detection with CT after injection of iodized oil. *Radiology* 1995; **197:** 377–80.
58. Ko S, Nakayima Y, Kanchiro H, *et al.* Liver transplantation for hepatocellular carcinoma: consideration from the findings on autopsy. *Transplant Proc* 1996; **28:** 1691–2.

The (continued) importance of the hypothesis in surgical oncology research

David A. Rew

Senior Lecturer and Honorary Consultant Surgeon, University of Leicester, UK

Introduction

There remains much to be learned about the biology and therapy of human solid tumours, despite remarkable progress in science, medicine and technology in the past century. The well-formulated, explicit and testable hypothesis remains the key to progress in scientific medicine. It should precede experiment and provide landmarks by which experimental results may be judged. Primary clinical research takes three common forms.[1] Experimentation in surgical oncology often reports the measurement of a biological moiety in a laboratory model or series of clinical samples, and correlates it with actual or surrogate measures of outcome. Clinical trials in surgical oncology assess the value of an intervention, such as a modification in surgical technique, a new drug, or a new form of adjuvent therapy, on outcome. Surveys measure a parameter in a cohort of patients. Each form of research mandates a hypothesis.

The scientific problem which underpins the scientific hypothesis may be a general or global issue, or it may be a subsidiary issue within a field of study. Unfortunately, many studies in surgical oncology lack a clear foundation hypothesis. This devalues much research effort, is scientifically inadequate and leads to a waste of human, clinical, laboratory and fiscal resources. This paper addresses the centrality of the hypothesis in the context of modern research processes and technologies.

Definition and functions of the hypothesis

The scientific hypothesis may be defined as an educated guess about creative and logical solutions to the problem.

Correspondence to: D. A. Rew MA MChir FRCS, Senior Lecturer in Surgery, University Surgical Unit, Glenfield Hospital, Groby Road, Leicester LE3 9QP, UK.

The null hypothesis is a sound starting point for many projects. In applying the null hypothesis, the treatment under study is assumed to have no therapeutic effect, or the expression of the biological marker is assumed to make no difference to the biological behaviour or clinical outcome.

The explicit hypothesis serves important functions during scientific work. It reminds the researchers of the goal and obliges forward planning. It imposes discipline and determines the allocation of resources. In a clinical trial, it persuades participating clinicians and patients of the value of the research, thus facilitating patient recruitment and completion of the study.

Creation of the hypothesis

Hypotheses arise because individual creative thought has a prior framework of knowledge upon which to build. Recognition of the problem which demands resolution will be conditioned by current knowledge, by prevailing attitudes, by previous research and by fresh insights. The personality, experience and abilities of its creator will be important. Factors will include a familiarity of the subject, motivation, conceptual skills and a critical faculty. The creative thinker will be able to suspend judgement during the formulation of the hypothesis and pending the interrogation of the experimental data.

The hypothesis should be formulated in advance of the research project, although it may be modified later in the light of new evidence. Thorough and critical analysis of the literature in advance of a project allows cross-checking of the hypothesis against published work and may avoid wasted research effort. Modern computer databases and search engines also reduce the risk of oversight of other published work. In the competition for recognition and resources, it is important not to overlook, understate or minimize the contribution of other groups to a field of knowledge and to the construction of the hypothesis.

Much intellectual energy has been devoted to the philosophical precepts underpinning the hypothesis, and the nature of absolute or relative truth which it addresses. The hypothesis may emerge from a curious observation, from a logical process of creative thought, in which case it is held to be inductive, or from observation of available data, in which case it is held to be deductive. The philosopher Karl Popper is closely associated with the advocacy of deductive thinking, and the process of falsification of the hypothesis by experimental testing. Abstract and semantic arguments about the nature of the hypothesis should not obscure the very practical and important functions which the well-founded hypothesis subserves.

The process of creation of hypotheses is one of continual change and maturation. As one hypothesis is resolved, so new hypotheses may emerge, producing a virtuous and continuing cycle of intellectual and technical progress.

The hypothesis in surgical oncology and pathology

Clinicopathological research in surgical oncology commonly encompasses the study of the expression of molecular moieties in clinical samples of human tissues and tumours. Such studies may be prospective, but are more often retrospective. Prognostic studies seek correlations between measurable parameters in a cohort of samples and clinical outcome data. Common examples of biomarkers selected for such research include DNA content, for S-phase fraction and DNA ploidy measurements; nuclear proteins such as p53 and c-myc; indices of apoptosis; membrane antigens such as growth factor receptors, adhesion molecules, and the drug resistance-associated glycoprotein pumps.

The clinicopathological hypothesis should be founded upon consideration of the known biological role of the moiety under study, and its relevance to the clinical question posed. For example, the study may be investigating clinical outcome in relation to the expression of a particular marker by a particular tumour. In this case, the researcher should consider the biological role of the molecule and decide whether it could reasonably be expected to have an influence on the biological aggressiveness of the tumour.

