CHAPTER 12

INFORMATION TECHNOLOGY IN PLANT DISEASE EPIDEMIOLOGY

ADRIAN NEWTON, NEIL McROBERTS AND GARETH HUGHES

12.1 INTRODUCTION

> "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point." (Shannon, 1948).

Plant disease epidemiology is a discipline that is very often concerned with communicating a message: When is a crop at risk from disease? What are the spatio-temporal dynamics of a new disease? What is the optimum time to take action to limit epidemic development? What are the weather conditions that indicate a high risk of disease development? The answers to these sorts of questions (and many others) contain the messages that epidemiologists routinely attempt to communicate. In order to send meaningful messages epidemiologists are required frequently to deal with large volumes of data generated by, and describing, complex biophysical processes. Given this demand for communication of messages derived from complex processes it is not surprising that there is a long-standing interest in the use of Information Technology (IT) among epidemiologists; the first computer-based simulators of plant disease, for example, were produced in the late 1960s (Waggoner, 1968; Waggoner and Horsfall, 1969). As in many other disciplines, the complexity of use of IT in plant disease epidemiology has grown in step with the opportunity to add complexity offered by advances in personal computer hardware and software over the last 40 years. Whether the increase in available complexity has led to a general increase in the quality of information provided to farmers and growers remains a moot point (Clark, 2000; Henley, 2000). The question here is whether the increasingly sophisticated IT infrastructure available to plant disease epidemiologists has actually been focused on addressing Shannon's (1948) *fundamental problem*.

In this chapter we will discuss two areas of relevance in the application of IT to plant disease epidemiology: (1) data acquisition, organisation and retrieval; (2) the application of knowledge (derived from data) in constructing decision tools in epidemiology. Our discussion will concern the role that IT has played in overcoming, and sometimes contributing to, the fundamental problem as it occurs in plant disease epidemiology. We also look forward, in the light of lessons learned to date, and highlight some areas that require attention in on-going efforts to get the important messages of plant disease epidemiology from their points of origin to points of delivery. At a strategic level we examine some issues concerning the

B. M. Cooke, D. Gareth Jones and B. Kaye (eds.), The Epidemiology of Plant Diseases, 2nd edition, 335–356.

availability and accessibility of information relevant to epidemiologists in the light of an exponentially-increasing volume and diversity of data and data sources. In this discussion we consider both empirical data collected from experiments, surveys etc., and information in publications and databases which has already been processed to some extent. Much of what we say applies equally to both. At the more practical end of epidemiology, in forecasting disease (see also Chapter 9), we highlight recent uses of Bayesian methods to evaluate the performance of decision tools (modern examples of which often have a heavy reliance on IT infrastructure). One of the main advantages of these methods, as we will show, is that they allow epidemiologists to identify those diseases for which the potential exists to change grower behaviour by providing disease or disease risk forecasts. The unifying theme across these areas of discussion is the use and flow of information in plant disease epidemiology under uncertainty.

12.2 DEFINITION OF INFORMATION TECHNOLOGY IN PLANT DISEASE EPIDEMIOLOGY

The definition we have adopted is ‘the use of computer-based technology to collate, process and disseminate information and knowledge for application to the study of plant disease epidemiology’. This includes existing technologies such as computer databases, statistical and field trial design packages, (e.g. Genstat, SAS, S-Plus, Agrobase), automated quantitative diagnostic systems, Laboratory Information Management Systems (LIMS) which may utilise bar-coding tools, and the newer ones referred to below, but the definition is broad enough to allow the inclusion of new developments not yet envisaged. Fig. 12.1 summarises the processes involved in the use of IT to develop epidemiological knowledge from data. Collation, processing and dissemination of information may operate separately or in an integrated manner depending upon the purpose. It is important to note that empirical data will consist of a mixture of facts (i.e. things that are true) and errors (i.e. things that are not true). The extent to which the information comprises only facts will depend on the quality of the mechanism by which the data are collected and the quality of the processing (if any) that is applied to the data to filter out errors. Some databases may simply collate information such as disease incidence or pathotype distribution. Both knowledge arising from the analysis of information (for example by statistical methods) and the information itself might be made available for practical use through some IT medium. In either case validation by successful use will ultimately separate those IT-based systems which are useful from those which are not. In theory, the process of validation-by-use should provide a means for further filtering of errors from facts and an improvement (if that is possible in the particular context) in the quality of available information and knowledge. Clearly, the content of databases and the way in which they are accessed have important consequences for the quality of information and knowledge that results from their use.

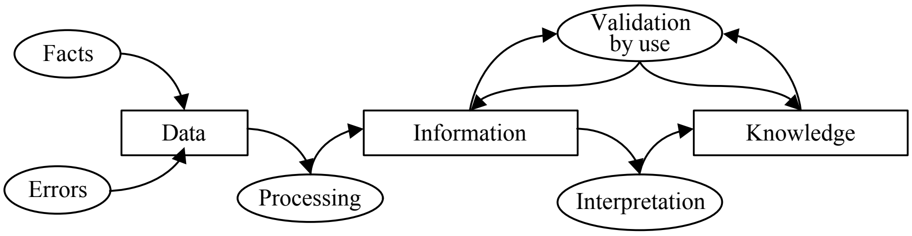

Figure 12.1. The process of knowledge development from data in plant disease epidemiology.

12.3 THE WORLD ACCORDING TO 'GOOGLE'

> "Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information upon it. When we enquire into any subject, the first thing we have to do is to know what books have treated of it. This leads us to look at catalogues and the backs of books in libraries" (Boswell, 1960, p 558).

Increasingly, people approach acquiring information and knowledge by searching the internet. Search engines are powerful and valuable tools. However, like most powerful tools, they can be dangerous in inexperienced hands for the following reasons: (1) searching databases is a skill in which few people have training or experience, and requires an understanding of the database content and search software to be used efficiently (for reasons noted above); (2) the information in the search engine database is likely to be a biased sample of the information available, if for no other reason than because; (3) it is limited to that available on public access servers on the world-wide-web, and (4) there will be a variable level of quality control over on the information, ranging from none to excellent, but probably with little or no indication of what level of credence the user should give to the source. There are many more reasons why information obtained from web searches should be treated cautiously, but we hope that it is apparent that search engines, by themselves, are not a way to obtain an objective assessment of any subject. Perhaps most importantly, whether information is available from direct web searches will depend on a complex set of interacting factors which balance the institutional and personal objectives of the responsible scientist(s) for knowledge transfer with the precedence of peer-reviewed publication and the need for the costs of publication to be recovered.

Much epidemiologically valuable data is made available through the institutions of scientific publishing. Whilst the knowledge published in many epidemiology papers will have been paid for by public funding, and a scientist's usual aim is to disseminate the information free to as wide a constituency as possible, the publishing process has costs which must be met. Furthermore, the requirement to protect originality and precedence in the primary scientific literature leads to time-lags between data being collected and being made widely available. Of course, some of these restrictions are overcome by the use of subscription services for journal access and/or database access. However, even in cases where such access is available, users should bear in mind that reservations, (1) and (2), about public access databases still apply, and further that: (3) the information will still be limited but this time to specific sources, and (4) data quality is defined by the data source.

