DESIGN FOR BIOMEDICAL ENGINEERS John G. Webster and Ramón Pallás-Areny

CONTENTS

1 DESIGN PROCESS

- 1.1 Defining design
 - 1.1.1 Design as problem solving
 - 1.1.2 Design environment: business, legal, social
- 1.2 Design process
 - 1.2.1 Establish the user need
 - 1.2.2 Design specification and project planning
 - 1.2.3 Conceptual design
 - 1.2.4 Detail design
 - 1.2.5 Production
 - 1.2.6 Sales
 - 1.2.7 Removal
- 1.3 Product development process
- 1.4 Designing biomedical products
 - 1.4.1 Interfaces with living systems and operator
 - 1.4.2 Specifications, recommendations, standards, codes and regulations
 - 1.4.3 Ethics
- 1.5 Health care provision schemes
- 1.6 Societal costs
- 1.7 Review questions
- 1.8 References

2 ACQUIRING AND PRESENTING THE IDEA

- 2.1 Information sources
 - 2.1.1 People
 - 2.1.2 Institutions
 - 2.1.3 Databases
 - 2.1.4 Publications
 - 2.1.5 Web
 - 2.1.6 Conferences and Expositions
- 2.2 Presenting the idea
 - 2.2.1 Keeping engineering notebook: Design records

- 2.2.2 Communication
- 2.2.3 Technical report writing
- 2.2.4 Oral presentations
- 2.2.5 Software for presentations, planning, design, manufacturing and analysis
- 2.3 Review questions
- 2.4 References

3 MINIMAL CRITERIA FOR DESIGN

- 3.1 Establishing health care needs
- 3.2 Efficacy
 - 3.2.1 Valid scientific evidence
 - 3.2.2 Health care technology assessment
 - 3.2.3 Efficacy vs. efficiency
- 3.3 Safety
 - 3.3.1 Biological safety
 - 3.3.2 Chemical safety
 - 3.3.3 Electrical safety
 - 3.3.4 Mechanical safety
 - 3.3.5 Nonionizing radiation safety
 - 3.3.6 Ionizing radiation safety
 - 3.3.7 Software safety
 - 3.3.8 Thermal safety
 - 3.3.9 Ultrasound safety
 - 3.3.10 The risk-assessment matrix
- 3.4 Biocompatibility
 - 3.4.1 Host reaction to biomaterials
 - 3.4.2 Body reactions to common biomaterials
 - 3.4.3 Deterioration of biomedical materials
 - 3.4.4 Implant encapsulation and sterilization
- 3.5 Environmental impact
 - 3.5.1 Environmental assessment
 - 3.5.2 Environmental impact statement
- 3.6 Maintainability and cost of operation
- 3.7 Ethics in biomedical engineering design
 - 3.7.1 Conflicts of interest
 - 3.7.2 Resource allocation
 - 3.7.3 Proper testing
 - 3.7.4 Animal research

- 3.7.5 Human experimentation
- 3.8 Biomedical product liability
 - 3.8.1 The legal bases for liability
 - 3.8.2 Negligence
 - 3.8.3 Strict liability in tort
 - 3.8.4 Breach of warranty
- 3.9 Societal costs of biomedical technology
- 3.10 Review questions
- 3.11 References

4 PRODUCT CHECKLIST FOR THE TEAM

- 4.1 The Quality Function Deployment (QFD) technique
- 4.2 Health care needs
 - 4.2.1 Functional performance requirements
 - 4.2.2 Environmental requirements
 - 4.2.3 Time requirements
 - 4.2.4 Regulatory requirements: product classification
 - 4.2.5 Cost requirements
- 4.3 Market research and analysis
- 4.4 Competing products: benchmarking
- 4.5 Patents, trademarks and copyrights
 - 4.5.1 The U.S. patent system
 - 4.5.2 Patent classification
 - 4.5.3 Patent search
 - 4.5.4 Patent infringement
 - 4.5.5 Patent application
 - 4.5.6 Trademark registering
 - 4.5.7 Copyrights
 - 4.5.8 Intellectual property treaty arrangements of the General Agreement on Tariffs and Trade (GATT) negotiations
- 4.6 Benefit/cost analysis
 - 4.6.1 Product life cycle
 - 4.6.2 Capital expenditure
 - 4.6.3 Manufacturing costs
 - 4.6.4 Sale analysis, reimbursement
- 4.7 Company resources
 - 4.7.1 Financial resources
 - 4.7.2 Manufacturing capability
 - 4.7.3 Service capability

- 4.8 Review questions
- 4.9 References

5 WRITING THE DESIGN SPECIFICATION

- 5.1 Explain the biomedical device or process
 - 5.1.1 Function
 - 5.1.2 Performance specification
 - 5.1.3 Benefits
- 5.2 Marketing rationale
 - 5.2.1 Market research results
 - 5.2.2 Intended users
 - 5.2.3 Product distribution
 - 5.2.4 Service
 - 5.2.5 Product retirement
 - 5.2.6 Approval procedure required
- 5.3 System description
 - 5.3.1 Functional blocks and subsystems
 - 5.3.2 Packing specifications
 - 5.3.3 Accessories
 - 5.3.4 Compatible devices
 - 5.3.5 Consumable materials and supplies
- 5.4 Design details
 - 5.4.1 Technical fields involved
 - 5.4.2 Expertise required
- 5.5 Schedule
 - 5.5.1 Define tasks, their objective and resources needed
 - 5.5.2 Planning charts (CPM, PERT, timelines, Gantt)
- 5.6 Cost estimates, budgeting
- 5.7 Review questions
- 5.8 References

6 CONCEPT DESIGN

- 6.1 Medical product design team, team building
 - 6.1.1 Group characteristics: roles, norms, cohesiveness
 - 6.1.2 Biomedical product design group roles
 - 6.1.3 Group development
 - 6.1.4 Decision making methods
 - 6.1.5 Reaching consensus

- 6.2 Develop engineering requirements
 - 6.2.1 Functional requirements
 - 6.2.2 Compatibility requirements with existing products
 - 6.2.3 Manufacturing requirements
 - 6.2.4 Cost requirements
- 6.3 Generate concepts, consider and discuss all possibilities
 - 6.3.1 Functional decomposition
 - 6.3.2 Generate concepts from functions and combine them into a complete design
 - 6.3.3 Methods for generating ideas
- 6.4 Conceptual design evaluation
 - 6.4.1 Feasibility judgment
 - 6.4.2 Technology readiness assessment
 - 6.4.3 Go/no-go screening
 - 6.4.4 Decision matrix (Pugh's method)
 - 6.4.5 Controlled convergence technique
- 6.5 Generate design records: The importance of design records
- 6.6 Summary of design records: design history file (DHF)
- 6.7 Iteration and convergence to final design
- 6.8 Review questions
- 6.9 References

7 DESIGN EVALUATION

- 7.1 Biomedical products design trade-offs
- 7.2 Human factors in medical products
 - 7.2.1 User-centered design
 - 7.2.2 The user interface
- 7.3 Design for compatibility
 - 7.3.1 Functional compatibility
 - 7.3.2 Electromagnetic compatibility
 - 7.3.3 Compliance with safety standards
- 7.4 Design for manufacturability
 - 7.4.1 Manufacturing processes: types, cost factors and design principles
 - 7.4.2 Design for assembly
 - 7.4.3 Subcontracting
- 7.5 Design for testability
 - 7.5.1 Hardware design for testability
 - 7.5.2 Software design for testability
- 7.6 Design for reliability

CONTENTS

7.6.1 Causes of product deterioration

7.6.2 Failure modes, mechanisms and causes

7.6.3 Failure analysis

7.6.4 Reliability analysis and prediction

7.6.5 Reliability in product design

7.7 Evaluate energy and information processes

7.7.1 Power supply

7.7.2 Thermal management

- 7.7.3 Hardware-software partitioning
- 7.8 Optimal design
 - 7.8.1 Functional relationships in optimal design

7.8.2 Unconstrained design optimization

7.8.3 Constrained design optimization

8 DETAILED TECHNICAL DESIGN AND DESIGN VERIFICATION

- 8.1 Dividing up the work
- 8.2 Computer aids for design
- 8.3 Performing biomedical device or process design
 - 8.3.1 Design management
 - 8.3.2 Electrical/electronic design

8.3.3 Mechanical design

8.3.4 Fluid system design

8.3.5 Electromechanical design, motion control

8.3.6 Material selection

8.3.7 Thermal design

8.3.8 Software design

8.3.9 Package design

8.3.10 Industrial design

8.4 Breadboard

8.5 Laboratory testing

- 8.5.1 Safety testing
- 8.5.2 EMC testing
- 8.5.3 Environmental testing

8.5.4 Accelerated life testing

8.5.5 Software validation

8.6 Biomedical product design review

8.6.1 Design review requirements

8.6.2 Design review check list

8.7 Review questions

8.8 References

9 DESIGN VALIDATION

9.1 Prototype

- 9.2 Biomedical design validation
 - 9.2.1 Product evaluation by users
 - 9.2.2 Functional performance evaluation
 - 9.2.3 In situ safety and compatibility evaluation
 - 9.2.4 Biological evaluation of medical devices

9.3 Animal testing

9.3.1 Animal models and alternatives

9.3.2 Design of experiments

9.3.3 Analysis of experimental results

9.4 Clinical trials

9.4.1 Clinical studies

9.4.2 Design of the clinical trial

- 9.4.3 The protocol
- 9.4.4 Clinical trial conduct
- 9.4.5 Clinical trial analysis
- 9.5 Final design review
- 9.6 FDA marketing clearance
 - 9.6.1 Premarket notification (510(k))
 - 9.6.2 Investigational device exemption (IDE)
 - 9.6.3 Premarket approval application (PMA)
 - 9.6.4 Product development protocol (PDP)
 - 9.6.5 Humanitarian device exemptions (HDE)
 - 9.6.6 PMA/510(k) expedited review
- 9.7 Review questions
- 9.8 References

10 PILOT PRODUCTION

- 10.1 Concurrent engineering
 - 10.1.1 Time to market
 - 10.1.2 Concurrent engineering of static products
 - 10.1.3 Concurrent engineering of dynamic products
- 10.2 Documentation release
 - 10.2.1 Design history file (DHF)
 - 10.2.2 Production documentation

- 10.2.3 Test documentation
- 10.2.4 User manuals
- 10.2.5 Device master record (DMR)
- 10.3 Production interface: design transfer
 - 10.3.1 Selection and evaluation of components and suppliers
 - 10.3.2 Specifications for manufacturing processes
- 10.4 Process validation
- 10.5 Review questions
- 10.6 References

11 VOLUME PRODUCTION

- 11.1 Quality assurance systems
 - 11.1.1 Quality control
 - 11.1.2 Product design assurance
 - 11.1.3 Total quality assurance
 - 11.1.4 Statistical process control
 - 11.1.5 ISO 9000 series of international QA standards
- 11.2 FDA Current Good Manufacturing Practices
- 11.3 Engineering change orders
- 11.4 Service and maintenance
- 11.5 Field failure analysis statistics: Medical Device Reporting (MDR)
- 11.6 Medical device tracking
- 11.7 FDA product recalls
- 11.8 Postmarket surveillance
- 11.9 Review questions
- 11.10 References
- Appendix 1: Desirable criteria for topic selection
- Appendix 2: Possible topics for design
- Appendix 3: Acronyms used in the book
- Appendix 4: Web sites of interest in Biomedical Engineering Design
- Appendix 5: Teaching styles for biomedical engineering design

1

Design Process

John G. Webster and Ramon Pallas-Areny

1.1 DEFINING DESIGN

"Engineering design is the process of devising a system, component, or process to meet desired needs. It is a decision-making process (often iterative), in which the basic sciences, mathematics, and engineering sciences are applied to convert resources optimally to meet a stated objective. Among the fundamental elements of the design process are the establishment of objectives and criteria, synthesis, analysis, construction, testing, and evaluation. It is essential to include a variety of realistic constraints such as economic factors, safety, reliability, aesthetics, and social impact." (ABET criteria).

1.1.1 Design as problem solving

Why is the design process for biomedical engineering distinctively different from that for nonbiomedical engineering design? Although the design process is similar to that for a nonbiomedical engineering project, other procedures, such as animal trials, FDA approvals, and clinical testing are unique to biomedical products.

The radiologist takes biopsies (samples) from the liver. Because the liver is a very vascular organ, after the biopsy needle is withdrawn, there is excessive bleeding. He asks you to solve this problem. The radiologist is your client and provides you information about the problem. To solve the radiologist's problem you need to proceed systematically through design steps. You (1) write objectives and review them with your client, (2) search the literature and talk to many people about the problem, (3) brainstorm with your team members, (4) prepare a list of possible solutions, (5) analyze these possible solutions with given constraints, (6) select the best solution and write a detailed specification, (7) build your system, (8) write a protocol with the radiologist and submit to the animal subjects committee. After approval, (9) test your system on a live pig. Your test results suggest changes to your system. (10) Make the changes and (11) apply for a patent. You have now accomplished biomedical engineering design.

1.1.2 Design environment: business, legal, social

What are the business functions of the partners involved in biomedical engineering design and how do they interface? A radiologist treats patients, a hospital provides an effective and safe environment for patient treatment, a medical device company manufactures effective devices, a biomedical engineer develops improved knowledge about medical products, processes and procedures to help mankind.

All of these businesses operate within legal constraints. The radiologist must have training, certification, licensing and consider possible malpractice lawsuits. The hospital must be accredited, select qualified staff, provide adequate facilities and equipment, and consider possible malpractice lawsuits. The medical device company must have new devices approved by the Food and Drug Administration (FDA), protect its intellectual property such as patents and trade secrets, and consider possible patent and product liability litigation.

Many engineers gravitate to biomedical engineering because it is a helping profession. Biomedical engineers help others by solving problems of disease and affliction. In order to solve biomedical engineering design problems, engineers must work effectively with health care personnel, patients, and medical device company personnel. They must be effective in their communication with persons not trained in engineering.

1.2 DESIGN PROCESS

The design process consists of a series of systematic steps to achieve an optimal design. Table 1.1 shows these steps as practiced within a biomedical engineering design class. Biomedical engineers working in industry would perform similar tasks.

1 1 1	4 т	י ויר	• •	1 •		
Table L.	11	Biomedical	engineering	design	process	steps
I GOIC II		JIOIIICalcal	engineering	acoign	process	Deeps

Design process step	Instructional action		
Acquire problems	Ask medical, veterinary, dental, health science personnel to suggest problems that biomedical engineers might solve		
Select problem	Advisors screen problems, students select problems and start notebook		
Form teams	Students form teams of four to six, select leader, communicator, web creator		
Meet with client	Develop questions prior to meeting client. Acquire specific information to clearly define the problem.		
Product design specification (PDS)	Write in engineering terms the problem, goal, function, constraints.		
Conceptual design	Search for information, brainstorm alternative solutions, evaluate, weight requirements to achieve best solution		
Make prototype	Order materials. Acquire tools, workspace, storage space, make prototype		

Test prototype	Obtain animal or human subjects approval, test subjects,
	evaluate, revise and retest
Present results	Give oral presentation, write written report, consider
	manufacturability, patents

1.2.1 Acquire the problem

Biomedical engineering design courses promote the best learning when there is a client who needs and wants a problem solved. Then the problem is real and not contrived. The student learns how to interact with the client in facilitating the client's precise description of the problem. Since there is an interested client with a real need, when the prototype is ready for test, the student has a supportive client who will facilitate animal or human testing. Thus we assume client-based design in this book.

How are clients with real problems acquired? A month before the beginning of a biomedical engineering design course, the advisor sends an e-mail solicitation to members of departments in nearby medical schools, veterinary schools, dental schools, health science personnel, hospitals or medical device companies requesting them to suggest problems that biomedical engineers might solve. Appendix A contains a typical solicitation. Workers in industry can similarly solicit problems from health care personnel.

1.2.2 Select the problem

First the advisor must screen the problems suggested and reject those that are trivial or are research oriented, instead of for a design process. Next, a week before the first class, the selected problems are placed on the course web site and an e-mail notice is sent to students asking them to review the projects and select their 1st, 2nd, and 3rd choices. At the first class, students sign up on posted signup sheets. Where too few students sign up for a project, the students negotiate among themselves to form teams of about four to six.

1.2.3 Form teams

The teams meet to get acquainted and select roles. The selected leader organizes activities and sends a weekly report to the client, advisor, and team members. The communicator provides the only conduit for contacting the client with or for information so the client does not receive multiple messages. The web creator creates a web site for posting contact information, weekly reports, schedule, budget, PowerPoint presentations, written reports, links to useful information, etc. If the client wishes, the web site can be password protected.

Each student maintains a bound, page-numbered design notebook. This notebook should be a dated ink written record of all information acquired and ideas developed during the design process. It should contain contact information, client meetings, brainstorming, literature searching, sketches, calculations, action items, etc. It should provide legal documentation of your design in case a patent is pursued. It is best to keep a separate loose-leaf notebook for bulky items such as articles, patents, component specifications, e-mail correspondence, etc.

1.2.4 Meet with the client

It is important to obtain clear information from the user (client). For example, if the radiologist inserts a biopsy needle, he is the user and can provide you the most information and can define the need. If a nurse has problems understanding how to operate a medical device, she can best define the need. If a wheelchair user has problems operating a wheelchair, he can best define his need. However if the wheelchair user develops pressure sores, you will need to talk with cushion researchers and manufacturers because the wheelchair user may not understand the origin of pressure sores.

The student team should develop questions prior to meeting the client. Search the literature to gain an understanding of the problem. Typical questions might include (1) "What is the end goal to be solved?" (2) "What are the medical constraints? Should we be concerned with patient confidentiality?" (3) "What are the technical constraints? Can we use metals and plastics? Must it be splash-proof?" Acquire information to define the problem.

1.2.5 Product design specification (PDS)

Create a Product Design Specification (PDS). The PDS is a **comprehensive** document, which contains all the facts relating to the product outcome, and should contain all the realistic constraints to be imposed upon the design by the client. Write in engineering terms the problem, goal, function, constraints.

Items in the PDS should be as **quantitative** as possible (e.g., the device must weigh less than 2 kg, the device must fit in a $1 \text{ m} \times 1 \text{ m} \times 1$ m space), and be ranked in order of importance. The PDS is a **dynamic** document that should evolve as the project scope develops. This is because frequently at the start of a project it is not always clear what is achievable and to what extent certain parameters are essential.

CONTENTS OF PDS

Title: The PDS should have all team members names listed, as well as the title of the project. It should also be dated, to avoid conflicts arising from different versions.

Function (a general statement of what the device is supposed to do): The PDS should begin with a brief, concise paragraph describing (in words) the overall function of the device. In the initial stages, this will be the problem statement, and will become more specific as you decide on a final design.

Client requirements (itemize what you have learned from the client about his/her needs): Briefly describe, in bullet form, the client needs and responses to your questions. For example: specific information on customer likes, dislikes, preferences, and prejudices should be understood and written down.

Design requirements: This device description should be followed by list of all relevant constraints, with the following list serving as a guideline. (<u>Note</u>: include only those relevant to your project):

1. Physical and Operational Characteristics

a. *Performance requirements*: The performance demanded or likely to be demanded should be fully defined. Examples of items to be considered include: how often the device will be used; likely loading patterns; etc.

b. *Safety*: Understand any safety aspects, safety standards, and legislation covering the product type. This includes the need for labeling, safety warnings, etc. Consider various safety aspects relating to mechanical, chemical, electrical, thermal, etc.

c. *Standards and Specifications*: international and/or national standards, etc. (e.g., Is FDA approval required?)

d. *Accuracy and Reliability*: Establish limits for precision (repeatability) and accuracy (how close to the "true" value) and the range over which this is true of the device.

e. *Life in Service*: Establish service requirements, including how short, how long, and against what criteria? (i.e. hours, days of operation, distance traveled, no. of revolutions, no. of cycles, etc.)

f. *Shelf Life*: Establish environmental conditions while in storage, shelf-life of components such as batteries, etc.

g. *Operating Environment*: Establish the conditions that the device could be exposed to during operation (or at any other time, such as storage or idle time), including temperature range,

pressure range, humidity, shock loading, dirt or dust, corrosion from spilled fluids, noise levels, insects, vibration, persons who will use or handle, any unforeseen hazards, etc.

h. *Ergonomics*: Establish restrictions on the interaction of the product with man (animal), including heights, reach, forces, acceptable operation torques, etc..

i. *Patient-related concerns*: If appropriate, consider issues which may be specific to patients or research subjects, such as: Will the device need to be sterilized between uses?; Is there any storage of patient data that must be safeguarded for confidentiality?

j. *Size*: Establish restrictions on the size of the product, including maximum size, portability, space available, access for maintenance, etc.

k. *Weight*: Establish restrictions on maximum, minimum, and/or optimum weight; weight is important when it comes to handling the product by the user, by the distributor, handling on the shop floor, during installation, etc.

1. *Materials*: Establish restrictions if certain materials should be used and if certain materials should NOT be used (for example ferrous materials in MRI machine).

m. *Aesthetics*, *Appearance*, *and Finish*: Color, shape, form, texture of finish should be specified where possible (get opinions from as many sources as possible).

n. *Competition*: Are there similar items which exist (perform comprehensive literature search and patent search)?

2. Production Characteristics

a. Quantity: number of units needed

b. Target Product Cost: manufacturing costs; costs as compared to existing or like products

3. Problem Statement

It is important to state the problem in engineering terms. It is not sufficient to state the problem as, "We want to stop the bleeding after taking a liver biopsy." Instead provide specific engineering details, objectives, and constraints as follows:

- Use a 14 gauge biopsy needle
- Make minimal modification to existing equipment
- Use biocompatible materials
- Preserve the tissue sample

- Minimize procedure time
- Make needle disposable
- Make needle sterilizable
- Maximize patient comfort
- Make easy to operate

Write a detailed product design specification (PDS) so all involved in the project can understand it and can offer suggestions for improvement.

Estimate time and cost to solve the problem. Estimations need to include the number of personnel, both their time and cost. Personnel might include design engineers, technicians to fabricate hardware, and animal handlers to perform tests on pigs. Cost estimates of hardware and supplies (including pigs) are necessary. Provide a schedule of activities to reach the specified goal of a tested prototype.

Example of a PDS for a Cauterizing percutaneous biopsy needle (Appendix B)

The procedure currently used in order to perform a biopsy involves inserting a sharp rod (stylet) that is housed inside a hollow cylindrical tube (introducer needle) through the skin and into the organ. The stylet is then removed from the introducer needle and body. Next, a biopsy needle is inserted into the introducer needle (which is still in the body). This biopsy needle is longer than the introducer needle. Therefore, it extends beyond the end of the needle into the organ. The end of the biopsy needle contains a special tray in which the tissue will be collected. A trigger is pulled, projecting the sheath forward over the sample tray and trapping the tissue sample inside the tray. The biopsy needle and tissue sample are then withdrawn from the patient. Finally, the introducer needle is removed from the body. Several factors must be taken into consideration before successfully integrating RF ablation into the biopsy procedure.

It is desirable to prevent bleeding associated with biopsies. The client proposed utilizing radiofrequency ablation in order to stop bleeding associated with biopsies. Radiofrequency ablation has been used for decades in medical procedures. When a conducting probe is connected to a radiofrequency generator at 500 kHz, the energy will flow through the probe, creating ionic agitation and friction in the nearby tissue. This friction results in the heating of the surrounding tissue, and ultimately a sufficient temperature (approximately 50 to 60 $^{\circ}$ C) is reached that kills the target tissue. There are several constraints:

1 Minimal modifications to existing equipment

Additional equipment is undesirable. The optimum solution would modify the existing equipment in order to perform the desired function. The device must be integrated into the current procedure as opposed to creating a new procedure.

2 Temperature of surrounding organ tissue

The ablation should increase the temperature of the surrounding liver tissue to between 65 °C and 80 °C while the tract is being cauterized. A temperature of 65 °C is necessary to ensure that the tract is cauterized completely. A temperature above 80 °C will cause excessive damage to the tissue. Temperature feedback would also be a desirable characteristic of the device.

3 Biocompatibility

The materials used in the device must be biocompatible and able to survive sterilization procedures such as autoclaving and radiation.

4 Preserve tissue sample

The sample of tissue that is trapped in the biopsy needle must not be damaged in any way.

5 Time for removal

Removal of the device from the patient should be as quick as possible in order to maximize patient comfort. It currently takes only a few seconds to retract the introducer needle.

6 Compatible with generator

The device must be compatible with various radiofrequency (RF) generators. Table 1.2 lists the specifications of the RITA Model 1500 RF Generator (the generator used to test our device).

Specification	Value
Power	0–150 W
Frequency	460 kHz
Power delivery accuracy	$\pm 10\%$
Power supply	110–240 V, 50–60 Hz, universal power supply
Impedance range	10–999 Ω, ±20%

 Table 1.2. Characteristics of the generator used in testing.

7 Disposable

The introducer needle must be disposable.

8 Temperature increase

The increase in temperature should be uniform throughout the tract created by the introducer needle. This will create a uniform lesion.

9 Insulation

If a large portion of the introducer needle is left uninsulated, the tissue sticks to the needle when the radiofrequency current is applied. Insulating the entire needle except the very distal end of it can prevent this from occurring.

10 The following characteristics are important for insulation used in this type of procedure [6–10]. The coating must:

- Be tightly adhered to substrate without voids at the interface
- Be of uniform thickness
- Have little impact on the dimensions of the device (no thicker than $100 \ \mu m$)
- Have a low coefficient of friction, preferably less than 0.2
- Be pinhole free
- Be sterilizable
- Be biocompatible and biostable (typically as a United States Pharmacopeia (USP) class VI material.)

- Meet industry standards
 - o Association or the Advancement of Medical Instrumentation (AAMI)
 - o International standards

1.2.6 Conceptual design

First, generate as many potential alternative solutions to the problem as possible. Brainstorm all kinds of ideas yourself and with others. Use your imagination to think of as many independent design concepts as possible (nothing is out of bounds yet). Do not reject any idea as unworkable. At this point, there should be no evaluation of design yet, just idea generation. Sketches are VERY important here to illustrate your ideas (and should be in your design notebook). For example we know that one way of stopping bleeding is by heating blood until it clots to solve the Cauterizing percutaneous biopsy needle problem. Alternatives from idea generation are heating the needle electrically or by hot fluid or heating adjacent tissue by passing electric current through it.

After generating many potential alternative solutions, concept evaluation begins. Evaluate each potential solution against the objectives and constraints. Consider weighting the importance of each of the objectives and constraints to help evaluate them. Students evaluated the Cauterizing percutaneous biopsy needle problem and decided to use electrosurgery to pass electric current through adjacent tissue.

Figure 1.1 places the above specifications into a weighted objectives tree. The respective weights of each of the above objectives are placed in parenthesis after the objective. This tree summarizes the objectives and gives their relative importance in the design with weighted values.

DESIGN PROCESS



Figure 1.1 Weighted objectives tree with weights given in parenthesis on a scale of 0 to 1, with 1 being the highest in priority. Note that all terminal weights sum to 1.0.

Alternative solutions

Two solutions proposed that integrate RF ablation into the biopsy procedure in order to cauterize the tract are detailed as follows:

"Without stylet"

All but the distal and proximal tips of the introducer needle are insulated. The distal tip is left uninsulated in order to apply energy to the surrounding tissue, and the proximal end of the introducer needle is left uninsulated in order to leave room for a connection to the RF generator. The needle is connected to an RF generator, and the introducer needle itself is thus used as the ablation catheter. Unfortunately, this design introduces concentrations of heat (hot spots) at the end of the needle because of its sharp edges left at the distal tip of the needle. This would create inconsistent lesions in those areas of the tissue.

"With stylet"

In order to alleviate the heat concentrations that could occur in the previous design, a rounded stylet can be placed through the introducer needle prior to applying the RF energy. The portion of the stylet through which the energy would be applied does not have sharp edges and energy concentration would be greatly reduced. However, placing the stylet through the needle introduces another step and additional time, albeit little, into a procedure where time is to be minimized. This design also provides a means for simple integration of temperature feedback into the design by using a thermocouple as the stylet. After the biopsy needle is removed, a thermocouple can be inserted into the introducer needle. The thermocouple thus would serve a dual purpose: to act as a stylet, alleviating heat concentrations, and to allow the user to obtain thermal feedback.

Insulation

In order to implement the above designs an appropriate insulation must be chosen. Possibilities for insulation include the following:

Heat shrink technology

- Examples are:
 - Polyvinylidene fluoride (PVDF)
 - Low-density polyethylene (LDPE)
 - Blend of polyolefin and zinc or sodium partially neutralized ethylene acrylic acid copolymer
 - High-density polyethylene (HDPE)
 - Fluorinated ethylene propylene (FEP)
 - Polyvinyl chloride (PVC)
- Heat shrink technology was quickly ruled out due to the following downfalls:
 - Process of heat-shrinking is subject to the operator's skill
 - Only LDPE, HDPE, FEP with minimum wall thickness of 410 μm were able to pass the HF18-1986 requirements (AAMI requirements)
 - No guarantee that the insulating material is well adhered to the substrate and that there are no air voids at the interface
 - Sterilization techniques such as autoclaving could increase the number of voids at the interface.

Parylene is currently used in electrosurgical devices. Parylene has many advantages in that it:

- Has a low coefficient of friction. Its lubricity approaches that of Teflon. (Coefficient of friction for Parylene C (static and dynamic) is 0.29)
- Is able to penetrate cracks and crevices because it is deposited in the gas phase
- Is polymerized at room temperature. No solvents or high temperatures are needed for polymerization, and therefore there is no thermal or mechanical stress introduced during the application process
- Is free of pinholes and other defects and deposits

- Has excellent uniform thickness, which is therefore controllable (from 10 nm to hundreds of micrometers)
- Can survive various sterilization procedures, including autoclave, radiation, and ethylene oxide
- Has excellent dielectric properties
 - Excellent electrical insulator
 - o Dielectric constant and dielectric losses are low
 - Dielectric constant and losses are unaffected by the absorption of water vapors
- Is expected to survive continuous exposure to air at 100 °C for ten years, meaning that it will be able to survive the temperatures associated with radiofrequency ablation
- Can be removed from the metal substrate if desired
- Is a biocompatible barrier to chemicals, moisture and biofluids
- Is recognized as a Class VI polymer by the FDA

Numerical evaluation matrices

Numerical evaluation matrices (decision tables) are useful for selecting one from a set of design alternatives. These have columns listing specifications, their weight with scores for each design alternative. Table 1.3 shows a numerical evaluation matrix for all of the specifications for insulation that are important in this design. Each is given a weight between 1 and 3 (with 3 being the most important). Table 1.3 provides a comparison between heat shrink tubing and Parylene by using the weight provided for each specification and the amount to which each insulation fulfills that specification.

Specification	Weight (1–3)	Heat shrink tubing	Parylene
Superior dielectric properties	3	3	3
Tightly adhered to the substrate without voids at the interface	2	1	2
Uniform thickness	2	2	2
Must have little impact on the dimensions of the device (must be very thin)	3	1	3
Low coefficient of friction	V	2	2
Resistant to wear and abrasion	1	1	1
Pin-hole free	2	2	2
Seal microporosity of the substrate	1	1	1
Sterilizable	3	3	3
Must meet USP Class VI and industry standards (biocompatible)	3	3	3
Total	23	19	22

 Table 1.3 Numerical evaluation matrix comparing heat shrink tubing and Parylene.

1.2.7 Prototype Development

Compare each design against the items in the PDS to see which idea best meets specifications. At this point, you may combine the positive aspects of different designs to form a single, final design. Once a decision has been made, go back and re-evaluate your choice—remember it is much easier to change things on paper than if something has been built! As you fill in the details of the design, you should continually evaluate your design choice. Here, you will consider items such as: (a) dimensions (b) materials, fasteners, etc. (c) analysis (loads, flow rates, etc.) (d) more sketches and drawings

It is also useful at this point to build simple models. This is useful for understanding spatial relationships of the design—how different parts fit together, etc. Be creative in using everyday materials for model building; you can use anything from popsicle sticks to coat hangers and anything else that works.

Once a single design has been chosen, there should be continual evaluation throughout the entire process; before building look over the design in its entirety.

Example: Cauterizing percutaneous biopsy needle

In vivo testing

The introducer needle was coated with parylene (courtesy of Vitek Research Corporation) before *in vivo* tests were performed. 4 mm of the distal tip of the needle was left uninsulated. The needle was coated with a layer of Parylene that is approximately 1 mil in thickness (1 mil = 25.5 μ m = 0.001 in.). Vitek Research Corporation estimates that it would cost approximately \$13. per introducer *needle* in order to coat introducer needles in bulk quantity. This number is acceptable because it would only increase the cost of the introducer needle by approximately 10–15%.

A thermocouple was also added in order to monitor the temperature of the tissue. After the biopsy needle was removed from the patient, a thermocouple for an Endocare CRYOcare[™] System was run through the introducer needle. This allowed the doctor to maintain a constant temperature as the introducer needle was withdrawn from the patient. (See Fig. 1.2)



Figure 1.2 Insulated introducer needle with thermocouple.

As used previously in *ex vivo* testing, a RITA® Model 1500 Electrosurgical Radiofrequency Generator was used as the source of RF energy.

1.2.8 Test prototype

Field test device under conditions it will encounter in practice.

Example: For a cauterizing percutaneous biopsy needle, initial tests were performed on ex vivo cow liver to save the expense of in vivo experiments. The two designs were tested via three sessions (See Figure 1.3 for experimental set up). Insulation, power, time, and the introducer needle with or without the stylet were tested for optimum performance by holding three of the four variables constant while varying the final variable and observing the affects of this variation.



Figure 1.3 Experimental setup for *ex vivo* testing. A wire was connected between the RF probe of the generator and the introducer needle to allow the RF energy to be delivered directly from the introducer needle to the tissue. The connections were made using alligator clips.

The first testing session verified that the insulation should cover most of the introducer needle (except for the tip). The larger surface results in less energy concentration. The final design will therefore have as small amount as possible of the distal end of the needle left uninsulated.

Heat transfer in the tissue is a function of (1) the amount of heat that is applied and (2) the rate at which heat flows. Therefore a higher power will give a better burn, and thus the highest possible power that does not destroy too much tissue will be utilized in the final design.

More testing must be performed *in vivo* in order to determine the optimum length of time that the needle must be left in the patient before and while it is withdrawn. The testing showed that longer lengths of time resulted in better burns, and therefore a compromise will need to be determined as to how much time the physician must hold the needle in the patient while still obtaining a satisfactory burn.

No conclusive results were drawn for the differences between the device with or without the stylet. More testing (preferably *in vivo*) must be performed before it is determined whether the device performs better with or without the stylet.

A pig was used to test the described system *in vivo*. Two trials were performed. One trial involved inserting the insulated introducer needle and thermocouple into the liver and the other trial involved inserting the insulated introducer needle and thermocouple into a kidney.

The insulated introducer needle and thermocouple were withdrawn from the pig without heating so as to obtain a control group. In the experimental group, the introducer needle was held in place in the pig while heating until the temperature rose to 80 °C. The needle and thermocouple were then retracted from the pig while attempting to maintain a constant temperature of 100 °C. The times for retraction were recorded, and the blood lost was soaked with a gauze pad and weighed.

The above trials were performed before and after the pig received Heparin (a blood thinner). The liver was also studied after clamping the inferior vena cava exiting the liver in order to stimulate a pressure build up in the liver.

In general, more bleeding occurred in nonablated control tracts than in ablated tracts. This implies that applying RF energy via an introducer needle while monitoring the temperature shows promise as a method for safely stopping bleeding after biopsies. More studies must be performed in order to further quantify the precise parameters (power and time) that must be observed during ablation in order to define a more quantitative difference between ablated and nonablated tracts.

1.2.9 Present results

Final Design Report (See example in Appendix B)

Prepare a final report, which should include the following:

- 1. Title, team members, client, advisors, date on a title page.
- 2. An abstract of your work, summarizing in less than 150 words the problem, your design solution, etc.
- 3. A concise statement of the design problem.
- 4. A summary of the information you have gathered concerning the design problem. Refer to sources of information (people, books, web, other).
- 5. A discussion of the most important design constraints, with the full PDS included as an appendix.