The form of the clinicopathological hypothesis will also influence the study design. The hypothesis might take the form: 'the expression of protein N as measured by method M has no significance for clinical parameter or event P in samples in fixative F of tumour class T'. This simplistic format proposes a null hypothesis which addresses the biological function, the technical environment and the constraints upon the study.

Inadequacies in clinicopathological hypotheses

The lack of a clear hypothesis is a feature of many studies in clinicopathological research. The measurement of biological parameters can become an end in itself, and a retrospective search for statistically significant correlations within random data sets can substitute for careful thought and critical observation. That a proper hypothesis has not been constituted in advance is often clear in submitted manuscripts.

Failure to formulate a hypothesis leads to a process which may be regarded as a 'trawl' for positive correlations and publishable 'results'. The trawl may take a number of forms, including the following:

The Sample Trawl, which starts with the premise: 'we have this collection of tumours, or clinical cases; let's see what data we can obtain from them'. This approach is not founded on a clear hypothesis and is unlikely to lead to valuable insights.

The Technology Trawl, which may seek to utilize a particular piece of research equipment without a clear hypothesis. It may then emerge in the form 'we have this piece of equipment E, let's see what we can measure with it'.

The Literature Trawl, or failure thereof, may produce excessive duplication and repetition of published studies. This is a common problem. Some duplication is of value to the scientific community, because it allows for the cross-checking of results and ideas. However, many studies continue to be submitted long after the subject has been exhausted in print without adding new insights, or long after papers, reviews and articles have addressed the limitations and problems with a technique.

The Data Trawl is the consequence of abuse of the process of statistical analysis. Statistical software packages have considerable power to analyse research data. A particular problem lies in their misuse to examine data from all perspectives in the hope of discovering significant differences. Correlations may be found which are entirely spurious epiphenomena, and which may be used to reconstruct the paper and its title.

The Socio-geographic Trawl is a repetitive search for data among particular geographic, racial, national or socio-economic groups. Many studies appear to use a different geographic location, country of origin or racial group as the only justification for a study. While it may be appropriate to seek differences in tumour biology from one population to another in order to deduce genetic, dietary or environmental causes, the researchers must offer valid scientific and hypothesis-led justification where such manuscripts are proferred. Such work is otherwise unlikely to produce new insights.

Testing the hypothesis

Hypotheses must be testable by experimentation using available technologies, by clinical observation, or by using epidemiological data. Hypotheses should be tested to logical destruction before physical resources are committed. One way to do this is to use the 'so what?' test. By asking 'so what?' of each assumption and statement in the tentative hypothesis, it is possible to refine the hypothesis and to establish a clear biological or clinical justification at the outset.

The tools for the task

The relationship between the hypothesis and technological progress is reciprocal. Science evolves largely because new technologies allow new approaches to problems and new

ways to test hypotheses. Consider, for example, the contribution of the development of the microscope and its components such as lenses, to the subject of histopathology. Conversely, a hypothesis whose testing is outwith the range of existing technology may allow the specification and development of new technologies and equipment.

New technologies have opened up a whole range of opportunities for the construction and testing of new hypotheses in surgical oncology in the late twentieth century. At the human population level, new techniques in computer communication, in electronic data collection, exchange and analysis, and in applied statistics (for example, cluster analysis), allow new forms of epidemiological study and multicentre clinical trial.

At the whole body and organ level, new forms of imaging (CT, MRI scanning) and endoscopy allow new forms of population screening for cancer to be designed. At the tissue level, new forms of stain technology, immunohistochemistry and fluorescence microscopy offer new ways to study normal and diseased tissues.

At the cellular level, quantitiative analysis using tools such as flow cytometry and laser scanning cytometry allow rapid quantitative studies of prognostic and therapeutic markers and of population heterogeneity.

At the subcellular level, electron microscopy and confocal microscopy allow the detailed study of static and dynamic architecture in the cell. At the genetic level, a host of techniques, including fluorescence *in situ* hybridization (FISH), Polymerase Chain Reaction (PCR), and DNA sequencing allow the definition of the genetic abnormalities in malignant cells down to the single codon error.

Model systems

Model systems have many advantages in cancer research. They simplify the complexity inherent in human tumours and provide for controlled experimental conditions. They allow sufficient studies to be undertaken to account for biological variation, at the time and convenience of the research team.

The principal model systems in applied cancer research are cell cultures and animal models of tumour biology. Cell line and animal models provide simple and controllable systems which can be designed and obtained to order to model particular genetic, physiological or functional traits and defects.

Unfortunately, the simplicity with which model systems can be selected or engineered to order to perform specific experiments disguises the extreme difficulties in translating experimental data to human tumours, which are immensely more complex than are experimental systems. In the case of cell lines, experiments cannot reflect normal physiology by the nature of culture systems. In small animal models, metabolism, physiology, immunology, scale and the nature of the experimental tumour, either induced or explanted, do not usually reflect the true biological environment and behaviour of human tumours *in vivo*.

