chapter ten

User evaluation in pervasive healthcare

Tim Adlam, Roger Orpwood, and Teresa Dunn
Bath Institute of Medical Engineering, United Kingdom

Contents

10.1 Introduction

Pervasive technologies are moving into health and social care. This move is supporting the transfer of care from structured institutions to the community. The increased ability to provide patients with effective and complex homecare is delaying and reducing institutional admissions and allowing earlier discharge of patients.

This dispersal of care from the institution to the community places new demands on care staff. Patients are monitored in their own homes using simple monitoring devices and the resulting data must be collected and analyzed. Risk reduction technologies[1] are being installed in homes, needing assessment and installation. Information systems make medical records available when and where they are needed.[2] Other devices and systems enhance communication between vulnerable people, their friends, and their families.[3]

Within healthcare institutions, pervasive technology is making care providers more efficient through the fast distribution of medical information. Pervasive technologies are also being used for tracking surgical instruments and other artifacts within operating theaters and around hospitals.[4]

Designing for healthcare presents a greater challenge than designing for industry because many of the intended users of a healthcare device are people with physical, neurological, or cognitive disabilities. Because people are complex and unique, it is not possible for a designer to simply imagine what it is like to be a user to be able to design appropriately for that user. This can be especially true when designing for impaired populations. The designer must obtain a thorough understanding of the users, their disabilities, their environments, and their problems. The greatest challenge to the designer is not solving the problem but understanding the problem. User evaluation is an essential tool for obtaining understanding.

Technology developed for use by people has an interface that is composed of the parts of the device or software that the user interacts with, and enables the user to communicate with the device. This user interface should be designed to ensure that it is accessible, usable, and useful to its intended users.

> **Accessibility**: n. The ease with which something is approached or entered.

It is essential that the user interface be accessed by its users. For example, a text-only interface is inaccessible to blind users.

> **Usability**: n. The effectiveness, efficiency, and satisfaction with which users can achieve tasks using the product.

High usability means a system is (1) easy to learn and remember; (2) efficient, visually pleasing, and fun to use; and (3) quick to recover from errors.[5]

> **Usefulness:** n. The degree to which a device or system enables a user to perform a previously difficult or impossible task that the user desires to perform.

Useful devices are used by users. We all have tools that are useful and are well designed for their purpose.

10.2 The roles of user evaluation in pervasive healthcare

User evaluation is used in several distinct contexts. These are:

1. As a tool for determining users' preferences and capabilities within the design process
2. As a means of assessing usability within a formal regulatory evaluation
3. As part of an investigation into an adverse incident due to human error

Different techniques and methods are employed in each. In particular, the design context requires several different methods of user evaluation depending on the depth and type of information required.

In order to discover how an interface should be designed, it is necessary to consult potential users and conduct evaluations to determine and validate the design. Evaluations within a design context are diverse in their objectives and in the methods used, ranging from an informal discussion to the use of formal outcome measures, questionnaires, and interviews.

Products are often tested and evaluated by regulatory bodies such as the Food and Drug Administration (FDA) in the United States and the NHS Purchasing and Supply Agency (NHS PASA) in the United Kingdom. The NHS PASA runs the Medical Device Evaluation Service. Devices are evaluated and tested to ensure that they are safe, comply with standards, and work as advertised by the manufacturer. Additionally, the Medicines and Healthcare Regulatory Agency (MHRA) investigates adverse incidents, and the National Institute for Clinical Excellence (NICE) evaluates new clinical treatments for efficacy. These evaluations where technologies and devices are assessed formally for their clinical effectiveness and safety are not the primary subject of this chapter. Therefore they are considered briefly at the end of the chapter.

Poor interface design can lead to accidents when controls and feedback are misunderstood by the user. When such incidents occur, an investigation is required to determine the cause of the incident. User evaluation may form a part of such an investigation. In these cases (in the U.K.) evaluation is carried out by the MHRA or is subcontracted by the MHRA to other organizations.

This chapter is primarily concerned with user evaluation as part of an iterative design process, rather than its use for clinical evaluation, technology assessment, or postmarket monitoring.

10.3 Evaluation in user-centered engineering design

A thorough evaluation is not the sole objective of user evaluation in design. User evaluation is a very powerful tool for discovering required elements of a design by incorporating input from intended users. In user-centered design, the design process is continually modified by the refinement of a specification through successive user evaluations. As such, conventional engineering design techniques are not applicable in this context.

10.3.1 Design methods

Engineering design presents special challenges and requirements when compared to other evaluation contexts. Evaluation enables the discovery of the nature of the problem being solved, followed by the evolution of a solution in response to increased understanding of the problem. The purpose of evaluation in this context is *not* for the gathering of knowledge for its own

sake, but rather for finding out the essential nature of the problem the user experiences so that it can be mitigated effectively by an appropriately designed device.

Engineering design is a complex, creative process that explores an identified problem and then develops a technological solution. Effective design depends on the definition of the problem to be solved: if the problem is not understood, the designer will solve the wrong problem or solve the correct problem inadequately. Discovery of the true nature of the problem is essential.

10.3.2 The conventional engineering design method

Conventional engineering design has three stages:

Stage 1—Development of a specification. The specification is used to clearly define and quantify the problem that needs to be solved in a way that the designers can understand. It also describes and quantifies the functions that the device must be able to perform. It is obtained through investigating the problem prior to design and must contain all the information the designer needs to make design decisions once the design process has begun. Before beginning design, the specification is "frozen" so that the designer has a fixed target to achieve.

Stage 2—Design of the device. This is a creative process of problem solving and is guided by the specification. Several concepts will be created and evaluated against the specification for their ability to solve the problem. A single concept will be selected for further development. The engineer will develop the selected concept into a detailed design, incorporating features that fulfill the original specification. Frequent reference will be made to the specification as design progresses, ensuring that the developing solution is compliant with the frozen specification.

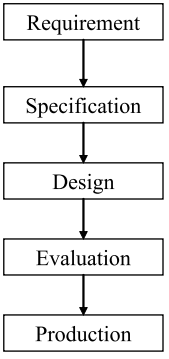

Figure 10.1 The conventional engineering design process.

Stage 3—Design completion. After the first prototype has been built, an evaluation of the prototype is required to confirm that the device solves the problem and meets the functional requirements outlined in the specification. The performance of the device will be compared to the specification and changes made to the design where necessary. These changes are likely to be implemented in a production version or, in the case of a custom-built one-off machine, in the prototype.

In this context, design and evaluation are carried out against a predefined specification that does not change as it completely describes the requirement. This design process is not appropriate for the design of healthcare devices. The design of a medical device presents special challenges because of the complexity of the human mind and body.

10.3.3 User-centered design

A human being is a complex adaptive system that is not easily defined or measured. Our materials are nonlinear composites and our systems are complex and interdependent. Each person is an individual with his or her own unique set of behaviors, reactions, and needs. For these reasons it is not possible to use the conventional engineering design method and define a complete specification for a device that interacts with a person, prior to device design and evaluation.

Designing a healthcare device is a large task that requires an iterative process involving engineers, users, and clinical professionals. User-centered design aims to design a device that identifies and meets users' needs as fully as possible.