- 6. A description of alternative solutions to the problem that you have developed based on your creative thinking. These should include ideas that involve different basic principles and concepts, rather than variations of a single principle or concept. Include a sketch of each of your ideas and a brief explanation of how the device or system would operate. You can use black ink or dark pencil to sketch, then scan it, and then insert it into a word processing program. Those who know a drawing program can use that.
- 7. A preliminary evaluation of each of your ideas, keeping in mind the requirements and specifications. This should be more than just a listing of the advantages and disadvantages of each option. The evaluation should take into account the relative importance of the different evaluation criteria.
- 8. The solution and a summary of your reasons for making this choice.
- 9. A concise presentation of the details of your chosen design—this should include items such as dimensions, materials, cost estimates, etc. Assembly and detail drawings should be included (these can be attached as an appendix or as additional files). Also, if you have built a prototype, you should include photographs.
- 10. Conclusion—this must include a section describing the ethical issues surrounding your project. Also include suggestions for future development of your design.
- 11. References (properly cited)
- 12. An appendix including your latest Product Design Specification. Appendices with additional information as necessary.

When preparing your project reports, please keep in mind the purposes of these reports. Certainly the reports serve to inform the reader of the specific facts dealing with the projects, i.e. the problem being addressed, the proposed solution that was pursued, and the results of the activity.

In addition to a straightforward, clear, presentation of these facts, however, the project report should provide the reader with some additional insight into your thinking on the project. For example, it is most likely important for the reader to know what critical decisions you had to make along the way and the reasons for choosing the directions that you pursued. Certainly, literally hundreds of decisions had to be made when developing your designs. Some of these decisions were relatively trivial and had no major bearing on the outcome of the project. Other decisions, however, were of much greater importance and had very significant effects on the result of your effort. It is important for you to identify these critical decisions and discuss the basis for the decisions that you made. Furthermore, it is likely that, as the project proceeded to a conclusion, you gained additional information and insight (as a result of the design process) that would lead you along a different path if you were to tackle the same problem again. It is certainly important for you to make this clear to the reader.

As a result of your work, you hopefully have also gained new insight into and knowledge about the particular problem that you have been dealing with. This insight and knowledge would most likely be useful to someone who was interested in following up on your work. It is important for you to convey this in your report.

In addition to the knowledge and insight that you have gained with respect to your particular project, you have also hopefully developed your design skills. Evidence of your understanding of the fundamentals of the design process as well as your growth along these lines should also be present in the report.

DESIGN PROCESS

Finally, the preceding comments should be seen as guidelines, and not a set of rules. Since you have done the work, you are in the best position to provide a full and reliable accounting of this work. Certainly you should keep in mind the ever present norms of good taste and high quality that are expected of engineers throughout your preparation of the report, but you should also exhibit the creativity in your writing that we expect in your work.

Prepare the report as a Word document with inserted and numbered figures with legends beneath them. When the final report is due, post it on your web site, send it as an e-mail attachment, and give a hardcopy to your client(s) and to your advisor.

You will be expected either to make a poster presentation or a lectern presentation based on your final report. You should inform your client of the time and place of the presentation, and invite them to attend. Appendix C contains poster presentation instructions.

1.3 PRODUCT DEVELOPMENT PROCESS

To be written

1.4 DESIGNING BIOMEDICAL PRODUCTS

To be written

1.4.1 Interfaces with living systems and operator

To be written

1.4.2 Specifications, recommendations, standards, codes and regulations

To be written

1.4.3 Ethics

To be written

1.5 HEALTH CARE PROVISION SCHEMES

To be written

1.6 SOCIETAL COSTS

To be written

1.7 REVIEW QUESTIONS

- 1.1 List and explain the differences between a lecture course and a design course.
- 1.2 List the possible sources of medical device design problems.
- 1.3 If you were team leader, describe how you would organize your team. Describe how you would promote their effective interaction.
- 1.4 List the requirements for a design notebook. Explain why these requirements are needed.
- 1.5 Distinguish between functions and objectives.
- 1.6 Your team has selected the following problem. List questions you will ask your client. "Persons with disabilities need access to all forms of modern health care, including dental procedures, health care check ups, and diagnostic procedures such as mammography. Unfortunately, barriers are common for persons with disabilities because of patient positioning, comfort and ease of use. A platform device is desired that enables wheelchair users access to health care procedures. The device should have two-degrees of freedom (rotation of 360°, and vertical translation from 3"-9" above the floor." (Suggested by National Design Competition: Innovations in accessible medical instrumentation (wheelchair), John Enderle and Jack Winters)
- 1.7 Write a product design specification for problem 1.6.
- 1.8 Create a weighted objectives tree for problem 1.6.
- 1.9 Create a numerical evaluation matrix for problem 1.6.
- 1.10 Create a schedule and a budget for problem 1.6.
- 1.11 Develop a list of five patents applicable to problem 1.6.
- 1.12 Describe a standard applicable to problem 1.6.
- 1.13 Describe a new medical device that does not exist and needs to be designed.

1.8 REFERENCES

Center for Devices and Radiological Health, http://www.fda.gov/cdrh/ 2003.

- Dym, C. L., Engineering Design: A Project-based Introduction, New York: John Wiley, 2000.
- King, P. H, and Fries, R. C., *Design of Biomedical Devices and Systems*, New York: Dekker, 2003.
- Laeseke, P. F., Winter, III, T. C., Davis, C. L., Stevens, K. R., Johnson, C. D., Fronczak, F. J., Webster, J. G., and Lee, Jr, F. T., Postbiopsy bleeding in a porcine model: reduction with radio-frequency ablation--preliminary results, *Radiology* 227, 493-499, 2003.
- Sawyer, D., Do it by design, http://www.fda.gov/cdrh/humfac/doit.html, 2003.

3

Minimal Criteria for Design

John G. Webster and Ramón Pallás-Areny

What are the next steps after the health care need has been identified? Developing precise questions for gathering pertinent information provides an effective roadmap. Prior to searching for technical solutions, search for regulations pertaining to the client's proposed product and ethical guidelines for biomedical engineers.

Medical products must be safe and effective for the intended use. Are there impediments? Discovering whether they exist and have an impact on the safety, efficacy and cost of the client's proposed product determines your acceptance or rejection of the client's request.

3.1 ESTABLISHING HEALTH CARE NEEDS

The first step of developing a list of specific questions builds your roadmap. Initially the information you gather from user interviews (groups or marketplace) will be expressed in user terms. Your task is to translate the user terms into precise engineering terms, therefore defining the problems to be solved (Section 2.2). However, before you begin, define the problem to be solved by answering the following design input questions (Lowery et al., 1996):

- 1. What is the real need for the new product?
- 2. Where will the new product be used?
- 3. Who will use the new product?
- 4. How will the new product be used?
- 5. With what devices will the product be used?
- 6. How long will the new product be used?
- 7. Other questions related to the specific product to be developed

The real need for the new product should not be judged only from the presence or absence of competitor products in the market. Good engineering design can improve the performance of existing products or reduce their cost. The lack of competing products may mean that the health care need is not met yet, but can also be a warning signal that the perceived need does not lead to any viable product. The judgment about the need should not be limited to the particular user or group considered in the initial search. A further search may lead from the group of users posing a need to the finding of other groups that can also potentially benefit from that product.

The place where the product is to be used sets different engineering, marketing and legal requirements. From the engineering point of view, the environment where a product works is quite different in a hospital, physician's office, ambulance or home. Achieving the same functionality and safety may require different engineering solutions according to the case. Furthermore, in a hospital, for example, laboratory products and products for a diagnostic, therapeutic or surgical unit are subject to different regulations. Buying cycles and procedures are different for private organizations and government agencies, and this results in different time requirements for developing a product and also different sale costs. Fitness and some sport (nonmedical) products concerning physiological parameters are better considered as consumer products. Biological research instruments must be designed according to guidelines for general scientific instruments. Regulations for drug manufacturing processes are different from those for medical devices.

The intended user also influences product design. Trained persons, particularly if they are routine users, operate equipment more efficiently and safely than unskilled or occasional users. Trained users can also recognize failures and malfunctions, and will be less error-prone than untrained users. Devices accessible to a wide public must not confuse the user about how to operate them, and should not pose great risks when used incorrectly.

The way the product is to be used determines several aspects related to the mechanical and electrical safety of a product, both for the user and the patient. The design of tools and devices applying energy to the patient must consider human factors engineering (Section 7.2). Implantable products must be biocompatible and highly reliable.

The compatibility with other products used simultaneously or successively with the proposed product also influences its design. Products that complement other products already available can reduce costs and be quickly accepted. Product compatibility must be achieved by considering energy, matter and information transfer interfaces, and by ensuring that the new product does not introduce environmental changes (electromagnetic fields, material emissions) that thwart the operation of other products.

The lifetime of the intended product influences material selection, the use of specific energy sources, the actual cost of the product in terms of operating expenses and the number of patients benefiting from it. The design of products including software components must consider future upgrades. The design of products intentionally incompatible with previous products with similar or complementary functions is ethically questionable.

Finally, many products pose specific questions, some of which can arise only in a knowledgeable design environment. Quality System regulations for medical devices depend on device classification (Section 4.2). Specific medical devices, such as contact lens or surgical gloves are subjected to particular requirements. Products using radionuclides are subjected to additional safety regulations. Devices using radiofrequency energy either for communication (medical telemetry) or for diagnosis, surgery or therapy, have operating frequency selection.

Answering the above questions leads to a complete and nonambiguous specification of a safe and effective product that addresses the actual need of the user and the patient.

Example: Design of a glucose sensor (Establish need)

In normal physiology, the glucose concentration in the body is actively managed within limits by a complex feedback control system including the pancreas, which releases the peptide hormone insulin in response to increases in blood glucose. Diabetes opens or disturbs this feedback control permitting glucose to rise to toxic levels (hyperglycemia). After years of hyperglycemia, slowly accumulating organ damage then results in illness or death from early heart disease, kidney failure, blindness and limb amputation.

In an effort to control their disease, patients can lance their finger to obtain a blood sample to measure glucose concentration with a small portable instrument. With this information they can then, in principle make informed adjustments to diet and self-injected insulin dosing. Unfortunately, achieving close to normal average glucoses with this noncontinuous finger prick method is very difficult to achieve even measuring glucose 5 to 10 times daily. Statistically, Americans with diabetes measure glucose less than once per day on average. In addition too much insulin can produce dangerously low glucose levels (hypoglycemia) resulting in disabling insulin reactions and even coma and death. As a result of these factors patients tend to underinsulinize to avoid the short term risk of insulin overdose and then ignore the long term consequences of high mean glucose levels. Patients and physicians agree that a continuous glucose sensor providing glucose level and alarm functions on an almost effortless basis is needed to give patients the information to optimally control their diabetes.

3.2 EFFICACY

Efficacy is the power to produce an effect. Biomedical innovations receive much publicity in the media and are usually seen as beneficial, so that there is some risk in having them accepted before their efficacy has been proven. Innovation in other products (food for example) are not so easily accepted by distributors and consumers.

A device is considered effective when its proper use provides clinically significant results in a significant portion of the target population. The evidence submitted by the manufacturer to the Food and Drug Administration (FDA) to substantiate the safety and effectiveness of the device, may take any form, but the FDA relies upon only "valid scientific evidence" to determine if that device is safe and effective for its conditions of use.

3.2.1 Valid scientific evidence

Valid scientific evidence is "evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use" (FDA, 2003). It follows, thus, that it is not enough to propose a plausible product, because its distribution is tied to a proven efficacy. We must show that we will be able to prove efficacy before engaging in detailed design.

The "evidence" is not determined by the manufacturer itself but by "qualified" experts. Therefore, it will be important to carefully specify the product characteristics and conditions of use, and to document in detail the reports submitted to permit scientific evaluation.

Usually, the valid scientific evidence shall consist principally of "well-controlled investigations," which means investigations that use a test device with standard design and performance, and which are performed according to the principles, stated in (FDA, 2003), which include:

- 1. A clear statement of the objective of the study.
- 2. A method for adequate selection of the subjects, their assignment to test groups, assuring comparability between test groups and control groups.
- 3. An explanation of the methods utilized to observe and record results, including variables measured and steps taken to minimize any bias of subjects and observers.
- 4. A comparison of the results of treatment or diagnosis with a control in such a fashion as to permit quantitative evaluation. There are four general types of comparisons:
 - a. No treatment. The results in a group of treated patients are compared to those in a group of comparable untreated patients.
 - b. Placebo control. The results of using the device are compared with those when using an ineffective device under similar conditions.
 - c. Active treatment control. The results in a group are compared to those in another group receiving a different therapy.
 - d. Historical control. The results of using the device are quantitatively compared with prior experience in comparable groups who received no treatment or an effective treatment.
- 5. A summary of the methods of analysis and an evaluation of the data derived from the study, including any appropriate statistical methods utilized. No specific statistical studies are prescribed, but those used must be thoroughly documented.

For products intended to treat less serious illness or provide relief from symptoms, the common method to show effectiveness relies on a clinical trial comparing the product to a placebo, not to any other product. In order to constitute an adequate showing of effectiveness, this clinical trial must be duplicated either by a second study or by a well-designed multicenter study. Section 9.4 describes the design of clinical trials.

Well-controlled investigations are not necessary when there is already valid scientific evidence available, or when their requirement is not reasonable applicable to the product, as determined by the FDA. Hence, it is not a manufacturer's option to decide whether or not to undertake an investigation showing the product effectiveness. The time, cost and resources required for that investigation must be considered to judge the product viability.

3.2.2 Health care technology assessment

The evaluation of safety and effectiveness prescribed by the FDA are only two aspects of technology assessment. Other relevant aspects are the cost, cost-effectiveness, cost-benefit and legal, ethical and social implications, both in absolute terms and in terms of alternative existing technologies. Some of these factors are not easily amenable to quantitative analysis, which may explain the disagreement in the interpretation of the available data. The biomedical engineer

must be aware of these aspects as the ultimate success of a product will be judged inside this broad framework, aiming to evaluate the overall impact of medical technology on society.

From the point of view of market evaluation by the designer, it is important to know the relative use of different medical procedures. This information is published under the names practice guidelines, practice standards, practice recommendations, practice options and clinical indicators. Their aim is to assist the health care practitioner to decide which diagnostic, therapeutic, or other clinical procedures are suitable under specific clinical circumstances. In Scannell et al. (1992) there are 533 classified references on practice guidelines spanning the years 1985 to 1992.

3.2.3 Effectiveness, Efficacy and Efficiency

In medical terms, effectiveness is the extent to which a drug or other agent or number of treatments achieves its intended purpose. Efficacy is the effectiveness of a drug or treatment or other intervention under ideal conditions. In health service planning, efficiency is the extent to which resources are converted into service.

In general engineering terms, an efficient product avoids loss or waste of energy. In biomedical products to be implanted or worn by the patient, energy efficiency is a coveted quality. For those products applying any form of energy to living tissues to elucidate a response or seeking a particular effect, efficiency means a minimal damage to nearby tissues.

In general terms, effectiveness requires that a biomedical product also have minimal side and long term effects. The relatively short periods for product testing do not often allow us to definitely determine those effects. However, it is a designer's responsibility to devise the product to reduce not only those undesirable effects that can be easily detected during testing, but also any effect whose existence may be suspected from whatever information is available to the engineer because of his or her involvement with the product.

3.3 SAFETY

Medical products must be safe for the patient, health care providers, people involved in their design, manufacture and distribution, and for the environment, including other products. However, if "safe" is understood as "free from risk or danger," we promptly realize that nothing is really safe, and neither are medical products.

According to (FDA, 2003), which describes medical device classification procedures, "there is reasonable assurance that a device is safe when it can be determined, based upon valid scientific evidence, that the probable benefits to health from use of the device for its intended uses and conditions of use, when accompanied by adequate directions and warnings against unsafe use, outweigh any probable risks." Therefore, "safe" is understood here as posing an acceptable risk, risk being a measure of the probability and severity of harm. Medical products must be so designed that they do not pose any unreasonable risk of illness or injury when used as prescribed.

Product safety must be considered from the onset of product design, rather than as a feature to add once the product has been designed to achieve the desired functionality. The reference to valid scientific evidence in FDA regulations implies that product safety must be

evaluated and documented. Safety testing may require nonclinical investigations including in vitro studies (Section 8.5), investigations using laboratory animals (Section 9.3), and investigations involving human subjects (Section 9.4).

The willingness to accept a risk is highly dependent on individual attitude and therefore the same situation can be deemed safe or unsafe by different persons. The safety of medical products is of special concern because the patient may be unaware of the risk, unable to react to it, or be especially vulnerable; health care providers and manufacturing workers can endure prolonged exposure to hazards; and some products are used in environments that include lifesupport devices.

Medical products pose hazards because they use different forms of energy, which are able to damage living tissues, may provide a path for infection or may trigger adverse responses from those tissues. This section describes biological hazards, hazards resulting from the interaction of different forms of energy with living tissues and the corresponding methods to reduce risks, summarized in Table 3.1. Section 3.4 discusses the host response to materials contacting living tissues and the degradation of implant materials.

Table 3.1 Safety hazards in medical products and basic methods for risk reduction

Agents	Hazards	Risk factors	Risk reduction
Biological	Infection	Agent nature	Primary containment
		Agent transmission	Secondary
			containment
		Contact nature	Decontamination
Chemical	Physical	Material	Material selection
		characteristics	
	Health	Exposure parameters	Protective equipment Good laboratory
Flectrical	Electric shock	Current nathway	Fauinment grounding
Lieethear	Burns	Current density	Double insulation
	Fire explosion	Current frequency and	Isolation transformers
	i ne, explosion	waveshape	isolution transformers
		Current duration	Overload and
			overcurrent protection
Mechanical	Sharp edges and	Exposed body parts	Machine guarding
	points		
	Weight	Patient mobility	Anchoring
	Fluid pressure	,	Cabinet design
Nonionizing	Electric shock	Distance to the	Limit access
radiation		radiating source	
	Burns	Exposed tissue	Personnel protective devices
	EMI	Radiation intensity	Limit exposure time
		Radiation frequency	Design for EMC
		Exposure time	
Ionizing	Any ionizing radiation	Dose absorbed	Limit exposure
radiation	is a hazard		
		Exposure time	Restrict access to
		Exposure dose rate	Personnel monitoring
		Exposure dose rate	equipment
		Tissues irradiated	Environmental
		1165 des infadiated	monitors and alarms
Software	Physical devices controlled	Intended use	Design review
	Mis/disinformation provided	Novelty	Documentation
	1	Specificity	
		Transparency	
		Capabilities offered	
Thermal	High contact	Time-temperature	Reduce surface
	temperatures	exposure	temperatures

	Hot spots in electrosurgery electrodes	Tissue irrigation	Large contact electrode area
Ultrasound	Thermal damage	Output power	Thermal index < 2
	Cavitation	Exposure duration	Mechanical index <
			0.3
		Tissue irradiated	ALARA principle

The CDRH established the Device Experience Network (DEN) to collect information on medical devices which may have malfunctioned or caused a death or serious injury. The DEN compiles reports received under both the mandatory Medical Device Reporting Program (MDR), and the voluntary FDA MedWatch Program. There were 600,000 reports submitted through August 1, 1996, available from the FDA home page at: www.fda.gov/cdrh/mdrfile.html. The FDA uses Safety Alerts to communicate specific problems with medical devices that have resulted in death or serious injury with a high probability of recurrence, and issues Public Health Advisories and Safety Notices to inform about lower-risk cases. They are accessible from: www.fda.gov/medwatch/safety.

The FDA Enforcement Report Index (available at: www.fda.gov/opacom/Enforce.html) lists all product recalls, often because of safety-related problems.

3.3.1 Biological safety

Biological hazards

Biological hazard or biohazard means those infectious agents presenting a risk of death, injury or illness. Those agents can come into direct or indirect contact with the patient or health care providers, including laboratory personnel performing diagnostics or other screening procedures on potentially infectious materials, such as human (or animal) fluids and tissues. The Hospital Infection Program of the National Center for Infectious Diseases (NCID) has published guidelines for the prevention and control of nosocomial infections, available from the Centers for Disease Control and Prevention (CDC) home page at www.cdc.gov/ncidod/hip/hip.htm. 29CFR1910.1030 regulates the occupational exposure to bloodborne pathogens. The National Academy of Sciences (1989) has published guidelines for the safe handling of infectious agents, the safe disposal of infectious laboratory waste and safety management.

Table 3.2 shows that risks from biological hazards depend on the nature, degree, frequency and duration of the contact between the device and the human body. The ISO-10993 standard (ISO, 1997) considers three body contact durations: limited exposure (less than 24 h), prolonged or repeated exposure (24 h to 30 d) and permanent contact (more than 30 d). It also distinguishes between surface, externally communicating, internal and implant devices.

Exposure			Contact nature			
Limited	Prolonged	Permanent	Surface devices	External communic ating devices	Internal devices	Implants
< 24 h	24 h to 30 d	> 30 d	External surfaces Mucous membranes Breached surfaces	Indirect contact with blood path Contact with tissue, bone or dentin	Contact circulating blood	Contacting blood Contacting bone or tissue

Table 3.2 Biological hazards depend on the duration (exposure) and nature of the contact.

Surface devices are categorized depending on whether they contact intact external surfaces only (e.g. electrodes, monitors, external prostheses), mucous membranes (e.g. contact lenses, anesthesia breathing circuits, endotracheal tubes, flexible fiber optic endoscopes, intravaginal and intraintestinal devices, respiratory therapy equipment and urinary catheters) or breached surfaces (e.g. occlusive patches and ulcer and burn tissue dressings). Health care providers face the risks of percutaneous and mucous membrane exposures to clinical material.

Externally communicating devices make an indirect contact with the blood path, or contact with tissue, bone or dentin. Indirect contact with the blood path means a contact at one point, for example to infuse fluid into the vascular system (e.g. blood administration sets and solution administration sets). Dental cements or fillings also belong to this category.

Internal devices are those that contact circulating blood, for example dialyzers, dialysis tubes and accessories, intravenous catheters and oxygenators. Implant devices are divided into those contacting blood (e.g. heart valves, internal drug delivery catheters, permanent pacemaker electrodes) and those contacting bone or tissue (e.g. bone prostheses and cements, replacement joints, pacemakers, implanted drug pumps, neuromuscular stimulators, breast implants and intrauterine devices).

Principles of biosafety

Laboratories handling infectious agents use the term "containment" to describe safe methods to reduce human and environmental exposure to potentially hazardous agents (Richmond and McKinney, 1993). Primary containment aims to protect personnel and the immediate laboratory environment. Primary containment relies on good laboratory practices and techniques (personnel competence, training and awareness, and operations manuals), and on appropriate safety equipment, described as primary barriers (e.g. biological safety cabinets, enclosed containers such as safety centrifugal cups, and personal protective equipment). Secondary containment aims to protect the environment external to the laboratory. Secondary containment relies on facility design, described as secondary barriers (e.g. separation of the laboratory from public access, decontamination and handwashing facilities, ventilation and air treatment systems) and on operational practice.

There are four biosafety levels which consist of combinations of different primary and secondary containment techniques. Each laboratory must rely on practices, safety equipment and facilities appropriate for the biosafety level recommended for the safe handling of the specific agents used in the laboratory. Richmond and McKinney (1993) list recommended biosafety levels depending on the organism considered.

Biosafety level 1 (BSL1) containment techniques are appropriate for laboratories handling microorganisms not known to cause disease in healthy adult humans, such as bacillus subtilis, escherichia coli K12, sacharomyces cerevisiae, and nonprimate cells. BSL1 relies on standard microbiological practices and handwashing.

Biosafety level 2 (BSL2) containment techniques are appropriate for clinical, diagnostic, teaching and other facilities, including those dealing with human blood, body fluids and tissues, which handle agents posing moderate risk, such as staphylococcus sterptococcus, measles, polio, enteric and bloodborne pathogens. BSL2 relies on good microbiological techniques, personal protective equipment such as gowns, gloves and face protectors, splash shields, and even biological safety cabinets or safe centrifuge cups in procedures with aerosol or high splash potential, and handwashing and waste decontamination facilities. See Figure 3.1.



Figure 3.1 A Class I Biological Safety Cabinet has an inward face velocity of 75 linear feet per minute) which provides some containment for laboratory workers and the immediate environment from infectious aerosols generated within the cabinet. HEPA stands for high-efficiency particulate air. Adapted from Richmond and McKinney (1993).
Biosafety Level 3 (BSL3) containment techniques are appropriate for clinical, diagnostic, teaching, research, or production facilities handling agents with a potential for respiratory transmission, and which may cause serious and potentially lethal infection, such as coxiella burnetti, mycobacterium tuberculosis and St. Louis encephalitis and yellow fever viruses. BSL3 requires a biological safety cabinet or other enclosed equipment to perform all laboratory manipulations, and reinforced secondary barriers such as controlled access to the laboratory and a specialized ventilation system that minimizes the release of infectious aerosols from the laboratory.

Biosafety level 4 (BSL4) containment techniques are applicable when handling dangerous and exotic agents which pose a high individual risk of life -threatening disease, such as Ebola, Marburg and Congo-Crimean hemorrhagic fever viruses. BSL4 relies on stringent primary and secondary barriers (separated or isolated buildings with special ventilation and waste management facilities).

Decontamination: disinfection and sterilization

Decontamination means the use of physical or chemical methods to remove, inactivate, or destroy microorganisms on a surface or item to the point where they are no longer capable of transmitting infectious particles and the surface or item is rendered safe for handling, use, or disposal. Medical products must withstand decontamination procedures used to reduce risks from biological hazards. The frequency and method of decontamination depend on the product materials and risks. The Hospital Infection Program of the NCID provides guidelines for the sterilization and disinfection of different groups of products. They are available at www.cdc.gov/ncidod/diseases/hip/sterilgp.htm.

Disinfection is a chemical decontamination process that eliminates virtually all recognized pathogenic organisms from inert surfaces, but not necessarily all forms of microbial life on inanimate objects. Depending on the kind of microorganisms eliminated, there are three levels of disinfection: high, intermediate, and low. High-level disinfection kills all organisms, except high levels of bacterial spores, using a chemical germicide cleared for marketing as a sterilant by the FDA (list available at www.fda.gov/cdrh/ode/germlab.html). Intermediate-level disinfection kills mycobacteria, most viruses, and bacteria with a chemical germicide registered as a "tuberculocide" by the Environmental Protection Agency (EPA). Low-level disinfection kills some viruses and bacteria with a chemical germicide registered as a hospital disinfectant by the EPA. Reusable devices that contact mucous membranes should be at least disinfected before each use. In the web site of the American Institute of Ultrasound in Medicine (AIUM) at www.aium.org/stmts.htm there are specific recommendations for the cleaning and preparing of endocavitary ultrasound transducers between patients. Reusable devices that do not touch the patient or contact only intact skin (e.g. blood pressure cuffs, crutches) must be disinfected when used in patients infected with drug-resistant or highly virulent microorganisms. Otherwise, they need to be disinfected only if grossly soiled with blood or other body fluids.

Sterilization is a decontamination process that destroys all microbial life, including highly resistant bacterial endospores. Some medical devices must comply with sterility controls established in the respective section of 21CFR880 (General hospital and personal use devices).

Sterilization agents can be physical or chemical (Bruch, 1990). The most common sterilization agent in hospitals is moist heat by steam autoclaving. Those devices that are too

large to fit into an autoclave or unable to resist their relatively high temperature (121 °C) or pressure, are sterilized by ethylene oxide (EtO), even though it is neurotoxic and carcinogen. Small, heat-resistant devices are sterilized by dry heat in an oven, at 140 °C. Other physical sterilization agents are gamma and electron beam radiation, commonly used for surgical sutures and drapes, syringes and knee and hip prostheses, and intense pulsed light. Other chemical sterilization agents are gaseous hydrogen peroxide, peracetic acid, ozone and chlorine dioxide.

The efficacy of a sterilization process is assessed by determining the sterility assurance level (SAL), defined as the probability that a given device will remain nonsterile after being subjected to a given sterilization process. The number of viable microorganisms on the product before sterilization is termed bioburden. The ST series ANSI/AAMI standards describe different sterilization techniques, their application and methods to determine their SAL. The SAL required for the devices which contact normally sterile areas of the body is 10^{-6} .

3.3.2 Chemical safety

Chemical hazards

Chemicals used in medical products or in their manufacture pose physical and health hazards to the patient, health care providers, manufacturing workers and the environment. Occupational chemical hazards are subjected to the Federal Hazard Communication Standard (HCS) (29CFR1910.1200), which also defines different hazards. Toxic and hazardous substances are covered either by reference (29CFR1910, sections 1000 to 1047) or by definition. The HCS does not apply to medical products themselves. The American Conference of Government Industrial Hygienists (ACGIH) (www.acgih.org) publishes the "Threshold limit values for chemical substances and physical agents and biological exposure indices," which is a common reference for establishing guidelines and work practices. The Toxic Substances Control Act of 1975 authorizes the EPA to issue rules concerning any chemical substance or mixture that "may present an unreasonable risk of injury to health or the environment."

Physical hazards are "chemicals for which there is scientifically valid evidence that it is a combustible liquid, a compressed gas, explosive, flammable, an organic peroxide, an oxidizer, pyrophoric, unstable (reactive) or water-reactive." The HCS defines the exact meaning of these terms.

Health hazard means a chemical for which there is statistically significant evidence based on at least one study conducted in accordance with established scientific principles that acute or chronic effects may occur in exposed normal living tissues. Health hazards include carcinogens, toxic or highly toxic agents, genotoxins, irritants (causing local inflammation), corrosives (destroy tissue at the site of contact), sensitizers (induce allergic response), hepatotoxins, nephrotoxins, neurotoxins, agents which act on the hematopoietic system, and agents which damage the lungs, skin, eye, or mucous membranes. Stine and Brown (1996) describe toxic effects in different physiological systems.

Some caustic agents and mineral acids can damage tissues directly. Heavy metals, poisons and venoms disrupt important enzymatic reactions. Chemicals can destroy cell membranes, decrease the intracellular pH, activate lysosomal enzymes which damage cell structures, cause oxygen deficiency leading to insufficient ATP production and consequent loss of capacity of the sodium–potassium pump, which results in water and ion transfer into the cell, and eventually cell death. Gases such as carbon monoxide, for which hemoglobin has higher

affinity than for oxygen, nitrogen oxides, sulphur dioxide, formaldehyde, chlorine, and others, also damage cells. Ozone irritates mucous membranes and its inhalation can result in pulmonary edema. 21CFR801.415 considers adulterated or misbranded any medical device that generates more than 50 nL of ozone per liter of air circulating through the device, either intentionally or incidentally.

The Registry of Toxic Effects of Chemical Substances (RTECS) is the toxicological database developed by the National Institutes for Occupational Safety and Health (NIOSH), and lists more than 130,000 substances alphabetically, with numerical toxicity endpoints in six fields of toxicity data. The Hazardous chemicals data base at http://ull.chemistry.uakron.edu/erd/ contains information for several hazardous chemicals based on a keyword search. Table 3.3 shows the National Fire Protection Association (NFPA) has developed a system for indicating the health, flammability and reactivity hazards of chemicals. Figure 3.2 shows the NFPA four color system.

Table 3.3 NFPA rating summary for chemical hazards. www.orcbs.msu.edu/chemical/nfpa/nfpa.html

Health	4	Danger	May be fatal on short exposure. Specialized protection
(Blue)			equipment required
	3	Warnin	Corrosive or toxic. Avoid skin contact or inhalation
		g	
	2	Warnin	May be harmful if inhaled or absorbed
		g	
	1	Caution	May be irritating
	0		No unusual hazard
Flammability	4	Danger	Flammable gas or extremely flammable liquid
(Red)	3	Warnin	Flammable liquid flash point below 37.8 °C (100 °F)
		g	• • • • • • • • • • • • • • • • • • •
	2	Warnin	Combustible liquid flash point of 37.8 °C to 93 °C
		g	(100 °F to 200 °F)
	1	Caution	Combustible if heated
	0		Not combustible
Reactivity	4	Danger	Explosive material at room temperature
(Yellow)	3	Warnin	May be explosive if shocked, heated under
		g	confinement or mixed with water
	2	Warnin	Unstable or may react violently if mixed with water
		g	5
	1	Caution	May react if heated or mixed with water but not
			violently
	0	Stable	Not reactive when mixed with water
Special notice	W	Water reactive	
key (White)	Oxy	Oxidizing agent	



Figure 3.2 The NFPA system to indicate health, flammability and reactivity hazards of chemicals uses four colors and a rating from 0 (no hazard) to 4 (danger) for each hazard (Table 3.x+1).

Chemical hazards from medical devices depend on the chemical characteristics of device materials and on the nature, degree, frequency and duration of the body exposure. Product categorization according to the nature of the exposure is the same described for biological hazards. Section 3.4 discusses material–tissue interactions.

Biological evaluation of medical devices

Medical devices must be evaluated to determine the potential toxicity resulting from contact of the device component materials with the body, either directly or through the release of the material constituents. The device materials should not produce adverse local or systemic effects, be carcinogenic or produce adverse reproductive and developmental effects. The biological evaluation of medical devices is currently governed in the United States by FDA blue book memorandum G95-1, a modification of ISO 10993-1, which is also accepted by the EU and other industrialized countries. Section 9.2.4 describes biological evaluation tests.

3.3.3 Electrical safety

Physiological effects of electrical currents

The human body contains ionic media and electrically-excitable cells and tissues. In addition, some cells communicate by electric signals. As a result, the flow of electric charge across the human body is a hazard. The risk depends on the current pathway, intensity, frequency and duration, and on body weight. IEC (1984) reports the effects of dc and ac currents from 15 Hz to 100 Hz on the human body. IEC (1987) reports the effects of ac currents above 100 Hz and currents with special waveforms and single impulse currents. Reilly (1992) details the mechanisms involved.

Electric currents are perceived when they excite nerve endings in the skin. The minimal current intensity that can be perceived is termed *threshold of perception*, and depends on

individuals, waveshape, duration, repetition pattern and polarity. This threshold is about 0.5 mA at 60 Hz, and 2 mA to 10 mA at dc, for intact, dry skin when grasping a large electrode with one hand and the return electrode contacts a foot or both feet.

Higher currents stimulate nerves and muscles, leading to pain and fatigue. The involuntary and unexpected contraction of muscles may cause secondary accidents (contusions, collisions, object droppings). If the current results from a hand grasping an electric conductor, both forearm and finger flexor and extensor muscles are stimulated, but forearm flexors are usually stronger, which keeps the hand closed, thus prolonging the contact. The intensity level preventing the hand from opening voluntarily is termed *let-go current*, and for 60 Hz current is about 10 mA.

Still higher currents can involuntarily contract respiratory muscles, leading to asphyxiation if the situation lasts. Nerve stimulation is painful, and strong muscle contractions fatigue the subject. Currents above 10 A cause burns at entry points and internal body burns because of resistive heating, and if prolonged, violent and lasting muscle contractions (tetanus) able to break ligaments and bones, internal hemorrhages and destruction of tissues, nerves, and muscles, and eventually death.

When the current path includes the heart (hand to hand, hand to leg, front to back), there are lethal effects even for relatively low intensities, because of the high susceptibility of the heart. If the current stimulates part of the myocardium, the affected tissue will go into the refractory period after repolarization, hence not reacting if the normal depolarization wave reaches it; the stimulus may propagate to nearby zones, thus disrupting the normal propagation of electrical activity in the myocardium. A large enough disruption affecting the ventricles leads to a high heart rate, inefficient blood pumping and death within minutes. This situation is termed *ventricular fibrillation*, and does not revert when removing the current that triggered it. The threshold for ventricular fibrillation is from 75 mA to 400 mA, increasing for large body weight because a large volume means a reduced current density.

If the current density through the heart is large enough, the myocardium contracts altogether and, when the current is stopped, the normal heart beat may resume. This is the working principle for defibrillators. The current required ranges from 10 A to 60 A (larger for large body weight, to yield enough current density). It turns out, therefore, that a large current can revert a dangerous situation resulting from a smaller current.