To epidemiologists, available on-line databases of real-time and historical epidemiological information such as disease incidence, severity, control treatment, host genotype, crop history and other such details are of great value. Similarly, crop advisory information, linked efficiently with its data source and resulting recommendations allows real-world pictures of disease management to be re-constructed, linking theoretical epidemiology with practice. As with research data, crop survey and management databases can be either subscription-access resources or open access. The cautionary comments above notwithstanding, web searches will generate many useful starting points for epidemiologists looking for sources of data. For example, entering the terms ‘crop, scouting, disease, weekly, data, status’ into Google in early 2005 resulted in approximately 39,500 hits. Some selective searching through the first 100 links resulted in the following data sources that give an impression of the sorts of information that can be obtained. Readers are reminded that the sample is selective and intended for illustrative purposes only: www.nysipm.cornell.edu/lfc/index.html – Cornell University’s IPM programme giving access to current and historical information on crop and disease status in the New York State area; www.aragriculture.org/news/cropreport/ – regular reports on crop and disease status published by the Arkansas Cooperative Extension Service; www.oardc.ohio-state.edu/wheat2004/ – web pages reporting results, including large tables of mean results for agronomic and disease resistance scores, of the wheat variety screening trials carried out by Ohio State University; www.sbrusa.net/ – the USDA web site for monitoring soybean rust risk giving up-to-date results from scouting bouts aimed at detecting this new (for the USA) disease. The bias towards USA sites, reflected in the small sample given here should be noted. We suggest two main reasons for this. First, the size and technological sophistication of the USA agricultural science infrastructure is likely to lead to a high representation of USA sites in the searchable population of sites. Second, in the particular context of publicly accessible information on plant disease, the Land Grant mission of relevant USA universities means that there is not the singular emphasis on peer-reviewed publication and impact factors that exists elsewhere, and a corresponding effort on other forms of knowledge transfer (such as web-sites) is made.

Crop Monitor – www.cropmonitor.co.uk/decisiontools/decision.cfm – is a regularly-updated crop disease reporting service for England and Wales which also links to online disease decision tools and risk forecasting systems. In the UK a levy raised on cereal crops is used to support research and knowledge transfer for the benefit of the industry. The levy body (Home Grown Cereals Authority, HGCA) web site has several useful open-access resources including a searchable database of variety performance data – www.hgca.com/varieties/rl-plus/index.html – and historical results from a UK-wide network of trials recording aspects of barley crop development www.hgca.com/BGS/. In Scotland, SAC provides access to these and other similar resources via its web site – www.sac.ac.uk/consultancy/cropclinic/clinic/Adoptacrop. These types of resources are typical of those available in many countries. Most web sites dedicated to plant pathology are run by national plant pathology societies listed by the International Society for Plant Pathology (ISPP) – www.isppweb.org – seeking to serve the needs of their members. These sites could, via the ISPP, serve as a focus for the integration of information technology resources in plant pathology. There are already efforts to provide integrated access to multiple data resources. One example is Plant Management Network International (PMNI) – www.plantmanagementnetwork.org – which is a co-operative body, involving industry, universities and professional bodies such as APS and the Canadian Phytopathological Society. The PMNI web site offers, among other things, the capacity to search a database, dedicated to crop science, constructed from the pooled resources of the partner organisations. Simple searches are free to all users; more detailed searching and full access to the site requires a subscription.

While using on-line search engines does not allow direct access to the primary epidemiology literature, it will, of course, provide the IT-aware epidemiologist with the web sites for journals where the literature can be found. In addition to the major international plant pathology research journals such as *Australasian Plant Pathology*, *European Journal of Plant Pathology, Journal of Phytopathology, Phytopathology, Plant Disease* and *Plant Pathology*, epidemiologists publish their work in a wide range of other journals too numerous to list here. On-line resources such as the ISI Web of Knowledge provide searchable databases of the primary literature allowing information to be retrieved by using keywords independently of the source of the publication. Having promoted the use of on-line databases, we raise the point (often made and often ignored) that they do not generally include information on literature that is more than, typically, 30 years old. Information dating from more than 30 years ago must be obtained by more manual searches, and in this context the final sentence of Samuel Johnson's words remains as true today as when Boswell noted them down in 1775.

For those who may wonder exactly what such a manual search involves, and to highlight the dispersed nature of information relevant to a particular topic, we offer the following quotation, taken from a review of the genetics of horizontal resistance carried out by the late Norman Simmonds (Simmonds, 1991):

> "…it will be appreciated that the literature search needed for the present review presented formidable problems. Some titles were available from reviews but my own reading showed that some valuable papers went uncited. Trials with abstracting journals were not encouraging because of the idiosyncratic and confused terminology adopted and, sometimes, even the apparent failure of authors to understand just what they had done. In the outcome I adopted what might be called a 'brute force' approach. I scanned runs of journals back to about 1965 (about 300 volumes in all) and trusted to secondary sources to identify papers I had failed to find. I might, with advantage, have scanned runs back to the 1930s, but with a declining probability of identifying useful works."

12.4 REAL WORLD DATA CAPTURE

The two crucial factors which initiated the explosion in use of IT are the wide availability of powerful personal computers at very low cost and the growth of the 'internet'. All the IT developments above require standard computers, normally connected to the internet for full operation. This is getting easier with small notebook/laptop computers, wireless connection to the internet, linkage through mobile phones, Personal Digital Assistants (PDAs) etc. Linkage with Global Positioning System (GPS) technology and Geographic Information Systems (GIS) which can assemble multi-layered facts, and associated software such as 'FarmWorks', provide easier access to on-line data in the field allowing synthesis of historical data and current data in real-time rather than relying on going back to an office. Linked to these developments there are now many electronic aids for recording field data automatically with greater speed, volume or accuracy. Examples of instruments include:

- Multi spectral scans (e.g. GreenSeeker and Cropscan) and continuous spectrum scans.
- Remote sensing, e.g. rapid estimation of crop areas, rapid disease level assessment over large areas by satellite spectral/x-ray/infrared etc. analysis.
- Weather stations (for forecasting).
- Photosynthetic measurement systems.
- Digital plant canopy imager and other portable area meters.
- Discriminatory image analysis, e.g. lesion/colony area, space fill.
- Weighing and moisture measurements on harvesters.

In general, digital data capture and storage is now routine. Furthermore, software standardisation on Microsoft® compatible or Open Source formats (and improved transparent translation between these formats) has also increased the speed of integration. The use of remote sensing in phytopathometry is discussed in Chapter 2.

12.5 INFORMATION ACCUMULATION OR DISSEMINATION?

The unplanned use of IT presents the same hazards as the misuse of other technologies. Powerful IT has the potential to allow users to do more than was previously possible. However, this is not a good thing if we do more of the wrong sorts of things. In particular, the ease with which information gets stored rather than being read, analysed, interpreted and then put to further use is a frequent problem associated with easy access to advanced IT facilities, as is unnecessary use of graphics 'because they are there'. Good information systems should always allow easy initial sorting of data into 'discard', 'store for later appraisal' or 'analyse or make full appraisal' categories. Information in the latter category requires more to be revealed for further assessment. However, information classified as 'store for later appraisal' needs to be very carefully labelled, sorted and stored. One point of view is that developments in IT have made it too easy to collect and store data which are then put to no use. Our contention is that this state of affairs is not an argument for not collecting the data, but rather an argument for improving the quality and use of meta-data (i.e. data about data – keywords are a widely used example of meta-data) so that stored data can be easily retrieved and used if the need arises in the future. Particularly in the case of field data, the world does not allow us the chance to make exact repetitions of conditions so the data storage capacity of modern IT should be exploited to allow us to 'go back in time' whenever the need arises. However, we stress that the value in adopting such an approach depends on the quality of the meta-data which describe the contents of the archives of primary data, in much the same way that the quality of databases of publications depends on the quality of the indexing and searching tools which are employed in their construction and interrogation.

