested. Initially, we plan developments in the directions discussed below. Further developments will be informed by experiments in this work package and findings from other strands of the programme. In particular, the computational model will be adapted to cope with cognitive factors relevant to the situated use of medical devices (WP4). This will allow more detailed situational contexts to be explored.

The existing salience model involves several parameters such as different types of cues and cognitive workload. The interplay of different cues is formally modelled by ordering overall salience levels based on individual rules for each cue type rather than just their visual salience. This idea is novel and central to our approach because it enables the detection of non-resilient interactions by simultaneously considering the influence of various factors, such as workload and interruptions, and underlying cognitive mechanisms. The first development will involve extending our user model with parameters to deal with the factors explored in parallel experiments, such as perceived similarity, task interruptions, mental rehearsals of actions, time pressure and other stressors.

Our existing user model (Rukšenas et al., 2008; Rukšenas et al., submitted) focuses on well-defined interactions such as reactive user behaviour guided by interface prompts. Another development will be extending the range of interactions captured by our model. It will involve the modelling of more opportunistic, yet still goal-directed, interactions such as error avoidance behaviour in the presence of interruptions. Our pilot work (Rukšenas et al., 2008b) indicates that this can be achieved by a game-based user model that consists of a collection of agents; we will research this further. An agent models one aspect of user behaviour: e.g., reactive agent. Such an architecture can be seen as a game of two agent coalitions (so-called angelic and demonic). The angelic (resolved in the best way) decision about the relevance of actions embodies the human ability to act according to the situation. The demonic (resolved in the worst way) choice between relevant actions embodies the human propensity to act differently in the same situation. Different levels of task knowledge and situated use can then be captured by appropriate instantiations of the generic model and suitable choice of agent coalitions.

Laboratory experiments have furthered the understanding of the cognitive mechanisms used to execute procedural knowledge. Our work is informed by, and contributes to, such work. Our experiments will manipulate factors such as cognitive workload; position, duration and nature of interruptions during task performance; device designs (e.g., introducing timeouts) and task scenarios. Such factors will be carefully controlled and manipulated to foster an understanding of the conditions under which errors are or are not made. Manipulating complexity in this way is not possible in the real world.

We will use experimental instruments that simulate existing medical devices (or are re-designs of such systems). For early studies, we will make use of an existing simulation of a Graseby 3400 syringe driver. The potential for errors with this device has already been identified by using an interactive walkthrough method (Thimbleby, 2007a), so these studies will serve as empirical tests of the findings from that analytical method. Experiments will be devised that evaluate the extent to which the interface design has the potential to cause systematic slip errors. Where instances of systematic errors are found, design manipulations will be tested to probe the underlying cognitive causes of the errors and to identify designs that minimise the potential for error.

Subsequent studies will focus on factors identified in other work packages as contributing to error (such as capture errors for similar devices, interruptions, poor task mappings and timeouts), and will exploit a variety of experimental
instruments (all based on medical device simulations, developed in WP2) that are adapted to support the studies. Each series of studies will systematically manipulate variables to develop a rich understanding of the particular phenomenon under consideration: typically 3-5 experiments per study focus. The computational model will be adapted as necessary to deal with these issues.

Our experiments will be designed so that participants who are non-domain experts (e.g., recruited from the UCL Psychology subject pool) can be trained in a 30-minute session. After training, they will be tested to ensure they understand the task and will only proceed to the experimental procedure if they perform well. Since the details of later experiments will depend on the outcomes of earlier ones, and on findings from other strands of the programme, we will not describe every experiment in detail here. Rather, we outline one exemplar series of experiments to illustrate the approach we are taking.

One series of experiments will investigate designs that support users in managing interruptions (which might occur at any point in an interaction). This demands a rich understanding of how people respond to and recover from interruptions. For example, people may prepare by delaying their response to the interruption until they have created an external reminder of where they are in the task, or by mentally rehearsing the next action. However, previous studies (Li et al., forthcoming) show that people often omit required procedural steps when returning to a primary task following an interruption. One experiment will study how participants may be encouraged to reflect on task requirements to better manage interruptions. Controlled elements of the task environment will be manipulated with regards to the onset of secondary tasks (which frequently occur during the situated use of medical devices, and will be investigated in WP4) alongside subtle variations in procedural routine that simulate device appropriation. By understanding the situational context in which individuals develop strategies to avoid error, such as being given the opportunity to reflect on previous actions, or using sensory cues as markers indicating the next step to be performed, we can identify features of the device design or task environment that enable interruptions to be better managed. Time to complete tasks and opportunities to reflect on task requirements will be manipulated in a version of the interface where participants can set visual reminders to themselves to complete selected steps. The experimental design will be mixed factorial, with a within subject factor of performance before and after an opportunity to reflect and a between subject factor of time to complete tasks.

**WP3.2 & WP3.3: PhD projects**

People often have to ensure that they are completing a task correctly while monitoring other situational events in the background (e.g., ensuring that a device has all the patient information required before beginning an infusion while monitoring the state of the patient). It is anticipated that one PhD student will conduct empirical studies on a selected aspect of how clinicians develop appropriate attentional strategies (WP3.2).

