INTRODUCTION

In this book, we are concerned mainly with the drug discovery and development process, proudly regarded as the mainspring of the pharmaceutical industry. In this chapter we consider the broader context of the human environment into which new drugs and medicinal products are launched, and where they must find their proper place. Most pharmaceutical companies place at the top of their basic mission statement a commitment to improve the public’s health, to relieve the human burden of disease and to improve the quality of life. Few would argue with the spirit of this commitment. Nevertheless, we need to look more closely at what it means, how disease is defined, what medical therapy aims to alter, and how – and by whom – the effects of therapy are judged and evaluated. Here we outline some of the basic principles underlying these broader issues.

CONCEPTS OF DISEASE

The practice of medicine predates by thousands of years the science of medicine, and the application of ‘therapeutic’ procedures by professionals similarly predates any scientific understanding of how the human body works, or what happens when it goes wrong. As discussed in Chapter 1, the ancients defined disease not only in very different terms, but also on a quite different basis from what we would recognize today. The origin of disease and the measures needed to counter it were generally seen as manifestations of divine will and retribution, rather than of physical malfunction. The scientific revolution in medicine, which began in earnest during the 19th century and has been steadily accelerating since, has changed our concept of disease quite drastically, and continues to challenge it, raising new ethical problems and thorny discussions of principle. For the centuries of prescientific medicine, codes of practice based on honesty, integrity and professional relationships were quite sufficient: as therapeutic interventions were ineffective anyway, it mattered little to what situations they were applied. Now, quite suddenly, the language of disease has changed and interventions have become effective; not surprisingly, we have to revise our ideas about what constitutes disease, and how medical intervention should be used. In this chapter, we will try to define the scope and purpose of therapeutics in the context of modern biology. In reality, however, those in the science-based drug discovery business have to recognize the strong atavistic leanings of many healthcare professions¹, whose roots go back much further than the age of science.

Therapeutic intervention, including the medical use of drugs, aims to prevent, cure or alleviate disease states. The question of exactly what we mean by disease, and how we distinguish disease from other kinds of human affliction and dysfunction, is of more than academic importance, because policy and practice with respect to healthcare provision depend on where we draw the line between what is an appropriate target for therapeutic intervention and what is not. The issue concerns not only doctors, who have to decide every day what kind of complaints warrant treatment, the patients who receive the treatment and all those involved in the healthcare business – including, of course, the pharmaceutical industry. Much has been

¹The upsurge of ‘alternative’ therapies, many of which owe nothing to science – and, indeed, reject the relevance of science to what its practitioners do – perhaps reflects an urge to return to the prescientific era of medical history.
written on the difficult question of how to define health and disease, and what demarcates a proper target for therapeutic intervention (Reznek, 1987; Caplan, 1993; Caplan et al., 2004); nevertheless, the waters remain distinctly murky.

One approach is to define what we mean by health, and to declare the attainment of health as the goal of all healthcare measures, including therapeutics.

What is health?

In everyday parlance we use the words ‘health’, ‘fitness’, ‘wellbeing’ on the one hand, and ‘disease’, ‘illness’, ‘sickness’, ‘ill-health’, etc., on the other, more or less interchangeably, but these words become slippery and evasive when we try to define them. The World Health Organization (WHO), for example, defines health as ‘a state of complete physical, mental and social wellbeing and not merely the absence of sickness or infirmity’. On this basis, few humans could claim to possess health, although the majority may not be in the grip of obvious sickness or infirmity. Who is to say what constitutes ‘complete physical, mental and social wellbeing’ in a human being? Does physical wellbeing imply an ability to run a marathon? Does a shy and self-effacing person lack social wellbeing?

We also find health defined in functional terms, less idealistically than in the WHO’s formulation: ‘…health consists in our functioning in conformity with our natural design with respect to survival and reproduction, as determined by natural selection…’ (Caplan, 1993). Here the implication is that evolution has brought us to an optimal – or at least an acceptable – compromise with our environment, with the corollary that healthcare measures should properly be directed at restoring this level of functionality in individuals who have lost some important element of it. This has a fashionably ‘greenish’ tinge, and seems more realistic than the WHO’s chillingly utopian vision, but there are still difficulties in trying to use it as a guide to the proper application of therapeutics. Environments differ. A black-skinned person is at a disadvantage in sunless climates, where he may suffer from vitamin D deficiency, whereas a white-skinned person is liable to develop skin cancer in the tropics. The possession of a genetic abnormality of haemoglobin, known as sickle-cell trait, is advantageous in its heterozygous form in the tropics, as it confers resistance to malaria, whereas homozygous individuals suffer from a severe form of haemolytic anaemia (sickle-cell disease). Hyperactivity in children could have survival value in less developed societies, whereas in Western countries it disrupts families and compromises education. Obsessivity and compulsive behaviour are quite normal in early motherhood, and may serve a good biological purpose, but in other walks of life can be a severe handicap, warranting medical treatment.