Earlier comments notwithstanding, the power of IT for data storage and recovery in the context of publications can lead to the impression that academic publication libraries where hard-copy journals are classified and filed on shelves are obsolete. Hard copy journals are available on-line but retain their page format using 'Portable Document Format' (.pdf) files and must be accessed by viewers such as 'Acrobat'. Ideally, the advent of on-line publishing should have enhanced the value of publications considerably, by allowing much greater access to: libraries of colour graphics; data generated on demand; interactive software tools; active forward and backward links to other information and databases; video images or even sound. However, the vested interests and established system of the big publishing houses simply means more colour images, easier on-line access, and occasionally additional information on the publisher's web site. Indeed on-line publishing rarely even results in faster publication times, which could be from 5 to 9 months less than for 'conventional' journals since individual papers could be available almost immediately they are accepted and formatted. The British Society for Plant Pathology (BSPP) does publish 'New and Unusual Records' of plant diseases on-line as soon as they are accepted, followed by annual hard-copy publication in Plant Pathology. A few other journals are doing likewise. However, the best 'deals' often involve site licences which have to be organised more formally, such as for access to the major bibliographic databases.

Forward citations require the journal to regularly update from citation indices that refer to the article, thereby positioning the publication within a network of information allowing highly flexible citation. Furthermore, software tools can be applied to data published in electronic format which enable each reader to read the paper effectively in a different way by applying further data analyses, informed by more recent data than the author had access to at the time of writing. The use of such tools might lead to cultural changes in publishing allowing the concept of original publication to be preserved, but enabling rapid re-interpretation in the light of more recent data. Raw, or partly processed, data are rarely attached to published articles except theses, but the new technologies open up opportunities to do so in appropriate cases. This would allow further use of the data to the benefit of the original and other authors.

Will textbooks (like this one) be obsolete soon? Should there be a new type of 'electronic text book'? (by 'text book' we mean an assimilating tool giving access to a subject from its fundamental principles; dict: 'manual of instruction, standard book in a branch of study'). Take, for example, someone requiring information on one of the best known plant pathogens, *Phytophthora infestans*, the cause of epidemics of potato late blight. Potential sources of information and resources of interest may be: books (like this), scientific papers, bulletins, crop intelligence reports, conferences, published statistics and surveys, agrochemical control measures available, practical and 'anecdotal' information from agricultural advisors, agricultural policy, germplasm collections of potato, its wild relatives and *P. infestans* and its relatives etc., genetic maps/databases of host and pathogen such as the EUCABLIGHT consortium database of characterised isolates across Europe (www.eucablight.org) scientists working in the field or current work being carried out on research grants. Clearly, to make optimal use of this diversity of information there is the need for a structured information system whereby the various sources can be linked together in a logical and helpful format. However, this demands 'good old-fashioned' experts to invest time and effort to create such resources and this means that the economic case for such resource investments has to be made.

12.6 BRINGING TOGETHER DISCIPLINES

Although the examples above largely bring together knowledge within a discipline, the construction of such web sites involves many challenges. More challenging still, and another dimension where the attributes of IT can be effectively exploited in epidemiology, is in bringing together related, disparate, but relevant, disciplines and making their resources accessible to new users. For example, an important aspect of understanding the epidemiology of plant-microbe interactions is to understand how abiotic stress caused by climate, e.g. drought and heat stresses, can affect the susceptibility of a plant to infection. A web resource called DRASTIC (Database Resource for the Analysis of Signal Transduction in Cells; www.drastic.org.uk) may not at first seem like an epidemiologist's favourite site. Likewise, one of its main resources, a database of plant gene expression data which provides valuable information on the potential interaction between biotic and abiotic stresses, may be difficult at first to relate to epidemiology. However, DRASTIC does not simply provide data from microarrays. The database contains information from a wide range of published papers on whether plant genes are up- or down-regulated in response to various biotic and abiotic stresses. Much of the information is based on experiments with the model plant *Arabidopsis thaliana* and because the database uses Arabidopsis Genome Initiative (AGI) numbers it is possible to be confident about which gene within a family of genes is actually being regulated.

On the DRASTIC web site is a good example of how new knowledge can be obtained by assembling data from disparate sources visually, namely the 'Metabolic pathways of diseased potato' wallchart. Although not an epidemiological example, this was assembled to assist in understanding the complex biochemical changes that can take place in diseased potatoes, but as a wallchart it is fixed both spatially and temporally. Whilst the chart itself is useful it still only touches the surface of making the information within it readily accessible and more importantly, more easily understood by non-specialists. To exploit the power of IT, ideally the information it contains should be stored in a database and drawn on demand from a set of queries. The graphics should not only demonstrate in an accessible and attractive style associations not hitherto observed, but also allow further interaction through an intuitive interface. This is still technologically challenging but becoming practical.

The DRASTIC database contains many examples of *Arabidopsis* genes that are regulated by both abiotic stresses (drought, cold etc.) as well as by infection with a range of fungal, bacterial and virus pathogens. It thereby provides valuable clues as to the molecular basis of how such interactions might occur. The site also brings together information on resistance-inducing compounds that are, or could be, exploited as crop protectants. Current developments centre around text mining as a very powerful tool capable of finding and relating documents on specific subjects to bring together ideas and information from diverse sources in answer to a specific query. We predict that the use of such tools will increase over the next decade as the drive to synthesise diverse sources of information increases. Of course, powerful as such software is, it can only work on data that have been placed on a computer somewhere. Not all of the issues raised by Simmonds (1991) will disappear with new software tools for data mining and management. However, the challenges in the future will be to understand how to construct useful queries and how complex information can be presented in an easily understood format. Indeed, the task of simplifying the presentation of complex information is at the heart of many problems related to biology in general, and epidemiology in particular and is one of the drivers of the development of simple forecasting tools described below, and of the use of methods for the evaluation of those tools.

Making knowledge useful requires synthesis and presentation. For example, the wheat breeding community have done this for marker-assisted breeding tools (maswheat.ucdavis.edu) and a similar site is planned for barley. Another example of more epidemiological relevance is the Cereal Pathogen Resistance Allele Database (CPRAD – www.scri.ac.uk/cprad). This catalogues information on the pathogen resistance alleles reported to be found in genotypes of barley and, in future, wheat and oats. It is fully searchable on a range of criteria and initial query results are displayed in a table. From this table there are links to the source reference or to a pedigree chart with all that is known about the inheritance of the selected allele. It can also show all that is known about all pathogen resistance alleles in any of the genotypes in the pedigree chart. The selected genotype is hypertext linked to display the probabilities that these and other alleles present in the pedigrees are expressed in the genotype. In this way, probabilities that these genes are present in the selected variety but are masked by other genes, for example the *mlo* allele in barley, are displayed. When this information is associated with virulence frequency and virulence combination (race) data, predictions about the likely durability of a variety can be made.