Design begins by discovering the users' requirement and the compilation of a functional specification, which may be revised later. Users sometimes ask for a particular type of device that they think will solve a problem, when in fact careful investigation will show that the real problem may not be solved by such a device at all. The "problem" is the users' perception of the

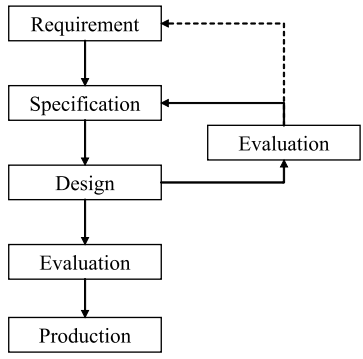

Figure 10.2 The user-centered design process.

need that is to be addressed by a device. The "requirement" is the true nature of the users' need, and it is not always immediately obvious. The process of requirement discovery reveals the true nature of the problem to the designer and sometimes to the user as well. The user's expression of the problem may not be detailed and may be expressed in inexact terms by a nontechnical user. The expressed problem is not usually sufficient to begin design. Further investigation is required to identify which needs the user is seeking to have met by the proposed device. A preliminary evaluation by users may be used at this stage and is often conducted with a simple prototype representing an aspect of the function of the proposed device. This method of requirement discovery may provide information that could not easily be obtained through questioning.[6]

Having obtained a specification that describes the users' need, work on the design can begin. At this stage the designer is aware that she or he does not fully understand the users' needs. Therefore the first prototype will be designed quickly and with no unnecessary complexity as it is likely to be heavily modified during subsequent iterations. Evaluation of the first prototype with potential users will yield large amounts of information. This can be a very exciting stage of the design process as this may be the first time users will be able to independently perform a difficult task or believe there is a technical solution to their problem.

Information from users will take many forms: some structured and anticipated, some unstructured and unexpected. Examples of structured information include results from questionnaires and interviews, formal measurements of the ability to perform a task, and physiological measurements. Unstructured information includes comments made by users after an evaluation and observations of a user attempting a task without the prototype or using the prototype in a way unintended by the designer. Structured information provides the details of the specification such as dimensions for components and operating procedures. It allows detailed design to proceed. Unstructured information provides insight into the problem. It suggests alternative solutions where no existing solution seems appropriate. It can help develop a stronger relationship with a user, leading to more candid user feedback in the future.

The design of the second prototype incorporates what has been learned from the first evaluation. It is likely that the first prototype will be discarded as many shortcomings will now have become obvious. It is for this reason that the designer should not put more effort into the first prototype than necessary to create a functional and safe design. The second evaluation will yield further details about the problem and the effectiveness of the second prototype. The users will have become more familiar with the evaluation process and, having a stronger relationship with the evaluating team, are more likely to be frank and honest about their opinions on the device when interviewed or given questionnaires.

Third and subsequent iterations will yield greater detail and more precise assessments of the usability and accessibility of the device. At some

point the designer must decide when the benefits of new information and insights from user evaluations no longer outweigh the time and material costs of conducting the evaluations. At this point the designer will freeze the design and proceed to production development. However, before committing the design to production, it is advisable to carry out a final preproduction evaluation.

10.3.4 Implementing user evaluation in design

Conducting a user evaluation within healthcare involves many diversely skilled people working together as a team. To be successful, this team requires strong technical and clinical skills, interpersonal skills, knowledge of statistical methods, knowledge of organizational structures, and the ability to adapt and react quickly to changing situations.

The following description and analysis of the process of pervasive healthcare device and system design has been broken down into these eleven sections:

1. Forming a team
2. Recruiting and selecting evaluators
3. An initial user survey
4. Validated outcome measures
5. Intellectual property protection
6. The first evaluation
7. Installing systems in users' homes
8. Supporting users and devices
9. Interim evaluations
10. The final evaluation
11. Preproduction evaluation

These sections generally follow in the order in which they would occur during the design and evaluation process, although as will be seen this order should not be taken as a completely rigid structure.

10.4 Forming a team

Evaluation is nearly always a team or collaborative activity because of the diverse skill and knowledge set required by the participants. Typical members of a team for a pervasive healthcare application might include software, electronics, and mechanical engineers; an occupational therapist, doctor, nurse, or other clinician depending on the application; a researcher; and, of course, the user evaluators themselves. It is important that the team members are able to work together and that communications between the team members are unimpeded.

Although there are many variations to the composition of a team, a common one is as follows:

- **The engineers** are responsible for the design of the device or system. They must be able to integrate incoming data from the evaluations into the design process. It is also likely that they will provide technical support to the users during the evaluation, which requires good interpersonal skills.
- **The clinicians** provide expertise relating to the disability or health problem being addressed by the proposed device. They are often the originators of a project. Initially they provide the interface between the users and the technical members of the team. Good interpersonal and observational skills are essential, as they will work closely with both the users and the engineers. Depending on the type of device being designed, they may be physicians, surgeons, nurses, or therapists.
- **The researcher** often coordinates the evaluations and acts as a personal point of contact for the users. This is particularly helpful in a larger project with many users; however, this role may be fulfilled by the clinicians and engineers in smaller projects or where access to a patient is limited to frontline care or clinical staff because of the needs of the patient. The researcher may also be responsible for statistical analysis of evaluation results. The researcher will do much of the face-to-face work with the users by visiting and telephoning them. A social care or psychological background can be valuable because it provides the researcher with the excellent interpersonal skills needed to make the most of contact with users.

Having formed the team for the design and evaluation work, the project can begin. One of the first tasks is the recruitment of evaluators.

10.5 Recruiting and selecting evaluators

Evaluators are the members of the team who will try out iterations of prototypes and provide feedback to the designer on specific design aspects as well as on the overall usefulness and user satisfaction of the device. Although these are the members of the team that are the most difficult to recruit and retain, they are also the most important and will provide information about the usability, accessibility, and usefulness of a prototype. Recruitment is not always a straightforward process and the selection of evaluators should not always be at random from the population of potential users. Although random selection will recruit a more representative population of evaluators, not all evaluators will be suitable to work in the iterative healthcare design process. For example, where initial evaluations are carried out using crude first-stage prototypes, some technical knowledge or background can enable an evaluator to ignore the minimal attention given to cosmesis and concentrate on analysis of the functionality of the prototype.

One method to reduce recruitment difficulties is to form a relationship with a preexisting group that represents the population of potential device

users. For example, if developing systems for people with stroke or dementia, it would be advantageous to work with the local group of the Stroke Association or the Alzheimer's Society. If they feel that the work the designer is doing is to the advantage of their members, such associations can forward questionnaires to their members and print letters and articles in their newsletters. The circulation of such newsletters may not be large, but they are very well targeted at the potential user population. Because of the great benefits to be had, a collaborative relationship with a user association should be formed as early as possible to facilitate initial survey work and begin the process of informing potential user evaluators about the proposed work. As development progresses, the association can provide updates on progress to its members, provide a forum for comment, and be used to recruit additional evaluators if necessary. In the authors' experience, once such a working relationship has been established, recruitment of evaluators is not difficult.

Alternatively, evaluators can be recruited through existing medical or therapeutic channels. Medical and therapeutic practitioners with whom the team already has contact can be asked to approach their patients or clients to ask if they would be willing to take part in an evaluation. Obviously where practitioners have longstanding relationships with their clients, recruitment is easier as the practitioners are trusted to a greater extent. An advantage to this recruitment method is that once the group is established, communication to the group is not difficult. However, expanding the size of the group can be difficult as there may not be many people within a clinician's patient population that are appropriate for any given evaluation.

A final means of recruitment is to use local media such as newspapers and radio. This type of appeal for user evaluators is much broader and reaches many more people than the previous two examples. A disadvantage is that the recruitment request comes from a stranger with whom the potential user evaluator has not built any trust, resulting in a much lower response rate.