Electric currents and electrostatic discharges in medical equipment pose fire and explosion dangers in atmospheres that contain flammable gases or vapors, such as anesthesia locations. Prolonged dc currents can cause chemical burns resulting from the electrolytic decomposition of sweat or electrode jellies, which yields free ions.

Electric shock

Electric currents through the human body constitute an *electric shock*. Section 14.2 in (Webster, 1998) describes different parameters concerning the severity of electric shocks. One relevant parameter is the point where the current enters the body. Externally applied currents constitute a *macroshock* while currents applied to the heart constitute a *microshock*. The hazardous currents levels for microshock are quite low: 80 μ A to 600 μ A can cause fibrillation. Hence, patients with conductive pathways to the heart (electrodes, catheters) are especially vulnerable.

Electric shock occurs when the body becomes a part of an electric circuit. The current depends on the circuit voltage and impedance, Figure 3.3(a). The effect depends on the current

density and tissues affected. The electric impedance of the human body depends largely on skin integrity and moisture, because the outermost skin layer, made up of dead cells, is an electrical insulator. The impedance of 1 cm² of dry, intact skin, ranges from 15 k Ω to 1 M Ω , depending on the part of the body and the moisture and sweat present. The impedance of wet or broken skin can be only 1 % of the value for dry skin. On the other hand, the internal impedance between any two limbs is only about 500 Ω . Hence, any medical procedure reducing or bypassing the impedance of the skin poses a higher risk of electric shock. Skin trauma also increases the risk when initially intact skin is broken, leading to an electric contact with internal tissues.



Figure 3.3 (a) Patients contacting a conductor close to a voltage source may suffer an electric shock. (b) The electric shock circuit involves the patient impedance Z_p connected to a voltage source through source-to-patient impedance Z_{sp} and patient-to-ground impedance Z_{pg} . Z_p is usually resistive but Z_{sp} and Z_{pg} are often capacitive.

Protection from electric hazards

Figure 3.4 shows three basic ways of preventing electric shock: (1) reduce the source voltage, or use batteries, (2) increase equipment insulation (e.g. using isolation amplifiers, isolation transformers, double insulation and nonconductive enclosures) and (3) connect to ground all accessible metal parts in order to drain any leakage or fault current away from the patient (green or green/yellow grounding wire). Electric equipment using these protective measures are classified, respectively, as class III, class II and class I in IEC 60601-1 (IEC, 1988) and UL 2601-1 standards (not to be confused with medical device classifications in the FDA and EU regulations). Figure 3.5 shows symbols for protective earth ground and Class II equipment. The 601-2-xx IEC standards establish additional safety requirements for several medical devices, including testing conditions. The design of biomedical equipment must consider compliance with electrical safety standards (Section 7.4) and electrical safety testing (Section 8.5).



Figure 3.4 Basic protections from electric shock rely on (*a*) low supply voltage (v_s^*), increased insulation (Z_{sp}^*) or (*b*) equipment grounding through R_g .





The degree of protection achieved by medical electric equipment is evaluated by their leakage currents under normal and specified fault conditions. The acceptable limits for microshock protection are somewhat controversial (Webster, 1998, S14.6). The ANSI/AAMI ES1-1993 "Safe Current Limits for Electromedical Apparatus" standard, accepts patient-lead leakage currents of 10 μ A in normal conditions and 50 μ A or 100 μ A in single-fault conditions (e.g. open ground conductor), respectively for isolated and nonisolated patient leads. These currents are deemed unsafe by some authors. The IEC 60601-1 standard classifies applied parts from medical equipment into types B, BF (body floating) and CF (cardiac floating) according to the leakage current accepted. Figure 3.6 shows symbols for type B, BF and CF medical equipment. Table 3.4 shows some leakage currents accepted below 1 kHz For currents with frequency 1 kHz < *f* < 1 MHz, the limits are multiplied by *f*/(1 kHz).



Figure 3.6 Symbols for (*a*) Type B, (*b*) Type BF and (*c*) Type CF applied parts of medical electrical equipment.

The installation itself can offer additional patient and user protection from electric shock. Fuses, circuit breakers and ground-fault interrupters shut off the current in the event of a ground fault, overload or overcurrent (short circuit, for example). However, those devices are too slow and tolerate excessive currents for direct microshock protection. Besides, power interruption is unacceptable in many patient areas. Hence, the National Electrical Code (NEC 70-1996) includes provisions such as the use of "Hospital Grade" receptacles (Article 517-13), a maximal potential difference between two exposed conductive surfaces of 500 mV in general-care areas or 40 mV in critical-care areas (Article 517-15), and use of a patient-equipment grounding point (Article 517-19). Operating rooms and other locations where there are flammable anesthetics use isolation transformers and line-isolation monitors.

3.3.4 Mechanical safety

Mechanical trauma

Mechanical trauma is a bodily injury from mechanical stresses. The injury can range from skin bruise, abrasion, scratch, cut and wound to muscle sprain and bone dislocation or fracture. Head, limbs and the eyes are particularly vulnerable. Hearing organs are vulnerable to excessive noise. Sound pressures above 200 Pa (140 dB above the threshold of hearing, 20 μ Pa) are painful. Continuous exposure to noise reduces hearing and speech understanding ability. Figure 3.7 shows that the frequency response for hearing changes with loudness.



Figure 3.7. Normal thresholds of hearing and pain for sound pressure as a function of frequency (Standard, 1996).

Mechanical hazards

Medical devices providing essentially a mechanical function, such as patient support or transfer, mobility aid or restriction, or other mechanical assistance, are an obvious hazard source in case of malfunctioning, deterioration or abuse. Excessive pressure can lead to inflammation, tissue death and pressure sores. Stroking the skin can result in inflammation. Abrasion of implanted devices can release small particles into the body. Any points or other protrusions or sharp edges in equipment enclosures are a hazard. Equipment with unguarded moving parts poses a risk of entanglement, trapping or impact. Heavy equipment is a hazard because it can become unstable. Equipment using fluid pressure is another hazard. Other mechanical hazards arise from poor ergonomic design (Section 7.2), which may expose humans to cumulative trauma disorders, excessive vibration or noise, and from unsuspected factors such as undesired adherence.

The noise exposure factors are: sound intensity, frequency spectrum and exposure duration and distribution (Standard, 1996). Loudness depends on sound pressure and frequency. To account for the increased sensitivity of the human ear to the higher frequencies, sound level measuring instruments have a frequency-dependent sensitivity.

Protection from mechanical hazards

Safety from mechanical hazards must be achieved by design whenever possible. Dangerous parts should be enclosed or at least guarded, for example by barrier guards or electronic safety devices. The guard itself should not pose a high risk. Internal rotating parts can be guarded by an

enclosure which is interlocked with the driver, so that the part cannot rotate unless the guard enclosure is in position. 29CFR1910.212, describes general requirements for machinery safety and 20CFR1910.243 describes guarding of portable power tools. These regulations do not directly apply to equipment that must deliberately contact the patient. However, some safety provisions can also be applied to medical equipment, e.g. anchoring fixed machinery to prevent it from moving and designing tool controls in a way to prevent the unexpected operation of the tool.

The mechanical design of medical electric equipment cabinets affects their vulnerability to liquid spills, which is particularly important in wet hospital areas because those liquids can reduce electrical insulation. Equipment can be drop-proof, splash-proof or water-proof, depending on its capability to reject spilled liquids.

Acoustic noise can be reduced at source, by changing the sound path or at receiver. Noise can be reduced at source by decreasing the energy driving vibrating parts, changing mechanical couplings to acoustical radiating systems and changing structures. Noise reduction along the path can be achieved by enclosing the noise source, increasing the distance between the source and the receiver, and by covering ceiling, walls and floor with acoustic absorbents. The noise received can be reduced by using enclosures that isolate humans from the sound and personal protectors (ear plugs, earmuffs).

3.3.5 Nonionizing radiation safety

Electromagnetic energy is used in medical products for a variety of specific diagnostic, surgical and therapeutic purposes, and also for power supply, communication and other common functions. The electromagnetic spectrum is arbitrarily divided into frequency bands, which receive particular names. Gamma and x rays, optical radiation (ultraviolet—UV—, visible light and infrared—IR—) and radio-frequency (RF) waves, including microwaves, propagate as an electromagnetic field. Below 3 kHz the electric and magnetic field components are considered separately because the wavelength is too long (100 km) to constitute a radiated field. The energy of electromagnetic radiation with frequency *f* is $E = hf = hc/\lambda$ where $h = 6.62 \times 10^{-34}$ J/s is Planck's constant, *c* is the speed of light (300 Mm/s in free space) and λ is the radiation wavelength.

The energy absorbed by living tissues exposed to nonionizing radiation depends on its frequency and intensity (or amplitude), and on the duration of the exposure. The energy required to remove one electron from a particular chemical element is about 1.6 aJ (10 eV). Hence, electromagnetic radiation below the optical region ($\lambda > 160$ nm), does not have enough energy to ionize matter. Nevertheless, that radiation can excite atoms by raising electrons in the outer shells to higher orbitals. If the heat is dissipated, the effects cease when the radiation is interrupted. However, a high thermal stress can yield persistent injuries, such as erythema, cataracts or burns. Short-wave UV radiation can ionize matter, thus inducing photochemical reactions.

Electromagnetic interference (EMI) to medical equipment is a hazard, particularly for implanted and life-support devices. Medical devices must be tested for electromagnetic compatibility (EMC) before being distributed, which implies that they must be designed to achieve EMC (Section 7.4).

Infrared radiation safety

The permissible exposure to infrared radiation (770 nm $< \lambda < 3 \mu m$) is 10 mW/cm² for extended periods (> 1000 s) and $1.8t^{-3/4}$ (W/cm²) for t < 1000 s. For a near-infrared source (770 nm $< \lambda < 1400$ nm) producing a weak visual stimulus, the maximal spectral radiance for periods exceeding 10 s should be smaller than (0.6 W)/(α cm² sr nm), where α is the viewing angle.

Ultraviolet radiation safety

Ultraviolet (UV) light (40 nm $< \lambda < 400$ nm) is absorbed by thin layers of tissue and can seriously damage the skin and the eyes. Hence, they should not be directly exposed to direct or reflected UV radiation. Overexposure of the eyes results in inflammation of the conjuntiva, cornea and iris, and long term injury. Overexposure of the skin results in erythema, and can lead to keratosis, allergic reactions, premature skin aging and skin cancer. Adequate goggles, gloves, caps and other garments protect eyes and skin. Short-wave UV radiation able to trigger photochemical reactions, produces significant biological effects.

Laser safety

Lasers are devices that produce a highly collimated beam of optical radiation. Current usage, however, applies the term laser also to the radiation itself. As opposed to an isotropic radiation source that emits uniformly in all directions surrounding it, a source of collimated radiation emits in a highly directional pattern. The radiation intensity decreases as the square of the distance from an isotropic source, but more gradually for a collimated source. As a result, the energy of laser beams can be very intense, and because it is absorbed by thin tissue layers, the resulting heating can quickly destroy tissues, particularly the lens of the eye. Extreme heating can yield a hot plasma (ionized gas), leading to a shock wave when the gas expands, which has found therapeutic uses. Short-wavelength lasers (100 nm to 315 nm) can induce photochemical reactions.

Each class of laser must fulfill specific designation and warning requirements, such as the use of laser protective eyewear (classes IIIa and higher) or the need to avoid direct skin or eye exposure. Risks are reduced by controlling the access to areas during laser operation and by removing highly reflective surfaces from those areas. Lasers also pose electric shock risks because they operate at a high voltage. Lasers working above 15 kV can emit x rays.

3.3.6 Ionizing radiation safety

Ionizing radiation consists of subatomic particles or photons (electromagnetic radiation) that have sufficient energy to remove electrons from atoms when interacting with matter. Subatomic particles, usually alpha and beta particles, come from the nuclei of unstable atoms, termed radionuclides. High-energy photons come from either the nuclei of radionuclides, gamma rays, or from the electron shells of atoms, x rays. Ionizing radiation is used in several medical procedures, both for diagnostic and therapy, and also to sterilize medical devices.

Types of ionizing radiation

Alpha radiation is made up of high-energy alpha particles, which are helium nuclei, hence particles with positive charge. Because of both their charge and their mass, alpha particles rapidly lose energy when passing through matter, thus inflicting severe damage to the irradiated tissue in spite of their shallow penetration. Since alpha particles are completely absorbed by the epidermis, external alpha radiation is not a hazard. However, ingested or inhaled alpha-emitters are very harmful because they dissipate all their energy in the surrounding living tissue. Alpha emitters are typically used in smoke detectors.

Beta particles are positrons (positive electrons) or electrons coming from the nucleus. (Nuclei do not have electrons: they result from nuclear reactions which yield protons and electrons from neutrons.) Beta particles are lighter than alpha particles, which results in deeper penetrations in matter, but less damaging over equally traveled distances. High-energy beta particles can penetrate the skin. Inhaled or ingested beta emitters are harmful because their emissions will be absorbed by the surrounding tissue. Beta emitters are used for radiation therapy and imaging.

Gamma rays (high-energy photons) have neither mass nor charge, hence they are not attracted by either nuclei or electrons, which allows them to deeply penetrate matter. Gamma rays came from the nucleus, sometimes accompanying alpha or beta particles. Gamma rays can pass through the human body or be absorbed by tissue, thus constituting a radiation hazard for the entire body. Gamma emitters are used to evaluate the heart function by injecting a radionuclide into the blood stream, to trace labeled chemicals compounds, to kill cancerous cells, tumors, bacteria and germs, and in radioimmunoassays.

X rays are also high-energy photons but they come from the electron shell, not from the nucleus. X rays are emitted when electrons previously excited to high-energy levels return to their stable levels and release the excess energy as photons. X rays have the same properties than gamma rays, but usually have less energy, hence they are less penetrating.

Other ionizing radiations are made up of neutrons, protons and electron beams from linear accelerators or cyclotrons. They have less penetration power than photons, but the charge of protons and electrons, and the radiation following the interaction of neutrons with nuclei, makes them quite dangerous.

Protection against ionizing radiation

Ionizing radiation always produces biological effects, even though not necessarily healththreatening. Hence, the basic protection method is to avoid exposure whenever possible and this can be achieved in part by restricting access to radiation areas. Environmental radiation resulting from medical, academic and industrial activities is regulated in the U.S. by the NRC, the EPA and some States. The FDA regulates the manufacture and use of linear accelerators and the States regulate their operation.

The next step in radiation protection is to limit exposure around radiation sources. Performance standards for ionizing-radiation-emitting products set, for example, exposure rates allowed at a given distance from the source, leakage radiation limits, characteristics of beamlimiting devices and enclosures, test conditions and labeling requirements. Occupational exposure is controlled by personnel monitoring equipment, such as film badges, pocket chambers, pocket dosimeters, or film rings. Environmental monitors and alarms warn when radiation levels exceed a given threshold.

3.3.7 Software safety

Software devices, components and accessories

Software which meets the definition in Section 201(h) of the Food and Drug Administration Act is a medical device (Section 1.3), and hence it is subject to applicable FDA medical device regulations, requiring the proving of the safety and efficacy of the product. Software intended only for educational purposes or for storage, retrieval and dissemination of medical information, and general purpose software, e.g. for accounting, communication and word processing, is not regulated by the FDA.

Software can be a standalone device, or it can be a component or part incorporated into another device, or distributed separately for use (as an accessory to) with another device. Most of the software subjected to FDA medical device regulations is incorporated into another medical device as a software component. Software accessories can have a reduced inherent risk as compared to the parent device, or, quite to the contrary, may introduce new capabilities for it and increase the risk. Standalone medical software devices are programs whose input is medically related data and that output the results of its function to health care personnel, for example, software which analyzes potential therapeutic interventions for a particular patient and hospital information systems.

3.3.8 Thermal safety

Thermal trauma

High temperature increases the permeability of cell membranes. Tissue damage begins at temperatures above 45 °C, mainly because of the coagulation of blood in vessels and protein denaturation. Below 30 °C, the skin loses heat faster than it can be restored metabolically. Thermal safety in medical products concerns mostly high temperatures and skin damage. The epidermis is thin and ischemic, so that heat diffuses from the surface to the dermis without temperature reduction by blood convection. When the temperature increases, the stratum corneum, made up of desiccated cells, chars or blisters. The subsequent epidermis layers, with live cells but no blood flow, desiccate. In the dermis, well irrigated by microvessels, there is hemorrhage and thrombosis. The burn region can extend into deeper dermis layers and underlying muscle tissue, affecting any peripheral nerves and blood vessels in the area. Neuron sensitivity to temperature is not uniform; certain neurons undergo irreversible changes at 45 °C; others at 50 °C.

Thermal injury depends not only on the total energy delivered to the tissue but also on heat flux rate. The time-temperature relationship to produce a graded degree of thermal injury is exponential, $t = t_0 \exp(T_0 - T)$, where the references t_0 and T_0 reflect the relative intensity of the injury (Diller, 1985). Skin will burn at 45 °C after several minutes, but in a few seconds at 60 °C.

Thermal trauma can result from heated parts, surfaces or substances and also from electric currents, either by direct contact or induced by electromagnetic fields, and from ultrasound. Thermal design of medical electrical equipment must ensure that there are not any hot spots on the cabinet or accessories (e.g. ultrasound transducers and applicators). IEC 60601-1 (IEC, 1988) limits the temperature of parts of medical devices (not intended for heating) to 41 °C. Since skin temperatures range from 31.5 °C to 35 °C, contact temperatures above 35 °C can induce erratic movements because of the contact reaction. Electrosurgical units can produce burns at the dispersive electrode if the contact impedance is high. Ensuring a large contact area reduces this risk.

3.3.8 Ultrasound safety

Interaction between ultrasound and living tissues

Ultrasound is a mechanical radiation with f > 16 kHz, which displaces and accelerates particles when propagating in a medium, thus inducing forces and stresses, and heating the medium. As ultrasound propagates in tissue, the radiation amplitude decreases, mostly because the acoustic energy associated with the radiation heats the tissue at a rate that depends on local acoustic properties and ultrasound field characteristics. Heat generation by sound is called acoustic absorption. Although diagnostic ultrasound transducers rely on transmitted power as low as 10 mW, the power per unit area, or intensity, can be large enough to yield appreciable localized heating. Typical diagnostic ultrasound equipment uses duty factors lower than 1 %, which means that the time-average intensity is very low as compared to the pulse-average intensity. The intensity is not uniform across the radiation beam.

3.4 BIOCOMPATIBILITY

Biocompatibility refers to the biological and chemical interactions of medical devices with the patient's body. The regulation of medical devices for biocompatibility in the United States become rigorous with the Safe Medical Devices Act of 1990 and the Medical Device Amendments of 1992. For example, 21CFR814.20 (Premarket approval) requires a section containing results from nonclinical laboratory studies including (among others) biocompatibility tests. The EU directives 90/385/EEC and 93/42/EEC, which concern medical devices, also include biocompatibility as an "essential requirement". Safety evaluation must consider the risks posed by device materials, either directly or from the release of any of its constituents or residuals in them. A device is deemed biocompatible when the benefits of its use outweigh any deleterious response. Bio-incompatible materials do not deserve consideration when designing those parts of medical devices contacting the human body.

3.4.1 Host reaction to biomaterials

A biomaterial is "any substance other than a drug, or a combination of substances, synthetic or natural in origin which can be used for any period of time as a whole or a part of the system which treats, augments, or replaces any tissue, organ or function of the body." Since biomaterials come in direct contact with living tissues, they must be biocompatible.

Biomaterials can be either biopassive or bioactive. Biopassive, or inert, materials are not intended to interact with the host's biological system, for example, metal alloys used in joint replacements. Bioactive materials aim to elicit a specific beneficial host response, for example, suture materials (polymers) which degrade in the body yielding products which are biologically eliminated.

Nevertheless, all biomaterials interact in some degree with the surrounding tissues. This interaction is minimal for biomaterials contacting the skin or mucous membranes but intense for implants. The host is affected by and affects biomaterials. Host reactions can be tissue-, organand species-dependent. Table 3.9 summarizes biomaterial–tissue interactions (Anderson et al., 1996). There are short-term effects, such as skin irritation, allergy reactions, skin erythema, edema and necrosis, systemic toxicity, genotoxicity, reproductive and developmental toxicity, thrombogenicity, and pyrogenicity; and long-term effects such as tumorigenesis and material degradation.

Effect of the implant on the host	Effect of the host on the implant
Local (cellular):	Physical-mechanical effects:
Blood-material interaction	Abrasive wear
Toxicity	Fatigue
Modification of normal healing	Stress-corrosion cracking
Infection	Corrosion
Tumorigenesis	Degeneration and dissolution
Systemic and remote:	Biological effects:
Embolization	Absorption of substances from tissues
Hypersensitivity	Enzymatic degradation
Elevation of implant elements in blood	Calcification
Lymphatic particle transport	

Table 3.9 Biomaterial-tissue interactions (Anderson et al., 1996).

Tissue-material interactions

Injury to tissue, either accidental or intentional, such as needed for implant placement, triggers a reaction sequence intended to maintain homeostasis and heal wounds: acute inflammation, chronic inflammation, granulation tissue, foreign body reaction and fibrosis.

The inflammatory reaction is the immediate response activated by vascularized connective tissue adjacent to the injury. Nearby capillaries first constrict to stop blood leakage and then dilate because of the increased activity of the endothelial cells lining them and become more permeable. This permits proteins and cells present in blood but not in extracellular fluids to gain access to extracellular sites and repair damage and fight possible infection. Blood clots and there is risk of embolism. The increased blood flow due to the action of mediators, axon reflex

and increased hydrogen ion concentration, causes erythema. Plasma from the capillaries, migrating leukocytes and fluids from the damaged local lymphatics, all exude into the injured tissue causing an acute inflammation that lasts from minutes to days, depending on the extent of tissue damage and infection, if any.

Blood-material interactions

Blood-contacting materials must neither induce thrombosis or hemolysis, nor damage plasma proteins or enzymes or other formed elements of blood. Blood clots, or thrombus, form by platelet aggregation and stabilize by a fibrin meshwork. Receptors bounded to the external membrane of platelets mediate both platelet–platelet aggregation and platelet–surface adhesion. Platelets adhere to injured blood vessels exposing collagen. Platelets adhere to artificial surfaces by binding to proteins adsorbed on the surface (for example, plasma proteins in an implant). Adhered platelets release factors that recruit additional platelets, enlarging the aggregate. The surface of the aggregate platelets produce thrombin, which generates fibrin that stabilizes the platelet mass. That thrombin production overrides the anticoagulant mechanisms of blood. Thus, while in normal condition coagulation in blood vessels is localized in injured sites, coagulation in artificial surfaces can lead to excessive thrombus formation and exaggerated inflammatory response. In addition, clot formation in blood vessels takes from 7 min to 12 min, but only 5 s to 12 s when contacting a foreign surface.

3.4.2 Body reactions to common biomaterials

Very few materials are perfectly biocompatible. The main classes of biomaterials are: ceramics, glasses, metals, polymers, composites and biological (natural) materials. Some biomaterials have common names because of their extended use in other engineering fields. Nevertheless, the processing techniques and purity to achieve biocompatibility and the performance required are far more demanding than those valid for other applications. Greco (1994) and Silver (1994) provide detailed reviews of biomaterials used in a variety of implants.

3.4.3 Deterioration of biomaterials

The biological environment is highly aggressive as compared to the common external environment. Moderate temperature, neutral pH and homeostasis may suggest quite the contrary. However, the body often exposes implants to stressful mechanical conditions and is readily prepared to "fight" the implant by chemical and biological methods. These factors act synergistically to accelerate the degradation of biomaterials (Coury et al., 1996) which sometimes are expected to last during the patient's life expectancy.

3.4.4 Implant encapsulation and sterilization

Implant encapsulation

Implant encapsulation must consider both material deterioration in the internal body environment and any possible adverse effect on the body. Encapsulation must grant robustness and rigidity, without adding excessive weight or volume, using biocompatible materials. Coatings for electronic implants must preclude moisture or any fluid from reaching the internal circuits. The complete external surface must be covered by anticoagulant, otherwise exposed areas may elicit clotting. Some implants are impregnated by an antibiotic coating to minimize the risk of infection. External contours must be streamlined for effective cleaning and sterilization and to prevent undue forces or pressures deriving from sharp corners or protrusions.

Implant sterilization

Implants must be sterile to prevent infection. Dry sterilization uses temperatures that are too high for most polymers. Steam sterilization relies on high steam pressure that attacks many polymers. Hence, polymers are usually sterilized by exposure to gamma radiation, EtO or formaldehyde. Residuals from chemical sterilization agents may lead to deleterious effects. Excessive sterilization also deteriorates polymers.

Metallic prostheses are sterilized by gamma radiation, which is rapid and effective (Kowalski and Morrisey, 1996). Intraoperative sterilization of metallic devices is performed by autoclaving, which is simple, fast, efficient and leaves no toxic residues. Biomaterials must be tested after sterilization in order to assess any undesired effect resulting from the sterilization process.

3.5 ENVIRONMENTAL IMPACT

The National Environmental Policy Act (NEPA) requires all federal agencies, including the FDA, to assess the environmental impact of their actions which may significantly affect the quality of the human environment. Hence, the FDA regulatory process also considers the environmental impact, including the identification of the parts of the environment that may be affected, the evaluation of pertinent environmental data, and the consideration of alternatives to the action.

The "impacts" or effects considered, either beneficial or detrimental, can be direct, i.e. simultaneous with the action, or indirect, i.e. caused by the action but happening later or in a remote place. The effects can be ecological, aesthetic, historic, cultural, economic, social or health. Some effects can be cumulative.

According to 21CFR25.22(a)(18), a medical device cannot be approved unless the manufacturer (or distributor or dealer) has submitted an adequate Environmental Assessment (EA) in the applicable format in 21CFR25.31, or claims a categorical exclusion. After analyzing the EA, the FDA either issues a "finding of no significant impact" (FONSI), or requires the manufacturer to prepare a full environmental impact statement (EIS).

Therefore, the design of a medical product must also consider its environmental impact. 21CFR25.23 enumerates actions that do not require preparation of an EA, and 21CFR25.24 lists categorical exclusions. For example, 21CFR25.24(e)(7) excludes "Action on an application for an Investigational Device Exemption (IDE) or an authorization to commence a clinical investigation under an approved Product Development Protocol (PDP), if the devices shipped under such notices are intended to be used for clinical studies or research in which waste will be controlled or the amount of waste expected to enter the environment may reasonably be expected

to be nontoxic." A substance is considered to be nontoxic when it does not cause any adverse effect in a test organism species at expected environmental concentrations. Hence, toxicity is not restricted here to human health effects. 40CFR302.4 lists reportable quantities of hazardous pollutants.

3.6 MAINTAINABILITY AND COST OF OPERATION

Maintenance ensures the safety and efficacy of the medical product once it has been distributed. Since maintenance affects (and is affected by) design specifications, it must be considered before engaging in actual design. The need for maintenance results from possible defects in manufacturing, deterioration, abuse, misuse and accident. Preventive maintenance aims to prevent failures. Corrective maintenance, or repair, overcomes failures.

Maintenance is addressed by the FDA's quality system regulation. 21CFR820.100 requires each manufacturer to establish and maintain procedures for implementing corrective and preventive actions. Those procedures must include the identification of quality problems, the investigation of its causes, the actions needed to correct and prevent those problems, the validation of the efficacy of those actions, and the documentation of these activities. 21CFR820.181 requires manufacturers to maintain device master records (DMRs), which shall include, among other records, information on installation, maintenance, and servicing procedures and methods. 21CR820.200 requires that, "where servicing is a specified requirement, each manufacturer shall establish and maintain instructions and procedures for performing and verifying that the servicing meets the specified requirements" and that "each manufacturer shall analyze service reports with appropriate statistical methodology in accordance with [21CFR] 820.100."

3.7 ETHICS IN BIOMEDICAL ENGINEERING DESIGN

Advances in medical technology have deeply affected doctor-patient relationships, decision procedures and responsibilities in health-care delivery, posing new dilemmas that also affect biomedical engineers designing medical products. The general engineering concerns about protecting public health, safety and welfare are exacerbated when the result of the engineering activity is a product intended to be applied to human beings to create a beneficial action. Regulations about safety and efficiency seek to protect patients from any wrongdoing, but they unavoidably fall short when considering the multiple implications of decisions taken during product design. As discussed in Section 1.4, lawful does not imply ethical.

Saha and Saha (1997) review some ethical issues in biomedical engineering, with emphasis on biomedical implants. Whitbeck (1995) points out several categories of medical products with profound implications for decisions and responsibilities: medical information systems, which store patient data and help in problem solving; rehabilitation devices, which greatly improve a patient's life quality and dignity; drug delivery systems, which improve the safety and efficacy for administering medication; teaching devices, which reduce patient suffering and student stress in learning clinical skills; and assessment systems, which improve the benefit–cost ratio in health care delivery. The *IEEE Engineering in Medicine and Biology* magazine started a regular section on issues in ethics in 1991.

Some of the ethical issues posed by medical technology concern biomedical engineering design. Often, it is so difficult to provide clear-cut answers to the questions posed that we are tempted to think that most answers are necessarily relative to the circumstances. However, in some cases this would amount to accepting that human rights are also relative, contrary to the prevalent opinion that they are inherent to human beings. The ultimate disagreement is about which are the human rights and what is a human being. The Universal Declaration of Human Rights adopted and proclaimed in 1948 by the General Assembly of the United Nations (available at www.unhchr.ch) includes:

- 1. The right to life (Art. 3).
- 2. The prohibition of torture or cruel, inhuman or degrading treatment (Art. 5).
- 3. The right to medical care (Art. 25).

The following are some topics concerning ethics that may arise when designing medical products. When anticipating or confronting ethical dilemmas, the biomedical engineer should give thought to reaching a decision in conformance with a code of ethics, either personal or adopted by any appropriate professional organization.

3.7.1 Conflicts of interest

Conflicts of interest arise for instance when decision making and evaluation are not performed by independent individuals or groups: the decision maker may decide according to personal interests, perhaps relying on the acquiescence or connivance of those that must approve the decision. This may seem unlikely when there are governmental bodies involved in the approval process, but no country seems to be free from political scandals related to dispensing favors.

Conflicts of interest may appear at different steps in the design process. Many of them would not arise if decisions were exclusively based on technical grounds, but this is quite difficult. Assume for instance that as a board member of a charity that operates several centers that train children with cerebral palsy, a biomedical engineer identifies the need for a specific rehabilitation device. This engineer works for a company able to design and manufacture that device and, because of his/her relationship with the charity, could recommend its acquisition. But that engineer also believes that a modified competitor's device would probably achieve the functionality required. Is it fair to pursue the development of the new device? If the engineer informs his employer about the alternative solution but the employer decides to pursue his own development, should the engineer either quit the job or resign from the board position?

Technical decisions are sometimes conditioned by company policies. For example, some companies owned by large corporations may specify parts manufactured by sister companies. Under this policy, engineers may get used to selecting parts without evaluating their cost-benefit. In small companies, the product designer can be also responsible for its evaluation, which makes an unbiased decision more difficult. Biomedical engineers involved in developing standards (which often become the basis for regulations) may have to struggle between loyalty to their employer and commitment to work groups when the outcome of the standards development may benefit competitors.

3.7.2 Resource allocation

Medical resources are scarce not only throughout the world but also locally. A simple clinical thermometer, a commodity owned by many families in developed countries, is unaffordable by entire communities in Third World Countries. On the other hand, magnetic resonance imaging is so expensive that only some major hospitals in developed countries can offer it. Biomedical engineers share much responsibility in reducing the cost of medical products. Engineers risk being misled by the appeal of new features and options, without weighing the cost incurred against the benefit provided. Biomedical engineers must consider that the actual performance of medical products is measured only partially by engineering parameters. What actually counts is the benefit for the patient, as measured by years of life expectancy, work absence days saved and survival rates. Moreover, in contrast to simple repair of inanimate devices, the outcome in patient treatment must also consider the quality of life (not to be confused with happiness), in terms of relief from pain, autonomy, mobility and mental capacity (Saha and Saha, 1997).

Because not only individuals but the society itself cannot afford to develop all the feasible or necessary medical products, which products should preferably be developed? Should cosmetic surgery products have lower priority than therapeutic devices? Should a company develop a useful biomaterial for human benefit when this implies diverting resources from some more profitable markets? How much risk must a company assume that tries a different design approach in order to reduce the cost of a product or procedure instead of trying just to obtain a share of the market at current costs? Health care technology assessment (Section 3.2.2) can help to allieviate the pressure when making decisions. Expensive devices should be designed for reuse whenever possible. Design solutions based on proven technologies should prevail against novelty for the sake of prestige or the prospect of larger profits.

Biomedical engineers are not responsible for the use of the devices they design. However, sometimes it is possible to anticipate illegal or immoral use for those devices (e.g. torture), and therefore a biomedical engineer may refuse to cooperate in their development. Other techniques such as genetic screening, which is useful to test for genetic diseases can be also used for questionable reasons, sex selection for instance.

3.7.3 Proper testing

Testing is often wrongly considered as a burden imposed on product development and manufacturing, which requires unproductive activity, which adds cost without benefit. When this is the mood in a company, chances are high that testing will be "simplified" and expedited in order to not delay product delivery. If testing activities do not receive adequate consideration, there may be a shortage of testing manpower and equipment. Engineers testing medical products should not acquiesce in omitting too "expensive" tests that "probably" do not add any benefit. Section 8.5 discusses product testing in accordance with regulations.

Testing can also pose conflicts of interest: a design cannot be considered complete until it has been tested; but if the same person or team performs the design and test, there is the risk that the extra time often expended in design is compensated by a shorter testing time. Excessive self-esteem may also cause the designer to neglect testing. Testing products may reveal defects able to produce failure. If an engineer obtains evidence that a competitor's product can experience the same failure, should the competitor be informed either directly or through the regulatory agencies?

Biomedical engineers can help to reduce the cost of testing, for example by developing well-documented testing protocols that minimize the number of measurements or avoid the need of very expensive equipment. However, they should never try to reduce the cost by eluding any apparently insignificant or redundant test. Products for markets with lax regulations should not be tested differently from products for more demanding markets.

3.7.4 Animal research

Testing the safety and efficacy of medical devices such as implants and artificial organs requires the use of animal models because it would be unethical to try them directly in humans without first gathering enough information on their effects. Even the best computer models cannot substitute for the testing of devices in the complex environment inside living beings. Moreover, animals are the source for biomaterials, such as porcine heart valves, and tissues for xenotransplants. None of these, however, authorize the use and abuse of animals at will. On the contrary, there is a broad consensus in considering animal experimentation acceptable only when conducted under "the most humane methods available within the limits of scientific capability" (U.S. FDA Position Paper on Animal use in testing FDA regulated products). Consequently, there are several laws that regulate animal research which are available from the Animal Welfare Information Center home page at www.nal.usda.gov.awic.

The Laboratory Animal Welfare Act of 1966, amended and renamed Animal Welfare Act (AWA) in 1970, applied to all research facilities using species designated by the U.S. Secretary of Agriculture: guinea pigs, hamsters, gerbils, rabbits, dogs, cats, nonhuman primates, marine mammals, farm animal species and warm-blooded wild animals. Those facilities must be registered with the U.S. Department of Agriculture (USDA) and have Institutional Animal Care and Use Committees that review and approve animal procedures before applying them and review the facilities each semester. In 1976, rats, mice, birds, horses and farm animals were excluded from coverage under the AWA.

In 1985, the Improved Standards of Laboratory Animals Act required the minimization of animal pain and distress, by adequate veterinary care and the use of anesthetics, analgesics, tranquilizers, or if necessary, euthanasia. Also in 1985, the Health Research Extension Act mandated the promulgation of policies governing the use of research animals supported by U.S. Public Health Service (PHS) funds. The Public Health Service Policy on Humane Care and Use of Laboratory Animals (available from the NIH home page at

www.nih.gov/grants/oprr/phspol.htm) implements those provisions for the care of laboratory animals, and requires the filing and annual updates of an Animal Welfare Assurance which describes the institution's animal care and use program. The Guide for the Care and Use of Laboratory Animals, first published in 1963 (NIH Publication No. 86-23), provides information about the care and use of laboratory animals in ways determined to be professionally and humanely appropriate. The Good Laboratory Practices enforced by the FDA (21CFR58.43 and 21CFR58.45) include provisions for animal care, space allotment, feeding, handling, disease control and treatment, identification, sanitation, feed and water inspection, waste and refuse disposal, and pest control.