12.7 MODELS, EXPERT SYSTEMS AND DECISION SUPPORT SYSTEMS

There are many Decision Support Systems (DSSs) or software products that integrate diverse information into a decision-making process. The Management Advisory Package for Potato (MAPP: www.potatomapp.co.uk), for example, is a software package that advises growers on seed rates, irrigation strategies and harvest times in order to optimise financial returns based on graded yields. Other examples are 'PC-Plant Protection' in Denmark, 'Plant Diagnosis', 'MoreCrop', 'Orchard

2000' and 'California Pestcast'. There are also many plant and disease models for canopy and root architecture, virtual plants, and virtual crops. Probabilistic systems can mix elements from different systems, and there are many models built to aid understanding of epidemics. Older established examples include Epimul, Epipre (temporal data) and resistan (fungicide resistance dynamics). There are also many disease/pest risk assessment programmes and libraries of electronic images to aid identification (CABI, 'CyberPest' etc.). There are computer-based systems for identification of organisms that use multi-access keys, for example the Biolog bacterial identification system. Many statistical analyses require IT visualisation tools before an understanding of the processes can be gained. For example, techniques such as wavelet transform, degrade or enhance images to define significant features. There are many cladistics/phylogenetic packages such as Fillip, Genstat routines and Bionumerics. Similarly, there are many mapping packages which allow the use of geostatistical methods for interpolation of data from sparse samples and compositor indexing of GIS data by, for example, soil moisture, elevation, phosphorous or nitrogen.

Some resources are freely available as teaching aids. For example 'LateBlight' or 'Blitecast' which enable the effects of the major epidemiological parameters to be simply simulated and represented graphically, demonstrating the principles of epidemiology, interactively and effectively (Fry, 1990). 'DISTRAIN' is a programme for teaching people to assess disease accurately by asking them to estimate the percentage cover then giving feedback (Tomerlin and Howell, 1988; see also Chapter 2).

Decision support systems highlight a problem with many IT tools. As a knowledge delivery mechanism they can be: inflexible, prone to go rapidly out-of-date, expensive, and often insufficiently user-friendly, and therefore frequently fail to deliver effective solutions or live up to users' expectations (see below). Essentially, developers of IT tools face a choice between one of two general aims in developing a new tool: (1) they can opt for multi-objective tools which will integrate large amounts of information, relevant to one or more important management decisions, and act as comprehensive sources of information (traditional DSSs would fall into this category); while such tools provide the user with potentially a single integrated resource for helping with decision-making, they generally have high development and maintenance requirements, for example, to keep lists of approved pesticide products, application rates and costs up to date, (2) alternatively, they can opt for simple tools which focus on individual decisions and utilise only (or mostly) simple generic information; examples of this approach are disease risk predictors based on statistical decision rules (see below, also Yuen and Hughes, 2002). The pros and cons of this option are, roughly speaking, the inverse of those for option (1). Decision tools of type (2) tend to provide less in the way of background or supporting information, but have lower development and maintenance requirements and ought to be simpler to use.

One aspect of integrating IT into the decision-making process that is sometimes overlooked is that decisions are often made collectively by more than one person, or after the main decision-maker has consulted several other people. These people might collectively be termed the decision-maker's Decision Support Network (DSN). As an illustration of the diversity of the constituents of a DSN, McRoberts *et al*. (2000) found that up to eight different types of person (spouse, business partner, children, accountant, advisor, commercial representative, lawyer, employee) were consulted *occasionally* by a group of Scottish arable farmers when making crop protection decisions. No farmer who provided information for the study *never* consulted anyone when making a crop protection decision. The three most frequently reported types of person who were *always* consulted were: advisors (36% of respondents), commercial representatives (24%) and business partners (22%). Whether a new IT-based tool is of type (1) or type (2) it is likely to have to compliment, rather than replace, the activities of the DSN. Within the context of the DSN, DSSs might be better thought of as Discussion Support Systems. On this basis future developments might most effectively be targeted on type (2) tools which are intuitive to use and which focus on the key decision to be made, leaving the DSN to supply the additional, background information often contained in a DSS. These comments notwithstanding, there are, of course, successful examples of type (1) tools in use. We briefly describe some of these in the next section, before dealing with some generic issues of the efficacy of tools, using examples of type (2) tools in the final section.

12.8 SOME EXAMPLES OF DSS

So far in this chapter we have dealt in broad terms with sources and presentation of information relevant to plant disease epidemiology. In attempting to raise some issues concerning the availability and quality of information that can be obtained directly on the desktop of the modern epidemiologist we have largely concentrated on sources of information of use to epidemiologists and on the use of IT simply in making information available. In this section we report some proven or historical examples of IT and DSSs. In the following section we consider recent developments in epidemiology for situations where IT is used in attempts to put epidemiology into practice for the purpose of forecasting disease and making crop management decisions. Disease forecasting is also dealt with in Chapter 9.

In Denmark a DSS for crop protection called ‘PC-Plant Protection’ was devised motivated by a decision made in 1986 to reduce pesticide use in Denmark by 50%. The Danish Institute of Plant and Soil Science and the Danish Agricultural Advisory Centre chose to implement their research findings with a detailed use of threshold values to support decisions on treatment need, choice of pesticides and the appropriate dosage for the actual problem using a PC programme. This is not available on the internet as it is licensed to users. Even as far back as 1996, 2800 licences were in use at agricultural schools and with advisers and farmers. It was judged to be user-friendly and the model was considered to be reliable and to meet the requirements of advisers. In 1996, a farm survey of 488 farmers who had used the system in 1995 showed that the system has been well accepted by farmers, not only because of its reliable recommendations but also because of the good profitability that resulted from its use. It also enabled Denmark to make substantial progress in pesticide input reductions. Development continued with a weather module, taking into account detailed weather information and an environmental module, giving farmers the opportunity to choose pesticides according to their environmental risk, integration of the DSS with site-specific (precision) farming and multimedia presentation of biological and pesticide information.

The same developers, have produced an internet-based platform for the dissemination of up-to-date crop protection information (www.planteinfo.dk). This aims to disseminate new and reliable information on the development of pests and diseases in crops during the growing season and is of great value for farmers and advisory services. Up-to-date calculated risks are based on weather data for *Septoria* species on winter wheat, *Drechslera teres* and *Rhynchosporium secalis* on barley, *Oscinella frit*, *Dasineura brassica*, *Phytophthora infestans* and records available from national pest and disease surveys of all major pests and diseases on winter wheat, winter barley and spring barley. Information can be generated automatically and illustrative maps can be presented such as weather-based risk calculations on maps of Denmark divided into 40 x 40 km grids. The grids are coloured according to the calculated risk: green if no risk, yellow if close to risk, and red if at risk. In each grid a comparison with the developments in previous years can be retrieved in a sub-layer. Linked to the maps is further information on the pathogen or pest and comments from crop protection specialists.

Other examples include apple scab forecasting, a component of 'ADEM', a user-friendly programme for growers and advisors developed by HRI at East Malling, Kent, UK. It also models mildew, canker and fruit rot and fireblight, integrating weather data into simultaneous simulations. Cultivar-specific effects are included and it has been used to achieve mildew and scab control levels similar to those with conventional fungicide programmes but with less fungicide application.

Two products primarily aimed at researchers are the Descriptions of Plant Viruses (www.rothamsted.ac.uk/dpv), formerly 'Plant Virus Notebook', which displays the classification of plant virus genera and families as well as genome structure and organization. The features of virus sequences can be displayed in different ways, graphically or textually, so that the positions and relationships of the features can be examined more easily. A complete plant virus classification is included, together with sample genome maps of the type members of each genus (where available). A related plant virus IT resource is the VIDE (Virus Identification Data Exchange – image.fs.uidaho.edu/vide/refs.htm – database).