Health cannot, therefore, be regarded as a definable state – a fixed point on the map, representing a destination that all are seeking to reach. Rather, it seems to be a continuum, through which we can move in either direction, becoming more or less well adapted for survival in our particular environment. Perhaps the best current definition is that given by Bircher (2005) who states that ‘health is a dynamic state of wellbeing characterized by physical, mental and social potential which satisfies the demands of life commensurate with age, culture and personal responsibility’. Although we could argue that the aim of healthcare measures is simply to improve our state of adaptation to our present environment, this is obviously too broad. Other factors than health – for example wealth, education, peace, and the avoidance of famine – are at least as important, but lie outside the domain of medicine. What actually demarcates the work of doctors and healthcare workers from that of other caring professionals – all of whom may contribute to health in different ways – is that the former focus on disease.

What is disease?

Consider the following definitions of disease:

- A condition which alters or interferes with the normal state of an organism and is usually characterized by the abnormal functioning of one or more of the host’s systems, parts or organs (Churchill’s Medical Dictionary, 1989).
- A morbid entity characterized usually by at least two of these criteria: recognized aetiological agents, identifiable groups of signs and symptoms, or consistent anatomical alterations (elsewhere, ‘morbid’ is defined as diseased or pathologic) (Stedman’s Medical Dictionary, 1990).
- ‘Potential insufficient to satisfy the demands of life’ as outlined by Bircher (2005) in his definition of health above.

We sense the difficulty that these thoughtful authorities found in pinning down the concept. The first definition emphasizes two aspects, namely deviation from normality, and dysfunction; the second emphasizes aetiology (i.e. causative factors) and phenomenology (signs, symptoms, etc.), which is essentially the manifestation of dysfunction.

Deviation from normality does not define disease

The criterion of deviation from normality begs many questions. It implies that we know what the ‘normal state’ is, and can define what constitutes an alteration of it. It suggests that if our observations were searching enough, we could unfailingly distinguish disease from normality. But we know, for example, that the majority of 50-year-olds will have atherosclerotic lesions in their arteries, or
that some degree of osteoporosis is normal in postmenopausal women. These are not deviations from normality, nor do they in themselves cause dysfunction, and so they do not fall within these definitions of disease, yet both are seen as pathological and as legitimate – indeed important – targets for therapeutic intervention. Furthermore, as discussed below, deviations from normality are often beneficial and much prized.

**Phenomenology and aetiology are important factors – the naturalistic view**

Setting aside the normality criterion, the definitions quoted above are examples of the *naturalistic*, or observation-based, view of disease, defined by phenomenology and backed up in many cases by an understanding of aetiology. It is now generally agreed that this by itself is insufficient, for there is no *general* set of observable characteristics that distinguishes disease from health. Although individual diseases of course have their defining characteristics, which may be structural, biochemical or physiological, there is no common feature. Further, there are many conditions, particularly in psychiatry, but also in other branches of medicine, where such physical manifestations are absent, even though their existence as diseases is not questioned. Examples would include obsessive-compulsive disorder, schizophrenia, chronic fatigue syndrome and low back pain. In such cases, of which there are many examples, the disease is defined by symptoms of which only the patient is aware, or altered behaviour of which he and those around him are aware: defining features at the physical, biochemical or physiological level are absent, or at least not yet recognized.

**Harm and disvalue – the normative view**

The shortcomings of the naturalistic view of disease, which is in principle value free, have led some authors to take the opposite view, to the extent of denying the relevance of any kind of objective criteria to the definition of disease. Crudely stated, this value-based (or *normative*) view holds that disease is simply any condition the individual or society finds disagreeable or harmful (i.e. *disvalues*). Taken to extremes by authors such as Szasz and Illich, this view denies the relevance of the physical manifestations of illness, and focuses instead on illness only as a manifestation of social intolerance or malfunction. Although few would go this far – and certainly modern biologists would not be among them – it is clear that value-laden judgements play a significant role in determining what we choose to view as disease. In the mid-19th century masturbation was regarded as a serious disease, to be treated if necessary by surgery, and this view persisted well into the 20th century. ‘Drapetomania’, defined as a disease of American slaves, was characterized by an obsessional desire for freedom. Homosexuality was seen as pathological, and determined attempts were made to treat it.

A definition of disease which tries to combine the concepts of biological malfunction and harm (or disvalue) was proposed by Caplan et al. (1981):

*States of affairs are called diseases when they are due to physiological or psychological processes that typically cause states of disability, pain or deformity, and are generally held to be below acceptable physiological or psychological norms.*

What is still lacking is any reference to aetiology, yet this can be important in recognizing disease, and, indeed, is increasingly so as we understand more about the underlying biological mechanisms. A patient who complains of feeling depressed may be reacting quite normally to a bereavement, or may come from a suicide-prone family, suggestive of an inherited tendency to depressive illness. The symptoms might be very similar, but the implications, based on aetiology, would be different.

In conclusion, disease proves extremely difficult to define (Scully, 2004). The closest we can get at present to an operational definition of disease rests on a combination of three factors: phenomenology, aetiology and disvalue, as summarized in Figure 2.1.