In the design context there are several evaluatory roles that users carry out. It is important to select appropriate people for those roles. For example, it is not possible to evaluate the design of every prototype iteration with a full-scale randomized controlled trial. When the engineers are designing a system or device, it can be useful to have a small group of well-informed users work closely with the engineers and evaluate the ability of a device to perform its given function as the prototype design progresses. This group may have a technical background and are likely to be people that can see past crude prototypical casings and housings to the device inside, evaluating specific features when needed. Such people are not easy to find, but they can assist the designer immensely by providing rapid answers to questions of functionality and usability.

When evaluating the interface of a device, it may be more appropriate to evaluate it with naïve users who have little experience of information technology interface conventions. Although technically naïve users are easier

to find, they are harder to recruit to the project and need greater reassurance that their contributions will be useful and valued.

The authors have found that potential users like to feel they are involved in the design work, not just occasionally testing something they don't really understand for a person they don't really know. For this reason, evaluators are invited to join the evaluation team and be a part of a research project rather than be experimental subjects. Evaluators are included and involved in the process of research rather than just being instruments in the hands of the researcher. This involvement is not illusory. Evaluators make an essential and valuable contribution to the design of the device or system.

10.5.1 Ethics, informed consent, and people with cognitive disabilities

When recruiting people for an evaluation program, remember at all times that they are not employees and have *no obligations* to the evaluation team. It must be made clear to the evaluators that they are free to leave the evaluation at any time without question or justification; that they are not obliged to use the devices given to them; and that their opinions and personal information will remain confidential within the project.

It is important to obtain *informed consent* for an evaluation from users. (Indeed in some contexts it is mandatory. In the U.K., for example, if an evaluation is carried out using NHS employees or patients, the evaluation must achieve formal ethics approval from an NHS ethics committee.) The consent of users who are not fully informed about the program they are agreeing to participate in is not meaningful or sufficient for involvement to proceed. Obtaining informed consent requires a clear explanation of the evaluation program using terms users can understand, careful listening to users' questions about the program, truthful and complete answers to those questions, and the distribution of clear and comprehensive materials describing the evaluation program that potential evaluators can read and understand at their leisure. The evaluation team must be sure that the users understand what they are consenting to. Asking a potential user to describe her perception of the evaluation program and her role in it can reveal misconceptions that need to be gently corrected. It is helpful, and often necessary, to ask the potential user to sign a written consent form. This suggests to the user that the decision about consent is a serious matter worthy of proper consideration. The form can also set out the responsibilities and expectations of the evaluation team as well as the rights of the user evaluator within the context of the evaluation.

Special consideration should be given to recruiting and working with users with cognitive disabilities and memory impairments such as Alzheimer's disease, stroke impairments, or traumatic brain injury. Unlike any other disabled user group, users with a cognitive disability may have a compromised ability to grant or withhold informed consent. Where users have more severe cognitive disabilities, it is not possible for them to give

informed consent. When working with people with such disabilities, informed consent must be obtained from a person who is suitably qualified and empowered to make that decision. For example, a person with dementia may have granted enduring power of attorney to a relative or close friend when he was still able to make such decisions. This person will be somebody that the person with dementia trusts and who is considered to be acting in the best interests of the person with dementia.

Obviously, degrees of cognitive disability are not measurable in finite terms but span a continuum from none to a level that completely prevents a person from understanding the evaluation program. Whatever the degree of disability, the evaluation team should endeavor to enable users to understand as much as they are able to about the evaluation program and to allow users to give or withhold their own consent based on that understanding. Depending on the degree of the disability, a greater or lesser weight should be placed on the opinion of close carers when judging whether informed consent has been granted. Often when making these decisions it is important to consult professional and unpaid caregivers that are close to the users. People with severe brain injury or dementia are still people and deserve the dignity of being able to determine their own futures as much as they are able. It is in situations like these that the interpersonal skills often referred to in this chapter become absolutely essential for working with and understanding disabled people and their carers.

The user evaluators are an integral part of the project team and should be treated as such. When technical support is requested by an evaluator it should be provided as quickly as possible. Cultural sensitivities should be respected and taken into account. Things such as small cultural differences can have a big impact on the effectiveness of an evaluation. The ENABLE[6,7] evaluation described as a case study later in this chapter contained such an incident, where a slight difference in the meaning of the word "evening" resulted in a new version of software for a device being hastily developed.

The key to a successful evaluation is to make evaluators feel like the valuable members of the team that they are. It is vital to respect them and listen to them, giving them time and attention when needed.

10.6 An initial user survey

The initial user survey is a unique part of the user evaluation process. This is often the first contact the evaluators have with the design team and it is the step where the first bits of information about a new design are gathered. It is here at the beginning that first impressions are made and the building of relationships begins. If the design team intends to use formal outcome measures as part of their evaluation, then the decision as to which measures are to be used should be made here at the beginning, allowing appropriate baseline measures to be made as the study progresses.

The initial survey is designed to discover as much as possible about the users' perceptions of the identified problem and to narrow the focus of the design effort to specific clearly identified problems. The initial survey will not necessarily identify the true nature of the problem, just the users' perceptions of it. Commonly used tools for initial surveys are questionnaires and interviews. The design of questionnaires is covered in detail in many other books and papers and therefore will not be covered in detail here.

It is likely that the research and design team will want to use a customized initial user questionnaire to question a large group of users about their perceptions of the problem to be addressed and their experiences of living with this problem. The researchers will try to find out not only the details about what the problem is and how it manifests itself, but also its physical and relational context. The following list identifies information that can be obtained from an initial survey that will be useful to the design team at the beginning and throughout the project:

1. A description of the problem by the users
2. Tools and methods that the users have tried and whether these worked or not
3. People who help the users solve the problem
4. The type of solution the users expect the designers to produce
5. What the users expect the solution to enable them to do that they could not do before

This is not an exhaustive list. The project team should think carefully about what information they need to make a well-informed start on the project and about what baseline information they need when analyzing results later on.

An initial survey questionnaire should use short, neutral, and unambiguous questions that are easy to understand and do not presuppose an expected answer. It should also use formal outcome measures, where the questionnaire is structured, validated, and unchanged.

Numbered rating scales, known as Likert scales, can be useful in questionnaires, enabling the researcher to quantify users' opinions and feelings about various aspects of a problem. Respondents are asked to indicate a number that represents their feelings about the question asked. For example, a question about the users' existing solution to a problem might be presented as in Figure 10.3. In this example, the 2 circled by the user indicates that this user finds his existing solution is only marginally effective at solving his problem.

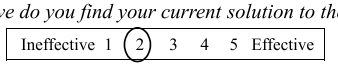

Figure 10.3 An example of a numbered rating scale.

Interviews take much longer to carry out than questionnaires and are usually used with a much smaller group of people. Interviewers must be patient and able to listen. The authors have found that the most effective interviews are semi-structured, with a list of topics to be covered in a conversational style. As with questionnaires, interviews that will form part of a formal outcome measure should be carried out as prescribed in the outcome measure method, to try and ensure that comparable data is collected from each queried user and that the data collected are valid for the outcome measure. Keeping the interview informal and mostly unstructured allows interviewees to relax and feel at ease with the interviewer. They are more likely to be honest about the problem in these circumstances, rather than giving the answer they think the interviewer wants. The interviewer must make it clear to the interviewee that the interview is not an examination or a test. The interviewees' opinions and feelings are sought and, as such, there are no correct answers.

Initial surveys are an important part of an evaluation. This starting point yields details for the design and establishes a way of measuring progress toward the final solution objective.