Similarly we also plan a computationally-centred PhD project. This will involve development of the computational model to involve modelling situated use in a basic setting of an individual and a single device. A possible focus for this project (WP3.3) will be on how design of work practices can help individuals to avoid slips in multi-tasking situations, but (like all PhD projects in CHI+MED) the detailed focus will be developed once the particular student has been recruited.

**WP3: Resources and deliverables**

This work package in planned to involve 2 PDRAs (an experimental expert at UCL and a formal methods expert at QMUL) for 54 months each, and two PhD studentships.

Deliverables from this work package will include:

- Papers in leading international conferences and journals, and also presented at workshops. Findings will be written up for multiple audiences, including experimental psychology, cognitive science, cognitive modelling, formal methods, computational reasoning and medical computing.
- Two PhD theses.
- A generic user model that will be made available for others to use, test and extend.
- Themed workshops (see Annex 6) on human error with medical devices and on computational reasoning about human error, aimed (respectively) at practitioner and researcher communities.
- Representations (e.g., digital stories, short videos, textual narratives; see, e.g., www.patientvoices.org.uk for the recognised value of these media) that can support stakeholder engagement work (developed in collaboration with colleagues in WP6).
Annex 4 (Work Package 4): Situated interaction

Work Packages 2 and 3 focus principally on medical devices out of context in order to better understand their design and use in a systematic way. However, that understanding is of little value unless it translates into the ways interactive medical devices are used in practice. WP4 thus focuses on use in practice. We will study the rich variety of settings in which medical devices (as defined in the case for support) are used: operating theatres, intensive care units, in-patient and day care wards, GP surgeries, homes and elsewhere. We will also be studying their use by different user populations (clinicians, patients, carers), including situations (such as wards) where multiple people may interact with the same device. WP4 will deliver accounts of the use of medical devices in practice that focus on the interaction design and the interplay between the interaction design and the situation of use. Note that we have proper access to patients, patient actors and clinicians, and that full ethical procedures will be put in place. The main investigations in WP4 (WP4.1) will be conducted by post-doctoral researchers; WP4.2, WP4.3 and WP4.4 will be supporting PhD projects.

WP4: Background

Observational studies in medical practice are predominantly focused on biomedical information systems and the behavioural skills of medical and nursing professionals rather than on use of medical devices as here. The area of observational work on biomedical informatics in context seeks to assess and improve the interaction of information systems in the medical field. For example, Bossen (2006) evaluated new electronic health records for patients and showed that medical staff had needs of the patient records that were not captured by the electronic version. Also, Reddy et al. (2003) studied the deployment of pagers for physicians and showed that the paging device broke down traditional communication channels between physicians of different seniority, which impacted on workflow. The second major area of observational work in the medical domain is on the assessment of the behaviour of medical staff which includes the recognition of behavioural markers of surgical excellence (Carthey, et al., 2003) and the assessment of the non-technical skills of surgeons (Yule et al., 2008), anaesthetists (Fletcher et al., 2003), scrub nurses (Mitchell and Flin, 2008) and surgical teams (Catchpole et al., 2007a). In this work, taxonomies of qualities such as situation awareness, decision-making, task management, leadership, communication and teamwork are used to structure observations and performance assessment.

Observational work with surgical teams has also investigated distractions and interruptions (Healey, Primus and Koutantji, 2007) or looked at the frequency of minor failures and how the system compensates for this (Catchpole et al., 2006; 2007b). Minor failures will normally be compensated for by actions of the individual and the team so overall performance is maintained (Catchpole et al., 2006; 2007b; Carthey, et al., 2001). Good practice looks to reduce the occurrence of minor failures to make the system safer.

Our observational work focuses on the design and interaction of medical devices in the socio-technical system rather than biomedical information systems or individuals’ behavioural skills. Particular features of our approach are that it focuses on device use and that it will take a Distributed Cognition perspective on understanding the situation.

Distributed Cognition, whilst not the only possible perspective for understanding team working, provides valuable structure for observation and analysis. The approach builds on our prior experience, i.e. in the development of DiCoT: Distributed Cognition for Teamwork (Blandford and Furniss, 2006; Furniss and Blandford, 2006). It provides a structured approach to organize ethnographic data in terms of five integrative models: an information flow model, a physical model, an artefact model, a social model and an evolutionary model. This structure allows the context to be analysed from a ‘complex cognitive system’ perspective (Flor and Hutchins, 1991). It is ‘complex’ because it involves a variety of different phenomena, e.g., communication, space and furniture layouts, representations, tools and artefacts, social hierarchies and historical developments of the context. It is ‘cognitive’ because it expands the computational metaphor of the mind to include activities outside of the head e.g., using a list to extend memory and Post-it notes to distribute attention. It is a ‘system’ because its perspective focuses on how cognition is made up of different interacting parts of the material and social environment. This perspective has been successfully applied to the London Ambulance Service control room (Furniss and Blandford, 2006) and has been extended to analysing agile software system development processes (Sharp et al., 2006). The application of DiCoT is ideal for socio-technical systems in the health domain and will facilitate the method’s development, as described in Annex 5.