Labelling human afflictions as diseases (i.e. ‘medicalizing’ them) has various beneficial and adverse consequences, both for the affected individuals and for healthcare providers. It is of particular relevance to the pharmaceutical industry, which stands to benefit from the labelling of borderline conditions as diseases meriting therapeutic intervention. Strong criticism has been levelled at the pharmaceutical industry for the way in which it uses its resources to promote the recognition of questionable disorders, such as female sexual dysfunction or social phobia, as diseases, and to elevate identified risk factors – asymptomatic in themselves but increasing the likelihood of disease occurring later – to the level of diseases in their own right. A pertinent polemic (Moynihan et al., 2004) starts with the sentence: ‘there’s a lot of money to be made from telling healthy people they’re sick’, and emphasizes the thin line that divides medical education from marketing (see Chapter 21).

**THE AIMS OF THERAPEUTICS**

**Components of disvalue**

The discussion so far leads us to the proposition that the proper aim of therapeutic intervention is to minimize the disvalue associated with disease. The concept of disvalue is, therefore, central, and we need to consider what comprises it. The disvalue experienced by a sick individual has
two distinct components\(^2\) (Figure 2.1), namely present symptoms and disabilities (collectively termed morbidity), and future prognosis (namely the likelihood of increasing morbidity, or premature death). An individual who is suffering no abnormal symptoms or disabilities, and whose prognosis is that of an average individual of the same age, we call ‘healthy’. An individual with a bad cold or a sprained ankle has symptoms and disabilities, but probably has a normal prognosis. An individual with asymptomatic lung cancer or hypertension has no symptoms but a poor prognosis. Either case constitutes disease, and warrants therapeutic intervention. Very commonly, both components of disvalue are present and both need to be addressed with therapeutic measures – different measures may be needed to alleviate morbidity and to improve prognosis. Of course, such measures need not be confined to physical and pharmacological approaches.

The proposition at the beginning of this section sets clear limits to the aims of therapeutic intervention, which encompass the great majority of non-controversial applications. Real life is, of course, not so simple, and in the next section we consider some of the important exceptions and controversies that healthcare professionals and policymakers are increasingly having to confront.

Therapeutic intervention is not restricted to treatment or prevention of disease

The term ‘lifestyle drugs’ is a recent invention, but the concept of using drugs, and other types of interventions, in a medical setting for purposes unrelated to the treatment of disease is by no means new.

Pregnancy is not by any definition a disease, nor are skin wrinkles, yet contraception, abortion and plastic surgery are well established practices in the medical domain. Why are we prepared to use drugs as contraceptives or abortifacients, but condemn using them to enhance sporting performance? The basic reason seems to be that we attach
disvalue to unwanted pregnancy (i.e. we consider it harmful). We also attach disvalue to alternative means of avoiding unwanted pregnancy, such as sexual abstinence or using condoms. Other examples, however, such as cosmetic surgery to remove wrinkles or reshape breasts, seem to refute the disvalue principle: minor cosmetic imperfections are in no sense harmful, but society nonetheless concedes to the demand of individuals that medical technology should be deployed to enhance their beauty. In other cases, such as the use of sildenafil (Viagra) to improve male sexual performance, there is ambivalence about whether its use should be confined to those with evidence for erectile dysfunction (i.e. in whom disvalue exists) or whether it should also be used in normal men.

It is obvious that departures from normality can bring benefit as well as disadvantage. Individuals with above-average IQs, physical fitness, ball-game skills, artistic talents, physical beauty or charming personalities have an advantage in life. Is it, then, a proper role of the healthcare system to try to enhance these qualities in the average person? Our instinct says not, because the average person cannot be said to be diseased or suffering. There may be value in being a talented footballer, but there is no harm in not being one. Indeed, the value of the special talent lies precisely in the fact that most of us do not possess it. Nevertheless, a magical drug that would turn anyone into a brilliant footballer would certainly sell extremely well; at least until footballing skills became so commonplace.

Football skills may be a fanciful example; longevity is another matter. The ‘normal’ human lifespan varies enormously in different countries, and in the West it has increased dramatically during our own lifetime (Figure 2.2). Is lifespan prolongation a legitimate therapeutic aim? Our instinct – and certainly medical tradition – suggests that delaying premature death from disease is one of the most important functions of healthcare, but we are very ambivalent when it comes to prolonging life in the aged. Our ambivalence stems from the fact that the aged are often irremediably infirm, not merely chronologically old. In the future we may understand better why humans become infirm, and hence more vulnerable to the environmental and genetic circumstances that cause them to become ill and die. And beyond that we may discover how to retard or prevent aging, so that the ‘normal’ lifespan will be much prolonged. Opinions will differ as to whether this will be the ultimate triumph of medical science or the ultimate social disaster.