12.9 DISEASE FORECASTING AND DECISION MAKING IN AN INFORMATION THEORY FRAMEWORK

Figure 1 in Shannon (1948) presents the fundamental problem of communication in diagrammatic form. Shannon's (1948) figure is reproduced, with adaptations, as Fig. 12.2 here. Shannon (1948) was primarily concerned with the mechanical aspects of communication (although he did explicitly state that "*The* destination *is the person (or thing) for which the message is intended*"). We have increased the emphasis on the human-mechanical interaction implied in his Figure 1 here to highlight the human elements in communication. Fig. 12.2 shows half of what is now generally called the Knowledge Transfer (KT) cycle. The missing half would mirror the first by illustrating the flow of information back from the receiver to the information source. The part of the KT cycle which we have not shown is often concerned with defining a problem which is to be solved and thus defining what the message is to be in the part which we have shown. For example, the problem may be a practical difficulty in knowing when to take action to control disease and the desired message is information/advice that resolves the problem. Clearly, it is possible for uncertainty (or noise) to enter the cycle of problem definition, message formulation and delivery at several points and from several sources. We have indicated some of these in Fig. 12.2, particularly the role that those acting as the Information Source and Destination play. As will become clear in the following discussion, the human element and uncertainty are important factors in the success of any IT-based forecasting system.

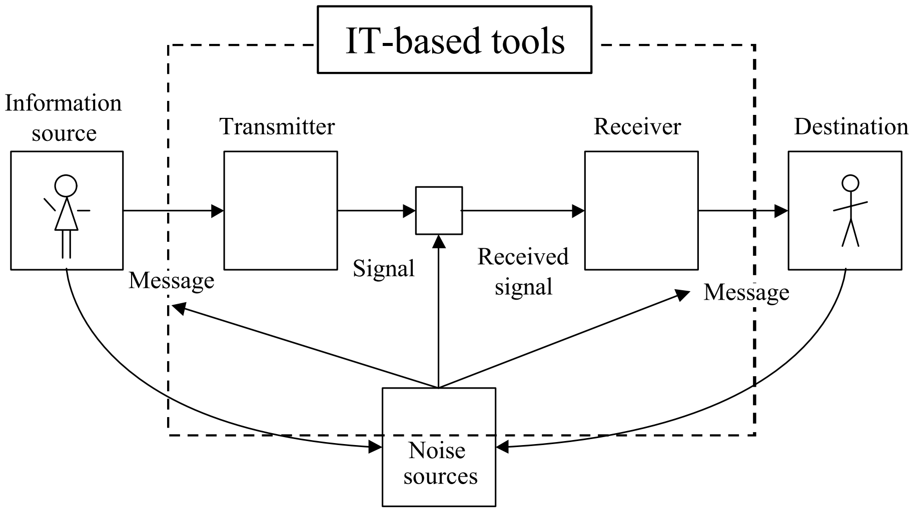

Figure 12.2. An adaptation of Shannon's (1948) schematic diagram of the process of communication illustrating the steps during which uncertainty might arise.

For practical purposes, a forecasting system must do two jobs. First, it must make a useful forecast of disease outbreaks. Secondly, it must present the results of the forecast in a useful format. Early in the development of IT-based forecasters the primary objective was most often to develop tools that would give growers warnings of disease outbreaks so that preventative action could be taken. More recently, concern over the damaging effects of overuse of pesticides on the environment, and a steady reduction in the profitability of crop production, have seen the avoidance of unnecessary pesticide applications become an important objective for disease forecasters. Irrespective of whether the aim of a forecasting system is to provide warnings of disease outbreaks (what might be called positive predictions) or assurances that outbreaks are unlikely (what might, correspondingly, be called negative predictions) the basic performance criterion that any system should meet was spelled out by Campbell and Madden (1990):

> "For a forecasting system to be successful, it must be adopted and implemented by growers. There must be the **perception** that the grower can realize specific, tangible benefits from using the forecasting system **that could not be realized in its absence**." (Campbell and Madden, 1990, p 424; our emphasis).

The quotation from Campbell and Madden (1990) highlights the important relationship that any forecasting system (IT-based or not) must have with the existing knowledge and decision making system of its intended users. The basic point underlying Campbell and Madden's (1990) perceptive statement is that in order to be useful a forecasting system must either: (1) provide new information to the user, or (2) extract previously unrecognized information from already available knowledge and data. Tools developed with these aims in mind might be better described as judgement-assistance or risk assessment tools than DSSs. The objectives for such tools would prioritise user attitudes and preferences, and therefore increase likelihood of not only use, but also feedback, leading to further improvement through active participation of the users in development. Such tools guide growers to a solution, which takes account of the conclusions from the research but which also makes use of growers' experience and their perceptions of risk. It is worth noting the obvious (but sometimes overlooked) point that the combination of objectives (1) and (2) above, constitutes an exhaustive summary of what it is possible for any forecasting system, constructed from empirical data and mathematical, statistical and logical relationships, to do; no such system can contain (or deliver) more information than the content of its components (Doucet and Sloep, 1992). Medawar (1972) made the same point more generally: *"No process of reasoning whatsoever can, **with logical certainty**, enlarge the empirical content of the statements out of which it issues*" (our emphasis).

This epistemic limitation of forecasters, tools, or DSSs holds, no matter how sophisticated the IT infrastructure is that is used in their development or deployment. Medawar's comment is particularly worth bearing in mind when the role of the IT-based forecaster or DSS in disease management is considered in more detail. In addition to the obvious first point (concerning the empirical limitation to the knowledge content of forecasters) it also makes a more subtle point about forecasting. Forecasting entails the loss of logical certainty because forecasts always seek to make statements beyond the empirical content of the information on which they are based. It is thus logically unavoidable that predictions are uncertain, even if we construct deterministic predictors of disease outbreaks that ignore that uncertainty. In the terminology of logical analysis we can say that forecasters are tools for *inductive* reasoning.

However, acceptance of the inductive nature of forecasts and forecasters offers the possibility to evaluate them within a coherent analytical framework using Bayesian analysis within an information theoretical framework. This approach leads to a simple and quantitative assessment of the conditions under which any given forecaster or DSS is likely to meet Campbell and Madden's (1990) '*success*' criterion. (i.e., the extent to which a grower is likely to perceive it as offering something tangibly informative). Evaluation of forecasters in this context is likely to result in more efficient use of IT. It could prevent pointless spending on the development of IT tools which will be useful only to a small proportion of the intended user-base or on tools for problems which cannot usefully be addressed through provision of forecasting systems or DSSs. In this way it might allow plant disease epidemiologists to get closer to solving Shannon's (1948) fundamental problem. The approach also explicitly includes users' perceptions within its analytical scope and thus offers some potential for uniting the technical and human elements of message delivery into one package.

The basic component of the analytical approach – the application of Bayes's theorem to calculate conditional probabilities of events (Howson and Urbach, 1989) – is widely used in medical diagnosis and decision-making. It has been discussed in a plant disease epidemiology context by Yuen and Hughes (2002), McRoberts *et al*. (2003), Yuen and Mila (2003) and Madden (2006). Bayes's theorem provides the means to calculate how the perceived probability of an event (say, the occurrence of a disease outbreak) will change in the light of new information (coming from a forecaster, for example). One version of Bayes's theorem useful in the evaluation of decision tools for control of plant disease, when what we have is an initial assessment of the need for action, is shown in equation 12.1:

$$\textit{Posterior odds}(D+ \textit{(given prediction)}) = LR_x \cdot \textit{Prior odds}(D+) \qquad (12.1)$$

Before describing the terms in equation 12.1, readers are reminded that the odds of an event, E, [odds(E)], is calculated as [P(E)/(1-P(E))]; i.e. the probability that the event occurs divided by the probability that it does not. From this it follows that the probability of an event can be found from the odds: P(E) = [odds(E)/(1+odds(E))]. Equation 12.1, then, states that the odds of need for action, D+ given the prediction, obtained by use of a forecaster with a known LR_x (where x is either + or -), is the product of LR_x and the odds(D+) before the forecast was made (based on whatever information was already available to the decision maker).