In contrast to individual cognition (WP3), there is little prior work similar to that which we are proposing on computational representation and reasoning for team working in rich resource-based environments. Early work was that of Zhang and Norman’s (1994) on computational representations of information forms for distributed cognition. The work of Doherty et al. (2008) provides a closely related precursor to our approach. Building on earlier work by Campos and Doherty (2006) that considered resource related analysis for structured tasks, they applied computational reasoning to an interaction scenario around process control involving mobile devices. They explored the use of an action-based analysis based on the explicit modelling of resources with an aim of determining whether situated actions are suitably resourced. As they note, our DiCoT analysis provides a potential wider framework for their approach. They also suggest the need for extensions, such as considering salience and cues, of the kind that will be supported by WP3. Their work shows the potential of computational reasoning to aid understanding of resource-based situations such as we plan. CHI+MED also goes further in exploring the benefits of this general approach in ethnographic work of real, complex situations.
WP4.1: The proposed work

As in the other core scientific work packages (WP2 and WP3), empirical data gathering and analysis will be complemented by computational representation and reasoning. This general approach is outlined in the Case for Support and described in more detail in Annex 3. The approach in WP4 will be similar to that in WP3, with insights from the data gathered used to model essential features of situations of interest. This will allow deeper understanding via exhaustive exploration of the consequences of the models. Insights resulting from the modelling will be fed back into the data gathering process for validation and to suggest areas for further observation. The reasoning in WP4 will differ from that in WP3 in that, as described in Annex 1 (WP1.2), the representations used will largely be developed at a higher level of abstraction, focusing on features of the people, devices and environmental factors that are central to their effective interaction. This will build on an approach developed in our previous work looking at strategies in electronic diary use that complemented empirical data gathering (Blandford et al., 2004). Features that we plan to model from the outset include information flows and roles such as ‘buffering’ and ‘information hub’ (Furniss and Blandford, 2006). Further areas for study will emerge as the project progresses and will be chosen on the basis of the likely insights computational reasoning could provide. DiCoT will be used as a framework around which to structure the modelling and reasoning activity.

This combination with computational reasoning will mean observations from empirical work are subject to a finer level of scrutiny and granularity. We do not propose to develop comprehensive simulations of the rich contextual activity of natural environments, but instead wish to explore how significant computational features and structures interplay to affect the efficiency and safety of the system. For example, a nurse interacting and monitoring one infusion pump might not pose a significant threat, but if that nurse has 5 infusion pumps to monitor, with alarms from other systems also demanding attention, then this might pose a different threat. To complicate issues further, as described in WP1, the use of different infusion pumps, often provided by the same manufacturer, on the same ward, can cause confusion in interaction. These issues would not present themselves in tests of individual devices away from the stresses and demands of practice, but will be a focus for the computational reasoning in WP4.

Since the greatest difference between WP4 and the other scientific work packages is in the data collection and analysis, we now focus on these issues.

Observational studies will be conducted in two modes throughout the research project: an exploratory mode where positive and negative issues of medical device use are explored with healthcare professionals and patients in practice; and a reactive mode where issues and questions that emerge from other work packages direct the focus of the observational studies. In both modes the stance of WP4 is one of listening to and observing what is actually happening in practice.

Fieldwork on medical devices will be conducted in operating theatres, intensive care units, in-patient and day care wards, GP surgeries and patients’ homes. In all of these contexts a socio-technical description will be developed using DiCoT as a structured filtering technique (see Annex 5). Appropriate engagement with these areas will necessitate the deployment of different data gathering and analysis techniques. Research strategies to improve the effectiveness and efficiency of data gathering will also be employed. These strategies will often only reveal themselves in response to the people and context under study but lessons will be incorporated from existing literature. Appropriate ethical approval will be sought and data gathering techniques will be worked through with medical and nursing staff, carers and patients.

Reports, interviews and observations form a framework for the data gathering techniques and methods that will be employed in this project.

• Reports will include structured literature reviews to orientate analysis to issues and contexts and the analysis of training materials on device use; they will also involve analysis of incident reports from sources such as the UK National Reporting and Learning System (NRLS), and encouraging medical and nursing staff, carers and patients to open new self-reporting channels using diary studies and other feedback routes. The effectiveness of the latter self-reporting channels will be trialled by providing lessons on user-centred design, and a vocabulary to describe device deficiencies and cognitive slips, so minor incidents with devices are more visible to the non-HCI professional.

• Semi-structured and structured interviews will also form a large part of the methodological approach. ‘Normal practice’ is failure free but specific incidents and near misses will be explored through the critical-incident approach (Flanagan, 1954) in the different contexts. Focus groups of patients and professionals will be used if they are deemed suitable to specific research questions. Indeed, scenario-based focus groups are recommended to elicit views of special groups (Mc Gee-Lennon and Clark, 2008). However, others criticise focus group approaches for more vulnerable people who may have travelling difficulties and feel less at ease away from their own home (Dickinson et al., 2003).

• Observations will form a major part of the methodological approach. Observations in operating theatres have in the past generally been structured observations but other approaches have been more reactive to the context employing ethnographic techniques and grounded theory to patterns and themes (Reddy, Shabot, and Bradner, 2008). Video analysis will be used where appropriate; e.g. Kaufman et al. (2003) use video analysis when carrying out user tests in people’s homes. Contextual inquiry will be used in work shadowing practice where it is appropriate to ask questions during tasks. The effectiveness of exploratory observations, which focus on wards rather than particular people and tasks, will also be investigated. These methods will allow us to detect certain behaviour patterns; e.g., as
described in the case for support, patients have been observed to turn off the alarms of their infusion pumps despite being told not to do so.