A particular consequence of improved survival into old age is an increased incidence of dementia in the population. It is estimated that some 700,000 people in the UK have dementia and world wide prevalence is thought to be over 24 million. The likelihood of developing dementia becomes greater with age and 1.3% of people in the UK between 65 and 69 suffer from dementia, rising to 20% of those over 85. In the UK alone it has been forecast that the number of individuals with dementia could reach 1.7 million by 2051 (Nuffield Council on Bioethics, 2009).

Conclusions

We have argued that that disease can best be defined in terms of three components, aetiology, phenomenology and disvalue, and that the element of disvalue is the most important determinant of what is considered appropriate to treat. In the end, though, medical practice evolves in a more pragmatic fashion, and such arguments prove to be of limited relevance to the way in which medicine is actually practised, and hence to the therapeutic goals the drug industry sees as commercially attractive. Politics, economics, and above all, social pressures are the determinants, and the limits are in practice set more by our technical capabilities than by issues of theoretical propriety. Although the drug industry has so far been able to take a pragmatic view in selecting targets for therapeutic intervention, things are changing as technology advances. The increasing cost and sophistication of what therapeutics can offer mean that healthcare systems the world over are being forced to set limits, and have to go back to the issue of what constitutes disease. Furthermore, by invoking the concept of disease, governments control access to many other social resources (e.g. disability benefits, entry into the armed services, insurance pay-outs, access to life insurance, exemption from legal penalties, etc.).
So far, we have concentrated mainly on the impact of disease on individuals and societies. We now need to adopt a more biological perspective, and attempt to put the concept of disease into the framework of contemporary ideas about how biological systems work.

**FUNCTION AND DYSFUNCTION: THE BIOLOGICAL PERSPECTIVE**

The dramatic revelations of the last few decades about the molecular basis of living systems have provided a new way of looking at function and dysfunction, and the nature of disease. Needless to say, molecular biology could not have developed without the foundations of scientific biology that were built up in the 19th century. As we saw in Chapter 1, this was the period in which science came to be accepted as the basis on which medical practice had to be built. Particularly significant was cell theory, which established the cell as the basic building block of living organisms. In the words of the pioneering molecular biologist, François Jacob: 'With the cell, biology discovered its atom.' It is by focusing on the instruction sets that define the form and function of cells, and the ways in which these instructions are translated in the process of generating the structural and functional phenotypes of cells, that molecular biology has come to occupy centre stage in modern biology. Genes specify proteins, and the proteins a cell produces determine its structure and function.

From this perspective, deviations from the norm, in terms of structure and function at the cellular level, arise through deviations in the pattern of protein expression by individual cells, and they may arise either through faults in the instruction set itself (genetic mutations) or through environmental factors that alter the way in which the instruction set is translated (i.e. that affect gene expression). We come back to the age-old distinction between inherited and environmental factors (nature and nurture) in the causation of disease, but with a sharper focus: altered gene expression, resulting in altered protein synthesis, is the mechanism through which all these factors operate. Conversely, it can be argued that all therapeutic measures (other than physical procedures, such as surgery) also work at the cellular level, by influencing the same fundamental processes (gene expression and protein synthesis), although the link between a drug’s primary target and the relevant effect(s) on gene expression that account for its therapeutic effect may be very indirect. We can see how it has come about that molecular biology, and in particular genomics, has come to figure so largely in the modern drug discovery environment.

**Levels of biological organization**

Figure 2.3 shows schematically the way in which the genetic constitution of a human being interacts with his or her environment to control function at many different levels, ranging from protein molecules, through single cells, tissues and integrated physiological systems, to the individual, the family and the population at large. For simplicity, we will call this the bioaxis. 'Disease', as we have discussed, consists of alterations of function sufficient to cause disability or impaired prognosis at the level of the individual. It should be noted that the arrows along the bioaxis in Figure 2.3 are bidirectional – that is, disturbances at higher levels of organization will in general affect function at lower levels, and vice versa. Whereas it is obvious that genetic mutations can affect function further up the bioaxis (as in many inherited diseases, such as muscular dystrophy, cystic fibrosis or thalassaemia), we should not forget that environmental influences also affect gene function. Indeed, we can state that any long-term phenotypic change (such as weight gain, muscle weakness or depressed mood) necessarily involves alterations of gene expression. For example:

- Exposure to a stressful environment will activate the hypothalamopituitary system and thereby increase adrenal steroid secretion, which in turn affects gene transcription in many different cells and tissues, affecting salt metabolism, immune responses and many other functions.
- Smoking, initiated as result of social factors such as peer pressure or advertising, becomes addictive as a result of changes in brain function, phenotypic changes which are in turn secondary to altered gene expression.
- Exposure to smoke carcinogens then increases the probability of cancer-causing mutations in the DNA of the cells of the lung. The mutations, in turn, result in altered protein synthesis and malignant transformation, eventually producing a localized tumour and, later, disseminated cancer, with damage to the function of tissues and organs leading to symptoms and premature death.