Regulations notwithstanding, there is still much room for ethical concerns. Biomedical engineers can help in devising alternative techniques (Bennett et al., 1994) which replace the actual use of animals, reduce the numbers used, and/or refine the techniques to minimize the potential for the animal to experience pain or distress. Some replacement techniques are: use of

living systems such as organ, tissue and cell culture (in vitro methods), invertebrate animals, microorganisms and plants; use of nonliving systems, such as immunochemical techniques and physical systems (e.g. mannequins); and use of computer simulations. The number of experimental animals can be reduced by sharing them, at least within the same institution, improving statistical design, relying on the least advanced species in the developmental scale (phylogenetic reduction) and using better quality animals. Refinement to reduce the pain and distress of the experimental animal can be achieved by decreased invasiveness, improved monitoring and analytical instrumentation, improved control of pain and improved control of handling and restraint.

3.7.5 Human experimentation

The information provided by animal models may not be enough to ensure the safety and efficacy of a medical product. Sometimes that information can even be misleading (Barnard and Kaufman, 1997). Therefore, there is a need for clinical investigations involving human subjects, which poses further ethical dilemmas.

The National Research Act signed into law in 1974, established the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. This Commission identified the basic ethical principles that should underlie the conduct of biomedical and behavioral research involving human subjects, and developed guidelines to be followed to assure that such research is conducted in accordance with those principles. The discussions of the Commission are summarized in the Belmont Report (available from the FDA home page at www.fad.gov/oha/IRB/appendh.html), which is the basis of 45CFR46, Protection of human subjects. Those basic principles are: respect for persons, beneficence and justice.

Respect for persons requires the investigator to acknowledge their autonomy and to protect those with diminished autonomy. This has led to the requirement of obtaining voluntary, written informed consent from potential subjects of clinical investigations.

The principle of beneficence (not to be confused here with simple kindness or charity) translates into two complementary rules: (1) do not harm and (2) maximize possible benefits and minimize possible harm. The benefits, either for the subject or in the form of knowledge gained, should always outweigh the risks.

The principle of justice means that benefits and risks should be fairly distributed. Therefore, subjects should be selected because of factors relevant to the research problem, avoiding considerations such as their availability, gullibility, social or economic position, or similar discriminatory factors or prejudices.

The World Medical Association adopted the Declaration of Helsinki (available at www.fda.gov/oha/IRB/appendg.html) in 1964 and has since revised it several times. The Declaration issues a series of recommendations as a guide to every physician in biomedical research involving human subjects. It includes basic principles and principles to be considered in clinical (or therapeutic) research and nonclinical biomedical research, i.e. research whose aim is purely scientific and without implying direct diagnostic or therapeutic value to the research subject. Biomedical engineers designing medical products may not be directly involved with patients, but they must be aware of those recommendations as they may be required to inform or assist physicians in their decisions.

Informed consent

The ethical principles in the Belmont Report applied to medical devices have led to 21CFR50, Protection of human subjects, which contains provisions for protecting the rights and safety of subjects involved in investigations. 21CFR50.20 to 21CFR50.27 regulate the informed consent from human subjects, and 21CFR50.40 to 21CFR50.48 regulate clinical investigations involving prisoners as subjects. These regulations require Institutional review boards (IRBs) formally designated by an institution to review biomedical research involving humans as subjects, to approve the initiation of and to review such research. 21CFR56.101 to 21CFR56.124 regulate IRBs, including their functions concerning research review.

No human being can be subject of an investigation without providing written informed consent, either directly or through a legal representative, using forms approved by the IRB. The investigator must seek that consent only in circumstances that do not coerce the prospective subject. Some exceptions are: life-threatening situations necessitating the use of the test article; inability to communicate with the subject; lack of enough time to obtain consent from the subject's legal representative; unavailability of alternative therapy; and some instances concerning emergency research. The basic elements of informed consent include:

- 1. A statement that the study involves research and its purposes, the expected duration, the procedures to be followed, and identification of any experimental procedures.
- 2. A description of any reasonably foreseeable risks or discomforts to the subject.
- 3. A description of any benefits to the subject or to others which may reasonably be expected from the research.
- 4. A disclosure of appropriate alternative procedures or therapies, if any, that might be advantageous to the subject.
- 5. A statement describing the confidentiality level of records identifying the subject.
- 6. For research involving more than minimal risk, an explanation of remedial actions and compensations, if any, available in case of injury.
- 7. An explanation of whom to contact for further information on the research and research subjects' rights.
- 8. A statement that participation is voluntary and that refusal to participate will involve no penalty or loss of benefits and that the subject may discontinue participation at any time without penalty or loss of benefits.

The following are additional elements of informed consent to be provided to each subject when appropriate:

- 1. A statement that the treatment may involve risks to the subject (or to the embryo or fetus, if the subject is or may become pregnant) which are currently unforeseeable.
- 2. Anticipated circumstances under which the subject's participation may be terminated by the investigator without regard to the subject's consent.
- 3. Any additional costs to the subject that may result from participation in the research.
- 4. The consequences of a subject's decision to withdraw from the research and procedures for orderly termination of participation by the subject.

- 5. A statement that significant new findings developed during the course of the research which may relate to the subject's willingness to continue participation will be provided to the subject.
- 6. The approximate number of subjects involved in the study.

All these detailed requirements reveal the great concern for human experimentation. Legal provisions certainly help to avoid abuses and to defend patient's rights but do not exclude ethical dilemmas. For example, what degree of understanding of explanations about medical procedures can be expected from lay patients? How is it possible to explain all facts involved when the experiment itself is intended to shine more light on some obscure matters? Can too much detail provoke a refusal to participate because of lack of understanding?

Risk/benefit assessment

The assessment of risks and benefits is difficult because the investigator may be biased towards an outcome that favors his interest, namely to engage the patient in the research. That bias can be inadvertently induced by the enthusiasm for the research work, perhaps derived from previous positive experience. The investigator, however, should first ascertain that there are no alternative ways of obtaining the benefits sought in the research.

The risks and corresponding benefits considered should not be restricted to physical and psychological aspects, but include legal, social and economic aspects too. Those risks and benefits concern not only the research subjects themselves but also their families and the society in general. In balancing risks and benefits, those affecting the individual subjects should carry more weight. However, it is acceptable to undertake an investigation that does not carry any particular benefit to the subjects involved, provided their rights have been protected. Nevertheless, risks must be always minimized, regardless of the excellence of the anticipated benefits and this is a criterion for the IRB to approve the research (21CFR56.111). Brutal or inhumane treatment of human subjects is never morally justified. This is the basis for example of the consensus in not using information gathered from Nazi experiments on prisoners, and in rejecting organ "harvesting" in jail execution chambers.

Subject selection

The selection of the research subjects is a matter of justice and concerns both the inclusion and the exclusion of subjects. Injustice arises when selecting minority groups for risky experiments, which may not even benefit them, and also when including unsuitable subjects in potentially beneficial experiments. Fielder (1993) cites an actual case where a clinical trial had to be stopped because of the death of two subjects that were included in the trial in spite of the evidence that they were not eligible. The "humanitarian" decision of including them harmed those who used the results of the trial.

Social justice may call for an order of preference in participating in potentially beneficial experiments. Distributive justice requires that different social groups share the burden of experiments, not only particular groups, such as users of public hospitals or citizens of less developed countries that, ironically, may not later benefit from the research outcome. 21CFR56.111 requires that the IRB determines that the selection of subjects is equitable in order to approve the research.

3.8 BIOMEDICAL PRODUCT LIABILITY

Liability is the quality or state of being liable, i.e. to be responsible or obligated according to law (or equity). Criminal law concerns responsibilities owed to society and civil law concerns responsibilities owed to individuals. Engineers' codes of ethics include acceptance of responsibility in making engineering decisions. Biomedical engineers are liable for their engineering work, either because of their professional practice as consultants or because of the products they design as employees.

3.8.1 The legal basis for liability

The legal causes of action stem from state or federal statutes and also from common law, meaning that responsibilities go farther than implied by regulations. The legal bases for liability in common law are torts and contracts. A tort is a wrongful act other than a breach of contract for which the injured part is entitled to compensation in civil law. This compensation is termed damage and may be based on personal injury, property damage or economic loss.

A contract is an agreement, which implies an obligation, but is not always enforceable at law. Formerly only those parties who had actually entered into a contractual relationship had legal rights and duties, but now subsequent users of products have these rights. As a result, buying a product is considered an agreement with the manufacturer through the distributor, wholesaler or retailer, and this agreement is law-enforceable. A contract breach is a failure to perform the duties or fulfill the obligations established in a contract. Direct damages from breach of contract are those that would be reasonably foreseen, when the contract was made, to follow naturally from the breach. Consequential damages are further damages resulting indirectly from the breach.

3.8.2 Negligence

Ordinary negligence is a usual tort in professional malpractice and product liability cases. Negligence is the failure to exercise a reasonable standard of care in a person's conduct as required by civil law. Negligence does not mean intentional wrongdoing, or wrongful conduct under the criminal law, as it often results from carelessness, neglect or unintentional mistake. Nevertheless, when acting carefully, imperfection does not imply negligence either. What is a "reasonable standard of care" often depends on the opinions of expert witnesses or on industry standards and guidelines, which change with time and place. Negligence per se is negligence that violates a statute or ordinance intended to protect people from the kind of harm that negligence caused. An example is failure to meet a compulsory standard.

The four major elements of tort action for negligence are (Fries, 1996, S14.1, Blinn, 1989, S7.2):

- 1. That a person or business owes a duty of care to another.
- 2. That the applicable standard to fulfill the duty was breached.
- 3. That the act or omission was a proximate cause of compensable injury.
- 4. That there were compensable damages to the plaintiff.

The duty of due care arises when the plaintiff should have foreseen that injury would result from the defendant's action. It is not necessary that the defendant foresees that injury too. In product liability, the manufacturer's obligation to exercise reasonable care embraces the design, manufacture, labeling, inspection, testing, packaging and provision of warnings concerning the use of the product. This obligation extends to parts that become components of other products and to products that use parts or subassemblies from other manufacturers. However, the Biomaterials Access Assurance Act of 1997 excludes suppliers of (nondefective) raw biomaterials from implantable product liability lawsuits. The standard of care for distributors, wholesalers and sellers is lax.

The plaintiff must establish each and every of the above elements to show defendant's negligence. Hence, it is not enough to provide evidence that the defendant has breached a duty of care due to the plaintiff, e.g. by finding a miscalculated component value, or a manufacturer recalling or removing a product because of a risk to health. It must be established that that breach caused the plaintiff's harm.

A negligent conduct by the plaintiff which is shown to produce or contribute to the alleged harm, prevents or reduces recovering damages based on the defendant's negligence. When the plaintiff voluntarily accepts the risk in advance, either explicitly or implicitly, the defendant is relieved of any obligations towards the consenting part ("assumption of risk"). However, neither of these are defenses in case of negligence per se.

3.8.3 Strict liability in tort

The concept of strict liability in tort, or liability without fault, applies to products, not to the defendant's conduct as in negligence cases. In order to recover damages, the plaintiff must produce evidence that:

- 1. The product is defective or is unreasonably dangerous to the user or consumer or to his property.
- 2. The product is expected to and does reach the user or consumer without substantial change in the condition in which it is sold.
- 3. The defective condition actually yielded the physical harm that caused the action.

A medical product is "unreasonably dangerous" if its risks outweigh its benefits or if it fails to perform safely when properly used. Risks are deemed unreasonable when:

- 1. Users are unaware that there is a risk.
- 2. Users are unable to judge the degree of risk even though they are aware of it.
- 3. Users are unable to deal with the risk.
- 4. The risk could be eliminated at a cost that would not price the product out of the market.

Users means here ordinary competent health care personnel. A manufacturer may be held liable even after exercising all necessary care in producing the product, i.e. if there is no negligence, but there was a defect and that defect produced the plaintiff's physical harm. The rationale is that manufacturers are assumed to have ample opportunity to detect and remove defects. Defects can be patent or latent. Patent defects occur in a series of units of the same product. Latent defects occur in a limited number of units of the same product. Patent defects can result from design or manufacturing. Manufacturing defects, either patent or latent, can be evidenced from similar harm to different users, complaints about performance and comparison with other units of the same product or similar products. Quality control programs are instrumental in reducing manufacturers' liability. Plaintiff's negligence, assumption of risk or misconduct, e.g. by using the product in an unanticipated way, and product alteration prevent damage recovery.

Design defects can result from negligence and lead to product liability. A design is deemed unsafe, for example, when it conceals dangers, does not include adequate safety features or does not specify parts with appropriate strength (Fries, 1996, S14.4). To avoid hazards, remove them if possible. If not, or if the removal costs outweigh the benefits when using state of the art solutions, guard against hazards. If this is not possible either, warn of the hazard. Failure to warn the user about a hazard where a reasonable person would do so, is negligence.

Since the responsibility for the use of medical products is with health care personnel, the manufacturer must provide them all the information needed to make the patient aware of the risks, particularly for "off label use", i.e. using the product for an application different from that it was approved for. Manufacturer's warnings must also describe the risks of misuse and the risks, if any, for the user itself (not the patient). Component manufacturers must warn the product manufacturer of any known risks posed by the component, so that the manufacturer can warn the ultimate user as necessary.

In the 1990s, while still protecting the consumer, several U.S. Congresses have attempted to change rules of evidence and standing to stop frivolous and trivial lawsuits, which sometimes awarded excessive, unpredictable, and often arbitrary damage awards. In June 1997, the Committee on Commerce, Science, and Transportation reported to the Senate the Product Liability Reform Act of 1997 to "establish legal standards and procedures for product liability litigation, and for other purposes." The bill included liability rules applicable to product sellers, renters and lessors, in addition to manufacturers and biomaterials suppliers, placed uniform time limitations on liability, and limited liability for noneconomic loss (pain, suffering, inconvenience, mental suffering, emotional distress, loss of society and companionship, loss of consortium, injury to reputation, and humiliation), which are quite subjective.

3.8.4 Breach of contract or warranty

A warranty is an undertaking or stipulation that a certain fact is or will be as it is expressed or by implication declared or promised to be. Express warranty is based on an affirmation of fact, by words or any other representation, relied on by the claimant. A "representation" requires an actual statement, not necessarily written or necessarily including the term warranty, a description of the goods or a sample or a model which are made a basis of the bargain. Estimates or opinions do not constitute a representation. Breach of an express warranty is a cause of action in both professional liability and product liability. Hence, product performance must never be overstated.

Implied warranties are inferred by law from the facts and circumstances without the need for any formal words or other representations. Such is the case, for example, in product sales where *merchantability* and *fitness for particular purpose* are two common implicit warranties. A product is merchantable, or salable, when it is of commercially acceptable quality. A merchant selling goods to a customer implies that the goods are reasonably fit for the ordinary purposes for

which they are used. According to Section 2-315 of the Uniform Commercial Code, the fitness for particular purpose warranty arises when at the time of contracting the seller has reason to know that a buyer intends to use the goods for a particular purpose, and that the buyer relies on the seller's skill or judgment to select or furnish suitable goods.

Sellers may explicitly limit their liability resulting from a sales contract by including disclaimers or modifications of warranties under the contract, provided they do not deceive the buyer. Specifications for parts or components to be used in a product must be precise enough as not to give suppliers too much leeway in meeting them. Any warnings or disclaimers about parts used in the product should be communicated to the buyer.

3.9 SOCIETAL COSTS OF BIOMEDICAL TECHNOLOGY

The steady growth in health care expenditures in the U.S. during the last decades is a major societal problem. Growth has coincided with medical technology expansion, which has sometimes been deemed the cause of rising health care costs. Technology and health care costs are linked indeed because no technology is cost free but there are additional factors to note: a growing population with needs, increasing administrative costs and the ever-increasing expectations of patients (Schwartz, 1992).

Aging populations, substance abuse and AIDS take a big share in health care systems. The assessment of health care needs for these social groups and clinical practice guidelines (Section 3.2) should guide biomedical engineers towards the development of safer and more effective devices and procedures suited to actual needs. Ethical responsibility in resource allocation (Section 3.7) should compel biomedical engineers to develop products with major patient benefit and minimal risk, even if that implies less profit, social prestige or recognizance. Adapting existing solutions often costs less than new approaches.

Medical malpractice liability, intended to compensate patients harmed by improper medical care, results in high direct and indirect costs. Direct costs include insurance, selfinsurance, awards and settlements, hospital legal fees and administrative costs. Indirect costs result from positive defensive medicine: added referrals, and tests and procedures not needed to treat the patient but ordered to reduce potential liability. Tort costs comprise a larger share of the Gross National Product in the U.S. than in any other developed country. The annual growth rate of direct costs from medical malpractice-related torts is larger than that of all torts, and even larger than the increase of health care costs. Section 3.8 discusses different ways of reducing product liability that can also help in reducing medical malpractice liability. A reform of the tort system aimed to lessen the frequency and size of malpractice damages would lower health care costs without depriving patients of needed care.

Positive defensive medicine also increases societal costs because the time it takes from health care personnel and equipment could be spent on patients that really need them. Negative defensive medicine, i.e. refusal to offer medical services because of fear of liability, and reduced access to health care because of its cost result in patients receiving less care. This may prevent diagnosis of a serious illness or let a minor problem develop into a major one, eventually costing more than initially saved. Biomedical engineers can help by developing affordable and reliable equipment intended for screening social groups at risk. Patient expectations have increased partly because of some successful biomedical products. Health care professionals feel compelled to use any existing means to fight illness, in spite of fear of malpractice suits because they feel urged to cure at any price. Yet, life expectancy in the U.S. is not better than that in countries with less expenditure on health care. Ironically, technology intended to improve the quality of life sometimes worsens terminal treatment by replacing tender family care with aggressive machine care.

Research engineers and scientists must carefully avoid the creation of misleading public expectations. Once there is a common belief that some product has a beneficial (or detrimental) effect, it is very difficult to convince everyone to the contrary. Trying to obtain citizens' support in order to secure research funds for a new product or study can be very risky if the information provided to the media is not carefully screened according to the scientific literacy of the public. Promoting a future product claiming its usefulness without first probing its efficacy, is irresponsible. Promotional literature must be accurate and avoid the hype accompanying some consumer products. Medical products must fulfill a need, not induce additional needs.

3.10 REVIEW QUESTIONS

- 3.1 Answer the design input questions in Section 3.1 for an implantable insulin pump.
- 3.2 Search and compare the regulatory requirements for a heart rate monitor using photoplethysmography and intended for a fitness machine with those for a heart rate monitor using electrodes and intended for home care and telemedicine.
- 3.3 List the basic principles to be met by a "well-controlled experiment" to establish valid scientific evidence.
- 3.4 Summarize the findings in the NIH Technology Statement on "Diffusion of ECMO Technology: Extracorporeal Membrane Oxygenation."
- 3.5 Summarize the findings of the NIH Technology Assessment Statement "Ultrasound screening: implications of the RADIUS study." (RADIUS stands for Routine Antenatal Diagnostic Imaging Ultrasound Study).
- 3.6 Compare the different body contact durations established in ISO 10993 with those in the "Tripartite Agreement for Biocompatibility Testing" between the U.S., Canada and the United Kingdom.
- 3.7 Outline the biosafety provisions for a laboratory running experiments to monitor the growth of Saccaromyces cerevisiae by measuring changes in electrical impedance.
- 3.8 Compare the relative advantages and shortcomings of sterilization by autoclave, ethylene oxide (EtO) and gamma rays.
- 3.9 Determine the occupational limit for ozone by searching the literature.
- 3.10 List the amplitude and corresponding effects of electric currents of increasing intensity applied to the body surface.
- 3.11 Calculate the maximal capacitance from live conductors in the power supply cord to ground in a type CF equipment (IEC60601-1) in order not to exceed the maximal ground leakage current in normal conditions.
- 3.12 For the electromagnetic spectrum, estimate the minimal frequency that causes ionizing radiation.

- 3.13 For each portion of the electromagnetic spectrum, describe the main interaction mechanisms between radiation and living tissues.
- 3.14 Compare the respective risks posed by alpha, beta and gamma emitters depending on whether they are inhaled or stay outside the body.
- 3.15 Discuss some relevant factors affecting the risks associated with software in medical products.
- 3.16 Define the Thermal Index (TI) and Mechanical Index (MI) for ultrasonic radiation.
- 3.17 Describe the stages of the host reaction to biomaterials.
- 3.18 Summarize blood-material interactions and list some methods to develop thromboresistant surfaces.
- 3.19 Describe potential adverse health effects from exposure to mercury from dental amalgam and list any reported effects.
- 3.20 Devise a maintenance plan for an anesthesia machine to be used in a hospital emergency room and discuss the potential benefits and shortcomings of alternative repair strategies.
- 3.21 Obtain an informed consent form for clinical testing of a medical device and discuss its implications for a biomedical engineer involved in designing that device.
- 3.22 Discuss the four major elements of tort action for negligence.

3.11 REFERENCES

- AIUM. 1997. *Conclusions regarding heat*. Official Statement, approved March 1997, Laurel MD: American Institute of Ultrasound in Medicine. Available at: www.aium.org/stmts.htm.
- Anderson, J. M., Gristina, A. G., Hanson, S. R., Harker, L. A., Johnson, R. J., Merritt, K., Naylor, P. T., and Schoen, F. J. 1996. Host reactions to biomaterials and their evaluation. In B. D. Ratner, A. S. Hoffman, F. J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.

ANSI. 1991. ANSIC95.1-1991, Standard safety levels with respect to human exposure to radiofrequency fields, 3 kHz to 300 GHz. New York: American National Standards Institute.

- ANSI. 1993. ANSIZ136.1-1993, American Standard for safe use of lasers. New York: American National Standards Institute.
- Barvard, N. D., and Kaufman, S. R. 1997. Animal research is wasteful and misleading. Sci. Am. **276**(2): 80-82.
- Bennet, B. T., Brown, M. J., and Schoefield, J. C. 1994. Essentials for Animal Research: A Primer for Research Personnel. 2nd Ed. Beltsville, MD: United States Department of Agriculture, National Agricultural Library. Available from the University of California at Santa Barbara home page at: research.ucsb.edu/pro/ess_idex.html.
- Blinn, K. W. 1989. *Legal and ethical concepts in engineering*. Englewood Cliff, NJ: Prentice-Hall.
- Bruck, C. W. 1990. Package sterilization methods. In J. D. O'Brien (ed.) *Medical Device* packaging handbook. New York: Marcel Dekker.
- CDRH. 1997. Information for manufacturers seeking marketing clearance of diagnostic ultrasound systems and transducers. Rockville, MD: Center for Devices and Radiological Health. Available at: www.fda.gov/cdrh/ode/ulstran.pdf.

- Coury, A. J., Levy, R. J., McMillin, C. R., Pathak, Y., Ratner, B. D., Schoen, F. J., Williams, D. F., and Williams, R. L. 1996. Degradation of biomaterials in the biologic environment. In B. D. Ratner, A. S. Hoffman, F. J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Diller, K. R. 1985. Analysis of skin burns. In A. Shitzer and R. Eberhart (eds.) *Heat transfer in medicine and biology*. New York: Plenum Press.
- FDA. 1996. Summaries from the CDRH Medical Software Policy Workshop (September 3–4). Rockville, MD: Center for Devices and Radiological Health. Available at: www.fda.gov/cdrh/ost/classif.html.
- FDA. 2003. Sec. 860.7 Determination of safety and effectiveness http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfCFR/CFRSearch.cfm?FR=860.7
- Foster, K. R. 1996. Electromagnetic field effects and mechanisms. *IEEE Eng. Med. Biol. Magazine*, **15** (4): 50–56.
- Fries, R. C. 1997. Reliable design of medical devices. New York: Marcel Dekker.
- Greco, R. S. (ed.) 1994. *Implantation biology, the host response and biomedical devices*. Boca Raton, FL: CRC Press.
- Gristina, A. G., 1996. Implant-associated infection. In B. D. Ratner, A. S. Hoffman, F. J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Hoffman, A. S. 1996. Biologically functional materials. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- ICRP. 1991. 1990 Recommendations of the International Commission on Radiological Protection. *Annals of the ICRP*, **21** (1–3), ICRP Publication 60. New York: Pergamon Press.
- IEC. 1984. *IEC 479 Effects of current passing through the human body. Part 1: General aspects.* Geneve (Switzerland): International Electrotechnical Commission.
- IEC. 1987. *IEC 479 Effects of current passing through the human body. Part 2: Special aspects.* Geneve (Switzerland): International Electrotechnical Commission.
- IEC. 1988. IEC 60601-1 (1988-12) Medical electrical equipment. Part 1: General requirements for safety. Amendment No.1, IEC 60601-1-am1 (1991-11). Amendment No.2, IEC 60601-1-am2 (1995-03). Geneve (Switzerland): International Electrotechnical Commission.
- ISO. 1997. ISO 10993, Biological Evaluation of Medical Devices, Part 1: Evaluation and Testing. Geneve (Switzerland): International Organization for Standardization.
- Johnson, R. J. 1996. Immunology and the complement system. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Kam, L. J. 1996. Understanding electromechanical engineering. Piscataway, NJ: IEEE Press.
- Kim, S. W. 1996. Nonthrombogenic treatments. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Kowalski, J. B., and Morrisey, R. F. 1996. Sterilization of implants. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Leveson, N. G. 1995. Safeware, system safety and computers. Reading MA: Addison-Wesley.
- Lowery, A., Strojny, J., and Puleo, J. 1996. *Medical device quality systems manual: a small entity compliance guide.* HHS Publication FDA 97-4179. Rockville, MD: Center for Devices

and Radiological Health. Accessible from:

www.fda.gov/cdrh/dsma/gmp_man.html#contents.

- Merritt, K. 1996. Systemic toxicity and hypersensitivity. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- National Academy of Sciences. 1989. *Biosafety in the laboratory (Prudent practices for the handling and disposal of infectious materials)*. Washington, DC: National Academy Press.
- NCRPM. 1992. *Exposure criteria for medical ultrasound: I. Criteria based on thermal mechanisms*. NCRP Report No. 113. Bethesda, MD: National Council on Radiation Protection and Measurement.
- Pathak, Y., Schoen, F. J., and Levy, R. J. 1996. Pathologic calcification of biomaterials. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Polk, C., and Postow, E. (eds.) 1996. *CRC Handbook of biological effects of electromagnetic fields*. 2nd Ed. Boca Raton, FL: CRC Press.
- Poole, C., and Ozonoff, D. 1996. Magnetic fields and childhood cancers. *IEEE Eng. Med. Biol. Magazine*, 15 (4): 41–49.
- Ratner, B. D., and Hoffman, A. S. 1996. Thin films, grafts and coatings. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Reilly, J. P. 1992. *Electrical stimulation and electropathology*. New York: Cambridge University Press.
- Richmond, J. Y., and McKinney, R. W. (eds.) 1993. *Biosafety in Microbiological and Biomedical Laboratories*, 3rd Ed. HHS Publication No. (CDC) 93-8395. Department of Health and Human Sciences, Public Health Services, National Institutes of Health. Centers for Disease Control and Prevention, Atlanta, GA. Available at: www.nih.gov/od/ors/ds/pubs/bmbl.
- Saha, S., and Saha, P. 1997. Biomedical ethics and biomedical engineering: a review. *Crit. Rev. Biomed. Eng.*, **25** (2): 163–201.
- Scannell, K. M., Miller N., and Glock, M. 1992. *Practice guidelines* (CBM 92-14). U.S. Department of Health and Human Sciences, Public Health Services, National Institutes of Health. Bethesda, MD: National Library of Medicine, Reference Section. Available from the NIH home page at: www.nlm.nih.gov/pubs/cbm/guidelin.html.
- Schoen, F. J. 1996. Tumorigenesis and biomaterials. In B. D. Ratner, A. S. Hoffman, F.J. Schoen, and J. E. Lemons (eds.). *Biomaterials science: an introduction to materials in medicine*. San Diego, CA: Academic Press.
- Schwartz, R. H. 1992. The physician and the social contract. Obstet. Gynecol., 79 (1): 1-4.
- Silver, F. H. 1994. *Biomaterials, medical devices and tissue engineering*. New York: Chapman and Hall.
- Standard, J. J. 1996. Industrial noise. In B. A. Plog, J. Niland, and P. H. Quinlan (eds.) *Fundamentals of industrial hygiene*. Itasca, IL: National Safety Council.
- Sterzer, F. 1980. RF therapy for malignancy, IEEE Spectrum, 17 (12): 32–37.
- Stine, K. E., and Brown, T. M. 1996. Principles of toxicology. Boca Raton, FL: CRC Press.
- Webster, J. G. (ed.) 1998. *Medical instrumentation: application and design*. 3rd Ed. New York: John Wiley and Sons.

Whitbeck, C. 1995. Biomedical engineering. In W. T. Reich (ed.) *Encyclopedia of bioethics*. Rev. ed. New York: Simon and Schuster Macmillan.

7

Design Evaluation

John G. Webster and Ramón Pallás-Areny

What defines design evaluation? Design evaluation means to first compare alternative solutions that meet the design specification before choosing the solution to be implemented. Section 6.4 discusses the evaluation of the conceptual design. It enables us to decide which solution(s) to pursue. Conceptual design also includes ongoing evaluation throughout the process. This chapter further analyzes the selected solution(s) and provides evaluation criteria for comparing alternative implementations to meet the particular requirements for medical products. Comparing options may require us to refine the conceptual design(s) we consider, at least those aspects deemed critical for the decision. Conversely, the fulfillment of the evaluation criteria influences the detailed technical design. Since the FDA mandates the evaluation of safety and efficacy of medical devices, design evaluation must also emphasize these aspects.

7.1 BIOMEDICAL PRODUCTS DESIGN TRADE-OFFS

Why must ideal designs vary from trial designs? Ideal designs yield maximal quality and performance at minimal cost in the shortest time possible. Actual designs reach a balance where these factors are optimized together. This trade-off between quality, performance, cost and time appears at different stages in the design process. At later design stages there are fewer options and also fewer trade-offs. Initial decisions are more relevant as they limit future alternatives.

Time, cost and performance requirements follow from the health care need (Sections 3.1 and 4.2). Quality requirements, including safety, derive from regulations. Higher performance and enhanced safety raise costs and lengthen development time, unless these rely on an engineering breakthrough. Hence, performance requirements must be realistic, and functionality or usability must prevail over any features that are added just because they are available. The best designs are those simple to manufacture, use and maintain. Unnecessary functions add cost and can yield an awkward product, which endangers safety and efficacy. Safety provisions must be commensurate with the state of the art, considering the benefits and risks involved. Cost reduction or time saving should never thwart safety but there is no need to overdo defensive design by including unjustified safety measures. Consensus standards provide a valuable reference for safety levels and provisions.

7.2 HUMAN FACTORS IN MEDICAL PRODUCTS

Human factors engineering, or ergonomics, studies human characteristics for the appropriate design of the living and work environment, aiming to improve the productivity, efficiency, safety and acceptance of the resultant system design (Kroemer, 1997). Human factors are essential in medical product design. The CDRH has published guidance on human factors for medical devices that summarizes basic concepts illustrated by actual examples, and provides some rules of thumb for design (Sawyer et al., 1996).

7.2.1 User-centered design

Many of the problems in medical products resulting from user limitations result from performance-oriented designs where the designer has assumed, often unconsciously, the role of average user or made unwarranted assumptions about the user's behavior. The designer tends to see the product as the implementation of an algorithm, a system enclosed by a cabinet or a process rooted in solid engineering or scientific principles, and to think about the user as an accessory that self-adapts to the product. The user plainly associates the product with its interface. Hence, since the user has limited ability to adapt to a product, product operation (and maintenance) must be simple and intuitive. The user's opinion is more important than the designer's feelings in deciding, say, which controls and displays are the most important and which operational procedure better suits the desired function.

Health care needs are basically stated in terms of patient needs and environmental conditions. Middendorf and Engelmann (1998) describe a four-step procedure useful to identify user (operator) needs:

- 5. Define the operational events: task analysis.
- 6. Identify the events more difficult for the operator to monitor or control.
- 7. Evaluate the environmental conditions.
- 8. Decide which operations can be automated, which can be combined, which will require special operator training, and which can be made easier to do by appropriate design of the equipment.

This procedure may require us to interview actual users, observe the use of similar products, build mockups, or work with focus groups of potential users, to evaluate alternative interface designs. Task analysis starts by identifying major tasks, their sequence and interactions, the information and accessories needed for each task, user actions and required decisions, and device reaction to them. Major tasks can be further subdivided into other tasks. The analysis of the user's body positions, movements and their frequency, and time spent in each task, and the relations and sequence between different tasks can help in arranging controls and improving the display of information and control menus.

7.2.2 The user interface

According to the FDA, user problems or errors are a common cause of reported adverse effects from medical devices, mostly because of faulty design of user interfaces. The subsection on design input in the Quality System Regulation (21CFR820.30) states: "Each manufacturer shall establish and maintain procedures to ensure that the design requirements relating to a device are appropriate and address the intended use of the device, including the needs of the user and patient." The user needs can be properly addressed by considering the user interaction with the product when operating it. 21CFR820.100, concerning corrective and preventive actions, requires manufacturers "to identify existing and potential causes of nonconforming product, or other quality problems." The FDA considers user error a nonconformity.

Therefore, design evaluation includes an assessment of the user interface requirements described in the design specifications and the fulfillment of those requirements by the proposed solutions. Design verification (Chapter 8) must confirm that the designed product fulfills the specified user interface requirements. Design validation (Chapter 9) must prove that the final design meets those user interface needs under the defined operating conditions. If necessary, the final users should hands-on test the product during its development stages.

User interfaces concern the device (hardware and software) and its labeling. The interaction between the user and the device depends on the user's physical, sensory (hearing, sight and touch) and mental abilities, as influenced by the environmental and personal circumstances when the device is operated or serviced. The design specifications must describe the user population, the operating environment and resources available to the user (written manuals, coworkers, telephone help lines).

Layout and design of controls and displays

Controls and displays must be arranged so that they can be easily identified, read and set. Figure 7.1 shows controls for sequential operation. Table 7.1 shows the minimum separation distance for controls.


Figure 7.1 Controls for sequential operations should be arranged according to the customary reading sequence (in the US, from left to right and from top to bottom), and each control should be beneath its associated display or directly related to it.

	Toggle switches	Pushbuttons	Continuous rotary controls	Rotary selector switches	Discrete thumbwheel controls
Toggle	19	13	19	<u>19</u>	13
Pushbuttons	13	13	13	13	13
Continuous rotary controls	19	13	25	25	19
Rotary selector switches	19	13	25	25	19
Discrete thumbwheel controls	13	13	19	19	10

Table 7.1 Minimum separation distance for controls. (Adapted from Human Engineering Guidelines and Preferred Practices for Design of Medical Devices, Association for the Advancement of Medical Instrumentation).

All distances are in millimeters.

To evaluate a design consider:

3.23 User capabilities in the operating environment. These include: vision, hearing, reach, strength, dexterity, memory and alertness, as affected by posture, garments (e.g., gloves, masks and goggles), lighting and acoustic noise level and pitch.