The different contexts under investigation will require appropriate methodological adaptation. Intensive care units, wards and GP surgeries will involve observations and interviews that will be more and less structured. The more structured will have organised analytic focuses in terms of observing a time, place or person, e.g. doing structured observations with work shadowing. The less structured will be more opportunistic and exploratory, and will involve informal observations. As an example, Carayon et al. (2005) observed the use of infusion pumps in wards during the morning and the evening as this was when medication was most likely to be administered.

Observational work in homes presents a special research challenge in terms of the efficiency, effectiveness, privacy and ethical issues of data gathering and analysis. However, this mode of study has importance in seeing how technology integrates with the home and makes patients feel more confident in a familiar setting (Dickinson et al., 2003). A combination of diary studies, interviews and video capture will be used to record minor incidents with medical devices. Participants will be loaned recording equipment to support their reporting of incidents. We also plan to compare new and experienced users of medical devices. New users will have teething problems whereas experts will have well developed routines and procedures. Learning the use of new medical devices raises issues of knowledge transfer and support in the wider socio-technical system. To elicit issues with medical devices there may be opportunity for holding focus groups with new and experienced users, which would also facilitate knowledge transfer. We will elicit people’s routine procedures, investigate deviations from this path, and compare this to normative interaction for the device.

WP4.2, WP4.3 & WP4.4: PhD projects

Our intention is that one PhD student (based at UCL: WP4.2) will investigate how the growing range of medical devices (for drug administration and monitoring) are used by professionals and cancer patients in the home and elsewhere. The focus will be on understanding how the devices are really used in practice: how they may serve as boundary objects between the patient and their carers (both lay and professional), how access to the devices are negotiated between these different user groups, the difficulties people experience in using them and how any errors are recovered from.

A second PhD student (based at QMUL: WP4.3) will develop ways of integrating an account of the environment and individual capabilities with a formal model of user behaviour. A possible focus for this project is on how to encode a representation in terms of distributed cognition that supports predictive reasoning about error. An alternative possibility would be to adopt and adapt the ideas of the resource analysis approach of Doherty et al. (2008) within the HUM framework (Rukšēnas et al., submitted) allowing the resource analysis to also take into account salience and cues as well as other aspects of the model. This could follow our approach combining HUM with GOMS (John and Kieras, 1996) to give an integrated timing and error analysis method (Rukšēnas et al., 2008d).

A third PhD student (based at Swansea: WP4.4) will investigate how a growing range of medical devices are used in clinical situations. Again, the focus will be on understanding how devices are used in practice: how they may serve as boundary objects between different members of the clinical team, how interruptions and timeouts are managed, the detailed causes of user difficulties and how any errors are recovered from.

WP4: Resources and deliverables

WP4 in planned to involve 2 PDRAs (at UCL and QMUL) for a total of 84 months, and three PhD studentships.

Deliverables from this work package will include:

- Papers in leading international conferences and journals, and also presented at workshops. Findings will be written up for multiple audiences, including safety, team working, human factors, computational reasoning and medical computing.
- Three PhD theses.
- Two themed workshops (see Annex 6) on situated use of medical devices, designed for research and practitioner communities.
- Representations (e.g. digital stories, short videos, textual narratives) that can support stakeholder engagement work (developed in collaboration with colleagues in WP6).
Annex 5 (Work Package 5): Method development and testing

One of the important features of CHI+MED is that we will be developing and evaluating mediating representations that support communication between science and stakeholders. The focus in WP5 is on tools and methods that support stakeholders in design or evaluation of medical devices. WP5 comprises three projects, each developing and testing methods from one of the scientific angles.

WP5: Background

Within CHI+MED, different approaches will be pursued for communicating both ways with developers and with other stakeholders. Mediating representations for engaging with other stakeholders are discussed in Annex 6; here we focus on the development of design and evaluation methods as a means of communicating with and empowering developers and other professionals with a direct interest in device design.

Design and evaluation methods are valuable tools for communicating theory to developers as well as delivering structured techniques for systematically developing or evaluating designs. Within other safety-critical industries, such as transport and power generation, there is an established tradition of applying rigorous design and evaluation methods to improve system safety. The number of techniques for reasoning about human error has mushroomed in recent years (e.g., Hollnagel, 1998; Cacciabue, 2004) to cover more and more of the space of concerns. The earlier approaches are typically probabilistic, and such approaches have a role in highly routinised settings where probabilities can be meaningfully established but not in variable situations such as those where medical devices are used. More recent approaches have been developed to support reasoning about human error within organisational contexts, but none support reasoning about the details of interaction design, particularly in relatively unstructured situations, or about cognitive slips. From another tradition, approaches such as Cognitive Walkthrough (Wharton et al., 1994), Heuristic Evaluation (Nielsen, 1994) and EMU—Evaluating Multimodal Usability (Blandford et al., 2008b), etc, have been developed to support developers in considering usability of their products, but these methods do not address safety properties directly nor are they generally analytic for use early in development. The wide range of existing methods will provide a useful foundation on which work in CHI+MED that focuses on safety in interaction design can build.