The pathogenesis of any disease state reveals a similar level of complexity of such interactions between different levels of the bioaxis. There are two important conclusions to be drawn from the bidirectionality of influence between events at different levels of the bioaxis. One is that it is difficult to pinpoint the cause of a given disease. Do we regard the cause of lung cancer in an individual patient as the lack...
The nature of disease and the purpose of therapy

Chapter

The nature of disease and the purpose of therapy

The nature of disease and the purpose of therapy

The nature of disease and the purpose of therapy

The nature of disease and the purpose of therapy

of control over tobacco advertising, the individual’s susceptibility to advertising and peer pressure, the state of addiction to nicotine, the act of smoking, the mutational event in the lung epithelial cell, or the individual’s inherited tendency to lung cancer? There is no single answer, and the uncertainty should make us wary of the stated claim of many pharmaceutical companies that their aim is to correct the causes rather than the symptoms of disease. The truth, more often than not, is that we cannot distinguish them. Rather, the aim should be to intervene in the disease process in such a way as to minimize the disvalue (disability and impaired prognosis) experienced by the patient.

The second conclusion is that altered gene expression plays a crucial role in pathogenesis and the production of any long-term phenotypic change. If we are thinking of timescales beyond, at maximum, a few hours, any change in the structure and function of cells and tissues will be associated with changes in gene expression. These changes will include those responsible for the phenotypic change (e.g. upregulation of cytokine genes in inflammation, leading to leukocyte accumulation), and those that are consequences of it (e.g. loss of bone matrix following muscle paralysis); some of the latter will, in turn, lead to secondary phenotypic changes, and so on. The pattern of genes expressed in a cell or tissue (sometimes called the ‘transcriptome’, as distinct from the ‘genome’, which represents all of the genes present, whether expressed or not), together with the ‘proteome’ (which describes the array of proteins present in a cell or tissue), provides a uniquely detailed description of how the cell or tissue is behaving. Molecular biology is providing us with powerful methods for mapping the changes in gene and protein expression associated with different functional states – including disease states and therapeutic responses – and we discuss in more detail in Chapters 6 and 7 the way these new windows on function are influencing the drug discovery process (see Debouck and Goodfellow, 1999).

**Therapeutic targets**

Traditionally medicine has regarded the interests of the individual patient as paramount, putting them clearly ahead of those of the community or general population. The primacy of the patient’s interests remains the guiding principle for the healthcare professions; in other words, their aim is to address disvalue as experienced by the patient, not to correct biochemical abnormalities, nor to put right the wrongs of society. The principal aim of therapeutic intervention, as shown in Figure 2.3, is therefore to alleviate the condition of the individual patient. Genetic, biochemical or physiological deviations which are not associated with any disvalue for the patient (e.g. possession of a rare blood group, an unusually low heart rate or blood pressure, or blood cholesterol concentration) are not treated as diseases because they neither cause
show how disease affects function at the level of cells, tissues, organs and individuals. The links between the two nevertheless remain tenuous, a fact which greatly limits our ability to relate drug targets to therapeutic effects. Not surprisingly, attempts to bridge this Grand Canyon form a major part of the work of many pharmaceutical and biotechnology companies. Afficionados like to call themselves ‘postgenomic’ biologists; Luddites argue that they are merely coming down from a genomic ‘high’ to face once more the daunting complexities of living organisms. We patient realists recognize that a biological revolution has happened, but do not underestimate the time and money needed to bridge the canyon. More of this later.

THERAPEUTIC INTERVENTIONS

Therapeutics in its broadest sense covers all types of intervention aimed at alleviating the effects of disease. The term ‘therapeutics’ generally relates to procedures based on accepted principles of medical science, that is, on ‘conventional’ rather than ‘alternative’ medical practice\(^6\). The account of drug discovery presented in this book relates exclusively to conventional medicine – and for this we make no apology – but it needs to be realized that the therapeutic landscape is actually much broader, and includes many non-pharmacological procedures in the domain of conventional medicine, as well as quasi-pharmacological practices (e.g. homeopathy and herbalism) in the ‘alternative’ domain.

As discussed above, the desired effect of any therapeutic interventions is to improve symptoms or prognosis or both. From a pathological point of view, therapeutic interventions may be directed at disease prevention, alleviation of the effects of existing disease, or permanent cure (i.e. restoration to a state of function and prognosis equivalent to that of a healthy individual of the same age, without the need for continuing therapeutic intervention). In practice, there are relatively few truly curative interventions, and they are mainly confined to certain surgical procedures (e.g. removal of circumscribed tumours, fixing of broken bones) and chemotherapy of some infectious and malignant disorders. Most therapeutic interventions aim to alleviate symptoms and/or improve prognosis, and there is increasing emphasis on disease prevention as an objective.

It is important to realize that many types of interventions are carried out with therapeutic intent whose efficacy has not been rigorously tested. This includes not

---

\(^6\)Scientific doctors rail against the term ‘alternative’, arguing that if a therapeutic practice can be shown to work by properly controlled trials, it belongs in mainstream medicine. If such trials fail to show efficacy, the practice should not be adopted. Paradoxically, whereas ‘therapeutics’ generally connotes conventional medicine, the term ‘therapy’ tends to be used most often in the ‘alternative’ field.
only the myriad alternative medical practices, but also many accepted conventional therapies for which a good scientific basis may exist but which have not been subjected to rigorous clinical trials.