- 3.24 Number and grouping of controls. The physical layout must be clear and intuitive. Arrange controls for sequential operations in sequence from left to right and from top to bottom (or in the customary reading sequence for the users). Assemble controls affecting a given family of parameters but do not cluster them. Consider multifunction keys instead of a large number of single-function keys. Avoid a large number of controls or options in emergency devices (there is no time to set them).
- 3.25 Control action. Provide tactile and/or acoustical feedback on the control action performed. For analog controls, use linear scales, so that there is a correspondence between the user effort and the resulting output. In devices intended to administer some form of energy or material, set the output to zero on power turn-on and use snap controls. Protect critical controls from inadvertent operation, for example by requiring two simultaneous actions with different hands or successive actions with separated controls.
- 26. Control-displays association. Ensure that displays and related controls are associated in a logical way so that they can be easily identified and remembered. When using color codes, avoid conflicts with industry conventions.
- 3.27 Displays. Ensure legibility for the viewing distance, viewing angle, brightness contrast and color contrast in the operating environment. In operating rooms consider emergency lighting. In emergency equipment, consider outdoor operation. Displayed text and symbols and their arrangement should agree with those shown in operating manuals. Provide analog displays for critical parameters rather than numbers alone. Figure 7.2 shows that the most important functions should be displayed in the so called primary display area, according to anthropometric data.



Figure 7.2 Important functions should be displayed in the preferred viewing area. (From Human Engineering Guidelines and Preferred Practices for Design of Medical Devices, Association for the Advancement of Medical Instrumentation.)

Component installation

The correct operation of some devices requires the user to install or replace accessories: tubes, valves, batteries, leads, cables, diskettes. Some common errors reported are: tubing connected to the wrong port; loose connections; accidental disconnection; electric leads inserted into an improper power source; batteries or bulbs inserted incorrectly; and valves or other hardware installed backwards or upside-down (Sawyer et al., 1996). The FDA reported that between 1985 and 1994, 24 infants or children received macroshock from unprotected lead wires or cables.

Five children were killed. To prevent any mishap with those accessories under the operating environment, in design evaluation consider:

- 1. Make the installation and removal of accessories easy for the intended user without requiring special tools or skills, yet secure enough. Make critical connections that lock mechanically or use angle (elbow-shaped) connectors. Design alarms to monitor patient's gas tubes, electrodes and cables. Store needed accessories in an accessible compartment. Make battery replacement handy.
- Accessories identification. Make the connecting ports for different accessories easy to identify, using color codes for instance. Use incompatible connectors for dissimilar outputs and for inputs and outputs. Number accessories to expedite in reordering and replacing defectives. Prevent batteries from being connected with the reverse polarity or provide an alert. Recognize and reject nonsystem diskettes.
- 3. Connectors and adapters. Use protected electrode lead wires and patient cables. Warn of hazards, such as loss of protection, when using unauthorized adapters, wires or cables that may fit the connectors provided.
- 4. Include accessories as a part of the system when testing.

Software design

Implementing functions by software adds features and flexibility but its impact on decisionmaking processes is not necessarily positive. The user must be in command of rather than subservient to the software, but the software must anticipate human errors and hardware failures. Software tends to reduce the number of controls and displays but may cloud the user's conception of a model about the device operation, burden the user's memory and try his or her patience.

Some common problems that can lead to errors are: illogical or cumbersome control sequences; unfamiliar language, symbols, or codes; inconsistencies among display formats; conventions contrary to user expectations; ambiguous or no feedback after input; functions hidden from the user; missing or ambiguous prompts, symbols, or icons; unsignalled resets or defaults; no status information; missing lock-outs or interlocks; and requirements for complex mental calculations (Sawyer et al., 1996). Software-induced errors can be difficult to remember, retrace or recreate because of the many options offered to the user. To evaluate the software-user interaction consider:

- 1. Make communication simple; ask for jargon–free dialogue; rely on common language and symbols, concise sentences, meaningful acronyms and abbreviations in command structures and menus when there are many commands to learn.
- 2. Make communication effective. Provide continuous feedback through status, error and help messages. Devise simple and reliable data entry methods with immediate user feedback and prompt in case of suspicious values or commands. Provide consistent, positive, constructive and selective displayed information. Provide information upon request. Group sequences of actions.
- 3. Decisions. Make the software help, not supplant users. Design checkout procedures and menus to build users' confidence. Design error-detection methods and simple procedures for

the user to correct errors and reverse actions. Make procedures for similar functions similar themselves.

Alarms

Medical devices use alarms to draw the user's attention. Acoustic alarms are pervasive and elucidate a fast response but can become annoying. Visual alarms provide more information and usually cost less than acoustic alarms but they are effective only in the user's line of sight. Hence, effective alarms combine acoustic and visual outputs. Stronger signals improve alertness. Some common alarm problems are: false alarms, delayed alarms, insensitive alarms, too sensitive alarms, imperceptible alarms, ambiguous meanings, inappropriate silencing, and accidental disabling (Sawyer et al., 1996). To avoid those problems, consider:

- 1. Critical alarms. Design visual and acoustic alarms and alerts for critical parameters from the patient and the equipment itself. For the equipment consider: high and low pressures, high and low flows, high temperature, power loss, battery charge level (or operation time), tilting and disconnected accessories. Rank critical alarms according to their priority.
- 2. Operating environment: background acoustic noise, lighting. Design alarms to be effective for the intended users in the different operating environments considered, including other possible equipment. Ambient noise levels in hospitals range from 50 dB in private rooms to 70 dB in operating rooms (Fries, 1997, Section 16.14.9), and is higher in ambulances and helicopters.
- 3. Alarm settings. Design a simple, unambiguous procedure to set alarm limits, inform the user about the actual settings and provide testing methods whenever possible. Design appropriate hysteresis to avoid unjustified recurrent alarm operation (for example because of EMI), but do not delay issuing the alarm. Recognize incompatible alarm settings and alert the user. Provide a test method for alarms with adjustable volume and/or brightness and color contrast, so that the user can verify that they are adequate. Prevent extreme limit settings, which might disable all alarms for a critical parameter. Design default limits for critical alarms to match the physiological limits for the intended use. Indicate the status of critical alarms, particularly for silenced critical acoustical alarms. Permit only temporary disabling of acoustical alarms, or require explicit acceptance of their prolonged disablement.
- 4. Alarm identification. Design alarms to be distinguishable from one another. To discern acoustic alarms from different equipment, provide pitch selection and place them to draw user attention towards the pertinent device. Table 7.2 shows the established convention for color codes: flashing red for high-priority alarms; flashing yellow for medium priority alarms; yellow for marginal condition (low-priority alarms); green for satisfactory conditions (e.g. power on); white for conditions without right or wrong implications (e.g. action in progress, alternative functions and transitory functions). Speech messages are deemed inadequate for alarms in medical products.

Table 7.2 Color coding for alarms. (Adapted from "Human engineering guidelines and preferred practices for design of medical devices", Association for the Advancement of Medical Instrumentation.)

Color Function Examples

Flashing red	High priority alarm	High or low pressure in
		fibrillation in ECG monitor
Red	Alert about inoperative system or	Displays such as "no-go"
Red	urgent action required	"error", "failure",
		"malfunction"
Flashing yellow	Medium priority alarm	Arrhythmia in ECG monitor
Yellow	Advise about a marginal condition,	Charging process in high-
	caution or attention required	voltage defibrillator, electrode
		lead failure
Green	Satisfactory condition, ready for	Displays such as "go-ahead",
	action	"in-tolerance", "ready",
		"power on"
White	Alternative or transitory conditions	"Input No. 1 in use", "Test in
	(not implying success or failure)	progress"

Labeling

Labeling is defined as all labels and other written, printed, or graphic matter on the device or any of its containers or wrappers, or accompanying the device. Labeling includes equipment labels, control labels, package labels, directions for use and operation, maintenance manuals, displays which provide instructions, prompts, cautions, or parameter identification information, and even advertising material. Labeling plays such an important role in the user–product interaction that is strictly regulated: 21CFR801 for general devices, 21CFR809 for in vitro diagnostic products, 21CFR812.5 for investigational devices, 21CFR820.120 for labeling design and manufacturing control and 21CFR1010 for electronic products.

The FDA requires that device labeling "bear adequate directions for use, operating and servicing instructions, and either adequate warnings against uses dangerous to health, or information necessary for the protection of users. All devices require directions for use unless specifically exempted by regulation." Adequate directions for use means "directions under which the layman can use a device safely and for the purposes for which it is intended." Subpart D of 21CFR801 lists the conditions for exemption from adequate directions of use, for example: prescription devices, other than surgical instruments, used by practitioners whose label bears (1) the statement "Caution: Federal law restricts this device to sale by or on the order of a [physician, dentist, veterinarian, etc.]", and (2) the method of its application or use; medical devices having commonly known directions; and medical devices for use in teaching, law enforcement, research and analysis, provided they do not involve clinical use.

Labeling is part of the design output and cannot be properly finished before technical design and product testing. However, in order to better meet the user needs and FDA regulations, it is advisable to generate the contents of labels and labeling during the design process. The following considerations may assist in developing and testing labels and labeling:

 Clarity. Words must correspond to the intended action. Avoid abbreviations, unusual acronyms or technical terms, jargon and ambiguous terms. Labels' position, size and color must be decided after considering the operating environment. Labels should be horizontal and read from left to right, usually above control knobs or switches. Vertical labels, if needed, should read from top to bottom. Mark receptacles and connectors according to their function.

- 2. Entirety. Include any symbol prescribed by applicable compulsory standards such as those shown in Figure 7.3 Warn about radiation, explosion, fire, shock, infection or any other hazards during operation or maintenance.
- 3. Consistency. Use the same terms in controls, displays, accessories, instructions, operation and service manuals (text and figures). Using all capitals can help to identify the titles of controls.
- 4. Integrity. All labels must remain in place and legible during the customary conditions of distribution, storage, use and maintenance. User instructions must remain legible during customary storage and use.



Figure 7.3. Standard symbols for non-ionizing radiation, ionizing radiation, explosion, fire, electric shock, infection, power off, power on.

Labeling devices for lay users demands special care as they lack the training and experience of medical personnel. The CDRH has published the booklet, "Write It Right: Recommendations for Developing User Instruction Manuals for Medical Devices Used in Home Health Care," to help manufacturers write effective manuals for home use devices.

7.3 DESIGN FOR COMPATIBILITY

Compatibility means the capability of working together with another device or system without modification. Incompatibility can arise from the different aspects of product interaction with its environment, including the user and the patient. Compatibility requires both an acceptable effect on the environment and an acceptable susceptibility to the environment. User compatibility is achieved by human factors engineering.

7.3.1 Functional compatibility

Even as different parts and subassemblies must be compatible in order to perform the desired function, the different devices in a given system must be compatible to fulfill the system needs. The system dimension can range from those of an implanted device to a health care facility. In the design of a single product, each considered part has a specific function and the design team selects the best part to implement the desired function. However, in the "design" of a system such as a surgical room prepared for a given intervention, for example, the resources used (devices) can have duplicate, complementary or even opposite actions and the time available to evaluate options and implement solutions to adapt different devices is quite limited. Medical products must be compatible in order to simplify system design. This can be straightforward for accessories of a given product but requires analysis in case of products of different manufacturers.

Electrical compatibility

Electric connections are intended for power supply and information communication. Power supply connections are characterized by maximal voltage and current (rms, peak and transient values), polarity, maximal power and by mechanical factors such as plugs–prongs–receptacle sizing and grounding connection. Figure 7.4 shows plugs for power cords. Common single-phase ac supply voltages are 105 V to 125 V and 210 V to 240 V, 50 Hz or 60 Hz. Common three-phase systems have 120 V, 240 V or 480 V between each live conductor and the neutral. Most airborne systems use 28 V, 400 Hz. Common dc supply voltages are 12 V, 24 V and 48 V.



Figure 7.4 Plugs for power supply cords (L, line; N, neutral; E, earth). (a) NEMA 5-15P–125 V, 15 A–United States, Canada, Mexico, Japan, Taiwan. (b) CEE7-V11–250 V–Continental Europe. (c) BS1363A–250 V–Great Britain, India, Singapore. (d) SEV 1011–250 V–Switzerland. (e) ASC112/NZSS198–250 V–Australia, New Zealand.

The physical compatibility of communication interfaces concerns electrical and mechanical parameters. The mechanical parameters are those relative to the connectors: dimensions, number of pins, their spacing and signal assignment (pin out). The main electrical parameters are voltage and current amplitudes (maximal and minimal values and logic levels), signal type [voltage or current, and single-ended (unbalanced), differential (balanced) or floating], load impedance, type of logic for digital signals (positive or negative), modulation/demodulation method and parameters, code, data speed, transmission mode and protocol. Figure 7.5 shows that there are two types of transmission: synchronous and asynchronous. In synchronous serial transmission, there is a clock signal available at both channel ends that permits the sender and receiver to count the ordered blocks containing groups of characters transmitted. Each block includes special characters, in addition to the data transmitted, intended to identify the start of the block and for error correction. This capability of error correction must be weighed against the need for data buffering for continuous transmission and the strict timing requirements. Asynchronous transmitters send data when available, not continuously under clock control. A controller circuit assembles the characters to be sent and adds control bits: start, parity and stop. Common standards for character coding are ASCII and EBCDIC. Serial interfaces use additional signals to coordinate data exchanges according to a set of conventions, or protocol, termed handshaking signals. Common asynchronous serial interfaces at the equipment level are EIA-232-E, EIA-422-A, EIA-423 and EIA-485. The I2C is a two–wire, synchronous, serial interface designed primarily for communication between integrated circuits. Hordeski (1995) describes physical and operational characteristics of several series and parallel interfaces.



Figure 7.5 Medical equipment that communicates data must use the same protocol. Protocols for synchronous serial transmission may be (a) byte-oriented or (b) bit-oriented. (c) Asynchronous serial transmission includes start and stop bits before and after each character.

Wireless local area networks (WLAN) permit the communication between portable medical devices, such as ECG monitors, and computer systems in healthcare environments. The IEEE 802.11 standard defines physical characteristics and rules for accessing the wireless medium for WLAN systems using infrared radiation or microwave signals (in the 2.4 GHz band), operating at 1 Mb/s and 2 Mb/s data rates.

Digital Imaging and Communications in Medicine (DICOM) is a standard in thirteen parts published by the NEMA from 1992 to 1996. It applies to point-to-point and networked communication of digital medical information, media storage and file format for media

interchange, aiming to ensure interoperability between digital imaging computer systems for diagnostic imaging and other clinical applications.

Mechanical compatibility

Devices to be connected to medical gases must be compatible with the respective outlets and use materials compatible with the gases involved. Some medical gases are: nitrous oxide, administered as an anesthetic agent during surgery; oxygen, administered for breathing therapy and as analgesic in a 50:50 mixture with nitrous oxide; piped medical air, used to administer drugs as a nebulized mist of air and medication; carbon dioxide, used to inflate body areas during keyhole surgery and also to stimulate breathing (mixed with air or oxygen); helium–oxygen mixtures, administered to patients with severe breathing difficulties; and several mixtures including nitrogen and noble gases used for diagnostic (lung function testing) or therapeutic (medical lasers) equipment, and high-pressure nitrogen for turbine tools.

Medical gases are supplied from either high-pressure cylinders or medical gas pipe systems, regulated by NFPA 99C (Gas and Vacuum Systems) and different standards from the Compressed Gas Association (CGA). Gas cylinders are designated with alphabet letters according to their size, A size being the smallest cylinder. Both cylinders and pipes are color coded, in the U.S. according to the CGA C-9 standard (Table 7.3). There are in addition, medical vacuum pipes for suction and drainage applications.

Table 7.5 Color code for medical gases.				
Gas	Unites States	International		
Carbon dioxide	Gray	Gray		
Cyclopropane	Orange	Orange		
Helium	Brown	Brown		
Medical air	Yellow	White and black		
Medical vacuum	White	-		
Nitrogen	Black	Black		
Nitrous oxide	Blue	Blue		
Oxygen	Green	White		

Table 7.3 Color code for medical gases.

To maintain a constant gas flow with varying supply pressures, there is a pressure regulator between the cylinder valve or central supply and the device that uses the gas. The connections on the low pressure side of this regulator are designed noninterchangeable by the Diameter Indexed Safety System (DISS). Each DISS connection involves a body, nipple and nut combination (Figure 7.6). The body has two concentric bores and the nipple two corresponding shoulders. The connection between the nut and the body is common to all gases, but the bore–shoulder sizes are specific for each gas. A faster connection method uses "quick couplers" (or "quick connects") whose male and female components are noninterchangeable between gases. Devices using medical gases must have the appropriate fittings.



Figure 7.6 Diameter Index Safety System. The bore (BB, CC) and shoulder (MM, NN) diameters are specific for each gas.

Software compatibility

Functional compatibility of software concerns the names and organization of data structures in device interfaces. This is a difficult problem because, on the one hand, computer programs cannot be copied but, on the other hand, ensuring interoperability requires an "agreement" between functional aspects of different programs, such as function names, parameters and data structures. For example, the input data format for a given program must be the same than the output data format of another program feeding data to the first program. In addition, using different names for similar functions (at the user level) in different programs may confuse the user and induce errors. Therefore, when interfacing programs from a different manufacturer, compatibility must be ensured without infringing any copyright, which calls for using only those "common" functions needed for interoperability. The merger doctrine denies copyright protection to the expression of ideas that can be expressed in only a limited number of ways (the "idea" and the "expression" are said to have merged, and since ideas are not copyrightable, none of the expressions can be copyrightable either). But displays and menus, for example, are copyrightable.

7.3.2 Electromagnetic compatibility

Electromagnetic incompatibility is usually more difficult to predict and recognize than functional incompatibility. However, the pervasiveness of electronic devices and the tendencies to work at higher frequencies and use low-power electronic circuits, which are more susceptible to interference, have rendered electromagnetic compatibility (EMC) a major issue in designing electric devices. Between 1979 and 1993, the FDA received reports of more than 100 suspected incidents of electromagnetic interference (EMI) with medical devices. The CDRH has reported the following accidents caused by EMI:

- 1. A monitor failed to detect a patient's critical condition.
- 2. A defibrillator failed to resuscitate a patient.
- 3. A wheelchair suddenly moved toward street traffic.
- 4. A laser beam moved into the audience area of a light show.
- 5. A radiation beam shutter did not close.

Magnetic resonance imaging (MRI) systems are a singular EMI hazard because they use a strong static magnetic field, pulsed gradient magnetic fields, and pulsed radio frequency (RF) fields. There are reported deaths of patients wearing pacemakers during MRI exploration. In addition, medical devices can distort images from magnetic resonance scanners.

EMI problem analysis

An EMI problem involves an electromagnetic disturbance (EMD) source, an unintended receiver, or victim, and a channel coupling the disturbance from the source to the receiver, Figure 7.7. Radiated disturbances couple through the air and conducted disturbances couple through cables, wires and power cords. A radiated disturbance can couple into a cable and yield a conducted disturbance. Conversely, radiated disturbance can result from disturbance or signal currents along a cable or power connection.



Figure 7.7 Elements of the EMI problem: energy from an EMD source couples to an unintended receiver adversely affecting its operation. Coupling can be wireless or by power cords and signal wires and cables.

Anticipated EMD sources for medical devices are some electromedical equipment, such as electrosurgery and RF hyperthermia units, x ray and magnetic resonance imaging systems, and common industrial EMD sources such as electromechanical switches, motors, pumps and power line disturbances. Unanticipated EMD sources include consumer electronic products and commercial and private radio transmissions such as mobile communication systems in vehicles and personal communication equipment. When recording weak bioelectric signals, power distribution lines (50 Hz, 60 Hz), computer monitors and fluorescent lights can become a nuisance too. Clock lines in digital circuits emit a broad spectrum of electromagnetic radiation. Users are a potential EMD source because of electrostatic charges accumulated on their bodies. Electrostatic discharges (ESDs), either by direct contact or through the air, which may be imperceptible, can nevertheless be fatal for a CMOS or BiCMOS integrated circuit.

EMC by design

EMI problems can be classified into internal and external. Internal EMI problems arise during the detailed system design and are consequently solved in order to achieve the desired functionality. External EMI problems must be anticipated and solved during the design process, not after finishing the design. This requires EMC awareness because external EMI problems do not necessarily become apparent during design but can result from EMD in the operating environment. Hence, the design input (Chapter 4) must include EMC requirements. Paul (1992)

provides a comprehensive analysis of EMI problems and solutions to achieve EMC. Ott (1988) offers a practical approach to EMC solutions.

EMI coupling can be reduced by increasing the distance to the source, e.g. removing active cellular phones from pockets close to pacemakers or using them far from critical equipment; by keeping connecting wires as short as feasible; and by using twisted and/or shielded wires. Electric shields shown in Figure 7.8 must be permanently connected to a constant voltage in order to prevent interfering currents from coupling to their internal wires. Common shielded cables are not effective in attenuating magnetic interference below 10 kHz, such as those from power supplies or CRT deflection circuits. Coaxial cables increase cost and assembly time.



Figure 7.8 Shielded cables. (a) Coaxial cable with braided shield. (b) Twin coaxial cable with braided shield. (c) Twin axial cable with common braided shield. (d) Mylar aluminum shielded two-conductor cable with tin-copper drain. (e) Two-conductor cable with spiral aluminum shield.

Power supplies are an important element for EMC. All power supplies produce EMI that can couple into internal circuits or propagate back into power distribution lines. At the same time, transient voltages in power lines can reach internal circuits through the power supply. In addition, switching noise from internal digital circuits, relays and electric motors can also travel into power lines through the power supply. The common cure for these problems is filtering, which is limited by safety requirements. In equipment with grounded enclosures, the most effective filters use large capacitors from power conductors to ground. However, large, grounded capacitors increase ground leakage currents. As a result, Figure 7.9 shows that medical-grade power line filters use small capacitors to ground and large series inductors, which are less effective because of the size, weight and cost of large-value inductors.



Figure 7.9 (*a*) The capacitance to ground in power line filters for medical equipment is small to reduce leakage current. (*b*) Power line filters for class II medical equipment do not include any capacitor connected to ground.

EMI from devices that intentionally radiate electromagnetic energy can be avoided by carefully planning the use of the electromagnetic spectrum: frequency used, RF power emitted and modulation method. 47CFR2 Subpart B regulates the allocation, assignment, and use of radio frequencies. 47CFR15 sets the regulations under which an unintentional (Subpart B), intentional (Subpart C), or incidental (47CFR15.13) radiator may be operated without an individual license. Table 7.4 lists some frequency bands designated by 47CFR18.301 for industrial, scientific and medical (ISM) applications and some common frequencies used. Even in sensitive frequency bands, modulation methods that radiate less than 0.5 s will seldom change the operation of an electronic circuit.

Table 7.4 Electromagnetic spectrum frequency allocation and uses for different industrial, scientific and medical (ISM) applications. In addition, the use of specific frequencies in the 460.6625 MHz to 465.8625 MHz band may be authorized, with 100 mW or less output power, to radio stations for one-way, nonvoice biomedical telemetry operations in hospitals, or medical or convalescent centers. 47CFR18.303 prohibits operation of ISM equipment within the following bands: 490 kHz to 510 kHz, 2170 kHz to 2194 kHz, 8354 kHz to 8374 kHz, 121.4 MHz to 121.6 MHz, 156.7 MHz to 156.9 MHz, and 242.8 MHz to 243.2 MHz.

DESIGN EVALUATION

Frequencies and bands	Application
1 MHz	Radar proximity sensors in cars
6.78 MHz ± 15.0kHz	ISM
13.56 MHz ± 7.0 kHz	ISM
27.12 MHz ± 163.0 kHz	ISM
40.68 MHz ± 20.0 kHz	ISM
150.775 MHz	Medical radio communication systems
150.790 MHz	Medical radio communication systems
152 MHz to 152.050 MHz	Medical radio communication systems
163.250 MHz \pm 12.5 kHz	Medical radio communication systems
174 MHz to 216 MHz	Biomedical telemetry (200 kHz, low power)
463.0625 MHz \pm 125 kHz	Medical radio communication systems
468.0625 MHz \pm 125 kHz	Medical radio communication systems
512 MHz to 566 MHz	Biomedical telemetry in hospitals
915 MHz \pm 13.0 MHz	ISM
2.450 GHz ± 50.0 MHz	ISM
5.800 GHz ± 75.0 MHz	ISM
10.25 GHz	Microwave radar
24 GHz	Microwave radar
24.125 GHz ± 125.0 MHz	ISM
25.25 GHz to 27.5 GHz	Data transmission from industrial and medical activities in
	space
34 GHz	Microwave radar
61.25 GHz ± 250.0 MHz	ISM (authorization needed)
122.50 GHz ± 500.0 MHz	ISM (authorization needed)
245.00 GHz ± 1.0 GHz	ISM (authorization needed)

EMC standards

A cost-effective method of addressing EMC in product design is by adhering to compliance with EMC standards, which are compulsory in the EU. CDRH encourages manufacturers of electromedical equipment to use the IEC 60601-1-2 standard: Medical Electrical Equipment, Part 1: General Requirements for Safety, Collateral Standard: Electromagnetic Compatibility - Requirements and Tests. This standard provides various limits on both emissions and immunity. While these limits are clear, the pass/fail criteria are not, and therefore it may be necessary to establish those criteria during design and testing. IEC 60601-1-2 refers to other international standards (IEC, CISPR) that detail the established requirements and test methods. These standards are subject to frequent revisions, whose status can be searched from the IEC home page at www.iec.ch.

7.3.3 Compliance with safety standards

A design technique to fulfill the requirements of a specific safety or EMC standard is the so called modular approach, which consists of building the system from certified parts that themselves fulfill the safety and EMC requirements set in the corresponding standards when working in the final environment where they will be used. In this approach it is important to recognize that medical safety standards are more stringent than industrial or commercial safety standards and, therefore, parts such as cables, power supplies and power line filters must be medical grade.

7.4 DESIGN FOR MANUFACTURABILITY

Alternative solutions to achieve a specified device functionality may have quite different manufacturing requirements and cost. Relegating the design of the manufacturing process until the product design has been defined can result in production delay, product redesign for manufacture and longer time to market. In addition, manufacturing costs can make a large part of product cost and must be anticipated. Hence, design evaluation must consider that devices have to be fabricated, packaged, labeled and sometimes sterilized. The simultaneous design of a product and its manufacturing process is termed concurrent engineering. Section 10.3 describes the production interface in biomedical product design. Here we consider those parameters influencing manufacturing complexity and cost.

7.4.1 Manufacturing processes: types, cost factors and design principles

Basic manufacturing processes can be described in terms of changes in geometry and material properties. Table 7.5 lists some basic mechanical, thermal and chemical processes for solid, granular, liquid or gaseous materials (Alting, 1994). A manufacturing process consists of basic processes, which can be grouped in three phases: phase I brings the material into a state suitable for the intended primary change in geometry or property; phase II produces the desired change; and phase III brings the component into the specified end state. The processes in phase II determine the pre- and postprocesses in phases I and III. Bralla (1986) details common manufacturing processes for metals, polymers, ceramics, glasses and other engineering materials, and provides tables for quick comparison of their advantages and limitations. Those processes include: casting, molding, machining, forming, forging, annealing, drilling, cutting, stamping, drawing, extrusion, joining (welding, soldering, brazing, adhesive bonding) and finishing (cleaning, brushing, grinding, blasting, polishing, plating, coating, passivation). Alting (1994) analyzes several processes by classifying them in mass-conserving, mass-reducing and assembly (or joining) processes.

Mechanical	Thermal	Chemical	
Elastic deformation	Heating	Solution/dissolution	
Plastic deformation	Cooling	Combustion	
Brittle fracture	Melting	Hardening	
Ductile fracture	Solidification	Precipitation	
Flow	Evaporation	Phase transformation	
Mixing	Condensation	Diffusion	
Separation			
Placing			
Transport			

Table 7.5 Some mechanical, thermal and chemical basic manufacturing processes (Alting, 1994).

Cost factors in design and process selection

Variable cost factors for producing common products are related to materials, direct labor, indirect labor, special tooling, perishable tools and supplies, utilities and invested capital

(Hornbruch, 1986). Packaging, labeling and sometimes sterilization add relevant costs to medical products.

The unit cost of materials is important when the compared alternatives involve different amounts or different forms of several materials. Yield and scrap losses can also be important. Materials or components with long lead times can delay production unless obtained earlier during design. Biomedical materials are bought in bulk quantities and their properties and performance, including biocompatibility, are strongly affected by processing techniques, particularly for plastics.

Direct labor costs depend on the design itself, the manufacturing process and workers' productivity. Direct costs increase for complex designs, closer dimensional tolerances, higher finish requirements and tooling involvement. A large number of manufacturing operations to complete a part increases direct (and indirect) costs, offers more chances for errors, and complicates production schedules. Some processes with low labor cost are: metal stamping and drawing, die casting, injection molding, automated machining and drilling, and automated processes in general. Conventional machining, investment casting and manual assembly involve high labor costs. Adjustments and calibrations are also labor intensive, hence expensive.

Design principles for manufacturability

Regardless of the manufacturing process, the following principles reduce production costs (Bralla, 1986) (Trucks, 1987) and should be considered in product design as far as they do not compromise product quality:

- 1. Simplicity. To reduce direct labor costs and improve reliability, design with few parts, parts with simple shape (easy to locate and handle), few precision adjustments and short manufacturing sequences.
- 2. Material, component and procedure standardization. Design using available materials and standard components to benefit from economies of scale, simplify purchasing, and reduce inventory, parts handling and indirect labor costs. Use automated assembly and standardized measurement and test methods whenever possible. Select materials for tissue replacement according to the physical properties of the natural material to be replaced and limit the range of candidate materials, which also must be biocompatible after processing and sterilization.
- 3. Design standardization. Specify the same materials, parts and subassemblies for similar products. Functional flexibility is less expensive when relying on software than when it is based on hardware modifications. Modular design and construction also facilitate disassembly for maintenance and removal.
- 4. Appropriate tolerances. Tight dimensional tolerances for mechanical parts (e.g. better than $\pm 25 \ \mu m$ for steel) is expensive because it takes more time (more finishing steps), precision equipment and skilled workers. Low-tolerance electronic components are quite expensive as compared to, say, 1 % resistors, 5 % capacitors and 5 % inductors. Precision components outside the common ranges (1 Ω to 10 M Ω , 10 pF to 10 μ F, 1 nH to 20 mH) are still more expensive. Integrated circuits for temperature ranges outside the commercial range (0 °C to 70 °C) are also more expensive.

7.4.2 Design for assembly

Assembly is a required part of the manufacturing process when two or more component parts must be secured together. Assembly methods affect production costs and their efficiency depends on the design. The assembly of reusable medical devices that must be disassembled and sterilized for each patient, must be simple and error-proof. Maintenance, repair, removal and recycling also require considering disassembly during design. The basic factors influencing assembly cost are the number of parts and the difficulty in handling, inserting and fastening those parts. Ullman (1992, S13.3) evaluates these factors in mechanical designs. Boothroyd et al. (1994, C6) analyzes printed-circuit board assembly.

Design for manual assembly

The efficiency of manual assembly is Ema = Nminta/tma, where Nmin is the theoretical minimal number of parts, ta is the average assembly time for a part posing no manipulation difficulties (usually 3 s), and tma is the estimated time to complete the assembly of the actual product (Boothroyd et al., 1994). Nmin is determined by combining separate parts in single parts unless one of those parts must move relative to the other parts, or is from a different material or must be separate from all other parts already assembled to enable assembly or disassembly.

Design for high-speed automated assembly

Automated assembly aims to reduce direct labor costs. It is not intended to merely replace difficult manual assembly. On the contrary, assembly processes which are difficult for humans are often difficult for machines. The human hand possesses many different motions, has touch sense and is controlled by all the senses. Machine perception and decision are coarse. Therefore, automated assembly may require product and manufacturing process design different from those for manual assembly. The economic benefits of automated assembly sometimes derive more from product redesign than from automation.

Design for robot assembly

Robots are common to produce printed circuit boards (PCBs) and to assemble electric motors, gear reducers and other mechanical parts. Robots are also useful for machine loading/unloading, material handling, spray painting, spot welding, arc welding, adhesive and sealant deposition and inspection, and forming seals and gaskets. Robots can work in hazardous environments and handle hazardous materials. Many of the design guidelines for manual and high-speed assembly also apply to robot assembly. Additional guidelines derive from robot capabilities: degrees of freedom of its arm and wrist movements, positioning systems, and sensors and end effectors available. For example, disparate parts may require different grippers and fixture rotation requires additional degrees of freedom for robot arms.

7.4.3 Subcontracting

Subcontracting or outsourcing means to contract to external suppliers one or more functions needed to bring a product into the market and service it. Some reasons for subcontracting are: cost reduction, shortening time-to-market and lack of in-house expertise. Cost reduction often derives from the large capital investment needed to perform some functions, that in addition may require equipment used only part-time or for a short term (e.g. sterilization equipment). Outsourcing can save time-to-market for example when the staff or equipment available are already working full time; outsourcing saves the time it would take to hire and train more staff or install more equipment and does not carry any future commitment. When there is no in-house expertise to achieve the required performance or quality, outsourcing may be better than achieving that expertise in-house, which takes time, particularly in areas outside the core competence of the company. At the same time, if that area is deemed of future interest, subcontracting may help in acquiring know-how. Hence, product designers can define solutions that rely on external resources.

7.5 DESIGN FOR TESTABILITY

Designs must be tested to verify that they work as expected; products must be tested to evaluate their reliability and safety and for maintenance. Design evaluation may also need testing in order to gather information about alternative designs. Testing compares some observable variables or results and their expected values. Testability can support fault detection and fault isolation in a confident, timely and cost-effective manner. Testability is not inherent to all designs. Rather, it must be incorporated into designs by identifying critical variables and, often, by including redundant components or modifying designs that otherwise yield the desired functional performance. Built-in test (BIT) is the integral capability of a system or equipment of automatically detecting, diagnosing or isolating failures. Designs difficult to test or inadequately tested can lead to defective products.

Testing poses different requirements for electrical and mechanical circuits and subsystems. A challenge in testing some mechanical systems is the lack of an interface (transducer) suitable for the measuring equipment available, essentially electronic. For instance, pressure, flow, force, torque, position, velocity, acceleration and temperature are usually easy to measure, but wear, fatigue and hardness are more elusive. Cost and space constraints do not normally permit the incorporation of redundant elements to monitor critical variables in mechanical systems. These limitations notwithstanding, product design must consider the accessibility needed for diagnosis, adjustment and calibration. Figure 7.10, for example, shows how adding an extra connection permits the calibration of a pressure gage without removing it.



Figure 7.10 The design in the left requires us to disassemble the pressure gage for calibration. The design on the right includes an otherwise redundant connection for calibration.

7.5.1 Hardware design for testability

Hardware testability depends on the least replaceable assembly in the system. This means that testability requirements during product development are different from those for functional testing or system troubleshooting at the different maintenance levels. Product development involves multiple tests for circuit continuity, supply voltage levels and signal monitoring. This requires probing many test points, injecting signals, temporarily removing components and so on. Breadboarding facilitates these tasks. However, once the design has been implemented, the accessibility of many circuit nodes and the capability of removing components are quite limited, particularly when using surface-mount technology (SMT) and multilayer PCBs.

Troubleshooting an intact circuit relies on causal behavior: the output of any circuit should correspond to its input. Unexpected outputs indicate that either the circuit behaves unexpectedly, because of internal faults or interference, or the input differs from that assumed. Simultaneous examination of input and output signals is not always enough to observe an internal fault. Figure 7.11(a), for example, shows an amplifier whose output stage is coupled through a high-pass filter. A 0 V output suggests a 0 V ac input, but regardless of the actual ac input, if the input stage is saturated or its output grounded, the output will be 0 V. To determine the actual situation we need to probe test point TP2. Similarly, if the output of the first inverter in Figure 7.11(b) is stuck at ground (or power voltage), the output of the second inverter will be 1 (0) regardless of the circuit input signal. These faults can be detected by observing the output when changing the input signal.



Figure 7.11 (a) A 0 V ac input yields a 0 V output, but an input amplifier whose output is saturated or shorted to ground also yields 0 V. Determining the actual situation requires probing TP2. (b) To test if the output of the first inverter is stuck at 0 or 1, we need to observe the output behavior for a changing input.

7.5.2 Software design for testability

Software design for testability means designing software that is prone to fail when it is faulty (Friedman and Voas, 1995), so that testing could reveal faults such as nonexecutable paths, infinite loops, incorrect logic and incorrect input/output processing. Software testability requires programs containing constructs able to yield an incorrect program state if the constructs themselves are incorrect, and programs able to propagate their incorrect states into software failures.