Blandford and Green (2008) present two scenarios that commonly lead to the development of new methods: a recognised need (in the context of CHI+MED, say to assess the vulnerabilities of a design to slip errors) or an emergent opportunity (e.g., the realisation that the process of rationally reconstructing a design yields valuable insights into the interaction structure). Both the Interaction Walkthrough (Thimbleby, 2007a) and DiCoT (Furniss and Blandford, 2006) methods represent emergent opportunities. From those starting points, the development of a method, like any other design problem, involves both creativity and critical evaluation and testing. Thus in this work package, we will be developing novel methods for design and evaluation from both directions: starting with existing resources (such as IW and DiCoT) and also developing methods in response to needs (e.g., to better support reasoning about cognitive slips).

As for any other product, methods need to be suitable for their context of use; in this case, they need to fit design or procurement practices (as identified in WP6). Methods also need to be evaluated. Many criteria and approaches have been identified for evaluating methods. Ultimately, what matters is what the costs and benefits of applying any particular method are. Costs include the time and effort it takes to learn a method and then apply it to a particular system; benefits include the insights obtained from applying a method—including the requirement that those insights should be valid. Other considerations include how well a method informs design practice (Wixon, 2003) and how easy it is for different evaluators to apply the same method consistently (Hertzum and Jacobsen, 2001). In the context of CHI+MED, methods will be developed and tested iteratively, applying the criteria of usability and usefulness.

WP5: The proposed work

Three projects will develop design and evaluation methods based on the three scientific strands of work. All will address recognised needs, and will exploit the scientific foundations being developed in Work Packages 2–4. Method development will proceed through iterative cycles of development and testing. Testing will consider usefulness, usability and ability to adapt to the intended contexts of use.

Development will take different starting points (as discussed below for each project). Methods will be tailored for different stakeholders—e.g., designers, evaluators, purchasers or investigators. Testing plans will be devised to be appropriate to the stage of development, the audience for the method, and the particular method/evaluation questions that are most pertinent at the time. Broadly, testing will involve developing tutorial materials describing and illustrating how to apply the method. These will usually combine written guidance, interactive resources and presentation materials. Tutorials will be tested formatively by being used in-house (see WP2.3), by organising training sessions with volunteers (e.g., UCL HCI MSc students, Swansea clinical students), and finally with professional stakeholder groups. Early studies will be organised by WP5 researchers, whereas later ones will be designed and conducted jointly with WP6 researchers. Data will be gathered from tutorial participants to measure both the usability and the perceived utility of the methods in order to inform redesign at each stage.

The methods to be developed build on different foundations from the three underpinning scientific work packages. In all cases, method development will be responsive to both developments in the science (WP2, 3 and 4) and findings from WP6 on practitioner requirements.
**WP5.1: Device-focused method**

One device-focused method will build on prior work on an Interaction Walkthrough (Thimbleby, 2007a). We will investigate means of making IW acceptable to and usable by practitioners, including by developing and testing a tutorial, building a repository of worked examples, and developing and testing tool support for the method (part of the API as discussed in WP2.1, and thus a facility manufacturers could build into actual device programs as an aid to rigorous development). Another approach we will investigate is the development of tools to generate precise illustrated standards, showing known (and desirable) interactive properties; for instance, it would be possible to run a user trace in a device simulator to generate a documented sequence of screen shots illustrating the scenario. We will develop advanced tools, like an updated and formalised expect (Libes, 1994), a tool that checks and helps maintain correct documentation by running the systems that can be used with typical medical devices. Such tools would leverage existing design skills in industry to use formal methods implicitly. [To illustrate the value of implicit formal methods: consider a web design tool that offers to find “orphan files” (web pages that a user cannot get to by following links from the home page but that perhaps have important content the designer has forgotten to refer to). Underneath, the tool is actually finding the complement of the transitive closure of the link relation, but a web developer does not need to know that to take advantage of it to design a better web site.] Suitably developed further, such tools would also be extremely valuable for design audits and regulatory approval purposes.

MAUDE, MDR, ISMP and other databases have a diverse collection of incidents. As we use these we will tag and annotate relevant ones in a local repository, but refer through to the originals: thus creating a web mashup to relevant incidents. For example we will specify action sequences (traces) to recreate the incident, and hence allow it to be demonstrated (and hence evaluated) on any of our device simulations using the tools that will be in the repository. This will result in an open resource useful for all medical error research as well as training purposes, including our stakeholder engagement in WP6.

A further investigation will be into the roles of user manuals. Given a well-defined device, we can generate precise user manuals and training resources—but currently we do not know whether such precision would be a help, even if the details affect error rates. Possibly, detailed manuals would be of minimal use to clinicians but useful for designers and technical authors who should (but currently cannot) know exactly what they are explaining. In earlier work (Ladkin and Thimbleby, 1995; 1997) we developed tools that could migrate initial precise (but typically long-winded) manuals that had been generated automatically to polished manuals that were rephrased by technical authors but which retained an audit trail back to their formal basis. Such tools would be very useful in iterative design, since obsolete sections of a manual can be identified automatically; we could also extend them to flag safety and other design requirements. Formative studies will be conducted with different stakeholder groups to investigate the roles of such tools.

**WP5.2: PhD project on an individual-focused method**

The development of a method to support reasoning about human error, particularly slips, will be assigned to a PhD project, drawing on prior work such as that described above, from both the human error and the HCI traditions, but will focus particularly on encapsulating insights from WP3. We anticipate that the method developed will be based on heuristics (e.g., extending and tailoring Leveson’s (1995) safety design rules) rather than modelling cognition (e.g., GOMS (John and Kieras, 1996)), but the exact decisions about the approach to take will depend on findings from WP6.1 as well as those from WP3.