### MEASURING THERAPEUTIC OUTCOME

**Effect, efficacy, effectiveness and benefit**

These terms have acquired particular meanings – more limited than their everyday meanings – in the context of therapeutic trials.

Pharmacological *effects* of drugs (i.e. their effects on cells, organs and systems) are, in principle, simple to measure in animals, and often also in humans. We can measure effects on blood pressure, plasma cholesterol concentration, cognitive function, etc., without difficulty. Such measures enable us to describe quantitatively the pharmacological properties of drugs, but say nothing about their usefulness as therapeutic agents.

*Efficacy* describes the ability of a drug to produce a desired therapeutic effect in patients under carefully controlled conditions. The gold standard for measurements of efficacy is the randomized controlled clinical trial, described in more detail in Chapter 17. The aim is to discover whether, based on a strictly defined outcome measure, the drug is more or less beneficial than a standard treatment or placebo, in a selected group of patients, under conditions which ensure that the patients actually receive the drug in the specified dose. Proof of efficacy, as well as proof of safety, is required by regulatory authorities as a condition for a new drug to be licensed. Efficacy tests what the drug can do under optimal conditions, which is what the prescriber usually wants to know.

*Effectiveness* describes how well the drug works in real life, where the patients are heterogeneous, are not randomized, are aware of the treatment they are receiving, are prescribed different doses, which they may or may not take, often in combination with other drugs. The desired outcome is generally less well defined than in efficacy trials, related to general health and freedom from symptoms, rather than focusing on a specific measure. The focus is not on the response of individual patients under controlled conditions, but on the overall usefulness of the drug in the population going about its normal business. Studies of effectiveness are of increasing interest to the pharmaceutical companies themselves, because effectiveness rather than efficacy alone ultimately determines how well the drug will sell, and because effectiveness may depend to some extent on the companies’ marketing strategies (see Chapter 21). Effectiveness measures are also becoming increasingly important to the many agencies that now regulate the provision of healthcare, such as formulary committees, insurance companies, health management organizations, and bodies such as the grandly titled National Institute for Health and Clinical Excellence (NICE), set up by the UK Government in 1999 to advise, on the basis of cost-effectiveness, which drugs and other therapeutic procedures should be paid for under the National Health Service.

*Benefit* comprises effectiveness expressed in monetary terms. It is popular with economists, as it allows cost and benefit to be compared directly, but treated with deep suspicion by many who find the idea of assigning monetary value to life and wellbeing fundamentally abhorrent.

Returning to the theme of Figure 2.3, we can see that whereas *effect* and *efficacy* are generally measured at the level of cells, tissues, systems and individuals, *effectiveness* and *benefit* are measures of drug action as it affects populations and society at large. We next consider two growing disciplines that have evolved to meet the need for information at these levels, and some of the methodological problems that they face.

### PHARMACOEPIDEMIOLOGY AND PHARMACOECONOMICS

Pharmacoepidemiology (Strom, 2005) is the study of the use and effects of drugs in human populations, as distinct from individuals, the latter being the focus of clinical pharmacology. The subject was born in the early 1960s, when the problem of adverse drug reactions came into prominence, mainly as a result of the infamous thalidomide disaster. The existence of rare but serious adverse drug reactions, which can be detected only by the study of large numbers of subjects, was the initial stimulus for the development of pharmacoepidemiology, and the detection of adverse drug reactions remains an important concern. The identification of Reye’s syndrome as a serious, albeit rare, consequence of using aspirin in children is just one example of a successful pharmacoepidemiological study carried out under the auspices of the US Department of Health and published in 1987. The subject has gradually become broader, however, to cover aspects such as the variability of drug responses between individuals and population groups, the level of compliance of individual patients in taking drugs that are prescribed, and the overall impact of drug therapies on the population as a whole, taking all of these factors into account. The widely used antipsychotic drug clozapine provides an interesting example of the importance of pharmacoepidemiological issues in drug evaluation. Clozapine, first introduced in the 1970s, differed from its predecessors, such as haloperidol, in several ways, some good and some bad. On the good side, clozapine has a much lower tendency than
haloperidol to cause extrapyramidal motor effects (a serious problem with many antipsychotic drugs), and it appeared to have the ability to improve not only the positive symptoms of schizophrenia (hallucinations, delusions, thought disorder, stereotyped behaviour) but also the negative symptoms (social withdrawal, apathy). Compliance is also better with clozapine, because the patient usually has fewer severe side effects. On the bad side, in about 1% of patients clozapine causes a fall in the blood white cell count (leukopenia), which can progress to an irreversible state of agranulocytosis unless the drug is stopped in time. Furthermore, clozapine does not produce benefit in all schizophrenic patients – roughly one-third fail to show improvement, and there is currently no way of knowing in advance which patients will benefit. Clozapine is also more expensive than haloperidol. Considered from the perspective of an individual patient, and with hindsight, it is straightforward to balance the pros and cons of using clozapine rather than haloperidol, based on the severity of the extrapyramidal side effects, the balance of positive and negative symptoms that the patient has, whether clozapine is affecting the white cell count, and whether the patient is a responder or a non-responder. From the perspective of the overall population, evaluating the pros and cons of clozapine and haloperidol (or indeed of any two therapies) requires epidemiological data: how frequent are extrapyramidal side effects with haloperidol, what is the relative incidence of positive and negative symptoms, what is the incidence of agranulocytosis with clozapine, what proportion of patients are non-responders, what is the level of patient compliance with haloperidol and clozapine?