Software testability decreases because of information loss, i.e. internal information used by the program which is not communicated in the program output. For example, software modules that use local variables may have errors that go undetected during functional testing because the value of those variables is not used during testing. This is analogous to the situation in Figure 4(a) where the output of the first amplifier is not observable from the circuit output.

Information loss can be implicit or explicit. Implicit information loss occurs when two or more different input parameters yield the same result. This is the case of boolean operators, trigonometric functions, and modulo n operations, among others. Explicit information loss occurs when variables are not validated during execution or at the end of execution. This is the case, for example, when using local variables in a software module.

To improve software testability:

- 1. Decompose the software specification to reduce the chances of data state error cancellation across software modules (routines or subprograms). Modules including subfunctions where different inputs produce the same output require more testing.
- 2. Minimize the reuse of variables by declaring more variables, even though this reduces performance and requires more memory.
- 3. Use more out parameters: insert write statements to print internal information; treat local variables as out parameters during testing in order to observe any parameter affecting the output; or insert assertions, which check internal information during program execution.

7.6 DESIGN FOR RELIABILITY

A product is reliable when it works without failure under specified conditions for a stated period. This period is given in terms of time for continuous operation or in terms of number of cycles for intermittent work, e.g. in switches, relays and rechargeable batteries. Even as for other qualities considered in this chapter, reliability must be designed into the product because it is not inherent in designs aimed to achieve functional performance alone. Adding reliability to an already designed product implies redesigning it.

7.6.1 Causes of product deterioration

Product malfunctioning can be partially avoided by considering during design causes that can deteriorate the product behavior. These causes must be anticipated to some extent and their effects controlled to render the product safe by design.

Operational stresses: chemical, electrical, mechanical, thermal

The design input specifies the operating environment for the product. The components and parts involved in the potential solutions being evaluated are subjected to stresses derived from the function they implement and from the environment resulting from the external conditions and from their own operation.

The actual environment can include: humid air, disinfectants, sterilants, cleaners, blood (even in noncontact devices), sweat and skin chemicals, different gases, spilled liquids (saline, parenteral solutions, coffee, soft drinks), dust, dirt, and so on. The electrical environment includes electric and magnetic fields, and high voltages and currents. The mechanical environment can include, for example, shocks from dropping, vibration (from tools, in ambulances and helicopters, during transportation), weight (e.g. from stacking) and high or low pressure. Thermal stresses derive from high or low temperatures, which can be locally different from the room temperature. Time is an inevitable factor, as any material ages even in the best environment.

Errors, misuse and tampering

The reliability of manufactured products depends on their storage, transportation, installation, maintenance and use. This is one reason that makes labeling (including instructions) and human factors so important. Products can be used in an unexpected way: patient cables knotted to prevent their movement, plastic tubing held by sharp wires, power cords extended by underrated wires or cords lacking a safety conductor, ventilation openings in equipment obstructed by other equipment or their manuals, controls operated by a hand tool (strong, sharp, sticky) or with excessive torque, duty cycles extended beyond specification, and so on. Operating procedures may be altered to save time, thus overlooking safety procedures. Maintenance may be careless, e.g. skipping schedules, using lower quality replacements or failing to make adjustments. Some patients misuse or tamper with products.

7.6.2 Failure modes, mechanisms and causes

A failure mode is a particular way in which a part can fail to perform its specified function. A failure mechanism is a process or deficiency causing the failure mode. A failure cause is the agent activating the failure mechanism. Medical device failures are quite common. The FDA determined that 9.3 % of the fatalities and 12.4 % of the serious injuries reported in 1991 were due to device failures. The bulk of the remaining incidents were due to user problems. As a result, medical devices contributed to an estimated 49 fatalities and 663 serious injuries in 1991 due to design-related problems in class II and III devices.

Chemical failures

Dirty surfaces make coating and bonding difficult. Polar and nonpolar chemical contaminants can corrode electronic circuits, contacts and connectors. The results are short and open circuits and current leakage. Polar contaminants, such as deposits from solder flux and finger salts dissociate into ions in the presence of moisture. These ions can migrate and reduce insulation. Ionic contaminants are particularly dangerous in implants because these work in a humid environment. Nonpolar contaminants include greases, oils and rosin, which can form insulating films, which lead to intermittent or open contacts. Fingerprints are sources of salt, grease and oils, and must be avoided by using gloves or cleaned. Oxide formation in connectors, plugs, sockets, switches, relays, solenoids, etc., leads to high resistance failures.

Electrical failures

The most common electrical failures are short and open circuits resulting from defective contacts or terminals, poor soldering techniques, and damaged and burned out components. Electric current can corrode junctions of dissimilar metals.

Fixed resistors fail open when overheated or overstressed, electrically by overvoltage or mechanically by shock and vibration. High value resistors "fail" when their terminals are close and dust, humidity or dirt provides an alternative path with lower resistance. Hot carbon composition resistors are flammable, thus posing a danger in the presence of explosive or inflammable gases. Wipers in potentiometers wear after use, leaving the contact open, and worn away particles can contaminate the circuit.

Capacitors fail when the dielectric breaks down because of overvoltage, excessive heating or chemical contamination. The common final result is a short circuit. High temperatures shorten capacitor life, and hence capacitors must be placed far away from heat-dissipating parts, such as power components. Electrolytic capacitors fail if the electrolyte leaks, for example because of a defective or overstressed package. Electrolytic capacitors with polarized terminals fail or perform differently when the voltage polarity is reversed. Some capacitors, particularly ceramic and glass, fail because of the thermal expansion of encasement materials used for environmental protection of the inner device and from moisture trapped between them. Ceramic capacitors are sensitive to vibration. Lead wires in some paper, mica and film electrodes can easily detach under mechanical stresses, leading to an open circuit.

ICs fail in a variety of modes, so that critical medical applications require screened components, i.e. components with 100 % inspection. Input and output IC terminals may be open or short circuited to ground or any power supply terminal. ICs are particularly vulnerable to

excessive temperature during soldering or operation, corrosion of lead wires by soldering flux residuals, electrostatic discharge and ionizing radiation. Fungi are known to attack polymides used in IC packages, releasing in addition corrosive waste products, such as hydrochloric and sulfuric acid. Some failure mechanisms in ICs are; diffusion and oxide defects, dielectric breakdown, electromigration, ionic contamination and alloy formation in metallic bonds.

Mechanical failures

Typical mechanical failure modes are: fatigue, deformation, yielding, creep, jamming, buckling, bonding loosening, imbalance, property change, wear, fluid leaks and clogs. Fatigue, excessive wear, static overload, and corrosion combined with high stress can all result in fracture, and subsequent unplanned load transfer to other components, which can lead to further failures. Deformation, yielding, creep and wear can misalign parts, hinder mechanical movements or render them imprecise, loosen joints and produce leaks. Fatigue leads to spring failure.

7.6.3 Failure analysis

21CFR820.30(g) in the FDA Quality System regulation, states "Design validation shall include software validation and risk analysis where appropriate." The following are some failure analysis techniques useful for evaluating the reliability (and safety) of alternative designs and for validating a specific design (Ireson and Coombs, 1996). Failure analysis seeks to identify design weaknesses that lead to risks and subsequently correct them or, if it is impossible, provide controls such as alarms and labeling to reduce the risk.

Fault tree analysis

Fault tree analysis (FTA) is a topdown process that examines the system in order to identify a critical failure (termed top event) and in order to determine the combination of faults in the next lower level in the system leading to the top event. The analysis is repeated for the next level until arriving at the basic events. Top events can be very broad, e.g. the alarm does not sound, the gas flow is interrupted, the tube leaks, electric shock. Basic events can include human errors and environmental factors in addition to component failures.

A fault tree is a graphical model that describes the combination of events (normal or faulty) that lead to the top event, within some predefined boundaries. Fault trees use a group of symbols for gates (e.g. AND, OR, inhibit gate—hexagon—) which describe the logical combinations of input events that result in a specific output event, and another group of symbols to describe events, which include: the rectangle (event represented for a gate), circle (basic event), diamond (fault events whose causes are not further studied—environmental, poor maintenance—) and oval (conditional event for inhibit gate).

Figure 7.12, for example, shows the fault tree for the event "No power supply during normal operation" in a system including a back-up battery supply. The tree is constructed by starting at the top: there is no power available when both the main and the back-up supply fail, and the failure can result from a missing energy source or a failure in any of the elements supplying the energy to the internal circuits from the power source. The detail level in a fault tree depends on the information available about the reliability of the different system components.

Henley and Kumamoto (1995) analyze fault trees for systems which include sensors, alarms and feedback loops.



Figure 7.12 Fault tree describing the combination of failures able to interrupt the power supply to equipment that includes a back-up battery supply.

Failure modes and effects analysis

Failure modes and effects analysis (FMEA) proceeds bottom-up: from the different failure modes of each individual part and their symptoms, to the system consequences of each failure. Since FMEA analyzes each component separately, with the other components assumed to work perfectly, FMEA is unable to identify critical combinations of component failures.

After investigating the effect of a given failure mode on the whole system, the next step is to analyze how critical is the resulting system failure, hence the name "failure modes, effects and criticality analysis" (FMECA). One method for such analysis follows the steps for risk assessment described in Section 3.3.9: failure frequency (using statistical data or a qualitative ranking: frequent, probable, occasional, remote, improbable); effect severity ranking (catastrophic, critical, marginal, negligible/minor); and risk index derived from the two precedent factors.

7.6.4 Reliability analysis and prediction

Reliability can be mathematically analyzed from its definition as the probability that an item performs a required function, under specified conditions, for a stated period. A high reliability means a high probability, i.e. close to 1, of performing as desired. In other words, units of that particular device seldom fail during the period considered. The failure rate λ is the number of failures of an item per unit measure of life (time, cycles), normalized to the number of surviving units. If in a time interval dt, Nf(t) units from a batch of N fail and Ns(t) survive, and life is measured in time units, the failure rate is

$$\lambda(t) = \frac{1}{N_{\rm s}(t)} \frac{\mathrm{d}N_{\rm f}}{\mathrm{d}t}$$
(7.1)

The reliability at any time t as a probability, is then

$$R(t) = \lim_{N \to \infty} \frac{N_{\rm s}(t)}{N}$$
(7.2)

N will always be finite in practice. Therefore, R(t) can only be estimated. Since at any interval between t = 0 and any time later t, units either survive or fail,

$$N = Ns(t) + Nf(t)$$
(7.3)

Substituting into (7.2), differentiating and applying (7.1) yields

$$\frac{\mathrm{d}R(t)}{\mathrm{d}t} = -\frac{1}{N}\frac{\mathrm{d}N_{\mathrm{f}}(t)}{\mathrm{d}t} = -\frac{\lambda(t)N_{\mathrm{S}}(t)}{N} = -\lambda(t)R(t)$$

Solving for R(t),

$$R(t) = \exp\left(-\int \lambda(t)dt\right)$$
(7.4)

Therefore, the reliability can be calculated from the failure rate, usually determined from experiments.

Failure models

The experimental study of the failure rate of many units of a device for extended periods, shows the trends in Figure 7.13, often termed the bathtub curve because of its shape, regardless of the

type of device. Units with obvious defects are not considered. Some units of the initial population fail shortly after they start to work, because of early failures, or break-in failures, leading to the so called "infant mortality." Early failures are usually due to microscopic defects in materials (cracks, dirt or impurities in insulation, coating, or structural materials, joints and connections) and to incorrect adjustments or positioning, that remain undetected after (inadequate) quality control. Since electrical, mechanical, chemical and thermal stresses in the operating environment sometimes exceed those during product test, normal units withstand those conditions but inferior units fail.



Figure 7.13 The failure rate of many different devices shows the same trend, which determines three stages in a product life: infant mortality, useful life and wear out stage. The causes of failure are different at each stage.

The flat segment in Figure 7.13 corresponds to the device's useful life. The failure rate during this period is almost constant (and hopefully small) and is due to chance failures (intrinsic or stress-related failures) which result from randomly occurring stresses, the random distribution of material properties and random environmental conditions. Chance failures are sometimes termed random failures, but early failures are random too, i.e. nonsystematic. Chance failures are present from the beginning but early failures predominate at that stage.

Some time after placing different units of a device in service, they start to fail one after another at an increasing rate. This is the wear out stage, wherein parts fail because of the deterioration caused by thermal cycles, wear, fatigue, or any other condition that causes weakening under normal use. In this stage there are still chance failures, but wear out failures predominate.

One approach to reliability estimation is to model the failure rate in Figure 7.13 using three separate statistical distributions for λ (Middendorf and Engelmann, 1998). If during the useful life, λ is assumed constant, then from (7.4)

1.

$$R(t) = e^{-\lambda t} \tag{7.5}$$

which means that for chance failures the reliability decreases exponentially with time. λ is commonly calculated from experiments that determine its reciprocal, which is called the mean time between failures (MTBF) and has units of time,

$$MTBF = m = 1/\lambda \tag{7.6}$$

For example, if 100 units are tested for, say, one month and after replacing early failures 5 units fail, MTBF = $(100 \times 30 \times 24 \text{ h})/5 = 14400 \text{ h}$ and $\lambda = 7 \times 10-5/\text{h}$. Tables for different components list λ in units of failures per million hours. Since the number of units tested and the operational or experimental conditions for determining the MTBF can be different, the values of λ for the same component in different data banks disagree, sometimes by orders of magnitude. The MTBF for a device must exceed its expected lifetime according to the desired reliability. For example, to achieve R = 0.9990, from (7.5) m = 103 t, i.e. the MTBF for a device expected to work (continuously) for, say, 5 years must be 5000 years.

During the break-in stage, defective units fail because of early failures, at a failure rate λd , and good units fail at a rate λ . One method to avoid early failures after product delivery is to force those failures by using "burn-in". This procedure subjects all the units to operational conditions (voltage, temperature, humidity) slightly more stringent than the rated conditions, until all the defective units have failed. The mean burn-in time (BIT) required is

$$BIT = \frac{1}{\lambda_d} \left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{N_d} \right)$$
(7.7)

where λd is the failure rate of the defective units and Nd is the number of defective units, neither of which is known. Nevertheless, BIT is relatively insensitive to Nd, so that a small number of units tested, which implies a small number of defectives, can be compensated for by increasing the burn-in time. If the decision is to remove all the units whose failure rate is $k\lambda$, then (7.7) estimates BIT from λ , and after a burn-in time several times higher (depending on k), no defective unit would probably remain. MIL-STD-883 "Test Methods and Procedures for Microelectronics" includes burn-in screening methods used by many semiconductor manufacturers.

The distribution of failure times in the wear out stage is approximately gaussian. Good design aims to delay the onset of this stage past the product lifetime. Preventive maintenance can identify parts that begin to deteriorate and replace them by good parts.

Another method of reliability analysis is to approximate the actual failure rate curve for the device by a single statistical distribution instead of three separate distributions as above. The Weibull distribution fits well to experimental failure data of mechanical devices and brittle materials (ceramics).

Product reliability calculation

The reliability of a product depends on that of the different parts it is built from. Failures in some parts may result in product failure whereas failures in other parts may not. If a failure in any of several independent parts causes the system to fail, those parts are functionally arranged in series, Figure 7.14(a), though they can be physically in parallel (e.g. the valves of different gas cylinders in anesthesia). The system will function only if each of the parts functions, and therefore the total reliability will be the joint probability of survival of the n parts,

$$R_{\rm s}(t) = R_1(t)R_2(t)\cdots R_n(t) = e^{-\left(\lambda_1 + \lambda_2 + \dots + \lambda_n\right)t}$$
(7.8)

Hence, the overall failure rate is

$$\lambda_{\rm S} = \lambda_1 + \lambda_2 + \dots + \lambda_n \tag{7.9}$$

meaning that a single element with a high failure rate makes the system unreliable. The same applies to a single part having different failure modes: when any of these failures occurs, the part fails.



Figure 7.14 (a) A system with functional parts in series fails when any of the parts fails. (b) A system with functional parts in parallel works as long as there is a working part.

If a system failure requires the simultaneous failure of several parts, then these parts are functionally arranged in parallel, Figure 7.14(b). The system functions as long as one of the parts functions. Considering the unreliability, i.e. the probability of failure, instead of the reliability, yields

$$U_{\mathbf{p}}(t) = U_1(t)U_2(t)\cdots U_n(t)$$

Since U(t) = 1 - R(t),

$$R(t) = 1 - \left(1 - e^{-\lambda_1 t}\right) \left(1 - e^{-\lambda_2 t}\right) \cdots \left(1 - e^{-\lambda_n t}\right)$$
(7.10)

For the particular case when all parts have the same failure rate λ ,

$$R(t) = 1 - \left(1 - e^{-\lambda t}\right)^n \tag{7.11}$$

Actual systems are combinations of series, parallel and other functional relations between parts, not always amenable to simple calculation. Some software packages for reliability analysis (mostly per MIL-STD-217E) are: ARM, CARP, Relex, Reliability PREDICTOR, Rel Plus, RL ORACLE and RPP.

7.6.5 Reliability in product design

Ensuring reliability by inspecting and testing the finished product for defects is inefficient because it does not prevent chance failures, is sometimes destructive and is expensive due to the required tests. Manufacturing process control is not enough to produce reliable products. The so called "1–10–100" law states that product defects detected during design can be corrected at 10 % of the cost of product defects detected during manufacturing; product defects detected during manufacturing can be corrected at 10 % of the cost of defects detected in a marketed product. Numerical precision aside, the point is that product design must consider reliability instead of relying on defect detection solely by testing.

Guidelines for reliable designs

The design specification includes the expected lifetime for the product. This determines the MTBF to achieve. In addition, we must avoid early failures. Some of the techniques to improve human factors, compatibility, manufacturability and testability also improve reliability, both by themselves and also because a product designed to be user-centered, compatible, easy to manufacture and test is obviously more reliable than a system lacking these qualities. It turns out, therefore, that reliable products are not necessarily more expensive, even without considering liability risks.

Design for maintenance

Maintenance is instrumental in preventing wear out failures which thwart safety and performance. Sawyer et al. (1996) report several problems often encountered by maintenance personnel. To prevent them, consider:

- 1. Label, code or number components clearly.
- 2. Design adequate self-diagnostic capability.
- 3. Facilitate visual and tactile part location.
- 4. Keep screws and other parts easy to reach or manipulate.
- 5. Arrange components logically.
- 6. Avoid the need for using uncommon tools.
- 7. Design for easy cleaning. Avoid the deleterious effects of corrosive cleaners.
- 8. Use durable materials for user interface.

7.7 EVALUATE ENERGY AND INFORMATION PROCESSES

In order to evaluate the processes that implement the functions and subfunctions outlined in conceptual designs, we need to consider alternatives for each process and criteria to compare them. The ultimate criteria are safety and efficacy, complemented by other criteria considered in Chapter 3 and in the sections above. This section discusses specific criteria which derive from basic engineering principles and the laws that sustain them.

Electromedical devices usually supply their internal components with dc voltages. Some displays and other actuators need ac voltages, commonly derived from specific ICs. The fewer the number of different supply voltages a system needs, the better. Many precision analog ICs need symmetrical power supplies ($\pm 12 \text{ V}, \pm 15 \text{ V}$) whereas most digital ICs need only standard single supply voltages (5 V and 3 V to 3.3 V). EMC is better achieved by using separate power supplies for analog, digital and power circuits, even if their supply voltage level is the same. There are two basic alternatives for supplying electromedical devices: power supplies and batteries.

The common power input for nonimplanted or portable medical devices is the ac voltage provided by the local electrical utility. Figure 7.15 shows that the ac input voltage is stepped down by a transformer, rectified by power diodes, filtered by capacitors, regulated and filtered again. The transformer also isolates the internal circuits from the power line because the power link between primary and secondary is through a magnetic field. Since the stray capacitance between primary and secondary limits that isolation, transformers for medical equipment perform better with separate windings on opposite sides of a magnetic core. The filter following the rectifier reduces the ripple. The voltage regulator accepts the filtered voltage and provides a smooth dc output by using an output filter.



Figure 7.15 Block diagram of a linear (top) and a switching (bottom) power supply.

Linear and switching power supplies

Figure 7.16 shows two basic methods for voltage regulation. The dc source represents the input voltage, actually a dc voltage with ripple. The output voltage is smaller than the input voltage and the regulator keeps it constant. In Figure 7.8(a), the regulator behaves as a variable resistance which changes in a way opposite to the load: when the load current increases, the regulator resistance decreases, and conversely. The power delivered to a resistive load is

$$P_{\rm L} = \left(\frac{V}{R+R_{\rm L}}\right)^2 R_{\rm L}$$

and the efficiency,



Figure 7.16 (a) Model for a linear regulator. (b) Model for a switching regulator based on a single switch, and (c) its control signal, whose duty cycle determines the output voltage.

The regulating element, a power transistor operating in the linear region, can be placed in series (as shown) or shunting the load, and dissipates the power not delivered to the load. Hence, the efficiency is low, particularly when the input–output voltage difference is large ($R_L \ll R$), and seldom exceeds 40 %. These regulators, termed linear, suit low-power loads, and loads requiring a voltage with very small ripple (e.g. most ICs, circuits for driving and biasing sensors, loudspeakers and CRT deflection systems). The bulky transformers and the filter capacitors needed to reduce the 120 Hz, or 100 Hz, ripple of the rectified ac voltage reduce the power density to less than about 0.12 W/cm2.

Figure 7.8(b) models a switching regulator based on a single switch. A signal with a controllable duty cycle turns the switch on and off, Figure 7.8(c). Assuming an ideal switch, the average voltage delivered to the load will be

$$V_{\rm L}(0) = \frac{V \times T_{\rm ON} + 0 \times T_{\rm OFF}}{T_{\rm ON} + T_{\rm OFF}} = V \frac{T_{\rm ON}}{T_{\rm ON} + T_{\rm OFF}}$$
(7.13)

Therefore, controlling the duty cycle yields the desired output voltage. If the switch were ideal $(RON = 0 \Omega, IOFF = 0 A)$, there would not be any power loss. In a real switch with drop in voltage VON when conducting, the power loss when closed is

$$P_{\rm ON} = V_{\rm ON} I_{\rm ON} = V_{\rm ON} \frac{V}{R_{\rm ON} + R_{\rm L}}$$
(7.14)

Assuming RON << RL, the efficiency is

$$\eta = \frac{P_{\rm L}}{P_{\rm L} + P_{\rm ON}} \approx \frac{V}{V + V_{\rm ON}} \tag{7.15}$$

hence high and independent of the difference between the input and output voltages. Therefore, there is no need for stepping down the input ac voltage to a value close to the output voltage. Instead, the power line voltage can be first rectified and regulated. An output transformer and filter then provide an isolated, constant dc voltage, and since both work at high frequency, they can be smaller and lighter than those in linear power supplies, where they work at power line

frequency and harmonics. The efficiency ranges from 75 % to 90 %, and the power density is higher than 3 W/cm2, but the output ripple is larger than in linear power supplies and the reliability is somewhat smaller. Since the frequency response of many electromechanical loads decreases with frequency, increasing the switching frequency improves the overall system. Nevertheless, power losses in electronic switches also increase with frequency, and transients coupled to power supply lines increase too, thus requiring an input line filter. Overall, the best performance is usually obtained from a switching power supply working in the range from 20 kHz to 1 MHz, followed by linear regulators formed from several ICs. The overall efficiency is from 60 % to 70 %.

Batteries

Batteries power implantable, portable and some hand-held medical devices. Batteries also provide back-up power for emergency supply, and supply clock, memory and alarm circuits which operate continuously. Batteries provide "clean" voltages and reduce power-line electrocution hazards, but have limited duration and capacity, and their disposal poses environmental concerns.

There are two basic battery types: primary (nonrechargeable) and secondary (rechargeable) batteries. Crompton (1996) discusses the technology, characteristics, theory, design and evaluation of many different batteries. Urquidi-Macdonald (1997) includes developments in lithium batteries. When selecting a primary battery, evaluate the following parameters against design specifications:

- 1. Battery discharge profile: initial voltage, normal voltage during discharge and circuit end voltage (i.e. minimum voltage to operate).
- 2. Current–voltage relationships: constant current, constant resistance, constant power.
- 3. Duty cycle: continuous, intermittent, continuous with pulses.
- 4. Storage and service life.
- 5. Environmental conditions in storage and operation.
- 6. Physical restrictions: dimensions and weight.

Uninterruptible power supplies

Uninterruptible power supplies (UPSs) provide power during power outages. Some specific areas in health care facilities are supplied by a mandatory emergency-power system that restores power after a 10 s power outage. This event must be considered in the design of electromedical equipment to be used in those areas.

Two common methods for UPS design rely on a battery bank charged from the main power supply. In the forward-transfer design, the load is normally connected to the power line; if power fails, an automatic transfer switch connects the load to the battery bank and an inverter (dc–ac converter) circuit. In the reverse UPS design, the load is always connected to the battery (charged from the power line) and inverter, and in the event of a battery failure, an automatic switch connects the load to the power line. Reverse UPSs are preferred because they provide immunity to power line transients.

7.7.2 Thermal management

The First Law of Thermodynamics states that energy is conserved. Hence, any difference between the energy entering and leaving a system accumulates in it, either as kinetic or potential energy, or as enthalpy (internal energy plus work). Changes in internal energy lead to changes in temperature. The accumulated energy can influence the system behavior and performance. For example, chemical reaction rates are proportional to temperature, biological samples do not survive outside a given temperature range, excessive kinetic and potential energy may stress a part beyond its rated strength, and so on.

Heat disturbs electronic systems. Many parameters related to accuracy and speed, such as offset voltage, bias currents and switching speed are temperature sensitive. Reliability decreases for increasing temperatures. However, since many electronic functions can be fully described in terms of operations on signals without involving energy concepts, there is some risk of overlooking thermal management.

7.7.3 Hardware-software partitioning

The partition between hardware and software functions largely depends on unit cost, development time and cost, power and volume requirements, and system compatibility. Usually, hardware solutions are more expensive but faster and cheaper to develop than software solutions. For large-volume production, hardware solutions are less convenient because each produced unit bears the cost of the components incorporated into it. However, the development cost of software solutions is shared by the produced units.

From a performance standpoint, hardware solutions are usually faster than software solutions, and analog functions faster than digital functions, yet sometimes use more expensive components. Software, however, is more flexible, can perform several tasks virtually in parallel and needs less power and space than hardware solutions.

7.8 OPTIMAL DESIGN

The design evaluation process considers several qualitative parameters not amenable to mathematical analysis. However, after selecting a given design alternative it is sometimes possible to determine the values for the involved parameters which yield the most favorable design, termed optimal design. If the design is described by a mathematical model, the optimal design is achieved through formal mathematical procedures. If the design is described by a physical model, the analysis of the effects of different parameter values can yield near-optimal designs, though sometimes at a high cost. Designs described by a computer model can be modified easily and may lead to acceptable results after several iterations. In any case, the separate optimization of different parts does not ensure that the whole system will be optimal. Nevertheless, separate optimal values may be a good start point for searching the globally optimal design.
7.8.1 Functional relationships in optimal design

Design optimization problems involve three different functional relationships between the design parameters (or variables) (Middendorf and Engelmann, 1998, C12). The criterion function, cost function or objective function is the mathematical relationship between the quantity to optimize and the remaining design variables. That quantity may be a single characteristic or a weighted combination of several characteristics. For example, the design of an implantable infusion pump can be optimized for minimal power consumption, but also for the minimal combination of power consumption, volume and weight. Nevertheless, there is no general method to optimize a design whose criterion function involves multiple variables. Nor is there a method to determine which characteristic should be optimized. The usual characteristics are cost, profit, yield, energy and reliability, but the design of some parts may require us to minimize variables such as temperature increase, volume, weight, or user effort (e.g. in rehabilitation devices).

The functional constraints or equality constraints are the equations which describe the physical laws implicated in the proposed design. Examples are the Law of Conservation of Mass, the First Law of Thermodynamics, Newton's First Law and Ohm's Law. The corresponding set of equations constitutes the mathematical model of the design. The number of equations must be equal to or less than the number of design variables. If there are as many independent equations as variables, we can find the corresponding value for each variable and the design is fixed. If there are more equations than variables, either some equations depend linearly on the others, or the formulation is inconsistent.

The regional constraints or inequality constraints define the acceptance limits for the design variables, expressed as inequalities. Design specifications impose regional constraints. Examples are acceptable stresses, volume, weight or cost. Physical values of parts or dimensions that must always be positive also impose regional constraints. There is no limit for the number of regional constraints.

Regional constraints can be used to search for a design which optimizes several characteristics: instead of using a criterion function involving different characteristics, select the most important characteristic for the criterion function and determine the effect of varying the limits for the other characteristics. The result is a set of optimal designs that reveal the trade-offs involved.

7.8.2 Unconstrained design optimization

When there are no constraints, the optimization problem reduces to finding the extrema (maxima and minima) of the objective function and the value of this function at the extrema.

The derivative method

The extrema can be determined by setting the first derivative of the objective function with respect to the independent variable equal to zero. The values of the objective function at the extrema will show which are maxima and which are minima, and, from them, the optimal design value. Since the range of values for the design variables is finite, it is convenient to also evaluate the objective function at the endpoints to verify whether or not the endpoints correspond to extrema for those ranges.

If there are more than one independent variable, then set the partial derivative of the objective function with respect to each variable equal to zero and solve the resulting system of equations. Whenever the design variables corresponded to component values which are standardized, first select the standard value close to the mathematical solution for the first variable solved, and then find the values for the remaining variables.

Search methods

Search methods can find optimal values even if the functional relationship cannot be expressed by an analytic equation. They are also appropriate when there are many design variables. Search methods seek the optimum by trying different values of the independent variable(s). They differ in the criterion to select the step size difference between trial points and search direction, and their efficiency depends on the problem. In problems involving a single variable, for example, the uniform search method, which spaces the trial points equally over the allowable range of values, is simple but inefficient. A Fibonacci search—the Fibonacci sequence is F0 = 1, F1 = 1, Fn = F n - 2 + Fn - 1—is quite efficient but requires us to decide in advance the number of trials n, which is difficult to guess if we do not know in advance the behavior of the function near the extremum. A golden section search, which places trial point pairs at 0.618 of each endpoint, is less efficient than a Fibonacci search but does not require any a priori knowledge or decision. An alternative method is to fit a polynomial to the function value at a few points and then find the extremum of that polynomial. The actual extremum will probably lie close to that.

The simplest method to find the search direction, for single- or multivariate problems, is the steepest descent/ascent (gradient) method, which follows the direction of maximum local slope until reaching the extremum. Some methods to improve its efficiency are: the conjugate gradient method, the modified Newton's algorithm and its modification by Marquardt (Arora, 1989, C5). Once the search direction is known, determining the step size is a single variable process.

7.8.3 Constrained design optimization

If, in addition to the objective function, there are functional constraints but no regional constraints, we can solve the functional constraints for each of the design variables and substitute them into the objective function. This procedure eliminates as many variables from the objective function as there are (independent) functional constraints. The derivative of the resulting objective function with respect to the remaining variables yields the extrema and one of these is the optimal value for the objective function. Replacing the values for the design variables at the corresponding extremum in the functional constraints yields the optimal values for the variables initially solved. An alternative method is to use Lagrange's multipliers (Arora, 1989, S3.4).

If there are functional and regional constraints, and the criterion function and all constraint functions are linear functions of the design variables, the optimization problem is termed a linear programming problem, which can be solved by algorithms such as Simplex or Karmarkar's (Edgar and Himmelblau, 1988, C7) (Arora, 1989, C4). If either the criterion function or any of the constraints is a nonlinear function, one alternative is to change the regional constraints into equalities by introducing additional, arbitrary variables termed slack variables, and then apply Lagrange's method. Since the Lagrange method involves taking derivatives, the slack variables are always introduced as squared quantities to ensure that they will not disappear

in the differentiation process. Another alternative to find extrema in nonlinear problems is to use one of the search methods above.

The NEOS Guide is a web site set at Argonne National Laboratory (wwwc.mcs.anl.gov/home/otc/Guide) which has information about optimization methods, software programs and packages by categories, case studies and test problems.

7.9 REVIEW QUESTIONS

- 7.1 Describe some trade offs in the design of a thermometer for measuring the temperature of the human body depending on whether it is for home or clinical use.
- 7.2 Search a picture or sketch of the front panel of a current intensive cardiac care unit monitor and compare its controls and displays with those of a similar monitor installed before 1980.
- 7.3 Suppose you are designing a device that includes a rotary snap control which also sets the volume for an alarm. Find how much torque the control should withstand.
- 7.4 Describe the tasks involved in recording a 12-lead ECG from a resting patient and discuss which procedures can be automated.
- 7.5 Report a recent case of adverse effect from a medical device attributable to user error and discuss design solutions able to prevent it.
- 7.6 Describe environmental data of interest in the design of a portable defibrillator for "911" emergency services.
- 7.7 Describe some criteria to evaluate the adequacy of accessories of medical devices to user needs in the operating environment.
- 7.8 List four important criteria in designing and testing labeling for medical devices.
- 7.9 Compare the power cords needed for electromedical equipment to be exported to Germany, Switzerland or the United Kingdom.
- 7.10 Describe the diameter, dimensions and designation of cardiac catheters.
- 7.11 Describe the Diameter Indexed Safety System (DISS).
- 7.12 Explain the merger doctrine.
- 7.13 Compare the manufacturer specifications of a power line filter intended to achieve EMC for medical equipment with those for industrial equipment.
- 7.14 Describe the relevant cost factors in the production of a disposable, sterile medical device.
- 7.15 List the requirements for a cleanroom of class 10,000.
- 7.16 List some guidelines to consider to simplify manual assembly.
- 7.17 Describe three situations which suggest subcontracting.
- 7.18 List three basic rules to improve software testability.
- 7.19 List some possible top events to consider in the fault tree analysis of an anesthesia machine.
- 7.20 A given system is made of two parts functionally in series. The first part is twice as complex as the second part, hence its MTBF is deemed half that of the second part. If the desired MTBF for the system is 2000 h, estimate the failure rate required for each part. [1/(3000 h) and 1/(6000 h)]

- 7.21 Estimate the reliability of a redundant standby system consisting of two similar units whose failure rate is 0.5 per million hours.
- 7.22 Discuss the difference between designing with a safety margin and derating.

7.10 REFERENCES

- Abramovici, M., Breuer, M. A., and Friedman, A. D. 1998. Digital systems testing and testable design. 2nd Ed. New York: IEEE.
- Alting, L. 1994. Manufacturing engineering processes. 2nd Ed. New York: Marcel Dekker.
- Arora, J. S. 1989. Introduction to optimum design. New York: McGraw-Hill.
- Boothroyd, G., Dewhurst, P., and Knight, W. 1994. Product design for manufacture and assembly. New York: Marcel Dekker.
- Bralla, J. G. (ed.) 1986. Handbook of product design for manufacturing: a practical guide to lowcost production. New York: McGraw-Hill. [To be replaced by the second edition entitled Design for manufacturability, due August 1998]
- Crompton, T. R. 1990. Battery reference book. 2nd Ed. Warrendale, PA: Society of Automotive Engineers.
- Dorsch, J. A., and Dorsch, S. E. 1994. Understanding anesthesia equipment: construction, care and complications. 3rd. Ed. Baltimore, MD: Willis & Wilkins.
- Edgar, T. F., and Himmelblau, D. M. 1988. Optimization of chemical processes. New York: McGraw-Hill.
- FDA. 1997. A primer on medical device interactions with magnetic resonance imaging system. Draft document. Rockville, MD: Office of Science and Technology. (Available from the FDA home page at: www.fda.gov.)
- Fowler, K. R. 1996. Electronic instrument design: architecting for the life cycle. New York: Oxford University Press.
- Friedman, M., and Voas, J. 1995. Software assessment: reliability, safety, testability. New York: John Wiley and Sons.
- Fries, R. C. 1997. Reliable design of medical devices. New York: Marcel Dekker.
- Henley, E. J., and Kumamoto, H. 1995. Designing for reliability and safety control. Englewood Cliffs, NJ: Prentice-Hall.
- Hordeski, M. 1995. Personal computer interfaces. New York: McGraw-Hill.
- Hornbruch, F. W. 1986. Economics of process selection. In J. G. Bralla (ed.) Handbook of product design for manufacturing: a practical guide to low-cost production. New York: McGraw-Hill. [To be replaced by the second edition entitled Design for manufacturability, due August 1998]
- Ireson, W. G. and Coombs, C. F. (eds.) 1996. Handbook of reliability engineering and management. 2nd Ed. New York: McGraw-Hill.
- Knatek, E. R. 1989. Design of solid state power supplies. 3rd Ed. New York: Van Nostrand Reinhold.
- Kroemer, K. H. E., H. J. Kroemer, and K. E. Kroemer-Elbert 1997. Engineering physiology: bases of human factors/ergonomics. New York: Van Nostrand Reinhold.
- Kroschwitz, J. 1986. Encyclopedia of polymer science: Degradation. Vol. 4. New York: John Wiley and Sons.