10.7 Validated outcome measures

This is an appropriate place to consider validated outcome measures. These are carefully designed and validated instruments that measure, for example, changes in quality of life, activity limitation, psychosocial impact, and so on. In a user-centered design context, especially in a healthcare context, they are one of the few tools available that provide a validated, quantified measure of whether a system or device is effective at solving the problem it was designed and built for. Validated outcome measures are not usually appropriate for use *during* the design process. They can require many hours of work to carry out and do not provide the detailed guidance required by a design team making decisions about specific aspects of a design. A validated outcome measure is more of an overall measure of the device. As such, it may assess whether a system has improved the quality of life of a user or whether users are able to perform a greater range of daily tasks more effectively with the device. However, validated outcome measures are too unwieldy and take too long to apply for them to be employed when making specific design choices such as deciding whether the symbology on a user interface is appropriate or whether a particular control button is in the right place.

There are many different measures available. Not all measures are appropriate for every project, so they should be carefully selected to fit the context in which they are to be used. Some can provide an absolute measure of a parameter, allowing comparison of outcomes between different projects or aspects of a project. Others measure only changes and do not give an absolute measure of a parameter. The designers of outcome measures

are usually willing to advise on the suitability of a particular measure for a particular purpose.

Two examples of validated outcome measures that may be appropriate for use in a pervasive healthcare technology design context are given below.

10.7.1 The Quebec User Evaluation of Satisfaction with Assistive Technology (QUEST 2.0)

QUEST 2.0 is a measure of users' satisfaction with a device, *not* of the performance or effectiveness of a device.[8] It provides a validated absolute measure of user satisfaction with an assistive technology device. QUEST 2.0 measures user satisfaction with the following features of a device: dimensions, weight, durability, comfort, adjustment, safety, simplicity of use, and effectiveness. It also measures user satisfaction with the following service features: service delivery, repairs and servicing, professional services, and follow-up services. The instrument produces a single score for a device and its accompanying service that is between one (dissatisfied) and five (very satisfied). QUEST 2.0 was designed for use in a face-to-face interview but it can also be used in a postal survey.

10.7.2 Psychological Impact of Assistive Devices Scale (PIADS)

PIADS, an outcome measurement instrument, measures competence, adaptability, and self-esteem.[9] It is self-reporting and is designed for adults and children above the age of ten. It is quick and can be completed by telephone. PIADS is essentially a quality-of-life measure that provides a standardized assessment of the psychosocial impact of a device. It can be used across a broad range of different devices and contexts and has some of the following advantages:

- It predicts device retention and abandonment.
- It can be used to assess stigma of use.
- It allows psychosocial impact to be distinguished from health and condition issues.
- It gives good agreement between self-reporters and caregivers.
- There is no charge for its use.

PIADS has some disadvantages:

- It can only be used on one device at a time.
- It measures abstract concepts, not the practical aspects of a device.
- It requires proxy reporting for children and people with learning and cognitive disabilities.
- It is influenced by user expectations.

There is no single outcome measure that is appropriate to measure all aspects of the impact of a device on a person's life. Measurement instruments appropriate to the required measurements must be chosen in each case.

Factors influencing this decision are the cost of the instrument materials, ease of application, and the appropriateness of the instrument to the population to be measured.

Validated outcome measures are powerful tools for assessing the effectiveness and impact of an assistive device. They enable comparison of effectiveness between different devices in an integrated system and quantify the impact of devices on key issues such as quality of life and the ability to perform tasks efficiently. If they are going to be used in a project, the decision about which measures to use should be made in the early stages of detailed project planning. Some outcome measures require an initial baseline measurement to be made, which must be carried out before the evaluator has sight of a proposed solution that may influence the initial perception of the problem.

10.8 Intellectual property protection

Once design begins, intellectual property will be generated and will need to be protected if the completed design is to be licensed to a manufacturer. If prototypes are evaluated in the community, they are disclosed, compromising any subsequent patent application. There is a conflict here. A design cannot be fully protected until it is finalized, but it cannot be finalized without an evaluation. Thus some evaluation will always have to preclude full protection. A decision must be made about when to protect and what to protect. It may be possible to protect the concept of the device at the beginning of the evaluation process and then protect the detailed design of the concept progressively as features are tested and confirmed. Internal device technology and concealed user interface features may be protected with a nondisclosure agreement between the design team and the evaluators until the design is finalized.

The degree of protection desired will vary from project to project and from team to team. Applying intellectual property protection to a device strengthens its value to a manufacturer, but it can make the process of doing the evaluation and building trust more difficult. The balance between these two conflicting desires must be found for each project and each team.

10.9 The first evaluation

After carrying out the initial survey, accurately identifying the scope and objectives of the project, and designing and building the first prototype, it is time to set up the first user evaluation. This evaluation will probably be carried out by a small group of users who have been identified through the initial survey questionnaire. The first evaluation is the most important. It will yield more data more quickly and will have a greater influence on the design of the device or system than any of the other evaluations. It is also critical relationally: if the users' trust and confidence in the project team is

compromised at this stage, further evaluations will yield compromised data or will not happen at all.

10.9.1 Beginning the first evaluation

Much is learned during the first evaluation by the users and by the researcher, clinician, or engineer conducting it. When the first prototype is delivered the user may feel many different ways: disappointment at its crude appearance, anticipation and excitement at the prospect of the problem being at least partially solved, and nervousness about testing a unique and perhaps expensive piece of equipment. The beginning of the first evaluation is a good time for the designer to accompany the researcher or clinician and meet the users, see their reactions, and hear their comments firsthand. Having the designer present also enables him to explain the use of the device accurately and answer any questions that might arise. The user is able to meet the person who is designing the device or system, strengthening the feeling of being part of the team. Some instant feedback will be given to the designer about the appearance and feel of the device. On some occasions the author has been to a user's house to deliver a device and then returned to his office with it on the same day: as soon as the user tried the device for the first time a serious deficiency in the design became apparent that had not been anticipated before the evaluation. The first evaluation was over within five minutes.

Often at first evaluations, there are family members and carers present. They sometimes try to speak and act on behalf of a disabled user. The team should try to interact directly with the user as much as possible. Although questions and explanations should be directed first to the user, it is necessary and appropriate to make sure carers and family members also understand the evaluation and have their own questions answered. After all, they are often supporting the user who is trying out a new and imperfect device.

By the end of the first meeting, the user has been further integrated into the team. She has a new device to try out and a new role to learn. The designer has met one of the people he is designing for and has firsthand information about the problem being addressed. At the end of the meeting, the project team should have made sure that the user has:

- The prototype device
- Full written and illustrated instructions on the safe and correct use of the device
- A sheet of contact numbers and addresses for the designer, technical support provider (if different), and the researcher

When handing over a prototype to a user, one of the authors, who is an engineer and provides technical support, also leaves his work, home, and mobile phone numbers with user evaluators. He makes sure that the user knows he is able to contact the author at any time. If the device is installed in a fixed place, a contact sheet is placed near or on the device.

10.9.2 Concluding the first evaluation

When the researcher and possibly the engineer return to the evaluator to collect the results of the evaluation, the evaluator will want to tell the researcher or engineer about his experiences with the prototype. In these circumstances evaluators often want to say what they think the researcher and engineer want to hear and are reluctant to criticize the device. The researcher and engineer should encourage the evaluators to be critical and candid about their experiences and opinions, explaining that they need the evaluator to be frank so that they can ensure that the design of the device is appropriate to the evaluator's needs.

First evaluations are an exciting and memorable part of the project. It is on occasions like these, where a user gains hope that their problem will be solved, that working in this field becomes so rewarding.

10.10 Installing systems in users' homes

Many pervasive healthcare devices will be handheld or at least portable, needing no infrastructure to be installed in the user's environment. Some systems, however, need more time-consuming installation with the consequential disruption of the user's household and life.