The ability of any forecasting system to discriminate between situations where action is needed, or not, can be captured in the likelihood ratio (LR) for the event in question. The LR can be calculated from data describing the frequency with which the forecaster makes correct and incorrect predictions (either positive or negative). This is most easily shown for the case where there is a simple binary choice (action needed or not, for example) and an equivalent binary outcome from the forecaster (prediction of need for action or no need for action). In this case a 2 x 2 decision table gives an exhaustive summary of the results from a set of situations in which forecasts are required and obtained. The frequencies of the four possible outcomes are shown in Table 12.1.

Table 12.1. Possible outcomes of applying a forecaster in a set of situations in which action to control disease is either needed or not needed

	Action actually needed (cases)	*Action actually not needed (controls)*
Forecaster predicts need (+)	a	b
Forecaster predicts no need (-)	c	d

The **sensitivity** of the forecaster is defined as a/(a+c); that is, the frequency of true positive predictions divided by the total number of situations where action was actually needed (so called 'cases'). The **specificity** of the forecaster is defined as d/(b+d); that is, the frequency of true negative predictions divided by the total number of situations in which action was not needed (so called 'controls'). The likelihood ratio for a positive prediction (LR_+) is defined as the sensitivity/[1− specificity] (*i.e.* the true positive proportion divided by the false positive proportion). The likelihood ratio for a negative decision (LR_-) is defined as [1− sensitivity]/specificity (*i.e.* the false negative proportion divided by the true negative proportion). In order to construct Table 12.1 it is necessary to have an independent means of distinguishing cases from controls; i.e., the forecaster must be assessed against an independent 'gold standard'. Murtaugh (1996) provides a useful discussion of this issue in the wider context of ecological indicators. The version of Bayes's theorem given in equation 12.1 allows the LR_x for a forecaster to be combined with an initial assessment of the need for action (the 'prior probability') to produce an updated assessment (the 'posterior probability').

In practical applications of Bayes's theorem in the evaluation of plant disease forecasters, the prior odds have often been based on the long-term prevalence of known cases, and decision makers' judgements of risk are accommodated by specifying different criteria for allocating individual tests to the categories of positive or negative predictions (Hughes *et al*., 1999; Yuen and Hughes, 2002). However, we note that the prior and posterior odds to which equation 12.1 refers may be based on a variety of formal and informal pieces of information and will, in almost every practical situation involving a real decision maker, contain some element of subjective judgement. Indeed, Howson and Urbach (1989; p39) argue that all probabilities (and hence odds) "*should be understood as subjective assessments of credibility, regulated by the requirement that they be overall consistent*". Some scientists may feel uncomfortable with the idea of allowing a definition of probability which uses the word 'subjective' but it should not be a difficult definition to accept for situations in which people are asked to make estimates of probabilities (or odds) without recourse to formal calculation. It should not, therefore, be too difficult to accept as the relevant description of probability for the context of the current discussion, where decision makers are using and assessing IT-based disease forecasters. We note that this view of probabilities as subjective judgements is in keeping with Campbell and Madden's (1990) comment on the importance of the *perceptions* of growers about the usefulness of forecasters in their adoption.

Consider the potential applications of a forecaster for a grower who has already obtained an initial assessment of the need for action. A positive prediction ought to increase the chance that action will be needed. In this case, LR_+ will be the relevant indicator of performance. A negative prediction ought to decrease the chance that action will be needed, in which case LR_- is of interest. It is not a reasonable expectation that any forecaster will give perfect performance. Equation 12.1 allows the limits for what should be expected from forecasters (given an initial assessment of the need for action) to be calculated for either positive or negative predictions (Yuen and Hughes, 2002; Yuen and Mila, 2003).

The following illustrative discussion combines related examples presented in Yuen and Hughes (2002), McRoberts *et al*. (2003) and Yuen and Mila (2003). Imagine a forecaster that has both specificity and sensitivity equal to 0.9. That is, it correctly distinguishes true positives from false negatives 90% of the time (or, in other words, only one in ten 'cases' would (wrongly) not receive treatment in the long run) and true negatives from false positives 90% of the time (only one in ten 'controls' would (wrongly) receive treatment in the long run). In this case, we have $LR_+ = 0.9/(1\text{-}0.9) = 9.0$ and $LR_- = (1\text{-}0.9)/0.9 = 0.11$. Referring back to Fig. 12.2, everything we have said so far about likelihood ratios would fall inside the central box. That is, the message about requirement for action to control disease is captured in the likelihood ratios, which summarise 'the empirical content' of the data, and it is this message that would be contained within the IT-based tool. We can also see that the development and evaluation of tools in this way involves all of the steps in Fig. 12.1 in an explicit manner that makes the development process transparent.

Consider a grower, A, who, based on previous experience, including personal subjective factors and the known long-term prevalence of disease, has a prior probability of 20% for the need for treatment, written as $P_{prior\_A}(D+) = 0.2$ (that is, $odds_{prior\_A}(D+) = 0.25$). Assume that in using the forecaster, A obtains a positive prediction of need for treatment. According to equation 12.1, A's posterior odds, $odds_{post\_A}(D+)$, are $LR_+ \cdot odds_{prior\_A}(D+) = 9 \times 0.25 = 2.25$, corresponding to a $P_{post\_A}(D+) = 0.692$. The conclusion from this is that A will have moved from a position of 'thinking' there was a 20% chance of need for treatment, to thinking that the chance is just under 70%. A second grower, B, has a prior probability $P_{prior\_B}(D+) = 0.8$ (corresponding to $odds_{prior\_B}(D+) = 4$). Assume that in using the forecaster, B obtains a negative prediction of need for treatment. According to equation 12.1, $odds_{post\_B} = LR_- \cdot odds_{prior\_B}(D+) = 0.11 \times 4 = 0.44$, corresponding to $P_{post\_B}(D+) \cong 0.31$. Summarising, we can say that B's position has changed from one in which the perceived chance of need for treatment was 80% to one in which it was just over 30%.

Returning to Shannon's (1948) fundamental problem, the messages that the forecaster can deliver are summarised in equation 12.1. We can paraphrase them in natural language as follows: (i) given a prediction of need for treatment, multiply your personal odds of need for treatment by LR_+ to find out what the odds of need for treatment are now, in the light of the message contained in the (positive) prediction; (ii) given a prediction of no need for treatment, multiply your personal odds of need for treatment by LR_- to find out what the odds of need for treatment are now, in the light of this (negative) prediction. Messages (i) and (ii) are quite simple and although this, in itself, is no guarantee of them being reproduced '*exactly or approximately*' at the point of delivery, it will not decrease the chances of this happening. Shannon (1948) pointed out that delivering a message with high fidelity says nothing about what that message means. A little thought quickly leads to the conclusion that the meaning of messages (i) and (ii) will depend on the user, for at least two important reasons. First, different users might start with different prior probabilities of need for treatment and will therefore have different posterior probabilities of need for treatment even when they obtain the same forecast (receive the same message). Secondly, even if we remove the variation in prior probabilities among users by using the known long-term prevalence of disease as a universal estimate of the prior probability (see Yuen and Hughes, 2002), the message will still mean different things to different people because different people interpret probabilities (and odds) in different ways, depending on their attitudes to risk.