- Kumamoto, H., and Henley, E. J. 1996. Probabilistic risk assessment and management for engineers and scientists. 2nd Ed. Piscataway, NJ: IEEE Press.
- Leveson, N. G. 1995. Safeware: system safety and computers. Reading MA: Addison-Wesley.
- Middendorf, W. H., and Engelmann, R. H. 1998. Design of devices and systems. 3rd Ed. New York: Marcel Dekker.
- O'Brien, J. E. (ed.) 1990. Medical device packaging handbook. New York: Marcel Dekker.
- Ott, H. W. 1988. Noise reduction techniques in electronic systems. 2nd Ed. New York: John Wiley and Sons.
- Paul, C. R. 1992. Introduction to electromagnetic compatibility. New York: John Wiley and Sons.
- Pease, R. A. 1993. Troubleshooting analog circuits. Boston: Butterworth-Heinemann.
- Porter R. F., and Kwon, O. 1997. Electronic device cooling. In C. A. Harper (ed.) Passive electronic component handbook. New York: McGraw-Hill.
- Salvendy, G. (ed.) 1997. Handbook of human factors and ergonomics. New York: John Wiley and Sons.
- Sanders, M. S., and McCormick, E. J. 1993. Human factors in engineering and design. 7th Ed. New York: McGraw-Hill.
- Sawyer, D, Aziz, K. J., Backinger, C. L., Beers, E. T., Lowery, A., Sykes, S. M., Thomas, A., and Trautman, K. A. 1996. Do it by design: an introduction to medical factors in medical devices. Rockville, MD: Center for Devices and Radiological Health. Available from the FDA home page at: www.fda.gov/cdrh/humfac/doit.html.
- Scheiber, S. F. 1995. Building a successful board-test strategy. Boston: Butterworth-Heinemann.
- Trucks, H. E. 1987. Designing for economical production. 2nd Ed. Dearborn, MI: Society of Manufacturing Engineers.
- Ullman, D. G. 1992. The mechanical design process. New York: McGraw-Hill.
- Urquidi-Macdonald, M. 1997. Batteries. In C. A. Harper (ed.) Passive electronic component handbook. New York: McGraw-Hill.

9

Design Validation

John G. Webster and Ramón Pallás-Areny

Validation means "confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use can be consistently fulfilled. Design validation augments the previous definition. Design validation means establishing by objective evidence that device specifications conform with user needs and intended use(s) (21CFR820.3(z))." Therefore, whereas design verification aims to assess whether the design conforms to its specifications, design validation aims to confirm the fulfillment of the user needs and intended use(s), and may reveal deficiencies in translating them into design specifications. Figure 9.1 shows the verification and validation sequence.



Figure 9.1 Design evaluation, verification and validation are sequential activities whose results may improve goal definition for previous activities.

The provision of objective evidence can rely on the published scientific literature and/or theoretical analysis based on current knowledge. However, when these do not provide enough information, preclinical and clinical testing under actual or simulated use conditions must complement laboratory testing (Section 8.5).

Preclinical testing aims to ensure that the device or process has a reasonable probability of being effective and does not cause undue risk to the patient. Preclinical testing may involve measurements in vitro and in vivo (in animals) to evaluate either the safety or the effectiveness, or both. Clinical testing involves a number of patients and it is performed by clinical investigators (usually physicians).

9.1 PROTOTYPE

The Quality System (QS) regulation mandates that "Design validation must be performed under defined operating conditions on initial production units, lots, or batches, or their equivalents."(21CFR820.30(g)). This poses a dilemma because, on the one hand, the manufacturer may wish to delay the investment necessary to mass produce the device until it has proved to be safe and effective, but on the other hand we need the device to assess whether or not it is safe and effective. This dilemma is often solved by first designing a prototype and, once validated, mass producing the final device.

Prototypes are essential in establishing the feasibility of new products and also in improving their design, because of the insight they provide about the interaction between the different parts that make the product, including the user interface. However, prototypes may not be entirely equivalent to the final device. This implies that the results from prototype testing should be analyzed knowing that there may be predictable differences in the final device, and that final devices may have to be tested again. Software devices and software which is part of a device must be tested once completed and after every change. Even minor software changes, though intended as improvements, can lead to adverse events.

9.2 BIOMEDICAL DESIGN VALIDATION

Design validation is a comprehensive process that looks back to the health care need that the designed product should meet. Hence, design validation requires us to reconsider the basic elements involved in establishing that need, relative to the patient, the user and the intended use environment (Section 3.1). Design validation must also consider those factors relative to materials, manufacturing processes, packing, sterilization and labeling that may influence the performance of the final product. Therefore, knowledgeable personnel involved in manufacturing and packing should also participate in design validation.

Validation must be planned during the design process. The design team must identify the functional and safety characteristics that cannot be assessed by laboratory testing, and subsequently establish validation methods and acceptance criteria for them. The validation methods needed depend on the particular product. 21CFR820.30(g) requires that "The results of the design validation, including identification of the design, method(s), the date, and the individual(s) performing the validation, shall be documented in the Design History File."

9.2.1 Product evaluation by users

Biomedical products involve three groups of people: customers, users and patients, whose needs are different, and not always complementary or necessarily compatible. Safety, effectiveness and cost concern the three groups, each with different emphasis. Customers, for example, loosely defined as those purchasing the product, are very concerned about acquisition, operation, maintenance and disposal costs. Users often emphasize safety for themselves and the patient. Patients expect effectiveness. For some biomedical products, the customer, user and patient are the same person. Usually, however, they are different and the user is the one in the best position to contribute to design validation.

9.2.2 Functional performance validation

The functional performance of biomedical products is validated from clinical studies based on tests performed in the intended use environment. Thus the product is exposed to the environmental conditions (e.g. temperature, humidity, shock and vibration, corrosive atmospheres) expected in normal use. These tests may be complemented by the available scientific literature and historical records, but cannot be replaced by them. Depending on the product, it may be tested on a simulator, on animals, on a small number of patients, or the CDRH may require a clinical trial. For example, external defibrillators are tested in dummy patients, but implantable cardioverter defibrillators (or any other new implants) need clinical trials. However, bench testing of implantable defibrillators using actual signals from patients can supplement clinical trials. Simulators are considered to be accessories to the device they test.

9.2.3 In situ safety and compatibility validation

The electrical, thermal, mechanical, chemical, radiation, etc., safety of devices usually can be determined by laboratory tests (Section 8.5). System compatibility validation may require in situ tests, for example when the device interfaces another device already installed in a health care facility. EMC validation requires in situ tests when the actual electromagnetic environment of the operating device is difficult to reproduce or simulate, for example health care facilities with installed wireless communication systems. Devices intended for use in the MR environment should be tested in that environment.

9.2.4 Biological evaluation of medical devices

Medical devices shall not include any toxic material contacting the body. Device materials whose chemical and physical characteristics are uncertain must be evaluated in order to demonstrate that the medical device is biologically safe. This evaluation requires us to select the tests to perform, conduct them and analyze the results. In product reviews, the FDA applies the ISO 10993-1 standard, which is based on toxicity evaluation principles:

Test selection for biological evaluation of medical devices

The selection of tests for biological evaluation of a medical device considers the chemical characteristics of device materials and the nature, degree, frequency and duration of its exposure to the body (Section 3.3.1). Common tests address acute toxicity, subchronic and chronic toxicity, irritation to skin, eyes and mucuous surfaces, sensitization, hemocompatibility, genotoxicity, carcinogenicity and effects on reproduction including developmental effects. Some specialized devices may need additional tests, e.g. for specific target organ toxicity, such as neurotoxicity and immunotoxicity. The FDA also recommends consideration of tests to detect any pyrogenic chemical component in device materials.

The FDA guidance to select the appropriate tests to evaluate the adverse biological response of medical devices is briefly designated as G95-1 (FDA, 1995a). The flow chart in Figure 9.2 helps to determine when the FDA requires a toxicity test. The approach to test selection uses a matrix which consists of two tables. This matrix is a modification of that in ISO 10993–1, hence the name FDA–modified matrix. Table 9.1 lists the initial evaluation tests for consideration. Some of the FDA modifications are the requirement of acute, subchronic, chronic toxicity and implantation tests for surface devices permanently contacting mucous membranes (e.g., IUDs) and of irritation, systemic toxicity, acute, subchronic and chronic toxicity tests for externally communicating devices, tissue/bone/dentin with prolonged and permanent contact (e.g., dental cements, filling materials etc.). Table 9.2 lists supplementary tests for consideration.



Figure 9.2 Flow chart to determine when the FDA requires a toxicity test for a medical device. Adapted from (FDA, 1995a)]

Table 9.1 Initial toxicity evaluation tests for consideration (FDA, 1995a). Contact duration: A limited, 24 h; B prolonged, 24 h to 30 d; C permanent, more than 30 d. Biological effect: 1 Cytotoxicity; 2 Sensitization; 3 Irritation or intracutaneous reactivity; 4 System toxicity (acute); 5 Subchronic toxicity (subacute toxicity); 6 Genotoxicity; 7 Implantation; 8 Hemocompatibility.

Device categories					Bio	logi	cal e	ffect		
H	Body contact	Cont	1	2	3	4	5	6	7	8
		act								
		durat								
		ion								
		А	Х	Х	Х					
	Skin	В	Х	Х	Х					
		С	Х	Х	Х					
		А	Х	Х	Х					
Surface devices	Mucous membrane	В	Х	Х	Х	0	0		0	
		С	Х	Х	Х	0	Х	Х	0	
		А	Х	Х	Х	0				
	Breached or compromised surfaces	В	X	Х	Х	0	0		0	
		С	Х	х	Х	0	Х	Х	0	
		А	Х	х	Х	х				Х
	Blood path, indirect	В	Х	х	Х	х	0			Х
		С	Х	Х	0	х	Х	Х	0	Х
		А	х	Х	Х	0				
External communicating devices	Tissue/bone/dentin communicating1	В	X	X	0	0	0	Х	X	
		С	х	х	0	0	0	x	х	
		Ā	X	x	x	x		0		х
								2		
	Circulating blood	В	Х	Х	Х	Х	0	Х	0	Х
		С	Х	Х	Х	Х	Х	Х	0	Х
		А	Х	Х	Х	0				
	Tissue/bone	В	Х	Х	0	0	0	Х	Х	
Implant		С	Х	Х	0	0	0	Х	Х	
		А	Х	Х	Х	х		Х	х	
	Blood	В	Х	Х	Х	х	0	Х	х	Х
		С	Х	Х	Х	Х	Х	Х	Х	Х

Notes: x = ISO evaluation tests for consideration. o = Additional test that may be applicable. (1) Tissue includes tissue fluids and subcutaneous spaces. (2) For all devices used in extracorportial circuits

Table 9.2 Supplementary toxicity evaluation tests for consideration (FDA, 1995a). Contact duration: A limited, 24 h; B prolonged, 24 h to 30 d; C permanent, more than 30 d. Biological effect: 9 Chronic toxicity; 10 Carcinogenicity; 11 Reproductive/developmental; 12 Biodegradable.

Device categories			В	Biological effect		
Body contact		Contact duration	9	10	11	12
	Skin	A B C				
Surface devices	Mucous membrane	A B C	0			
	Breached or compromised surfaces	A B				
		C A	0			
	Blood path, indirect	B C A	Х	X		
External communicating devices	Tissue/bone/dentin communicating	В				
	Circulating blood	C A B	0	Х		
	en e	C A	Х	X		
Implant devices	Tissue/bone	B C A	Х	X		
	Blood	B C	X	X		

Notes: x = ISO evaluation tests for consideration. o = Additional test that may be applicable

The immunotoxicity testing framework guidance (FDA, 1996a) specifically addresses immunotoxicity testing and complements G95-1. This guidance provides a flow chart (Figure 9.3) to help in establishing the need for immunotoxicity testing and three tables (Tables 9.3 to 9.5) to assist in deciding what specific testing should be performed. According to Figure 9.3, the FDA requires immunotoxicity tests whenever a device contains new potentially immunotoxic material not previously characterized for the intended use and population.

If immunotoxicity tests are necessary, Tables 9.3 to 9.5, used in sequence, lead to the type of testing which may help in evaluating product safety. Contact types and duration are classified as in Tables 9.1 and 9.2. Materials are placed in one of four broad categories or in an

additional category for other materials, such as low molecular weight chemical stabilizers or cross-linking agents for polymers and degradation products, or new materials.



Figure 9.3 The decision about the convenience of immunotoxicity testing depends on the data available about the considered material for the intended use. Adapted from (FDA, 1996a)

9.123

Table 9.3 Potential immunological effects of devices and constituent materials (FDA, 1996a). Contact duration: A limited, 24 h; B prolonged, 24 h to 30 d; C permanent, more than 30 d. Immunological effect: 1 Hypersensitivity, 2 Inflammation, 3 Immunosuppression, 4 Immunostimulation, 5 Autoimmunity. Materials are designed as follows: p = plastics and other polymers, m = metals, c = ceramics, b = biological materials, x = all above materials.

Device categories		Immu	nologic	al effe	cts	
Body contact	Contact duration	1	2	3	4	5
Surface devices						
- Skin	А	pmb				
	В	pmb				
	С	pmb				
- Mucous membrane	А	pmb	pmb			
	В	pmb	pmb	mb		
	С	pmb	pmb	mb	mb	mb
- Breached or compromised	А	pmb	pmb			
surface						
	В	pmb	pmb	mb	mb	mb
	С	pmb	pmb	mb	mb	mb
External communicating devices						
- Blood path, direct and indirect	А	pmb	pmb			
	В	pmb	pmb	mb	pmb	mb
	С	pmb	pmb	mb	pmb	mb
- Tissue/bone/dentin	А	pmb	pmb			
communicating						
	В	pmb	Х	mb	pmb	mb
	С	pmb	х	mb	pmb	mb
Implant devices						
- Tissue/bone, blood, other body	А	pmb	pmb	mb		
fluids						
	В	pmb	х	mb	pmb	mb
	С	pmb	х	mb	pmb	mb

Table 9.4 shows some common responses associated to each immunological effect in Table 9.3. The tests to be selected should address the identification of critical responses and of those noncritical responses deemed appropriate, for example when a critical test is positive. Table 9.5 shows some representative tests suitable to study the immune responses listed in Table 9.4.

Table 9.4 Classification of specific immune responses associated with potential immunological effects (FDA, 1996a). Immune responses: C = critical, NC = noncritical, NA = nonapplicable or not needed; T = T cells, NK = natural killer cells, M = macrophages, G = granulocytes (basophils, eosinophils, and/or neutrophils).

Immunological effects	Immune responses							
	Hysto- pathology	Humoral response	Cellular responses			Host resista	Clinical sympto	
							nce	ms
			Т	NK	Μ	G		
1 Hypersensitivity	NC	С	С	NA	NA	С	NA	С
2 Inflammation	С	NC	С	NA	С	С	С	С
3 Immunosuppression	NC	С	С	С	С	NA	С	С
4 Immunostimulation	NC	С	С	NA	NC	NA	NC	С
5 Autoimmunity	С	С	С	NA	NA	NC	NA	С

Immune responses	Functional assays	Phenotyping	Soluble mediators	Other
Histopathology	NA	Cell surface markers	NA	Morphology
Humoral response	Skin testing, Lymphocyte proliferation Plaque-forming cells	Cell surface markers	Antibodies, Complement, Immune complexes	
Cellular responses				
T - cells	Skin testing, Local lymph node assay, Lymphocyte proliferation, Mixed lymphocyte reaction	Cell surface markes (helper and cytotoxic T-cells)	Cytokine patterns indicative of T cell subsets (e.g. Th1 and Th2)	
Natural killer cells	Tumor cytotoxicity	Cell surface	NA	
Macrophages	Antigen presentation, Phagocitosis	MHC markers	Cytokines (IL-1, TNF-alpha, IL-6, TGF-beta)	
Granulocytes	Phagocyotsis, Degranulation	NA	Chemokines, Bioactive amines	Cytochemistry
Host resistance	Resistance to bacteria, viruses and tumors	NA	NA	
Clinical symptoms	NA	NA	NA	Allergy, Skin rash, Urticaria, Edema, Lymphadenopat hy

Table 9.5 Representative tests, indicators and models for the evaluation of immune responses (FDA, 1996a). NA = not applicable or not needed.

Tests for biological evaluation of medical devices

The FDA does not mandate any particular test method for each biological evaluation of medical devices. Nevertheless, even as for other nonclinical laboratory studies, these tests must be conducted in conformance with good laboratory practices (21CFR58). The ISO 10993 standard has the following parts that discuss the design and interpretation of particular tests aimed to analyze a given biological effect:

Example: Design of a glucose sensor (Validation)

In vitro validation is performed to measure response time, current output from standard glucose concentration steps, linearity of output from 0 to 22 mM glucose, freedom from electroactive interferences (ascorbic and uric acids, acetaminophen, amino acids), and insensitivity of sensor output at 22 mM glucose when solution oxygen concentration varies from 2.5×10^{-4} M to 5.0×10^{-5} M (*PO*₂ range of 150 to 30 mmHg). Sensors are checked periodically for 1 to 2 weeks to guarantee stability of performance.

9.3 ANIMAL TESTING

The need for testing product safety, particularly biocompatibility, and effectiveness in animals will not probably be eliminated in the foreseeable future. The common approach to design validation is to use in vitro testing, animal testing and clinical trials. Ethical concerns (Section 3.7) call for replacement of animals and reduction in the number used and for a refinement in experimental techniques aiming to minimize their suffering.

9.3.1 Animal models and alternatives

The use of animal models relies on the similarities in structure and function between different species. Species variability, however, precludes testing devices only in animals or even invertebrates. Animal models are appropriate only up to the extent that it is possible to identify some common mechanism of action for the device in animals and humans. When there is no apparent mechanism of action, tests should involve more than one species, and some regulations mandate that one of them be a nonrodent. When there is no valid animal model for the considered disease, device effectiveness may be tested in normal animals. Some tests are performed in inappropriate animals because of cost and lack of alternatives. Implants, for instance, are tested in dogs in spite of their platelets adhering more to foreign surfaces than human platelets.

9.3.2 Design of experiments

The design and analysis of an experiment should rely on statistical techniques. An inappropriately designed experiment can yield meaningless results, which is worse than a bad analysis of good results because this can sometimes be revised to infer the correct information. Forthofer and Lee (1995), Altman (1991) and Box et al. (1978), among many others, discuss statistical tools useful in experimental design and analysis. Selwyn (1996) provides a nonmathematical perspective and outlines the following planning steps:

- 1. Definition of study objectives.
- 2. Consideration and selection among different experimental designs.
- 3. Estimation of sample size.

- 4. Writing of the protocol.
- 5. Generation of a randomization plan or schedule.

The study objectives should be specific, realistic, limited in number and explicitly written. These needs arise from the implicit assumptions in scientific experiments discussed below.

Principles of statistical design

Scientific experiments rely on the principle of causality: every effect has a cause, and research aims to infer the cause from the observation of the effect and the relations connecting them. Hence, it is important to select the outcome variable, response variable or endpoint for the experiment which clearly indicates the presence of the effect. The outcome variable may be an event or a measurable quantity, but whatever the case, there is often more than one cause influencing it. When no single cause has more influence than other causes, the outcome variable seems not to depend on any particular cause and displays random values lacking any definite trend. But if a single cause dominates over the other causes, a variation in the intensity of that cause, termed explanatory variable, is reflected in a corresponding variation in the effect as assessed by the outcome variable. Nevertheless, the other causes present, termed nonexperimental or influencing variables, will still contribute to the effect, and the outcome variable may not have a single, defined value.

The basic assumption in experimental measurements is that if the outcome variable is correctly observed many times, and there is a single predominant cause for its value, then the result is the true or exact value for the outcome variable. The observation (or measurement) procedure, however, should be correct, which means performed according to established methods deemed acceptable for the intended purpose, i.e. conforming standard procedures when available. Moreover, to obtain the true value we should ideally observe all possible outcome values. Since this is impossible, we should at least measure the outcome variable "many" times, i.e. measure a sample set which is representative of the entire set of possible outcome values. By so doing we can expect to obtain a good estimate of the true value, but the true value itself is out of reach.

The random fluctuation of measurement results about the true value are termed measurement variability and it is inherent to any observation performed with enough discrimination ability or resolution. A large variability requires a large number of observations to infer the true value because it is more difficult for opposite influences to cancel each other. Hence, a first objective in experimental design is to reduce the variability of the results, for example by controlling nonexperimental variables. In vitro experiments, for instance, permit us to control the temperature, pH, nutrients, humidity and atmosphere for the culture, so that the outcome will not depend on variations in these parameters.

Nevertheless, an unrecognized nonexperimental variable may influence the results and yield a systematic difference between the measured quantity and its true value. Systematic means that the difference does not change from time to time, provided the influencing variable remains constant, and that the difference changes predictably when the influence variable follows a definite trend. Clearly, an undetected bias yields a wrong result. Bias can sometimes be identified by repeating the experiment after changing a single factor. A second objective in experimental design is to eliminate, or at least reduce, bias in order to obtain accurate

9.128

measurements. The basic techniques to control bias and variability are: randomization, blocking and blinding.

The selection of the test sample from the target population influences both bias and variability. Testing a device in only, say, males would be an obvious experimental flaw if that device was intended for both sexes, i.e. the sample would not be representative of the entire population and it would not be possible to infer valid conclusions for the population from the sample measurements. In animal testing we need to specify the species and line of descent (strain), the age range, often a weight range, and any other factors that may influence the results, such as diet and lodging. Other possible sources of bias and variability are: season of the year, time of the day and the variability inherent to biological and chemical assays. Pilot studies involving a small number of animals can help in quantifying the influence of some of these factors.

Randomization means to allocate experimental units or specimens (animal, organ, tissue or other) to differing treatment groups or conditions according to a stochastic law, e.g. computergenerated random numbers. Randomization aims to balance the different groups with regard to nonexperimental variables. Any difference in the observed outcome for each group can then be attributed to the different treatment or condition applied, i.e. the explanatory variable. In a uniformity trial, no treatment is applied to the groups, so that any significant difference in the outcome variable can be attributed to a bias source and needs further study.

A block is a set of units which are expected to respond similarly to the same treatment. Blocking is grouping the specimens according to a nonexperimental variable which may affect the outcome variable. This reduces the variability within each group. In animal testing, blocks are defined by sex and initial weight range. The specimens are then separately randomized within each block.

Blinding or masking intents to reduce bias, particularly in subjective evaluations. In a blinded animal test, the evaluator is not aware of whether the animal has received treatment or not, and therefore his/her observations are not influenced by the expected outcome.

Sample size estimation

A basic question in experimental design concerns the minimal number of samples to evaluate in order to infer a valid conclusion. Estimations based on more specimens are more precise than those based on fewer units but there are cost, time and ethical constraints to consider. The minimal sample size depends on the statistical context, which for biomedical experiments is normally either a point estimation, an interval estimation or hypothesis testing (Selwyn, 1996, C4) (Forthofer and Lee, 1995, C4, C7 and C9).

A point estimation aims to calculate a value, termed sample statistic, as close as possible to a parameter of the population to where the sample belongs. The true parameter value would be obtained by measuring the whole population. Since this is impossible, the sample statistic is just an estimate of the parameter. If we repeat the test, because of the influence of nonexperimental variables the measured sample statistic will probably differ from that measured for the first sample. However, measuring two samples should provide more information than measuring one sample. In fact, if the variance of the population is σ_2 and the statistic follows a Gaussian distribution, then the variance of the statistic is σ_2/N , where N is the sample size. This means that if we measured the whole population (N infinite) we would obtain the true value. If we measure a single sample of size N, then we will have a probability α , calculated from normal distribution tables, that the statistic calculated from that sample, x_s , will not differ from the true value x by more than a given amount δ . In probability notation,

$$\operatorname{Prob}\left[\frac{x-x_{s}}{\sigma/\sqrt{N}}\right] < \alpha \tag{9.1}$$

If the value corresponding to a tail probability α is Z_{α} , then

$$Z_{\alpha} = \frac{x - x_s}{\sigma/\sqrt{N}} = \frac{\delta}{\sigma/\sqrt{N}}$$

and the minimal sample size is

$$N = Z_{\alpha}^2 \left(\frac{\sigma}{\delta}\right)^2 \tag{9.2}$$

For example, if we wish a 95 % confidence that our estimate deviates less than 10Ê% (either by excess or by defect) from the true value, then $\alpha = (1 - 0.95)/2 = 2.5$ % and the tables of the normal distribution yield $Z_{\alpha} = 1.96$. From (9.2),

$$N = (1.96)^2 \left(\frac{\sigma}{0.1x}\right)^2 \approx (1.96)^2 \left(\frac{\sigma}{0.1x_s}\right)^2 = 384 \left(\frac{\sigma}{x_s}\right)^2$$

If, for example, $\sigma = 0.2x_s$, then N = 15.4 and the sample should have 16 specimens.

The sample average is a statistic that follows a Gaussian distribution and therefore N can be calculated as shown. Nevertheless, we need to know σ , either from the literature, our previous experience or a pilot study. If instead of the variance of the population, σ , we use the sample variance, s, then $(x - x_s)/(s/\sqrt{N})$ does not follow a normal distribution but a Student's tdistribution and the value t α corresponding to α in the equivalent to (9.1) must be searched in the respective probability tables. In this case (9.2) underestimates N, particularly for N < 30. Forthofer and Lee (1995), Box et al. (1978), and most books on statistics, detail the estimation of averages, variances and standard deviations from samples.

An interval estimation aims to determine an interval such as [L,U], termed confidence interval (CI), which has a given probability, termed confidence coefficient or confidence probability, of containing the true population parameter. Because of random influences, each sample leads to a different CI. If the estimated parameter x_s follows a normal distribution, then the upper limit minus the lower limit is

$$\left[x_{s} + Z_{\alpha}\left(\frac{\sigma}{\sqrt{N}}\right)\right] - \left[x_{s} - Z_{\alpha}\left(\frac{\sigma}{\sqrt{N}}\right)\right] = \delta$$

Therefore, the sample size needed is

$$N = 2Z_{\alpha}^2 \left(\frac{\sigma}{\delta}\right)^2$$

which is twice that in (9.2) for point estimation. The difference between two sample means, for example, follows a normal distribution and (9.3) gives the size of each sample in order for that difference to be smaller than δ with a confidence probability α . A large sample permits us to define a narrower CI (δ) and with a higher confidence (smaller α) than a smaller sample. For a

(9.3)

given sample size, the narrower the CI, the lower the confidence probability. If σ is unknown, the same comments above for point estimation apply.

Hypothesis testing starts by stating a hypothesis of interest H_0 , normally a hypothesis of no effect, hence termed null hypothesis, and an alternative hypothesis H_a , such that for the considered population either H_0 or H_a is true. The aim is to accept or reject H_0 on the grounds of data gathered by sampling the population. Some examples of null hypothesis are: "the proportion surviving two years in the treatment group is the same as the proportion surviving two years in the control (i.e. no treatment) group"; "the probability of a positive response is independent of treatment condition"; "the mean plasma concentration is the same in the treatment group and in the control group".

Two basic elements of hypothesis testing are the test statistic and the decision rule. The test statistic is a value calculated from the sample and which is indicative of whether or not the null hypothesis holds. An extreme value for the test statistic suggests that H_0 is false. The decision rule specifies which values of the test statistic, or function thereof, should be considered extreme.

Since H_0 is either true or false and as a result of the test we either accept or reject it, we respectively reject or accept H_a , and have the four possible outcomes in Table 9.6. If we accept the hypothesis that is true, the decision is correct. If we reject H_0 (i.e. accept H_a) when it is true, we make a Type I statistical error. If we accept H_0 (i.e. reject H_a) when it is false, we make a Type II statistical error. Sample size estimation seeks to minimize the chances of Type I and Type II errors. Usually, the probability of making a Type I error is designated α and the probability of making a Type II error is designated β . Common values for α , termed level of significance or alpha level are 0.01, 0.05 and 0.1. The statistical power of a test is the probability that H_0 will be rejected when it is false, i.e. the statistical power is $1 - \beta$. Common power levels are above 0.8. The decision rule depends on the allowable statistical errors.

Many scientific studies involving a hypothesis test report the p value of the test, defined as the probability that a result as extreme as (or more extreme) than that observed would occur by chance when the null hypothesis is true. A very small p suggests that the null hypothesis is unlikely to be true. When p is below an arbitrary cut-off value, e.g. 0.05 or 0.01, the result is called statistically significant, not to be confused with clinically significant or relevant. A result with p = 0.05, for example, means just that since that result would be obtained only once in 20 times if H0 were true, we opt for rejecting H0.

Table 9.6 Statistical errors in hypothesis testing. For the population considered, either the null hypothesis H_0 or the alternative hypothesis Ha is true.

Conclusion	Real situation				
	H ₀ true	H _a true			
Accept H ₀	Correct decision	Type II error			
Reject H ₀	Type I error	Correct decision			

Tests results about a condition or disease use the terminology in Table 9.7. The result of the test corresponds to the decision made in Table 9.6, and the null hypothesis is no disease. A false positive (FP) corresponds to a Type I error and a false negative (FN) corresponds to a Type

II error. The sensitivity of the test is the proportion of units with the condition or disease that have a positive test, TP/(TP + FN). The specificity of the test is the proportion of units without the condition or disease that have a negative test, TN/(TN + FP). Positive predictive value is the proportion of units with a positive test that have the condition or disease, TP/(TP + FP). Negative predictive value is the proportion of units with a negative test that do not have the condition or disease, TN/(TN + FP).

Table 9.7 In a test intended to ascertain the presence or absence of a condition or illness there are four possible outcomes, two of them correct (TN and TP) and two false (FP and FN).

Test result	Actual condition or disease			
	NO	YES		
Negative	True negative (TN)	False negative (FN)		
Positive	False positive (FP)	True positive (TP)		

The sample size needed to test an hypothesis depends on the level of significance and power desired. If we make it very difficult to reject H_0 in order to reduce Type I errors, then we risk not rejecting H_0 when it is false, i.e. making a Type II error. Conversely, if we wish a powerful test that rejects H_0 when the data are somewhat suspicious, then we risk rejecting H_0 when it is true, i.e. making a Type I error.

For a test comparing the means of two Gaussian populations with known and equal variances, σ , the size of each sample needed to test H_0 : $|m_1 - m_2| < \delta$, H_a : $|m_1 - m_2| > \delta$, is (Selwyn, 1996, Appendix B)

$$N = 2\left(Z_{\alpha} + Z_{\beta}\right)^2 \left(\frac{\sigma}{\delta}\right)^2 \tag{9.4}$$

where Z_{α} and Z_{β} are the values for the standard Gaussian distribution corresponding to, respectively, the tail probabilities of size α and β . For example, if $\alpha = 0.05$ and $\beta = 0.9$, $Z_{\alpha} = 1.96$ and $Z_{\beta} = 1.29$ and $N = 21.13(\sigma/\delta)2$. If the result of the test suggests that we accept H₀, then there is a 5 % probability that H₀ is false. If the result of the test suggests that we reject H₀, then there is a 10 % probability that H₀ is true.

For a test about the mean of a normal population, with H_0 : $|m| < \delta$ and H_a : $|m| > \delta$, the sample size needed is

$$N = \left(Z_{\alpha} + Z_{\beta}\right)^2 \left(\frac{\sigma}{\delta}\right)^2 \tag{9.5}$$

where Z_{α} and Z_{β} have the same meaning as in (9.4). Kanji (1993) provides a fine summary of statistical tests.

A final word of caution about N in the above equations. All the sample sizes calculated are estimates based on several assumptions, some of which may be difficult to prove. N is not an exact number. Hence, it may be convenient to modify some of the relevant factors involved in the equations in order to seek their effect on N. In any case, a high value for N indicates the need for pilot studies aimed to reduce the experimental variability.

Common experimental designs in biomedical engineering

The ideal experiment should be quick, inexpensive and involve a minimal number of specimens (animals, cultures) and suffering. These objectives, however, are sometimes in conflict with the aims of the study, so that no experimental design outperforms other alternatives.

In a completely randomized design each specimen is randomized to a single treatment condition. If this condition involves a single factor we have a one-way layout. The aim is to assess the response to different levels of a factor, such as treatment or no treatment (exposure or no exposure), different treatments each at a single level, or a single treatment with different levels. If the treatment condition involves two or more factors, we have a factorial design, which permits us to study the simultaneous effect of several factors, each of them with different levels (including absence) if needed. A study with only two factors, each with k levels is termed a 2k factorial design. Even for a small k, the number of combinations can be very high. Fractional factorial designs explore only those combinations deemed more relevant beforehand. Since each experimental unit receives a single treatment, completely randomized designs are quick. However, they involve a large number of units and comparison between treatments implies comparisons between different units, hence is dependent on between-subject variability, which is normally larger than within-subject variability.

In a randomized block design each block (or stratum) receives all treatments and treatments are compared within the blocks. The randomization procedure allocates specimens randomly within each block, but otherwise each block is expected to respond similarly to the treatment. Block designs reduce bias if the blocks are defined according to a suspected influencing factor, and can also help in analyzing the influence of the factor used to define the blocks. A disadvantage of block designs is that if each block receives several treatments sequentially, each of which needs a washout period between treatments, the study lengthens.

Crossover designs are particular randomized block designs in which each experimental unit constitutes a block and treatments are administered sequentially. The administering order is balanced across all the units, so that at a given period of time there is the same number of units receiving each treatment. To reduce carryover between treatments, it is necessary to wait for a washout time to elapse, not necessarily the same after each treatment, which lengthens the study. The advantage is that the inherent variability is very small because comparisons are made within each unit.

Sequential designs perform interim analysis of test data in order to decide to interrupt the study when there is enough evidence to reach a conclusion, i.e. accept or reject the null hypothesis, or to continue the study when it is not yet possible to reach a conclusion. The decisions about the hypothesis to test, the test statistic, the outcome variable and when to perform successive interim analyses, , are made before starting the experiment. To prevent the test from going on indefinitely, triangular test designs define a region for continuation (i.e. if the test statistic falls in that region the decision is to continue) which narrows as the number of units tested increases, hence that region is shaped as a triangle for one-sided alternatives. Sequential designs are quite common in clinical trials (Whitehead, 1997).

In group sequential analysis there are two to five interim analyses, fixed beforehand, and a final analysis. Each interim analysis decides whether to continue the experiment or not, and the final analysis either accepts or rejects the null hypothesis. Usually, interim analyses are conducted at equal intervals in terms of experimental units. The selection of a control or reference against which we compare the results is essential to derive conclusions from biomedical experiments. Historical controls are experimental units separated in time from the population under study. Concurrent controls are experimental units which receive an alternative treatment (including no treatment) in parallel with the treated units. Concurrent controls pose less bias risk than historical controls. The control can be either

negative, positive (active) or both, e.g. in genotoxicity studies using in vitro cultures. Negative controls receive a treatment that it is expected to produce a response indistinguishable from the background level for the considered population of experimental units. One method for negative control is to use each unit as his/her own control: the unit is observed before and after treatment, and significant changes after the treatment are attributed to it. This method, however, is not blinded. An untreated control does not receive treatment, but permits us to blind the observer. A placebo, or sham, control receives a treatment that matches the actual treatment except in the active ingredient. For example, in studying the thrombogenicity of an implant material, all animals undergo the same surgical procedure but one group of animals are implanted with the actual material and another group are implanted with a nonthrombogenic material.