10.10.1 Preinstallation survey

The installation of complex systems needs careful planning and sometimes an additional visit to the user's house to do a preinstallation survey of the installation site. The survey is used to:

- Check the availability of electricity and other required utilities.
- Identify and check the dimensions of the installation site.
- Check the location and compatibility of any other appliances that the system must work with such as the telephone or television.

After consent is given, the preinstallation survey can be done by the researcher during the evaluation setup visit. The most appropriate person to do the survey is the designer, who fully understands what is needed for a successful installation.

10.10.2 Installation guidelines

Some of the points made in this section may seem obvious; however, it is easy to allow standards of thorough preparation and conscientious working to slip when performing the same installation task many times. Maintaining high standards can save time and frustration, while instilling the user with a sense of confidence in a professional and competent research team.

Installation should take place at a time that is convenient for the evaluator, which may not be during normal office hours. The installation time

should be arranged well in advance, making sure that any additional team members needed during the installation are available. The evaluator may also wish to have a primary carer present.

When preparing for the installation, the installer should prepare a checklist of items needed for the installation. This list can also be used as an inventory when leaving to make sure that nothing is unintentionally left behind. Before leaving for the installation, the installer should go through the checklist and make sure that all the necessary maps, contact details, tools, materials, device components, passes, and identification and user information are present. If any of these items is missing, the installation may have to be aborted.

The authors have found the following guidelines to be helpful when installing in users' homes:

- On arrival at the installation site, the installer should check that consent is still granted before beginning work.
- Work should be carried out in a way that minimizes the amount of damage caused to the evaluator's house or apartment. It is the responsibility of the project team to rectify any damage caused during the installation.
- Questions about the installation should be answered truthfully and conscientiously.
- If, at any time during installation, user consent for the installation is withdrawn, the installer should remove what has been installed and leave the site in its original condition.
- Social engagement and conversation with the evaluator *during the evaluation* engenders trust and strengthens the relationship between the evaluator and the project team.

These are guidelines, not rigid rules. The most important thing is for the team to act with integrity and diligence and to treat all evaluators with dignity and respect at all times. It is easy at the end of a long and difficult day to become irritated by a person with dementia who is asking for the tenth time who one is, why one is there, and what this thing is for. Although the irritation cannot be prevented, speaking and acting irritably has no place in this context.

10.10.3 Installing systems in the homes of people with cognitive disabilities

When working with people with cognitive disabilities, special consideration should be given to the needs of the user evaluator. Some special measures may need to be taken to ensure that the installation is completed with a minimum of disruption to the user and the installers.

In addition to the points made above, the following guidelines[1] are presented for evaluation teams working with people with cognitive disabilities:

1. Always send two people to an installation. Sending two people allows one of the installers to leave the building while the other remains to prevent a loss of access to the site and work in progress. This can occur if the evaluator forgets who the installers are and decides to deny access to her home when the absent installer returns.
2. Be patient. People with dementia may ask the same question many times. This may be because they do not remember asking it previously or they cannot remember the answer given. Installers should answer each time as though it is the first time they have been asked. During evaluations in the U.K. and Europe, the authors found that evaluators with moderate dementia did sometimes remember the purpose of an installation after having it explained clearly to them six or seven times during the installation process.
3. Listen to the evaluator. As well as asking the same questions many times over, evaluators may tell the same stories many times over. Such life stories can provide insight later in the evaluation into why the evaluator is reacting to a situation in a particular way.
4. Be sensitive to the evaluator's state of mind. Try to observe cues that may indicate the evaluator's state of mind such as repeated actions, facial expressions, tone of voice, persistent questioning about a particular aspect of the installation, and so on. If the installers are unsure if the evaluator is distressed, they should ask the evaluator first, then a carer, if present also. If the evaluator is becoming distressed, stop the installation in time to allow the site to be returned to a fully functional state before leaving.

The effective evaluation of complex systems for people with cognitive disabilities by people with cognitive disabilities is possible and valuable; however, these evaluations are not as straightforward as evaluations by people with no such disabilities. Care and consideration of their needs must be taken into account at all stages of the project over and above what would normally be applied during an evaluation.

10.11 Supporting users and devices

Effective and rapid support of evaluators is important because, inevitably, prototypical devices are unreliable. This "feature" of prototypes must be clearly explained to evaluators before the evaluation. The evaluators may become dependent on a device, so rectifying its loss due to unreliability carries a high priority when allocating resources. Engineers providing technical support should be able to work with evaluators sensitively and efficiently. For a particularly troublesome prototype, they may be the evaluator's most frequent point of contact with the project team and as such should be prepared to collect and distribute any findings that are relevant to the evaluation.

Where possible, systems should self-report faults. If a system has even a low level of intelligence, remote reporting of faults and failures is easy to achieve with a mobile phone text-messaging module. Such modules are readily available from major mobile phone manufacturers, are inexpensive, and are easy to use and control with a simple USB or RS232 interface. They do not use the user's own telephone landline, providing a protected independent link to rapid and informed technical support. It should be noted that while text-messaging services are usually instantaneous, there can be delays in message delivery during peak times, therefore text messaging should not be used for safety critical messaging where a rapid response is essential.

If a technical fault cannot be rectified quickly onsite, then a swap-in/swap-out method provides the best service option for the user. The engineer responding to the identified fault brings a spare device to the evaluation site and substitutes it for the faulty device if it cannot be repaired immediately.

Users may also need support themselves during an evaluation. Occasionally the presence of a device addressing a fundamental human need can trigger strong emotional reactions. The project team should make sure that the evaluator's usual sources of emotional support are aware of the evaluation and are willing and able to provide support should it become necessary.

10.12 Interim evaluations

Subsequent evaluations usually have a less dramatic impact on the user. The evaluators should understand the operation and concept of the device and know the team well enough to feel uninhibited about criticizing the design. There may be communications between the evaluators and the design team during the interim evaluations for the purpose of clarifying the use or implementation of a new feature, for the evaluator to ask for urgent changes to be made, or for the designer to ask the evaluators' opinion on a potential new feature. In a design context, interim evaluations are less structured temporally than they would be for a formal product evaluation, with the duration of the evaluation depending on the needs of the user.

10.13 The final evaluation

Making the decision to conclude the design and evaluation process is not always straightforward. It is likely that not all the evaluators will be satisfied with the design. Some compromises will have been made for the sake of manufacturability or to accommodate the needs of the greater proportion of users at the expense of an unusual minority. Nevertheless at some point the decision to end the process must be made. This is likely to be when further design and evaluation is not producing significant changes to the design, when *most* evaluators are satisfied that the design meets their needs, and

when the device is deemed to be manufacturable at reasonable cost with some production development.

Before the design effort can be concluded, a decision must be made to freeze the design and proceed to production development. The final evaluation will be used more to validate a stabilized design than to investigate new features. It is often a short evaluation, as by this time the users are familiar with the device and can quickly assess whether it works well or not.

It is advisable to recruit some naïve evaluators at this stage who do not have previous experience with using the device. When the device becomes a product, new users will not have used it before, so it is important to check that the device is appropriate for new users as well as with those who are familiar with its operation and may have adopted ways of dealing with some of its shortcomings.

A formal outcome measure applied at this point in development provides validated proof that the device improves quality of life and has a positive impact on the users' lives. Such proof is persuasive when approaching companies for manufacture, sales, and distribution agreements.

10.14 Preproduction evaluation

The preproduction evaluation is not part of the design process, but it is essential for the validation of the design following modifications for ease of production. This is also an evaluation that should in part be done with naïve users to ensure that people who are unfamiliar with the device will benefit from its use.