Looking at the examples used above in relation to Campbell and Madden's (1990) success criterion, there is an intuitive sense in which both A and B have been given new information. This results from the fact that the forecaster in both cases suggested a state of affairs which was contrary to their expectations (based on their prior probabilities of need for treatment). In the case of A $P_{prior\_A}(D+)<0.5<P_{post\_A}(D+)$ and for B, $P_{piror\_B}(D+)>0.5>P_{post\_B}(D+)$. A probability of 0.5 is an intuitive point of interest in decision making because it represents the point at which, for an event with two possible outcomes, either outcome is as likely as the other. Thus, if a forecaster is able to move the user's perceived probability of need for treatment from one side of 0.5 to the other it is likely to offer the user something truly informative. This idea also provides us with the means to specify the required performance of our IT tools ahead of constructing them, thus potentially saving time, effort and money. We illustrate the approach by using the case where a negative prediction (i.e., one of no need for action) is of interest.

By fixing the value of the required posterior probability not to take action and re-arranging equation 12.1 (after substituting the relationship between odds and probability) we obtain the required values for LR_- (written LR_{-req}) which will convert prior probabilities to the required posterior probability, as shown in equation 12.2.

$$LR_{-req} = \left(\frac{P_{post\_req}(D+)}{1-P_{post\_req}(D+)}\right) \Big/ \left(\frac{P_{prior}(D+)}{1-P_{prior}(D+)}\right) \tag{12.2}$$

Two examples are shown in Fig. 12.3, where the required posterior probability (P_{post_req}) of need for action is either ≤0.2 or ≤0.1. Some general points about Fig. 12.3 are worth highlighting. First, note that the vertical axis in Fig. 12.3 shows the values as $\ln(LR_{-req})$. The curves in Fig. 12.3 are essentially a pair of calibration curves that translate a prior probability into a required LR given either of the two required (target) posterior probabilities. Note that at the P_{post_req} values (0.1 and 0.2 respectively) the $\ln(LR_{-req})$ value is 0, corresponding to a LR_- of 1 ($e^0 = 1$). Referring back to equation 12.1 we can see that this makes sense. If the prior probability already equals the target posterior probability (and therefore the prior and posterior odds are equal) then the value of LR_- in equation 12.1 must be 1. The behaviour of the $\ln(LR_{-req})$ curves at high P_{prior} illustrates the point, made on several occasions (Yuen and Hughes, 2002; Yuen and Mila, 2003; Madden, 2006), that forecasters must have very good performance in order to bring about large changes in perceptions (i.e. differences between posterior and prior probabilities), particularly when the prior is especially small or especially large. Note that the line corresponding to required posterior probability of 0.1 is lower than that for 0.2 across the whole range of prior probabilities, indicating that, for any P_{prior}, LR_{-req} must be smaller to achieve a P_{post_req} of 0.1 than 0.2. An example, based on an assumed P_{prior} of 0.9 is illustrated in Fig. 12.3. To obtain the value of LR_{-req} we begin by projecting vertically upwards from horizontal axis at 0.9, until the projected line intersects with the LR_{-req} curves. Reading across from the point of intersection to the vertical axis provides the desired value. In the example it can be seen that the $\ln(LR_{-req})$ values for P_{post_req} of 0.1 and 0.2 are, respectively, $\cong$ -4.5 and $\cong$ -3.5. The actual values LR_{-req} obtained from the mathematical relationship are, respectively, 0.012 and 0.028 to 3 decimal places ($0.012 = e^{-4.39}$ and $0.028 = e^{-3.58}$).

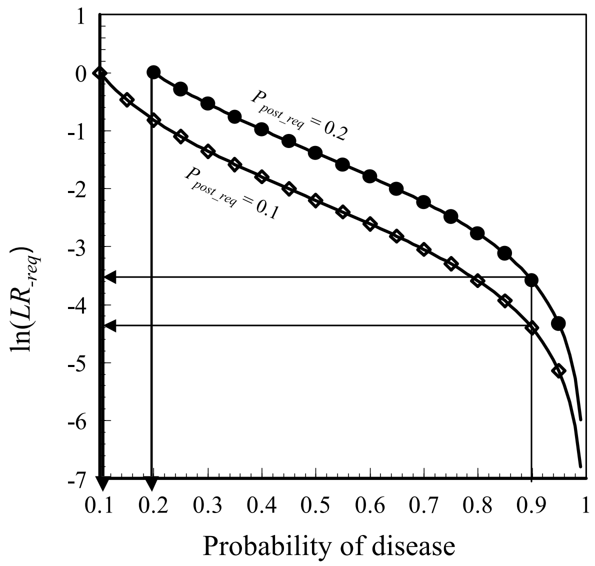

Figure 12.3. Likelihood ratios for the prediction of no disease (LR_-) required to achieve fixed posterior probabilities of disease occurrence as a function of the prior probability of disease. The curves for posterior probabilities of 0.2 and 0.1 are shown.

Recalling the definition of LR_- given above, (i.e. the false negative proportion divided by the true negative proportion) we can translate these values of LR_{-req} into an indication of what we are expecting from our IT-based forecaster in terms of correct and incorrect predictions. For $P_{post\_req} = 0.2$ ($LR_- = 0.028$), for example, the frequency of false negative decisions is (approximately) 3% of the frequency of true negative decisions; i.e., for every 100 correct predictions not to take action, only three incorrect decisions not to take action can be tolerated. The more stringent criterion of having the $P_{post\_req} = 0.1$ ($LR_- = 0.012$) requires the forecaster to make only one false negative prediction for every 100 true negative predictions. How do these standards compare with forecasters that are actually in use? Yuen and Hughes (2002) reported three alternative LR_- values for a *Sclerotinia* stem rot forecaster for use in oilseed rape in Sweden: 0.130, 0.274, 0.684, depending on the choice of risk point threshold for action. De Wolf *et al.* (2003) (see also Madden, 2006) estimated a LR_- of 0.200 for their risk prediction system for *Fusarium* head blight of wheat in the north-east USA. A web-based prediction system (Burnett and Hughes, 2004) for the need for stem-extension treatment for eyespot (caused by *Oculimacula* spp.) in winter wheat developed by SAC and the University of Edinburgh, UK has LR_- values of approximately 0.3 and 0.45 offering growers two levels of risk acceptance. Comparing the values of LR_- for actual forecasters with the value for our hypothetical system, we can see that none of the real systems mentioned has as good performance as our hypothetical system, and, consequently, none of them is capable of generating such large differences between P_{prior} and P_{post}.

12.10 WHERE NEXT?

Bottlenecks in the development of operating systems, computer processing speeds and memory, data storage, and most importantly, communications speed represent potential limitations for applications of IT in plant disease epidemiology. The technology and infrastructure has improved vastly in the last few years, demanding considerable investment by IT users to keep up with developments. The dilemma is now that some control of the quality and appropriateness of material on the internet seems desirable, for the sake of search efficiency. However, control is contrary to the conceptual assumptions of making information available. There is a strong drive towards accreditation of methods and facilities and thereby the data produced, i.e. a move towards Quality Assurance (QA) of data by standardised calibration of instrumentation according to international standards. Security of data and data accreditation and electronic signatures are major issues particularly for protecting intellectual property that can include raw data. As data become more readily available internationally, the same dissemination network should facilitate information validation. The danger lies in the most powerful providers dictating the standards.