Positive controls are helpful in validating negative findings: since the positive control assesses the validity of the experiment or test system, any negative finding must be considered the actual response to the explanatory variable, not the consequence of an interference. Positive controls are also an option when negative controls are unfeasible or inappropriate, e.g. for ethical reasons.

Dose-response experiments, such as toxicology studies, must consider how many dose levels to use. In broad terms, if the dose-response curve is known to be a n-degree polynomial, then we need n + 1 doses. Nevertheless, the decision depends on the objective of the experiment. Consider for example the LD50 test, which aims to determine the median lethal dose, i.e. the amount of a test substance able to kill half of the test animals. Toxic doses are assumed to follow a Gaussian distribution. In the conventional LD50 test, a number of animals is administered each dose level, which uses many animals. An up-and-down design uses a single animal per dose level. In the first trial, the animal is injected a guess dose according to its weight. If the animal dies, the next injection uses a dose decreased by a predetermined factor (e.g. 1.3), and the procedure is repeated. If the first animal survives, the next injection uses a dose increased by a predetermined factor (e.g. 1.3), and the procedure is repeated a predetermined number of times. As the lethal dose is approached, the survival-experiment curve oscillates (Figure 9.4). The tests are lengthy but the saving in animals and substances is considerable: whereas the classical LD50 determination kills from 40 to 50 animals, the up-and-down design uses only 6 to 9 animals per chemical (Ecobichon, 1992). An alternative design does not use the animal death as the endpoint but the observation of severe signs of toxicity. In 1991, regulatory agencies in the U.S., the EU and Japan dropped the classic LD50 as the required measure of acute toxicity.



Figure 9.4 Survival-experiment curve to determine a lethal dose: the animal survives in tests 1, 2, 3, 6 and 8, and dies in tests 4, 5, 7 and 9. The oscillation shows that the lethal dose is somewhere between those used in experiments 8 and 9.

Test protocol

A test protocol is the schedule of events defining the conduction, data collection and analysis for a test. The protocol should include (Selwyn, 1996):

- 1. Specific objectives of the study (specific and explicitly written, realistic and less than two or three).
- 2. Background.
- 3. Inclusion and exclusion criteria for experimental units.
- 4. Justification of the sample size.
- 5. Randomization process.
- 6. Description of study methods, including those of data analysis.

Some decisions about these points derive from statistical considerations described above. Other decisions result from regulations. Koëter (1991) compares several international, U.S., European and Japanese guidelines for different toxicity studies.

9.3.3 Analysis of experimental results

In order to extract the maximal amount of information from experimental results, the objectives, design and data analysis must be closely matched. The objectives influence the experimental design, and both determine which analysis method is more appropriate. Conversely, if a given study requires a specific data analysis method, then the experiment must be designed accordingly.

Analysis methods depend on whether the experimental data are discrete or continuous. Discrete data can be either dichotomous (alive/dead, tumor/no tumor, healthy/ill) or multilevel (e.g. disease state: none, mild, moderate, severe). Continuous data can take any value within a given range. Parametric analysis methods assume that the data belong to a population with known or presumed statistical distribution, e.g. Gaussian (usually), Poisson (for events which occur infrequently) or exponential (for survival data). Nonparametric analysis methods do not assume any statistical distribution for the data analyzed.

The first procedures in data analysis are checking and screening (Altman, 1991, C7). Data checking assesses the plausibility of the recorded values. This procedure is simple for discrete data. For continuous data, we can often define an expected range and then check that all recorded values fall inside that range. Nevertheless, for continuous variables it may be difficult to determine if there is any "impossible" value. Out-of-range values or outliers should not be discarded but carefully checked and eliminated only if they come from a mistake.

Data screening aims to determine the suitability of the available data for the intended type of analysis. Analysis assuming that the data come from a population with a Gaussian distribution, for example, can yield unexpected results if applied to a non-Gaussian population. Normal plots of cumulative frequency of data measure normality, though subjectively. The ShapiroĐWilk W test for normality provides an objective measure. In some cases, a transformation of the data, e.g. by taking logarithms, square roots or reciprocals, yields a distribution much closer to the Gaussian. This is often the case when the initial data include potential outliers.

Tables 9.8 to 9.10 list, respectively, some parametric and nonparametric tests for continuous data and tests for discrete data. Altman (1991), Box et al. (1978) and Forthofer and Lee (1995) detail those tests, and Kanji (1993) summarizes the procedure. Data analysis is currently performed by software packages, which are helpful because of their simplicity and low cost. Nevertheless, lack of judgment in the selection of the analysis function or carelessness in data entering invariably lead to meaningless results. Some statistical packages are BMDP, Minitab, SAS, SPSS, Stata, and Statgraphics.

Application	Data description	Statistical test
Test hypothesis about the	Data from a single sample	One sample t-test
mean of a single		
population		
Test hypothesis about the	Paired data	Paired t-test
difference in means		
Test hypothesis about the	Two independent samples	Two-sample t-test
difference in two means		
Test global hypothesis of	Two or more independent	Analysis of variance
difference among means	samples	(ANOVA)
Comparison of the means	More than two independent	Dunnett's test
of several samples with a	samples	
control or standard		
Pairwise comparisons	More than two independent	Modified t-tests: Bonferoni
	samples	method, Duncan's test
Testing for dose response	Several independent	Trend tests (linear,
	samples at different dose	nonlinear)
	levels	
Testing for treatment	Multifactorial or other	ANOVA with additional
differences	extended design	comparisons

Table 9.8 Parametric tests for continuous variables, assumed Gaussian. Adapted from (Selwyn, 1996).

Application	Data description	Statistical test
Test hypothesis about the	Data from a single sample	Wilcoxon signed rank test
location (mean) of a single		
population		XX7'1 ' 1 1 / /
Test hypothesis about the	Paired data	Wilcoxon signed rank test
difference in means		with paired data
Test hypothesis about the difference in two means	Two independent samples	Wilcoxon test
Test global hypothesis of	Two or more independent	Kruskal–Wallis test
difference among means	samples	
Comparison of the means	More than two independent	Dunn's test
of several samples with a	samples	
control or standard		
Pairwise comparisons	More than two independent	Multiple comparison
	samples	methods
Testing for dose response	Several independent samples at different dose levels	Jonckheere trend test
Testing for treatment	Multifactorial or other	ANOVA on ranks with
differences	extended design	additional comparisons as
	č	necessary

Table 9.9 Nonparametric tests for continuous variables. Adapted from (Selwyn, 1996).

9.1	137

Application	Data description	Statistical test
Test hypothesis about	Frequency distribution data	Chi-square test
frequencies in a single population	from a single population	
Test for independence in a 2 \times 2 table	Frequency data from a 2×2 table without all marginal totals fixed	Chi-square test
Test for independence in a 2 \times 2 table	Frequency data from a 2×2 table with all marginal totals fixed	Fisher's exact test
Test for trends in proportion data	Independent proportion in dose–response study	Cochran–Armitage trend test; exact permutation trend test
Test for independence in an $R \times C$ table	$R \times C$ contingency table of observed frequencies	Chi-square test; exact permutation test
Test for lack of association between rows and columns in one or more $R \times C$ tables	One or more $\mathbf{R} \times \mathbf{C}$ contingency tables of observed frequencies	Cochran–Mantel–Haenszel test
Testing general hypothesis in multiway contingency tables	Multiway contingency tables of observed frequencies	Loglinear models

Table 9.10 Hypothesis testing methods for discrete variables. Adapted from (Selwyn, 1996).

Example: Design of a glucose sensor (Validation)

After obtaining animal subjects committee approval, in vivo validation is performed in dogs. The glucose sensor is implanted transcutaneously behind the shoulder blades where the dog cannot bite the unit. A bolus intravenous infusion of glucose into the peripheral vein results in a circulation time to reach the sensor of about 15 s plus another 15 s before the sensor begins response. The response time is the following time delay to reach 90% of the steady-state and is 2 to 5 min. We measure slow drift of calibration, which indicates that recalibration should be performed every 4 to 12 weeks during lifetime.

9.4 CLINICAL TRIALS

The range in animal anatomy, physiology and biochemistry is very broad, which may call for human clinical studies to complement nonclinical studies. Thalidomide, for example, did not show any harmful effect in rats, but later painfully proved harmful to humans. Section 3.7.5 describes how ethical concerns about human experimentation lead to the creation of IRBs (21CFR56) and the requirement of informed consent (21CFR50). Medical devices subjected to clinical studies are termed investigational devices and regulated by 21CFR812, unless exempt per 21CFR812(c), e.g. investigations involving some lawfully marketed devices, custom devices and most noninvasive diagnostic devices (Section 9.6.2). The testing of investigational devices in

the U.S. is supervised by the Division of Bioresearch Monitoring in the Office of Compliance of the CDRH, which has issued several guides on clinical investigations, available at the FDA web site: www.fda.gov/oha/IRB/signific.html.

9.4.1 Clinical studies

Clinical studies are conducted by clinical investigators, usually physicians, who have agreed with a sponsor to conduct the study. A sponsor is a person or entity that initiates a clinical investigation of a device (or drug), not necessarily its manufacturer. A sponsor shall not conduct the investigation. A clinical investigator, however, may serve as sponsor–investigator. The sponsor is responsible for fulfilling any applicable regulation, must obtain FDA approval and report the results of the study to the FDA. Clinical investigators agree to FDA regulations by signing an FDA form that certifies that they have obtained IRB review and approval prior to conducting the study.

Pilot studies

Pilot studies, also termed feasibility studies, involve a small number of patients (say, less than 25) and their scope is limited. Some possible objectives are: to obtain information needed to plan a more complete clinical trial, to identify clinically meaningful endpoints to validate effectiveness, to optimize the design of the device, to validate new measuring tools for the outcome variables and the associated variances to estimate sample sizes, to evaluate factors that may introduce bias, a preliminary safety and/or performance evaluation, or to investigate a new modality or use for a device with proven safety. Pilot studies are not required for FDA approval but they are recommended to save time and money. In order to reduce the risks to humans, nonclinical studies must precede pilot studies.

Clinical trials

A clinical trial of a medical device is a controlled study involving human subjects, designed to evaluate prospectively its safety and effectiveness. Medical devices requiring premarket approval (Section 9.6) need one well-controlled trial to demonstrate safety and effectiveness. By contrast, medical drugs need at least two trials and involve three phase studies. Phase I studies involve a small number of healthy volunteers and place emphasis on safety. Phase II studies are in patients, but involve too few patients to draw valid conclusions on effectiveness. Phase III studies are double blind tests, involve a large number of patients and place a major emphasis on effectiveness. There are many books on drug clinical trials whose fundamentals apply to medical devices, e.g. Bulpitt (1996) and Pocock (1984). FDA (1996b) is a guide on statistical aspects relative to the design, conduct and analysis of clinical trials of nondiagnostic medical devices, which are discussed below.

Postmarketing studies

Postmarketing studies involve a small number of patients and, even as pilot studies, have a limited scope. Their aim is usually to obtain additional information about a specific factor concerning safety or efficacy, or to investigate an adverse effect or failure. They can also provide

information to substantiate claims or for comparative studies, for example in health care technology assessment.

9.4.2 Design of the clinical trial

The need for a careful design of the clinical trial stems from the basic concepts on experimental design discussed in Section 9.3.2, and it is reinforced by the involvement of humans as test subjects. The scientific objective of the clinical trial is to answer a research question. In order for this answer to constitute a proof of valid scientific evidence (Section 3.2.1), the study must fulfill the requirements in 21CFR860.7(f), which affect variables, subjects and comparison methods.

Variables involved

The outcome variables should be the most suitable to measure the clinical safety and effectiveness of the device, which implies that we must consider the intended use and target population identified in the design specifications. If necessary, we must perform nonclinical studies and pilot studies to identify which outcome variables are better related to the biological effects to assess and to quantify the relationship between treatment and outcome.

The next step is to identify as many influencing variables as possible. These variables, also termed baseline variables, confounding factors and prognostic factors may affect both the outcome variables or the relationship between the treatment and the outcome variables. This interference causes bias and measurement inaccuracy. These undesired effects may depend on the measurement method for the outcome variable. Measurements which provide a continuous result are preferred to those which offer discrete data, but many variables are not amenable to continuous measurement, e.g. pain level and quality of life.

Study and control populations

The statistical population in a clinical trial is the target population of the device. The sample extracted to perform the study is termed the study population, and must be representative of the target population if we wish to infer valid conclusions from the study. A very homogeneous study population results in a reduced sample size needed and less variability in the responses. However, it may not be representative enough to approve the treatment for the whole target population. Influencing variables which may bias the results may compel us to define blocks or strata, e.g. according to disease severity, concomitant disease, prognosis and/or demographic characteristics. The sponsor must define the inclusion/exclusion criteria. Guidance documents by the FDA suggest criteria for some devices, e.g. for Intrapartum continuous monitors for fetal oxygen saturation and fetal pH:

Patients are usually selected from a few sites, not from the entire target population, and therefore the sample is not truly random. However, if the sites are themselves representative of the variety of sites where one can encounter such patients, the test samples are in some way random and representative, provided patients are selected at random at each site. The FDA has issued several documents concerning different aspects of patient recruitment, such as payment to study subjects, assent of children and studies involving non-English speakers. They are available at the FDA web site: http://www.fda.gov/oc/ohrt/irbs/default.htm.

The FDA regulation on informed consent exceptions does not preempt State and local regulations, some of which prohibit clinical research without prior informed consent.

Specific trial designs

The completely randomized design (Section 9.3.2) is the more simple to perform and analyze, but the comparison between treatments is influenced by the variability between patients. Randomized block designs reduce bias from the prognosis factors we consider to define the blocks, but take longer and their analysis is more involved as the patient's response to any particular treatment may be influenced by previous treatments. Experimental designs which imply a long trial duration face a larger risk of patient dropout, which produce bias. If the cause of dropout is not related to the study itself, the only influence is because of the reduced sample size. But if the dropout is related to the study, the inferences for the target population may be questionable.

9.4.3 The protocol

The protocol describes the conduction of the trial and data collection and analysis, and must necessarily refine the information provided in animal tests (Section 9.3.2). FDA (1996b) recommends determination of the following points prior to starting the trial and inclusion of them in the protocol:

- 1. Background: description of previous scientific studies relevant to the research question.
- 2. Trial objective(s): stated clearly, specifying any medical claim and indication addressed by the research question, a clinically meaningful effect and associated outcome variables.
- 3. Complete description of the trial design: design type, method of data collection, type of control, method and level of masking, justification of sample size and method of treatment assignment (randomization, stratification, other).
- 4. Complete description of the study population: study site, inclusion and exclusion criteria for recruiting patients and type of patients (e.g. inpatient or outpatient). Relate the clinical and demographic characteristics of study subjects to the characteristics of the target population.
- 5. Description of intervention, including frequency and duration of application, and measures of physician and patient compliance.
- 6. Follow-up visits: schedule, description of the procedure for each visit, measurements to perform, information collected. Describe the handling of patient withdrawal and procedures to determine the health status of dropouts.
- 7. Data collection and analysis: methods to gather, validate and monitor data, statistical analysis methods and stopping rule for early trial termination.
- 8. Investigators (curriculum vitae), monitoring methods and trial administration techniques, including methods to identify and readjust the protocol.
- 9. Definition of relevant terms (clinical and nonclinical) to be used during the trial, including those related to entrance criteria and to the observation of outcome and influencing variables.
- 10. All informed consent forms and any provisions not included above but that may be required by the IRB.

9.4.4 Clinical trial conduct

The conduct of the clinical trial follows from the application of the protocol. Nevertheless, there is a need for contingency plans to deal with unforeseen problems arising during the study in a way that does not hinder the achievement of the trial objective.

The trial monitor must ensure that the study follows the specifications in the protocol relative to entering subjects, assigning the intervention, measuring the relevant variables as planned, and recording the data. Many studies of therapeutic devices require that the device be operated by the investigator or a subinvestigator according to the protocol.

9.4.5 Clinical trial analysis

The protocol determines which analysis can and cannot be applied to the data collected. Nevertheless, some deviations introduced during the study may alter the initial provisions and call for different analysis methods.

The first step in data analysis after checking and screening is to validate assumptions underlying the proposed analysis methods. For example, whether the target population is Gaussian, the equality of the variance of two comparison groups, or the independence between two variables.

The report must include the hypothesis tested, the statistical tests used and the underlying assumptions. Some FDA guidance documents suggest analysis methods for the specific devices. For example, the guidance for the above mentioned balloon valvuloplasty catheters specifies a survival analysis using actuarial life tables built from follow-up data at three and six months, to show the estimated probabilities of freedom from each postoperative complication at the end of each follow-up period. Life table results should be compared to controls, which may be the results of similar studies, using statistical methods such as the MantelĐHaenszel, one-degree of freedom, chi-square test.

Example: Design of a glucose sensor (Validation)

After obtaining human subjects committee approval, in vivo validation is performed in small numbers of humans. Large scale clinical trials require FDA approval.

9.5 FINAL DESIGN REVIEW

The requirements of the Quality System (QS) regulation outlined in Section 4.2 include design controls (21CFR820.30), which call for formal design review after appropriate stages in the design plan. These design controls affect most new devices and changes to the design of marketed devices. Specifically, the QS regulation requires us to control the design input and the design output. Consequently, for a complex system there is usually a review of design input requirements (Chapter 5), a conceptual design review (after design evaluation, Chapter 7), a review of the initial design following design verification (Section 8.6), a review of the final design after design validation and prior to pilot production, and a preproduction review after production process validation (Section 10.4), before running full-scale production. Postmarket

surveillance (Section 11.8) leads to design review following changes which address adverse events and for product improvement.

9.6 FDA MARKETING CLEARANCE

The FDA classifies all medical devices in one of three categories depending on the extent of control necessary to ensure their safety and effectiveness (Section 4.2). All medical devices are subjected at least to general controls, which include a premarket notification (PMN) application. Figure 9.5 shows that most class III devices require a premarket approval (PMA) application. Many low-risk devices, which make up to one third of medical devices categories, are exempt from PMN requirements. However, they remain subject to establishment registration, device listing and QS regulation, which include regular factory inspections, record keeping and adverse event reporting. Electronic medical devices, including those which emit radiation, have additional requirements, which may include: reports and records (21CFR1002), notification of defects and failure to comply (21CFR1003) and performance standards (21CFR1010, 21CFR1030, 21CFR1040 and 21CFR1050).



Figure 9.5 Steps in deciding which FDA marketing clearance procedure corresponds to a new edical device.

9.6.1 Premarket notification [501(k)]

Each manufacturer who wishes to market a new medical device must submit a premarket notification to the FDA at least 90 days before commercial distribution is to begin, unless exempt as a low-risk device. The FDA reviews these notifications to determine if the new device is "substantially equivalent" (SE) to a pre-Amendments device, i. e. a device that was legally marketed before the passage of the Medical Device Amendments to the Federal Food, Drug and Cosmetic Act (FD&C Act) in May 28, 1976. If the new device is deemed SE (within the meaning of section 513(i) of the FD&C Act) to a pre-Amendments device, it may be marketed immediately and is regulated in the same regulatory class as the pre-Amendments device to which it is equivalent. Since section 510(k) of the FD&C Act describes premarket notification, pre-Amendments devices are often termed "510(k) devices".

If the FDA determines that the new device is not substantially equivalent (NSE) to a pre-Amendments device, it is automatically placed in class III requiring premarket approval, which demands clinical testing. Under the FDA Modernization Act (FDAMA) of 1997, however, the sponsor can ask for immediate down reclassification into class II or class I, based on the risk level of the device, by requesting a de novo classification within 30 days of receiving an NSE determination (CDRH, 1998a). The FDA then has 60 days to respond to the request with a written order specifically classifying the device. If the FDA classifies the device into class I or class II, the product is then considered cleared and may be marketed, subject to other applicable provisions of the FD&C Act (Figure 9.6). If the FDA keeps the device in class III, the product needs either an approved PMA or a completed product development protocol (PDP).


Figure 9.6 The "de novo" classification procedure of the FDAMA, evaluates the automatic class III designation of new devices according to their risk (CDRH, 1998a).

Traditional 510(k) application

21CFR807.81 requires PMN submission for devices marketed for the first time, i.e. devices NSE to a pre-Amendments devices or a device reclassified into class I or II, and when the device is significantly changed or modified in a way that could significantly affect its safety or effectiveness. For example, PMN submission is required when there is a change in design, components, materials, chemical composition, energy source, manufacturing process, sterilization, conditions for use, patient or user safety features or intended use. Devices requiring

PMA do not need PMN. Rice and Lowery (1995) discuss 510(k) submission requirements codified in 21CFR807, Subpart E. The documentation required for pre-Amendments status is described in a FDA document available at www.fda.gov/cdrh/comp/preamend.html. Each nonexempt device needs a single PMN, even if marketed by more than one distributor.

To streamline the evaluation of PMNs the FDA has developed "The new 510(k) paradigm" (Figure 9.7), which adds two options to the traditional method of demonstrating SE: the "Special 510(k): Device Modification" option and the "Abbreviated 510(k)" option (FDA, 1998).



Figure 9.7 "The new 510(k) paradigm" (FDA, 1998) to demonstrate significant equivalence includes the traditional method and two optional paths: the "Special 510(k)" option applicable for some device modifications and the "Abbreviated 510(k)" option, which relies on the use of guidance documents, special controls and recognized standards to facilitate 510(k) review.

Special 510(k) application

The "Special 510(k): Device Modification" option utilizes certain aspects of the QS regulation (Chapter 11). Effective June 1, 1997, manufacturers of class II, class III, and certain class I devices (including all those automated by software) must follow design control procedures when originally developing devices and for subsequent modifications (Section 4.2). 21CFR807.81(a)(3), further detailed by (FDA, 1997), specifies that the following modifications of an existing device require a 510(k) submission: changes resulting from a recall or corrective action; labeling changes concerning indications for use, changes in warnings and precautions, contraindications addition or deletion; technology or performance changes; and materials changes.

Abbreviated 510(k) application

The "Abbreviated 510(k)" option in Figure 9.6 uses guidance documents, special controls and recognized standards to facilitate 510(k) review. Device-specific guidance documents identify the information recognized as appropriate for marketing authorization. Special controls, intended for class II devices (Section 4.2), are "those controls, such as performance standards, postmarket surveillance, patient registries, development and dissemination of guidelines, recommendations and other appropriate actions that provide reasonable assurance of the device's safety and effectiveness." The FDAMA authorizes the FDA to recognize all or part of national and international standards through publication of a notice in the Federal Register. Recognized standards can be cited in guidance documents or individual policy statements, or established as special controls that address specific risks associated with a type of device.

9.6.2 Investigational device exemptions (IDE)

An investigational device is a medical device which is the object of a clinical study to determine its safety and/or effectiveness. Clinical studies are necessary to support a request for premarket approval and for design validation, and are regulated by 21CFR812 in order to protect the subjects. An approved IDE exempts a device from certain sections of the FD&C Act that otherwise would impede those studies. For example, misbranding (section 502), registration, listing, and premarket notification (section 510), special controls (section 513), performance standards (section 514), premarket approval (section 515), banned devices (section 516), records and reports (section 519), restricted devices (section 520(e)); good manufacturing practices (section 520(f)); and color additive requirements (section 721). 21CFR809.10(c) exempts in vitro diagnostics to be used in investigations from otherwise enforceable labeling requirements.

21CFR812.2(c) exempts the following devices from the IDE regulations, with the exception of 21CFR812.119 (Disqualification of a clinical investigator):

- 1. A device, other than a transitional device, marketed before May 28, 1976, when used or investigated in accordance with the indications in labeling in effect at that time. (Transitional devices are devices regulated as drugs prior to May 28, 1976.)
- 2. A device, other than a transitional device, marketed on or after May 28, 1976, that FDA has determined to be substantially equivalent to a device in commercial distribution immediately before May 28, 1976, and that is used or investigated in accordance with the indications in

the labeling FDA reviewed under 21CFR807, subpart E, in determining substantial equivalence.

- 3. A diagnostic device, if the sponsor complies with applicable requirements in 21CFR809.10(c) and if the testing:
 - a. is noninvasive,
 - b. does not require an invasive sampling procedure that presents significant risk,
 - c. does not by design or intention introduce energy into a subject, and
 - d. is not used as a diagnostic procedure without confirmation of the diagnosis by another, medically established diagnostic product or procedure.
- 4. A device undergoing consumer preference testing, testing of a modification, or testing of a combination of two or more devices in commercial distribution, if the testing is not for the purpose of determining safety or effectiveness and does not put subjects at risk.
- 5. A device intended solely for veterinary use.
- 6. A device shipped solely for research on or with laboratory animals and labeled in accordance with 21CFR812.5(c).
- 7. A custom device as defined in 21CFR812.3(b), unless the device is being used to determine safety or effectiveness for commercial distribution.

.6.3 Premarket approval

Figure 9.8 shows the premarket approval review process used by the FDA to evaluate the safety and effectiveness of class III devices. Pre-Amendments class III devices are not required to submit a PMA application until 30 months after the promulgation of a final classification regulation or until 90 days after the publication of a final regulation requiring the submission of a PMA application, whichever period is later. In the meanwhile, pre-Amendments class III devices may be marketed through the PMN [510(k)] process. Post-Amendments class III devices deemed by the FDA to be SE to pre-Amendments class III devices, are regulated as the latter. Post-Amendments devices which the FDA finds NSE to either pre-Amendments devices or post-Amendments devices classified into new class I and class II devices, and transitional devices need an approved PMA application before commercial distribution. Otherwise they are considered adulterated under section 501(f) of the FD&C Act. 21CFR814 lists PMA application requirements and Park et al. (1997) explain them. Devices covered by IDE regulations do not need a PMA application.



Figure 9.8 Post-amendment class III devices that are not substantially equivalent to preamendment devices need a premarket approval from the FDA.

Premarket approval application)

According to 21CFR814.20, the applicant or an authorized representative must sign the PMA application. If the applicant does not reside or have a place of business within the United States, an authorized representative residing or maintaining a place of business in the U.S. must countersign the application.

PMA amendments and resubmitted PMAs

An applicant may amend a pending PMA or PMA supplement to revise existing information or provide additional information as specified by 21CFR814.37. The FDA may request the applicant to amend a PMA or PMA supplement with any information regarding the device deemed to be necessary to complete the review of the PMA or PMA supplement.

PMA supplements

After FDA approval of a PMA, an applicant must submit a PMA supplement for review and approval by FDA before making a change affecting the safety or effectiveness of the approved device, unless the change is of a type for which the FDA has advised that an alternative submission is permitted (21CFR814.39).

The FDA must review a PMA application within 180 days after receipt of an application that is accepted for filing and for which the applicant does not submit a major amendment (21CFR814.40). As a result of the review, the FDA will send the applicant an approval order under 21CFR814.44(d), an approvable letter under 21CFR814.44(e), a not approvable letter under 21CFR814.44(f), or an order denying approval under 21CFR814.45. The approvable letter and the not approvable letter will provide an opportunity for the applicant to amend or withdraw the application, or to consider the letter to be a denial of approval of the PMA application under Sec. 21CFR814.45 and to request administrative review under section 515 (d)(3) and (g) of the FD&C Act.

Postapproval requirements

The FDA may impose postapproval requirements in a PMA approval order or by regulation at the time of approval of the PMA or by regulation subsequent to approval (21CFR814.82).

Modular approach to PMA review

Part of the information required in a PMA concerning product design, preclinical (bench and animal) testing, clinical data and manufacturing is previously submitted in an IDE application and consequently reviewed. In order to increase efficiency and efficacy, the CDRH proposed a modular approach for data development, submission, review and closure in a PMA (CDRH, 1998b). Under this approach, a complete PMA will consist of a set of completed "modules", e.g., product design, biomaterials, bench/animal data, electrical safety, EMC testing. If the modules pertaining to a given device are agreed upon during the IDE stage, the manufacturer may submit the data for review as each module is completed, one at a time, rather than waiting for submission of the full PMA. If a completed module is submitted and reviewed during the IDE phase of product development, the data will be incorporated into the PMA by reference to the IDE, and will not need to be reviewed again, unless the final review of the other modules raises new issues.

9.6.4 Product development protocol

The Product Development Protocol (PDP) is an alternative to the two-step IDE and PMA review procedures for class III devices, though never used from 1976 to 1998 in spite of potentially being faster and less expensive. Nevertheless, the FDA has issued a guidance on the contents of PDP applications, expected actions and time frames in the development of a product under a PDP (CDRH, 1998c).

9.6.5 Humanitarian device exemptions

The FDA may grant an exemption from the effectiveness requirements of Sections 514 (performance standards) and 515 (premarket approval) of the FD&C Act, for a device that:

- is designed to treat or diagnose a disease or condition that affects fewer than 4000 individuals in the U.S.,
- is not available otherwise, and there is no comparable device available to treat or diagnose the disease or condition,
- will not expose patients to unreasonable or significant risk of illness or injury, and the benefits to health from the use outweighs the risks.

9.6.6 PMA/510(k) expedited review

The FDA may grant expedited review status to Premarket Notifications (510(k)) and Premarket Approval Applications and their supplements, concerning devices intended for more effective treatment or diagnosis of life-threatening or irreversibly debilitating diseases or conditions. The expedited review status means that the marketing application would receive priority review before other pending applications. Otherwise, they are subject to all other controls and requirements applicable to comparable applications in the standard review process.

9.7 REVIEW QUESTIONS

- 9.1 Discuss the main difference between design verification and design validation as meant by the FDA's Quality System regulation.
- 9.2 List the main tests used for the biological evaluation of medical devices.
- 9.3 Describe the criteria to establish the device categories in the FDA-modified matrix for toxicity evaluation of medical devices.
- 9.4 Explain the difference in objective between an acute toxicity test and a subchronic toxicity test.
- 9.5 Discuss the reasons that made synthetic polymers a main concern in the biomedical evaluation of medical devices.
- 9.6 Search the BjorkĐShiley heart-valves case and discuss some of its consequences for both patients and manufacturer.
- 9.7 Discuss the meaning of replacement, reduction and refinement when considering alternative methods in animal testing.
- 9.8 Describe the basic techniques to control bias and variability in experimental design.
- 9.9 Assume that serum albumin values in patients with primary biliary cirrhosis have a Gaussian distribution with a mean of 35 g/L and a standard deviation of 6 g/L. Determine the size of the sample required so that the 99 % CI is no more than 5 mg/L. (N = 39)
- 9.10 Calculate the sensitivity, specificity, positive predictive value and negative predictive value of a liver scan procedure which yields: TN = 54, TP = 231, FP = 32 and FN = 27. (Sensitivity = 90%, specificity = 63%, PPV = 88% and NPV = 67%)
- 9.11 Discuss the advantages and shortcomings of experiments that use randomized block designs.
- 9.12 Explain a triangular test design.
- 9.13 List the different methods for selection of controls in biomedical experiments.

- 9.14 Describe the classic LD50 and the underlying hypothesis about toxic doses.
- 9.15 Determine the range of the variables (in adults) to be used for checking the data recorded in the following clinical tests: arterial blood pressure, heart rate and cardiac output (in rest and during exercise).
- 9.16 Draw a diagram that shows the relationship between the sponsor, clinical investigator, FDA and IRB for an SR study.
- 9.17 Define a clinical trial and when is it required.
- 9.18 Describe a passive concurrent control in a clinical trial.
- 9.19 Describe a third party blinding clinical trial.
- 9.20 Find the inclusion and exclusion criteria recommended by the FDA for clinical trials of balloon valvuloplasty catheters.
- 9.21 Explain pre-Amendments devices.
- 9.22 Describe a reserved class I device and give some examples.
- 9.23 Define a predicate device.
- 9.24 Search which information must be included in a Premarket Notification [501(k)] summary.

9.8 REFERENCES

- Altman, D. G. 1991. Practical statistics for medical research. New York: Chapman and Hall.
- Bennet, B. T. 1994. Alternative methodologies. In Bennet, B. T., Brown, M. J., and Schoefield,
 J. C. Essentials for Animal Research: A Primer for Research Personnel. 2nd Ed. Beltsville,
 MD: United States Department of Agriculture, National Agricultural Library. Available from the University of California at Santa Barbara home page at:
 research.ucsb.edu/pro/ess_idex.html.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. 1978. Statistics for experimenters, an introduction to design, data analysis and model building. New York: John Wiley & Sons.
- Bulpitt, C. J. 1996. Randomized controlled clinical trials. 2nd Ed. Hingham, MA: Kluwer.
- CDRH. 1998a. New Section 513(f)(2) Evaluation of Automatic class III Designation, Guidance for Industry and CDRH Staff. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/modact/classiii.html.
- CDRH. 1998b. A modular approach to PMA review. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/pmat/modpmat.html.
- CDRH. 1998c. Contents of a Product Development Protocol. Draft Guidance. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/pdp/pdpguide.pdf.
- CDRH. 1998d. PMA/510(k) Expedited Review Guidance for Industry and CDRH Staff. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/modact/expedite.html.
- Ecobichon, D. J. 1992. The basis of toxicity testing. Boca Raton, FL: CRC Press.
- FDA. 1995a. General Program Memorandum (Blue Book Memo) G95-1. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/g951.html.

- FDA. 1995b. Significant risk and nonsignificant risk medical device studies. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/d861.html.
- FDA. 1996a. Immunotoxicity testing framework. Draft document. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/ode/frmwrk3.html.
- FDA. 1996b. Statistical guidance for clinical trials of nondiagnostic medical devices.. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/manual/statgde.html.
- FDA. 1997. Deciding when to submit a 510(k) for a change to an existing device. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/ode/510kmod.html.
- FDA. 1998. The new 510(k) paradigm: alternate approaches to demonstrating substantial equivalence in premarket notifications. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/ode/parad510.html.
- Forthofer, R. N., and Lee E. S. 1995. Introduction to biostatistics, a guide to design, analysis and discovery. San Diego, CA: Academic Press.
- Gad, S. C. 1997. Safety evaluation of medical devices. New York: Marcel Dekker.
- Hendriksen, C. F. M. 1991. Animal experimentation and the concept of alternatives. In Hendriksen, C. F. M. and Koëter, H. B. W. M. (eds.) Replacement, reduction and refinement: present possibilities and future prospects. New York: Elsevier.
- Hudson, V. M., and DeBerry, E. 1996. Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing, A Bibliography with Abstracts. Bethesda, MD: National Institutes of Health, National Library of Medicine. Available at the NIH web site: sis.nlm.nih.gov/alt9601.txt.
- Hudson, V. M., and DeBerry, E. 1997. Alternatives to the Use of Live Vertebrates in Biomedical Research and Testing, A Bibliography with Abstracts. Bethesda, MD: National Institutes of Health, National Library of Medicine. Available at the NIH web site: sis.nlm.nih.gov/alt397.htm.
- Kanji, G. K. 1993. 100 statistical tests. Newbury Park, CA: SAGE Publications.
- Koëter, H. B. W. M. 1991. Current guidelines and regulations in toxicological research. In C. F. M. Hendriksen, and H. B. W. M. Koëter (eds.) Replacement, reduction and refinement: present possibilities and future prospects. New York: Elsevier.
- Park, J., Melvin, M., and Barcome, A. 1996. Investigational device exemptions manual. HHS Publication FDA 96-4159. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/manual/idemanul.html.
- Park, J., Poneleit, K., and Parr, R. 1997. Premarket approval manual. HHS Publication FDA 97-4214. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/manual/pmamanul.pdf.
- Pocock, S. J. 1984. Clinical trials: a practical approach. New York: John Wiley & Sons.
- Rice, L. L., and Lowery, A 1995. Premarket notification 510(k): Regulatory requirements for medical devices. HHS Publication FDA 95-4158. Rockville, MD: Center for Devices and Radiological Health. Available at the FDA web site: www.fda.gov/cdrh/manual/510kprt1.html.
- Whitehead, J. 1997. The design and analysis of sequential clinical trials. 2nd Ed. New York: John Wiley & Sons.