10.15 Conclusions

The process of evaluation as a part of user-centered design is complex and multidisciplinary. Above all, however, it is a process of listening and understanding people and their needs and wishes. When users' needs and abilities are given their proper place at the center of the design effort, it creates devices and systems that are useful, usable, and accessible. This approach to the design of devices that interact with people in complex and sometimes intimate contexts is also commercially sound: it results in products that can be confidently marketed as meeting the correctly identified needs of the intended users.

It is also important to emphasize that user evaluation should not be abandoned even where the intended users are technically naïve or physically or cognitively disabled. A lack of knowledge about the internal workings of a device and modern digital interface conventions, or a disability that reduces the scope for human interaction with a device, should not impede user evaluation. If a person has a need that can be met with technology, then that person is capable at some level of interacting with prototypes of a device and influencing the design of the final product.

10.16 Case studies

This section contains four case studies of user evaluations that illustrate evaluation techniques that can be usefully applied to the design of pervasive healthcare devices and systems. Two of the studies (with the Home Energy Tutor and Turvy) are taken from applications outside of the field of pervasive healthcare; however, the techniques illustrated are very appropriate to the healthcare environment.

10.16.1 Case study: A cooker monitor for people with dementia

The ENABLE[7] project sought to measure how technology for people with dementia impacted quality of life.[1] It was active in five European countries: the U.K., Ireland, Norway, Finland, and Lithuania. Five devices were evaluated by people with dementia in their own homes and the Dementia Quality of Life Instrument (DQoL)[10] was used as the outcome measure to assess the impact of those devices on the evaluators' quality of life.

Some of the devices evaluated were finished products, while others were prototypes still under development. Although it was primarily an evaluation of the impact of the devices on quality of life, ENABLE unintentionally became an iterative design evaluation as well, because it identified and rectified faults and problems with the prototypes being evaluated.

One of the devices evaluated was a cooker monitoring system for people with dementia. It used smoke, heat, and natural gas sensors to detect potentially dangerous problems within a cooker. If a problem was detected, the system actively turned off the cooker knobs. If the problem did not go away after this intervention, the cooker monitor would isolate the cooker from its gas supply and then call for help from an external carer. The system had a feature that later proved valuable. It was able to self-check and report identified technical faults to the engineers directly using mobile phone text messaging.

In the U.K., the evaluations were set up by a researcher who was trained to work with people with dementia and had previous experience doing research with people with dementia. Evaluators were selected by the mental health service of the local social services department.

Evaluators and their primary carers were introduced to the project initially by the researcher and the evaluators' community mental health nurses or occupational therapists. A subsequent visit was attended by the researcher, an engineer, and possibly the mental health nurse or occupational therapist. At this second visit, the device and its functions were described in detail by the engineer, consent for the evaluation was requested, and, if consent was given, an initial site survey was carried out. Occasionally the site survey was carried out a few days later.

Installation of the system typically took six hours and was carried out by two installers who were engineers working on the design team. As was referred to earlier in this chapter, two engineers were sent after an early experience when a lone installer left the installation site and was, on his

return, refused access to the site by the evaluator. After a short negotiation, access was granted and work continued. Subsequently two installers attended installations and one always remained at the site.

Evaluations were structured to last for a year. The researcher visited the evaluators every month to discuss the progress of the evaluation, the evaluator's thoughts about the system, and to carry out a DQoL measure. The DQoL was designed to be used as a questionnaire, but it was found during the evaluation it was more successful when used conversationally. The DQoL questions were woven into a conversation rather than asked sequentially with no conversational structure.

The cooker monitor proved to be unreliable. Some of the installations needed frequent visits to rectify problems that occurred with the sensors and mechanisms that turned off the cooker knobs. Many of the evaluators of the cooker monitor dropped out of the evaluation because of the unreliability of the device or because they were transferred to hospital or residential care for unrelated reasons. The ENABLE cooker monitor evaluation led to the development of a simpler, lower-cost, easier-to-install version called the cooker minder. All the cooker monitors were removed at the end of the ENABLE project, except for one in Lithuania.

One of the U.K. evaluators continued to work with the designers and is now evaluating a cooker minder. This evaluation is a less formal, long-term evaluation that has been running for a year. The cooker minder uses voice messaging to prompt the user when a problem is detected. Although there is no active control over the knobs, the device can still isolate the cooker from the gas supply if the problem does not abate. The cooker minder is fitted with a data logger developed for the project that logs sensor, isolation, and reset events, time stamps them, and stores them in flash memory for future analysis. The data logger can be collected by a researcher or engineer when he visits the evaluator or it can be easily removed from the cooker minder and mailed to the designer by the evaluator.

10.16.1.1 Observations from ENABLE

The ENABLE evaluation project highlighted the need to make sure that prototypes intended for people with dementia are further developed than usual because of the difficulty this population has in adapting to change such as technical faults with equipment. When working with people with dementia, reliability is of great importance. This slows down development by increasing the time to first prototype and by increasing the design as well as the amount of internal technical evaluation time needed between evaluation iterations.

ENABLE also demonstrated the feasibility of working directly with evaluators with dementia, so long as those in contact with the evaluators relate to them sensitively and appropriately. Personal relationships have been the key to successful evaluations in the ENABLE project. Where relationships were nurtured, evaluators had better recall of the project team, which reduced the need for repetition and increased the amount and quality of feedback from the evaluators.

Some of the quality-of-life results from ENABLE were inconclusive, particularly device evaluations where the dropout rate was high. While some evaluations did show a positive impact on quality of life, one of the major outcomes of the project was the experience gained working with people with dementia and their carers in a device evaluation environment.

10.16.2 Case study: An evaluation of end user sensor installation in a domestic environment

This evaluation illustrates the use of very low-cost, low-technology proxy devices to evaluate the ability of users to install a distributed sensor network in their own homes.[6] Although this example is for an energy monitoring application, the technique described could usefully be employed in a healthcare context. Distributed sensor networks might be usefully employed to assess the well-being and disease progression of a person with Alzheimer's disease or a physically frail older person.

Evaluators were recruited by a market research firm and were screened to be representative of the target user group of the Home Energy Tutor. Evaluators with a technical background were not recruited. Each evaluator received $75 for participating. A package was delivered to users containing the Home Energy Tutor: this was a mock-up of an energy use monitoring system that used a wireless network of various sensors positioned in key places around the home. One-to-one scale models of different sensors were supplied along with instructions for their installation.

Two observers delivered the package to each evaluator's home. The observers remained to observe and assess the evaluators' reactions to the package and its contents and the accuracy of the evaluators' positioning of the sensors based solely upon the written instructions supplied. No help or coaching was given by the observers. Data were collected by the observers in the form of notes, photographs, and a questionnaire. After the installation was completed, the evaluators were interviewed to determine their level of understanding of the requirements of each sensor.

The evaluation yielded the key results quoted here:

- Make appropriate use of user conceptual models for familiar technologies: use appearances, interfaces, and controls that the user understands and is familiar with.
- Balance installation usability with domestic concerns.
- Avoid the use of cameras, microphones, and highly directional sensors if possible. Users are not good at positioning such sensors accurately.
- Detect incorrect installation of sensors and provide value for partial installations. Partial installations can still yield useful information about the capabilities of the evaluators and the design of the device. Make sure that the assessment tools used can capture this information.
- Educate the user about data collection, storage, and transmission.