**WP5.3: Situation-focused method**

We anticipate that the method to support situated reasoning will extend and tailor prior work on DiCoT (see Annex 4). DiCoT’s application to the London Ambulance Service control room (Furniss and Blandford, 2006) and an agile software systems development context (Sharp et al., 2006) has proved it to be useful as a tool for organising ethnographic observations from a cognitive systems perspective. It highlights computational influences of the wider socio-technical system through the application of five models with corresponding distributed cognition principles (Blandford and Furniss, 2006). The developments of DiCoT will include: applying the method to the range of settings in which medical devices are used and document case studies and challenges in these areas; second, developing the social and evolutionary models in DiCoT which have not been attended to thus far; third, consolidating and developing the distributed cognition principles that have been extracted from the literature and form part of the method; fourth, developing how the method’s representations are received by designers and non-specialists in this wider research project; and, finally, developing instruction material so it can be more easily adopted by and taught to others. Larger developments include exploring how DiCoT can be used as a basis for identifying markers of resilience, which has shown promise in preliminary post hoc analysis (Back et al., 2008). In this work, markers identify computational properties of the system that foster resilient behaviour, such that the system is able to recognise, react and cope with variance in what is demanded from it. In the behavioural studies referred to previously this was largely described as how individuals compensate to prevent minor failures developing into major failures (see Annex 4). However, behavioural markers at the small team level will include the wider computational socio-technical structure of the system that goes beyond the behaviour of individuals.
WP5: Resources and deliverables

The method development and testing work associated with this work package is currently allocated 90 PDRA months. One PhD student will be associated with this work package (focusing on method development associated with the cognition work package, WP3).

Deliverables from this work package will include:

- Methods and tools for exploring and analysing devices, ultimately useful for design, procurement, testing etc. Such tools would also be extremely valuable for design audits and regulatory approval purposes.
- Scientific papers in international journals, conferences and workshops reporting on findings on the methods being developed and evaluated.
- One PhD thesis.
- Tutorials, including tutorial notes, tools and example analyses that will be made available via the Internet as well as being delivered at conferences or through CHI+MED events, on the methods developed in this work package. The tutorials will be tested and adapted in the light of interactions with developers.
Annex 6 (Work Package 6): Stakeholder engagement

For our basic scientific research on medical device design and use to be transformational, it is essential that we engage in a dialogue with a wide range of stakeholders. This stakeholder engagement work package is not just concerned with disseminating the results of that scientific research. It is a programme of novel research in its own right, both about how to effect transformational change and about the social context around medical technology development and procurement.

We will build stakeholder dialogue starting with a programme of research, using methods from the social sciences, to extend current understanding of those stakeholders, their situations and values, and the barriers to and opportunities for change. This research will directly inform the work of the other work packages, ensuring that appropriate contextual factors are taken into account in prioritising studies (WP2, WP3, WP4) and developing methods and tools (WP5). We will further test this understanding by applying it in the form of different kinds of dialogue and engagement activity. In this way, in addition to gaining deeper understanding of the stakeholders, we will generate research results on the effectiveness of different techniques for developing dialogue for transformational change.

WP6 comprises 6 projects:

- WP6.1: understanding current development practices and the external factors that shape them;
- WP6.2: working with developers, investigating interventions in development practices to improve interaction design;
- WP6.3: understanding current policy and procurement practices and the external factors that shape them;
- WP6.4: working with the full range of stakeholders, investigating how to improve understanding of interaction design (this project has less of a technical focus than WP6.2);
- WP6.5: a PhD project will focus particularly on engagement with students, linking CHI+MED to the ongoing Computer Science for Fun PPE project at QMUL.
- WP6.6: systematic evaluation of the stakeholder engagement work.

Since the programme of work is highly integrated, the demarcations between projects will not be strong; rather, the difference lies in the emphasis, with WP6.1 and WP6.2 having a technical focus, WP6.3 and WP6.4 focusing on other professional stakeholders, WP6.5 focusing on the general public (particularly students) and WP6.6 considering evaluation across all the activities.

WP6: Background

There are many different stakeholders in the design, deployment and use of medical devices. As outlined previously, they include: manufacturers; technical writers; regulators and policy makers; incident investigators; procurement, training and device deployment staff; researchers; the clinicians, patients and carers who operate devices; and the designers and users of the future.

Like other work packages, the work package on stakeholder engagement will proceed through three overlapping phases, all of which involve direct dialogue with stakeholders: understanding the current situation; developing and testing ways to build an effective dialogue with stakeholders; and evaluating the effectiveness of approaches to engagement.

Studies of current development practices in other software industries have employed qualitative techniques such as observation (Robinson et al., 2007) or interviews (Furniss, 2008). While there are published accounts of how medical software development should accommodate human factors (e.g. Gosbee and Richie, 1997), there are no accounts of actual practices that give useful insight into the detailed practices and cultures of interaction design for medical devices. Similarly, while instructions on good procurement practices proliferate, there is minimal information on how interaction design is taken into account in procurement, regulatory on incident investigation practices. Current guidance on procurement makes minimal reference to interaction design; for example, a recent purchasing guide for insulin pumps (CEP, 2008) omits any discussion of interaction design but notes (p.9) that “Individuals with poor cognitive function should not be responsible for self-management of a pump”.