In summary, pharmacoepidemiology is a special area of clinical pharmacology which deals with population, rather than individual, aspects of drug action, and provides the means of quantifying variability in the response to drugs. Its importance for the drug discovery process is felt mainly at the level of clinical trials and regulatory affairs, for two reasons (Dieck et al., 1994). First, allowing for variability means of quantifying variability in the response to drugs. Its importance for the drug discovery process is felt mainly at the level of clinical trials and regulatory affairs, for two reasons (Dieck et al., 1994). First, allowing for variability is essential in drawing correct inferences from clinical trials (see Chapter 17). Second, variability in response to a drug is per se disadvantageous, as drug A, whose effects are unpredictable, is less useful than drug B which acts consistently, even though the mean balance between beneficial and unwanted effects may be the same for both. From the population perspective, drug B looks better than drug A, even though for many individual patients the reverse may be true.

Pharmacoeconomics, a branch of health economics, is a subject that grew up around the need for healthcare providers to balance the ever-growing costs of healthcare against limited resources. The arrival of the welfare state, which took on healthcare provision as a national rather than an individual responsibility, was the signal for economists to move in. Good accounts of the basic principles and their application to pharmaceuticals are given by Gold et al. (1996), Johannesson (1996) and McCombs (1998). The aim of pharmacoeconomics is to measure the benefits and costs of drug treatments, and in the end to provide a sound basis for comparing the value for money of different treatments. As might be expected, the subject arouses fierce controversy. Economics in general is often criticized for defining the price of everything but appreciating the value of nothing, and health economics particularly tends to evoke this reaction, as health and quality of life are such ill-defined and subjective, yet highly emotive, concepts. Nevertheless, pharmacoeconomics is a rapidly growing discipline and will undoubtedly have an increasing influence on healthcare provision.

Pharmacoeconomic evaluation of new drugs is often required by regulatory authorities, and is increasingly being used by healthcare providers as a basis for choosing how to spend their money. Consequently, pharmaceutical companies now incorporate such studies into the clinical trials programmes of new drugs. The trend can be seen as a gradual progression towards the right-hand end of the bioaxis in Figure 2.3 in our frame of reference for assessing the usefulness of a new drug. Before 1950, new drugs were often introduced into clinical practice on the basis of studies in animals and a few human volunteers; later, formal randomized controlled clinical trials on carefully selected patient populations, with defined outcome measures, became the accepted standard, along with postmarketing pharmacoepidemiological studies to detect adverse reactions. Pharmacoeconomics represents the further shift of focus to include society in general and its provisions for healthcare. A brief outline of the main approaches used in pharmacoeconomic analysis follows.

Pharmacoeconomics covers four levels of analysis:

- Cost identification
- Cost-effectiveness analysis
- Cost-utility analysis
- Cost–benefit analysis.

Cost identification consists of determining the full cost in monetary units of a particular therapeutic intervention, including hospitalization, working days lost, etc., as well as direct drug costs. It pays no attention to outcome, and its purpose is merely to allow the costs of different procedures to be compared. The calculation is straightforward, but deciding exactly where to draw the line (e.g. whether to include indirect costs, such as loss of income by patients and carers) is somewhat arbitrary. Nevertheless, cost identification is the least problematic part of pharmacoeconomics.

Cost-effectiveness analysis aims to quantify outcome as well as cost. This is where the real problems begin. The outcome measure most often used in cost-effectiveness analysis is based on prolongation of life, expressed as life-years saved per patient treated. Thus if treatment prolongs the life expectancy of patients, on average, from 3 years to 5 years, the number of life-years gained per patient is 2.
Comparing cost and outcome for different treatments then allows the cost per life-year saved to be determined for each. For example, a study of various interventions in coronary heart disease, cited by McCombs (1998), showed that the cost per life-year saved was $5900 for use of a β-adrenoceptor blocker in patients who had suffered a heart attack, the corresponding figure for use of a cholesterol-lowering drug in patients with coronary heart disease was $7200, while coronary artery bypass surgery cost $34,000 per life-year saved. As these drugs have reached the end of their patent life and become low-priced generic medicines, the cost difference changes in favour of their use. Any kind of all-or-nothing event, such as premature births prevented, hospital admissions avoided, etc., can be used for this kind of analysis. Its weakness is that it is a very crude measure, making no distinction between years of life spent in a healthy and productive mode and years of life spent in a state of chronic illness.