10.16.2.1 Observations from the Home Energy Tutor

This evaluation shows the value of a quick, low-technology evaluation performed early in the design process. It yielded useful results with little investment in time and materials. The information gained will guide the future development of domestic distributed sensor installations and showed that home installation by users is a realistic proposition provided that the users understand the purpose and nature of the sensors being employed, not just where they should be positioned. The initial naïvety of the evaluators was important here as naïve users were those most likely to experience problems with the installation. Recruitment actively selected naïve users in preference to those who were technically aware. Where the user will be a person with a cognitive disability, it may be more appropriate for a carer or healthcare worker to install the sensor network. It is still likely that this person will be technically naïve.

10.16.3 Case study: Evaluating Turvy—the Wizard of Oz method

This case study illustrates the "Wizard of Oz" method of evaluation admirably and for this reason it is included in this chapter.[11] The Wizard of Oz[12,13] method employs a human to simulate an intelligent artificial system in an evaluation of an interface between the user and the intelligent system. The substitution of direct human control of an interface for control by an artificial system means that usefulness and interaction methods of intelligent systems can be evaluated and modified before the development of the complex software needed to run them. Fundamental changes to the system's behavior can be implemented without having to rewrite software, and hardware development is minimized and confined to the interface itself.

Turvy was a simulation of an instructible software agent. It was designed to learn how to perform repetitive text-based tasks through user-guided examples and instructions that are input by the user though a speech recognition interface and mouse cursor pointing.

The objective of building and evaluating the simulation was to learn how users respond to a software agent and what interface paradigms would be appropriate in this context. Turvy, the simulation of the agent, was very carefully designed and defined before beginning the evaluation. Tasks were standardized and the amount of information introduced into the systems by the user was minimized by making the task purely manipulative of existing data. Turvy worked by observing an example, making a plausible generalization, and then revising it as further examples and instructions were given.

Users sat at a computer and worked on the text-based task, in this case reformatting bibliographic entries. Next to the user sat a facilitator, and nearby sat Turvy, played by the system designer, who was also equipped with a keyboard, screen, and mouse attached to the same computer as the user. The system designer was to one side and slightly behind the user and was not visible unless the user turned away from her own computer screen.

Having learned how to do the reformatting, the user was asked to teach Turvy how to perform the same tasks. At this point, the Wizard (the person playing Turvy) becomes active and interacts with the user through speech and shared control of the word processing software. Turvy's speech and grammar and responses were strongly structured and controlled, simulating the limited vocabulary, grammar, and decision-making ability that a software agent might employ. It was found that users quickly adapted to communicating with Turvy. They spoke differently to the agent, mirroring its clipped sentences, and referred to the Wizard and Turvy as separate entities.

10.16.3.1 Conclusions from evaluating Turvy

The evaluation of Turvy yielded a rich set of measured and observed data on user interaction with software agents. Users adopted Turvy's vocabulary and preferred to use speech to describe a new focus of attention rather than point with the mouse cursor.

The Wizard of Oz evaluation answered the designers' questions about user interaction with a software agent and also raised some additional questions about users' style of interaction with Turvy. Turvy's designers compiled guidelines for working with Wizard of Oz evaluations:

1. *Prior implementation experience is invaluable*. The experience of Turvy's designers suggests restrictions should be placed on the capabilities of the simulated intelligent system.
2. *The agent's behavior should be based on an algorithm*. This keeps the simulation honest and ensures consistent behavior and experimental repeatability.
3. *The agent's dialogue capabilities should be based on a constrained interaction model*. Interaction with the simulated system agent must be constrained by an explicit list of instructions and feedback that the agent can work with.
4. *It is possible to build real systems derived from studies of verbal Wizard-human discourse.*
5. *The designer benefits from becoming the Wizard.* The designer was trained by playing the role of agent. Revisions were strongly motivated by the designer's responsibility for the users' confusion and discomfort. Simulating an incomplete design reveals its deficiencies.
6. *Qualitative results are the most valuable.* By acting as Wizard, interviewer, and facilitator, the project team becomes immersed in the work and many important results become obvious.
7. *Interviews are essential and video recordings are useful.* Questioning and listening to users is the most efficient way of finding out what works, what doesn't work, and also the users' own internal models of how a system works. Video recordings of interviews allow analyses of users' speech and gestures.

10.16.3.2 Applying Wizard of Oz to the healthcare context

Turvy was implemented in the laboratory. Most pervasive healthcare devices will be evaluated in the field. This can be made possible with remote communications technology between a small call center and the evaluation site or by carrying out short-term snapshot evaluations by bringing the necessary equipment to the evaluation site and taking it away again after the evaluation. Examples of healthcare contexts in which the Wizard of Oz technique could be applied include the prompting of people with cognitive disability using a navigational aid or a responsive voice control interface for a profoundly physically disabled person.

10.16.4 Case study: Evaluating the CareNet display in situ

The CareNet Display is an ambient display showing a photograph of an elder and providing information about the elder's daily life to his local carers (the local members of his care network).[3,14] Devices were placed in the homes of key local care network members, providing them with updated and relevant information about the elder's daily life and needs. They could also have been placed in the elder's home, though this was not necessary for the carers' devices to be functional and was not done in the evaluation.

The CareNet Display consists of a display placed in a wooden picture frame, connected via a GPRS wireless modem to a remote Web server. The display shows a picture of the elder, surrounded by several icons representing different aspects of the elder's life. The display is touch sensitive and interactive. Additional information about the elder's life becomes visible when the icons are touched. The seven information categories displayed are meals, medications, outings, activities, mood, falls, and calendar. The icons change to indicate that something significant has happened in the category they represent.

The information represented on the CareNet Display would be collected by a distributed sensor network in the elder's house; however, because this evaluation was in the early stages of development, the information was collected by people. The designers spoke to experts on distributed sensor networks to ensure that the information displayed could realistically be collected by a sensor network now or in the near future. This evaluation was used to inform the future selection of sensors in the distributed sensor network.

Prior to the finalization of the design for the evaluation, the information to be displayed was determined through a roundtable discussion with seventeen care network members. Twenty types of information were ranked using a card-sorting exercise. The top seven types of information that could be gathered by phone call (the selected means of gathering information) were selected and incorporated into the CareNet Display devices. The elder who is the source of the information to be displayed chose which of the care network members could see which information as some of the information was potentially sensitive. The information shown on the display was

collected by researchers who telephoned the elders several times every day and updated the displays remotely.

The design team conducted several months of group discussions and interviews with their target users before finalizing the design and installing the displays for evaluation.

In the in situ evaluations, prototypes of the CareNet Display were used for three weeks. There were thirteen evaluators: four elders, and two or three family members per elder who provided regular care. Each family member had a prototype of the CareNet Display. All the evaluators were interviewed before and after the evaluation, and the nine family members also completed a questionnaire halfway through the evaluation.

Evaluators were recruited by several different methods: talks at geriatric care conferences, posters in day centers, and through working with local eldercare experts. In addition to the recruited elders and care network members, other family members (children and spouses) also contributed to the evaluation.

The displays were installed for three weeks at a time (this seemed to be the maximum commitment of time the already busy caregivers were prepared to make). There were no special verbal instructions given as to how to use them, though a help booklet was left for the evaluator to read. The displays were mounted in a custom beechwood frame. Elders were telephoned between three and six times a day and were always asked at the end of each call if it was okay to share the information gathered with the display users. The caregivers were paid $150 to participate and the elders and other data providers were each given between $75 and $300.

Evaluators were interviewed for between sixty and ninety minutes before and after the evaluation, which was also documented with researcher notes, questionnaires, audio recordings, and photographs.