Many different techniques have been developed for engaging with these different stakeholder groups. As Conradi and Dybå (2001) note, simple presentation of documentation is generally ineffective: there is a need to engage in meaningful dialogue with all stakeholders. In some situations, well established academic approaches (e.g., organisation of and participation in joint workshops, and dissemination through paper and internet-based publications) have been adopted and adapted. Web 2.0 presents new possibilities based around social computing paradigms (e.g. Preece and Maloney-Krichmar, 2003). Others have investigated innovative techniques such as theatre (Rice et al., 2007) or digital stories (Murphy, 2007) for engaging with technology users. Within the EPSRC cs4fn Partnership for Public Engagement project for engaging with school children, we have investigated the use of magic shows with each trick linked to computer science (including one used to explain how medical tomography works), computer games with a human error twist, and a written magazine that uses off-beat links to films and music, games and puzzles (Curzon 2007a, Curzon and McOwan, 2008). There are diverse possibilities in terms of ways of building up dialogues with stakeholder groups, but their efficacy is poorly understood. In CHI+MED, we will not only develop and apply a variety of approaches to stakeholder engagement, but also evaluate the approaches that are developed, in collaboration with those stakeholders.

Most work on evaluating public engagement activities (e.g. Rowe and Frewer, 2005; Abelson et al., 2003) focus on public participation in decision making, and identify criteria such as representativeness (i.e. whether the participants in
decision making are representative of the broader population) and efficiency as evaluative measures. From a different perspective, Cummings and Teng (2003) consider the initial factors that contribute to success in knowledge transfer. There are, as yet, no mature evaluative frameworks for assessing stakeholder engagement of the kind proposed in CHI+MED; one of the aims of this programme will be to develop such a framework.

**WP6.1 – WP6.4: The proposed work**

The first stage of research in this work package will be to understand the current stakeholder situation. We will perform empirical studies to develop a rich understanding of why medical devices installed in clinical settings are the way they are: for example, why manufacturers design the interactions with their medical devices as they do (WP6.1), and why procurement staff make the decisions they do (WP6.3). This will also give insight into the kinds of interventions that may be effective in delivering transformation of the overall situation. This initial work will form a sound basis for the later stakeholder engagement work (WP6.2, WP6.4), allowing it to be done from a position of understanding. Studies will be qualitative, including observational studies, interviews, focus groups and survey techniques, as well as analysis of relevant documents. These methods will be similar to those adopted for understanding the situation in WP4. We will use qualitative data analysis techniques such as Grounded theory (Strauss and Corbin, 1998), to develop a rich account of stakeholder practices and values, including their interdependences. This initial focus on understanding their context and the challenges they face will strengthen links with stakeholders early in the project.

In order to better understand what is possible, we will also review the literature on other safety-critical areas, such as aviation, where there has been a step change in the recognition of the role of interaction design in improving safety, and the development of a no-blame culture leading to a richer investigation of causes and subsequent preventative change (e.g. Dekker, 2003).

To change the climate of opinion, and so change the way things are done, we will need a range of techniques for interacting with stakeholders. These will include meetings for discussion (e.g. workshops, policy forums) and for dissemination (e.g. conferences, tutorials), publications (academic, practitioner and public) and web resources. We do not regard stakeholder engagement as one-way dissemination, but as a dialogue. Different techniques will be adopted for different kinds of stakeholder. We will work within a framework of action research (Avison et al., 1999). We will thus use the work as a case study to develop understanding of how to transform opinion and carry that over to transformative action. We will investigate which kinds of engagement work, how and why. Questions will include: how can one provide stakeholders with tools for thinking to help them understand the issues; how can the science best be made accessible; how can the scientists be guided by stakeholders’ understanding and values; and how can the underlying science of understanding human cognition be transformed into product improvement.

We will develop a web repository of resources including the simulations of real devices developed in WP2, case studies of device use in practice (from WP4) and a collection of issues and problems linked directly to the simulations. Videos of scenarios (e.g. run at the Health Information Research Unit for Wales) will illustrate specific issues in context. An area of the web repository will be dedicated to advice for patients and carers. We will actively investigate how such resources can be used effectively.

Specific interventions with manufacturers (WP6.2) will include participation in trade exhibitions, writing articles for trade magazines, and developing and delivering tutorials in collaboration with colleagues in WP5. We will also present work at medical device conferences such as Designing Medical Devices and MedTech UK. We will investigate ways of engaging more directly in development practices, building on findings from WP6.1. We will engage directly with organisations such as the Association of British Healthcare Industries, MediWales, the National Patient Safety Agency and national patient and carer support bodies (Beating Bowel Cancer, Cancerbackup, etc.) via one-to-one meetings to discuss project concerns relevant to the organisations. We will work with researchers on related projects (e.g. MATCH www.match.ac.uk) that are working in complementary areas so that each gains extra leverage from the work of the other. We will also organise policy forums and workshops to bring together both academics and practitioners, encouraging debate around the issues, increased understanding by academics of the practitioner context and increased involvement of practitioners in research. A provisional plan for the themed workshops that are proposed (2 per annum) is:

- Each of the scientific work packages (WP2, WP3, WP4) will co-organise two workshops with WP6 researchers, one focusing on engagement with other researchers and one on engagement with practitioners;
- A doctoral consortium on HCI for medical devices will be organised;
- Two policy forums will be convened; and
- Two workshops will bring together all stakeholder groups around focused themes (to be determined).