Cost-utility analysis is designed to include allowance for quality of life, as well as survival, in the calculation, and is yet more controversial, for it becomes necessary somehow to quantify quality – not an endeavour for the faint-hearted. What the analysis seeks to arrive at is an estimate known as quality-adjusted life-years (QALYs). Thus if the quality of life for a given year, based on the results of the questionnaire, comes out at 70% of the value for an average healthy person of the same age, that year represents 0.7 QALYs, compared with 1 QALY for a year spent in perfect health, the assumption being that 1 year spent at this level of illness is ‘worth’ 0.7 years spent in perfect health.

Many different questionnaire-based rating scales have been devised to reflect different aspects of an individual’s state of health or disability, such as ability to work, mobility, mental state, pain, etc. Some relate to specific disease conditions, whereas others aim to provide a general ‘quality-of-life’ estimate (Jaeschke and Guyatt, 1994), some of the best-known being the Sickness Impact Profile, the Nottingham Health Profile, the McMaster Health Index, and a 36-item questionnaire known as SF-36. In addition to these general quality-of-life measures, a range of disease-specific questionnaires have been devised which give greater sensitivity in measuring the specific deficits associated with particular diseases. Standard instruments of this kind are now widely used in pharmaco-economic studies.

To use such ratings in estimating QALYs it is necessary to position particular levels of disability on a life/death scale, such that 1 represents alive and in perfect health and 0 represents dead. This is where the problems begin in earnest. How can we possibly say what degree of pain is equivalent to what degree of memory loss, for example, or how either compares with premature death? This problem has, of course, received a lot of expert attention (Gold et al., 1996; Johannesson, 1996; Drummond et al., 1997) and various solutions have been proposed, some of which, to the untrained observer, have a distinctly chilling and surreal quality. For example, the standard gamble approach, which is well grounded in the theory of welfare economics, involves asking the individual a question of the following kind:

**Imagine you have the choice of remaining in your present state of health for 1 year or taking a gamble between dying now and living in perfect health for 1 year. What odds would you need to persuade you to take the gamble?**

If the subject says 50:50, the implication is that he values a year of life in his present state of health at 0.5 QALYs. An alternative method involves asking the patient how many years of life in their present condition he or she would be prepared to forfeit in exchange for enjoying good health until they die. Although there are subtle ways of posing this sort of question, such an evaluation, which most ordinary people find unreal, is implicit in the QALY concept. Figure 2.4 shows schematically the way in which quality of life, as a function of age, may be affected by disease and treatment, the area between the curves for untreated and treated patients representing the QALYs saved by the treatment. In reality, of course, continuous measurements spanning several decades are not possible, so the actual data on which QALY estimates are based in practice are much less than is implied by the idealized diagram in Figure 2.4. Cost-utility analysis results in an estimate of monetary cost per QALY gained and it is becoming widely accepted as a standard method for pharmaco-economic analysis. Examples of cost per QALY

---

*Imagine being asked this by your doctor! ‘But I only wanted something for my sore throat’, you protest weakly.*
As a result, cost–benefit analysis has been largely shunned as a practical approach for evaluating medicines, but may be unavoidable as more personalized and expensive medicines become available (Hunter and Wilson, 2011 and Chapter 22).

**SUMMARY**

In this chapter we have discussed concepts of disease and the aims of therapeutics, the needs newly introduced drugs have to satisfy, and the ways in which their ability to satisfy those needs are judged in practice. There are many uncertainties and ambiguities surrounding the definition of disease, and ideas are constantly shifting, but the two components that most satisfactorily define it are dysfunction and disvalue. Disvalue, which therapeutic interventions aim to mitigate, in turn has two main components, namely morbidity and prognosis.

We have described the bioaxis, which represents the various levels in the organizational hierarchy of living systems in general, and human beings in particular, and emphasized that disease inevitably affects all levels on the bioaxis. The drugs that we invent home in very specifically on one level, namely proteins, although the effects we want to produce are at another level, namely individuals. Furthermore, we emphasize that healthcare issues are increasingly being viewed from the perspective of populations and societies, and so the impact of drugs at these levels – even further removed from their primary targets – has to be evaluated. Evaluation of drug effects from these rather lofty perspectives, through the application of the emerging disciplines of pharmacoepidemiology and pharmacoeconomics, although fraught with problems, is an important trend the pharmaceutical industry cannot ignore.

Having taken this rather nervy look at the world about us, we turn in the next chapter to discuss the different therapeutic modalities on which the pharmaceutical and biotechnology industries have focused, before retreating to the safer ground at the left-hand end of the bioaxis, where the modern-day drug discovery business begins.

**REFERENCES**


McCombs JS. Pharmacoeconomics: what is it and where is it going? American Journal of Hypertension 1998;11:112S–9S.