The evaluation validated the effort put into discovering the nature of the problem being addressed by this device (isolation) before it was installed. The results were positive and encouraging for further development with real improvements in care and reductions in stress levels being achieved. For example, it was discovered that one of the elders was eating the same food day after day. She had diabetes and mild dementia, so a poor diet was particularly unhelpful for her. Another example of a finding that could only have been made through user evaluation was that in a darkened room the display ceased to be an ambient device fitting in unobtrusively into a room; it became obtrusive because of its brightness. The subtle differences in icon color were found not to be an effective means of communicating information as the changes were not often noticed. It was also found that most of the elders liked being telephoned several times every day by a researcher. This was an unexpected finding. Many caregivers wanted "human touch" information to be included on the display, which would be difficult to gather with a sensor network. A challenge for future work is to incorporate information such as for *whom* Mrs. Jones is knitting, not just that she is knitting.

10.16.4.1 Analysis

This evaluation was very successful and met its objectives of informing the design team about the expectations and abilities of the elders and their care networks. It also successfully informed the team about sensor network design, although some additional challenges were raised that had not been considered. The value of the effort put into the initial interviews and discussions with elders and caregivers became evident when the displays were installed and reviewed. There were no major design problems to overcome during the evaluation and no dropouts. All the evaluators said that they would use the displays if they were available and most said that they would buy one at reasonable cost.

This case study has not described all the findings in the paper referenced for the sake of brevity; however, this was a particularly effective evaluation of an early prototypical pervasive device that is worthy of further study.

10.17 Regulatory device evaluation and adverse incident evaluation

User evaluation is also used by test and evaluation centers funded by regulatory bodies such as the FDA, the NHS PASA, and MHRA. User evaluation is employed to check the usability of devices and systems. Reports on devices are written and published to guide healthcare purchasing decisions.

10.17.1 Product assessment

User evaluation and usability assessment against consensual "standards" of good usability design by experts are used as tools for the formal assessment of a product by healthcare agencies. Reports generated from such evaluations are used by healthcare providers to facilitate purchasing decisions. The evaluations must be thorough, unbiased, and provide the information that potential users and purchasers require to make decisions.

The validation process can also lead to design changes after the launch of a product. Manufacturers of medical devices are required to keep a vigilance file for each product. The identification of adverse incidents that have been the result of inappropriate design may lead to the device being iteratively redesigned and evaluated to ensure that the design fault has been removed. This is particularly so if the cause of adverse incidents was a poorly designed interface that led to human error rather than a technical fault with the design.

10.17.2 Adverse incident investigation

An important role for government-funded test and evaluation centers is to investigate adverse incidents where harm has been caused to a user or patient or where harm has been averted by exceptional user vigilance. The cause of the adverse incident must be determined, whether it is, for example,

operator error, device malfunction, or use of the device in an inappropriate environment. It is necessary, if possible, to determine what the root cause of the incident is and how future similar incidents can be avoided. If human error may be involved in the incident, user evaluations will be conducted to determine how people tend to use the device and how the error that led to the incident occurred.

Particular care is now being taken to observe how users interact with their environment, colleagues, the interface of the device, and the device as a component of a larger system, possibly including other devices. System-wide changes may be necessary to avoid recurrence of similar incidents. Examples of remedies include changes in means of communication between team members, changes in the design of user interfaces, changes to the environment in which the device is used, user training, and so on.

10.18 Conclusions

User evaluation is an essential part of the successful design of pervasive healthcare devices and systems. Evaluation is a wide-ranging activity requiring a team with many different skills, both technical and interpersonal. It can only be effective when the design team engages with the evaluators at a relational level while ensuring at the same time that evaluations are thorough and rigorous. The guidelines and case studies in this chapter illustrate that it is possible to evaluate effectively, even with people with cognitive disabilities.

Participating in user evaluation brings the designer and user closer together and changes both people for the better. Designers begin to understand the people they are designing for and also the users' needs, fears, and desires. The users become people with faces rather than a faceless specification. This is a strong motivating factor toward good design. Users become more acquainted with the challenges of solving technical problems and the huge amount of work that is put into a well-designed product. In learning to work together, designers, evaluators, and, ultimately, all product users benefit. User evaluation is not just a useful tool for a designer in pervasive healthcare but it is also key to the design and production of devices that work.

References

1. Adlam, T. et al. The installation and support of internationally distributed equipment for people with dementia. *IEEE Transactions on Information Technology in Biomedicine*, 8, 253, 2004.
2. Bardram, J. *Hospitals of the Future—Ubiquitous Computing Support for Medical Work in Hospitals*. UbiHealth 2003, Seattle, http://www.pervasivehealthcare.com/ubicomp2003/papers/ (accessed December 2005).
3. Consolvo, S., Roessler, P., and Shelton, B. The CareNet display: lessons learned from an in-home evaluation of an ambient display. In *Proceedings of the 6th International Conference of UbiComp*, N. Davies, E. Mynatt, and I. Siio, eds., Berlin, Germany: Springer, 2004.

4. Fishkin, K. et al. Ubiquitous computing support for skills assessment in medical school. *UbiHealth 2004*, Nottingham, http://www.pervasivehealthcare.com/ubicomp2004/papers/ (accessed December 2005).
5. Howe, D. The free on-line dictionary of computing. http://www.foldoc.org/ (accessed November 2004).
6. Beckmann, C., Consolvo, S., and LaMarca, A. Some assembly required: supporting end-user sensor installation in domestic ubiquitous computing environments. In *Proceedings of the 6th International Conference of UbiComp*, N. Davies, E. Mynatt, and I. Siio, eds., Berlin, Germany: Springer, 2004.
7. Gilliard, J. and Hagen, I. Enabling technologies for people with dementia—cross-national analysis report. European Commission document number QLK6-CT-2000-00653, August 2004, http://www.enableproject.org (accessed December 2005).
8. Demers, L., Weiss-Lambrou, R., and Ska, B. The Quebec user evaluation of satisfaction with assistive technology (QUEST 2.0): An overview and recent progress. In *Technology and Disability: The Assessment of Assistive Technology Outcomes, Effects and Costs*, G.J. Gelderblom and L. de Witte, eds., Amsterdam: IOS Press, 2002.
9. Jutai, J. and Day, H. Psychosocial impact of assistive devices scale (PIADS). In *Technology and Disability: The Assessment of Assistive Technology Outcomes, Effects and Costs*, G.J. Gelderblom and L. de Witte, eds., Amsterdam: IOS Press, 2002.
10. Brod, M., Stewart, A.L., Sands, L., and Walton, P. Conceptualization and measurement of quality of life in dementia: the dementia quality of life instrument. *The Gerontologist*, 39, 1999, p. 25.
11. Maulsby, D., Greenberg, S., and Mander, R. Prototyping an intelligent agent through Wizard of Oz. *Proc. InterChi93*, 1993, 277.
12. Gould, J.D, Conti, T., and Hovanyecz, T. Composing letters with a simulated listening typewriter. *Proc. ACM CHI '82*, 1982, 367.
13. Wilson, J. and Rosenberg, D. Rapid prototyping for human interface design. In *Handbook of Human–Computer Interaction*, M. Helander, ed., New York: North Holland, 1988.
14. Consolvo, S. and Towle, J. Evaluating an ambient display for the home. *Proc. CHI2005*, ACM, Portland, Oregon, 2005.

Bibliography

Orpwood, R. Design methodology for aids for the disabled. *J. Med. Eng. Technology*, 14, 2, 1990.
U.K. NHS Purchasing and Supply Agency Evaluation Service (NHS PASA). http://www.pasa.nhs.uk/evaluation/ (accessed April 2006).
U.K. Medicines and Healthcare Products Regulatory Agency (MHRA). http://www.mhra.gov.uk/ (accessed April 2006).
U.S. Food and Drug Administration (FDA) Device Evaluation. http://www.fda.gov/cdrh/ode/ (accessed April 2006).