The Advisory Group will also have an important role in terms of stakeholder engagement (see Management Annex). By ensuring that the Advisory Group represent the stakeholder groups we will also develop direct channels to engage with and transform the understanding and practices of those groups. Many members of the committee will be in positions to disseminate findings to their communities, as well as representing the views of their communities to CHI+MED.

**WP6.5: PhD project on engaging with students**

We will engage with the next generation of stakeholders in a range of ways: through schools talks and workshops, through activities at science festivals and also through written articles and activities disseminated through the EPSRC
funded cs4fn website (www.cs4fn.org). An example of how intervention may be effective is illustrated by our trial work in schools. A computer science talk on a scenario of developing computer support for a patient with locked-in syndrome was given to a mixed group of sixth form girls intending to be medics or IT professionals as part of a school careers fair. The talk’s aim was to change the audience’s perception of how computer science combines the understanding of devices with the understanding of people. The formative feedback was excellent, indicating that the approach could be a useful part of CHI+MED. A PhD student will explicitly investigate engagement with students and the general public over these issues.

**WP6.6: Evaluating the stakeholder engagement work**

The final strand within this work package (WP6.6) is that of evaluating its success. This will cover both the overall success of the stakeholder engagement work and the individual forms of engagement. Early in the programme, in collaboration with our external evaluator (Meagher), we will develop a comprehensive evaluation framework, which will be reviewed and refined throughout the project. The framework will guide all project participants in gathering data for evaluating the effectiveness of all project activities and will provide a common basis for review of impacts across the whole project. Evaluation will be done throughout the project and be used to assess the effectiveness of the whole programme and of the individual work packages and to deliver results to others interested in understanding how to achieve transformational impact. Both quantitative and qualitative data will be collected continuously by the team as part of their normal activities and interactions with stakeholders, along with periodic qualitative inputs from, for example, focus groups, interviews and surveys. This will allow for more intensive final evaluation during which the external evaluator will help capture successes, problems, impact indicators and lessons learned so that others can benefit from our experience.

Gathering hard evidence of broad social impact of research is challenging. Any such impact of the scale we propose will necessarily be long term. Also, as with any intervention of this type, the longitudinal assessment of success will be confounded by a range of factors, such as work on other projects, legislative change as a result of other causes, or major incidents leading to loss of life that change public opinion. Our evaluation will therefore be based around channels of impact for each stakeholder group. For each channel we will identify a series of proxy indicators that, while not in themselves demonstrating the ultimate impact of our vision, will instead show whether the situation is moving towards our end vision (Meagher et al, 2008). An example of a proxy indicator is the depth of relationships between researchers, manufacturers and procurement staff. Both qualitative and quantitative measures will be devised around such indicators. For example, impact channels for procurement will include: the extent to which procurement staff in our partner hospitals are willing to engage in studies; participation in joint workshops where different stakeholders engage in issues; and changes in hospital purchasing practices. Specific metrics for overall evaluation will include qualitative ones such as: changes in the language of official reports about device procurement, changes in the focus of incident reports to include interaction design where relevant, and an improvement in the interaction designs of medical devices. Quantitative metrics will include numbers of visits to the project web repository; numbers of downloads of materials such as simulations and papers and numbers taking part in events we host or organise.

Each kind of intervention will be evaluated for success. Specific criteria and evaluation methods will be used for each. For example, workshops and talks will primarily be evaluated via feedback forms, and web-based resources by a combination of feedback forms and web statistics. In each case the methods and specific questions will be based around the common evaluation framework. Evaluation of the scientific programme will be based on both reviewer feedback (qualitative) and bibliometrics (quantitative), drawing on experience of the final reviews of Interdisciplinary Research Centres such as EQUATOR.

**WP6: Resources and deliverables**

This work package in planned to involve 4 PDRAs across all three sites for a total of 180 person-months, and one PhD student at QMUL. Laura Meagher will be employed as a consultant to provide direction on both the conduct and the evaluation of the stakeholder engagement work.

Deliverables from this work package will include:

- Papers in leading international conferences and journals, and also presented at workshops. Early papers will present analyses of the stakeholder situation (e.g. what current development practices are, particularly pertaining to interaction design and evaluation). Later papers will present and critically evaluate the approaches being tested to stakeholder engagement.
- One PhD thesis.
- The organisation and delivery of themed workshops that bring together different groups of stakeholders, as described above. Workshop reports will disseminate the outcomes from workshops more widely.
- Reports on the evaluations of representations (e.g. digital stories, short videos, textual narratives and tutorials) from other work packages.
- Materials aimed at more generalist audiences (articles, schools talks, presentations at science fairs, etc.).
- Policy briefings to inform future policy on HCI for medical devices.
- A framework for integrating science and stakeholder engagement, presented in a form that can be taken up, tested and adapted by others involved in public engagement activities.
- A website for dissemination of programme resources.

**Other Attachment**

CHI+MED Technical Annexes
Annex 7: References