The vast improvement in the availability, accessibility and quality of graphics on electronic information systems is one of their great attributes. Although demand for large files increases institutional costs for storage, transmission and processing, efficiencies are also gained by replacing one type of information with others that are more appropriate. For example, images can allow multiple levels of technical description to be bypassed, enabling access by the non-expert into highly specialist fields. This is especially valuable for pest and pathogen identification. Linked with verification data such as geographic location, host etc., many of the pitfalls of such ‘picture book’ pathology can be avoided. This represents a new or enhanced use of IT, where the combination of information results in a better or a new product, not simply the same ones accessed in a different manner. It is such enhanced usages that represent the greatest potential of IT.

A conceptual change is required in order to best utilise IT. The volume of quality information can and should be maximised as the constraints of storage and access are removed. The major proviso is that all information should be appropriately tagged and it must include verification/authentication details to establish its status. Using text mining software can compensate for lack of appropriate meta-tagging. However, many find, for example, editing and proof reading on-screen far more difficult and error prone than on hard copy. Software to aid such processes should be used with discretion as it is seldom ‘smart’ enough. Similarly, authoring software is rarely as user-friendly as paper spread out on a desk even with today’s large, high-resolution flat screen technology. Perhaps the next major visualisation developments that replace screens with true ‘electronic books’, i.e. paper-thin flexible media, will herald the next major advance in IT application. Our attitude to IT should be to adapt our methodology and working practices to use it most effectively and to use established low-technology methods if and when they do the job better or complement the IT.

12.11 CONCLUSIONS

One of the biggest problems in developing information networks will be compatibility and accessibility of information resources. Web interfaces already allow hitherto incompatible data storage formats to be made accessible without having to access the software directly. The user is shielded from the complexities of variations in operating system, application software and hardware platform by the use of a common user interface. Options, preferences and permissions can therefore be tailored to users’ requirements, allowing different levels of access for different needs. There are several examples of information systems that go some way towards the objectives of a common epidemiology information system and point in the direction that such a development might go: the CABI Crop Protection Compendium; veterinary systems such as ‘EqWise’ and human medicine where UMLS (NLM) and Galan (EU) have considered conceptual integration of diverse information types.

However, broad information systems relevant to specific subject areas, such as plant disease epidemiology, are difficult to implement because of lack of funding for ends that are difficult to quantify. Furthermore, sustained development is essential since users’ needs and available resources will change over time. The key to success for such a project is likely to be the active involvement of users in the development of such systems, perhaps borrowing the ethos of the global user community effort in the development of the Linux operating system and associated Open Source software.

If IT is used in a ‘people-centred’ manner, it will add value to our work and increase our understanding of epidemiology. Use of evaluation methods for IT-based forecasting and decision systems should lead to a reduction in reliance of validation by end-use, and may lead in the long-term to a higher adoption rate for such tools among growers and extension workers.

ACKNOWLEDGEMENTS

We thank the Scottish Executive Environment and Rural Affairs Department for financial support and the British Society for Plant Pathology for when one of us (ACN) was actively involved in its IT activities. We thank Simon Oxley (SAC, Edinburgh) and Gary Lyon (SCRI) for valuable discussion and ideas.

REFERENCES

Boswell, J. (1960) *Life of Johnson*. J M Dent & Sons, New York (Originally published in 1793 in Vol II of Boswell's *Life of Dr Johnson*).

Burnett, F.J. and Hughes, G. (2004) *The development of a risk assessment method to identify wheat crops at risk from eyespot*. HGCA Project 2382 Final Report, HGCA, London, UK. www.sac.ac.uk/consultancy/cropclinic/cropadvice/hgcaresearch/eyespotmodel.

Campbell, C.L. and Madden L.V. (1990) *Introduction to Plant Disease Epidemiology*. John Wiley & Sons, New York.

Clarke, W.S. (2000) '*Problems of communication and technology transfer in crop protection': A practitioner's perspective*. Proceedings, The BCPC Conference – Pests & Diseases 2000, pp. 1185-1192, BCPC Publications, Farnham, UK.

De Wolf, E.D., Madden, L.V. and Lipps, P.E. (2002). Risk assessment models for wheat Fusarium head blight epidemics based on within-season weather data. *Phytopathology*, **93**, 428-435. – www. wheatscab.psu.edu/riskTool.html.

Doucet, P. and Sloep, P. (1992) *Mathematical Modelling in the Life Sciences*. Ellis Horwood, New York, NY.

Fry, W.E., Milgroom, M.G., Doster, M.A. *et al*. (1990) 'LateBlight': A disease management game. Cornell University, Ithica, New York, USA. www.apsnet.org/online/feature/lateblit/software.htm.

Henley, P.M. (2000) *Issues in the design and delivery of commercial software applications for crop protection.* Proceedings, The BCPC Conference – Pests & Diseases 2000, pp. 1199-1202, BCPC Publications, Farnham, UK.

Howson, C. and Urbach, P. (1989) *Understanding Scientific Reasoning: the Bayesian Approach*. Open Court, La Salle, IL.

Hughes, G., McRoberts, N. and Burnett, F.J. (1999) Decision-making and diagnosis in disease management. *Plant Pathology,* **48**, 147-151.

Madden, L.V. (2006) Botanical epidemiology: some key advances and its continuing role in disease management. *European Journal of Plant Pathology* (in press).

McRoberts, N., Foster, G.N., Sutherland, A. *et al.* (2000) *Do hunter-gatherers make good farmers?* Proceedings, The BCPC Conference – Pests & Diseases 2000, pp. 11935-1198, BCPC Publications, Farnham, UK.

McRoberts, N., Hughes, G. and Savary, S. (2003) Integrated approaches to understanding and control of diseases and pests in field crops. *Australasian Plant Pathology*, **23**, 167-180.

Medawar, P.B. (1972) *Induction and Intuition in Scientific Thought*. (3rd Edition). Methuen & Co. London, UK.

Murtaugh, P.A. (1996) The statistical evaluation of ecological indicators. *Ecological Applications,* **6**, 132-139.

Shannon, C.E. (1948) A mathematical theory of communication. *Bell Systems Technical Journal*, **27**, 37-423. – www.math.psu.edu/gunesch/Entropy/Infcode.html.

Simmonds, N.W. (1991) Genetics of horizontal resistance to diseases of crops*. Biological Reviews*, **66**, 189-241.

Tomerlin, J.R. and Howell, T.A. (1988) DISTRAIN: A computer program for training people to estimate disease severity on cereal leaves. *Plant Disease,* **72**, 455-459.

Waggoner, P.E. (1968) Weather and the rise and fall of fungi, in *Biometeorology* (ed. W P Lowry) Oregon State University Press, Corvallis, pp. 45-60.

Waggoner, P.E. and Horsfall, J.G. (1969) EPIDEM, a simulator of plant disease written for computer. *Connecticut Agricultural Experimental Station Bulletin*, 698, 80 pp.

Yuen, J.E. and Hughes, G. (2002) Bayesian analysis of plant disease prediction. *Plant Pathology*, **51**, 407-412. – krakatau.evp.slu.se/~evat/cgi-bin/sclerot.pl?first=1.

Yuen, J.E. and Mila, A. (2003) *Are Bayesian approaches useful in plant pathology?* Proceedings, Bayesian Statistics and Quality Modelling in the Agro-Food Production Chain, 95-103, Wageningen, The Netherlands – library.wur.nl/frontis/bayes/index.html.