Statistical planning and testing

Sound statistical principles should be employed at the outset to calculate the numbers of samples to be analysed in order to prove or refute the hypothesis. This will indicate that the hypothesis cannot be tested with the likely number of samples or in the time available. Modern software may allow modelling of the study to test the design further.

Particular care must be taken to avoid type I and type II statistical errors. A type I error implies the assumption of a significant statistical difference between data sets where no difference actually exists. A type II error is the missing of a true difference in data sets which are erroneously assumed to be similar.

Tissue processing and sampling

Tissue processing also handicaps the analysis of archival material. The type and duration of tissue fixation, and of pre-staining preparation, are critical. Formalin fixation may damage protein epitopes. In studies requiring sample disaggregation into suspension for laser cytometry, for example, considerable fragment artefact may be introduced during preparation.

Another problem with the archival histological sample block is the uncertainty over the degree to which will be representative of the original tumour. Unless the tissue archive was established for the express purposes of the particular study, it cannot be assumed that the sample is appropriate for the hypothesis posed. For example, a biopsy from the necrotic central area will be very different from one from the growth face of a tumour.

Controls and optimization

Studies need to be optimized for a particular marker, and appropriate controls selected. Biological moieties may be extremely sensitive to subtle variations in pH, temperature, exogenous signals and growth factors, and hypoxia. Appropriate and sufficient controls must be introduced to account for all variables. The use of normal tissue controls is often particularly important. This will normally be the tissue of origin of the tumour, for example breast epithelium, colorectal mucosa, or normal bone marrow. The expression of the moiety under study should be evaluated in both tumour and tissue of origin, to reduce the risk of misinterpretation of data, for example in relation to the assessment of function of mutant and normal key regulatory molecules in tumour cells.

Time as an impediment to testing

Static measurements of dynamic processes

All biological processes are dynamic, transient events and all proteins and enzymes have kinetics of activity and turnover. Key functions and dynamic cellular processes may not be apparent in histochemical sections. A static measurement may give a wholly misleading picture of biological events within the tissue. It will give no indication of the rate of a process, as a single satellite photograph will give no measure of the traffic speed on a highway.

The limitations of static measurements are also acute when applied to individual molecular species such as

oncoproteins, enzymes, signalling factors, membrane receptors and multidrug resistance pumps. All of these molecules subsume biological functions which are dynamic in time and space. The half-life of many proteins is short, and turnover and degradation is rapid. This renders their detection difficult. These molecular processes may not adequately be assessed by counting the number of cells expressing the moiety or even by counting the number of molecules present within each cell. Knowledge of the membrane turnover of the protein and its rate kinetics for transport of a particular drug is also necessary in order to infer the drug resistance characteristics of the cells. Moreover, some important molecules may be expressed at levels below the limits of conventional measurement techniques. Others may effect functions through conformational changes, molecular associations or metabolism such that the apparent cell content of the moiety is unchanged.

Three examples illustrate the problem of static sampling. One is the study of cell proliferation, which may be assessed by mitotic counting. However, mitosis is a very rapid process (2–3 h), and only a very small proportion of cells will be in mitosis. The mitotic index will thus be very low, even in a highly proliferative tissue.[2] Continuous dynamic tissue labelling with a thymidine analogue identifies a much larger proliferative compartment, as cells move into the S-phase over time and incorporate the analogue into their DNA.

Another salutary example is that of apoptosis.[3] The process of programmed cell death is critical to the modelling and growth of tissues and tumours, and yet it went unrecognized by pathologists until long after it was described because apoptosis itself is so rapid as to leave virtually no trace in tissue sections when viewed by conventional stains.

The third is that of sampling on the tumour growth curve. Tumours, like embryos, grow dynamically and variably in space and time. Measurement of one parameter at one point in the tumour growth curve may not be indicative of expression throughout the tumour's life, and may invalidate comparisons between tumours sampled at different stages in their growth curve. Cell proliferation studies are an example of this problem. During early development, the proportion of proliferating cells (and the growth fraction) may approach 100%. As the tumour expands and outgrows its blood and nutrient supply, so the growth fraction declines and processes such as necrosis supervene.

The biology of metastases and tumour aggression

The measurement of a parameter in a resected primary tumour may have little bearing on the behaviour of the metastases which determine clinical outcome. Most data derived from histochemical, flow cytometric or molecular biotechnological techniques describe specific features of malignant cells within primary tumours. Metastases are less commonly the subject of extensive clinicopathological studies than primary tumours. This reflects the impracticability in many cases of obtaining consistent histopathological material from metastases. Moreover, we cannot be sure that biological features which distinguish malignant from normal cells, such as aneuploidy, are those which determine biological aggressiveness. The metastatic cells may be a highly selected and genetically diverse offshoot of the primary tumour.

Technology and the clinicopathological hypothesis

The human eye and brain remain unsurpassed in many of the attributes needed to study the biological complexity inherent in tissue and tumour samples. However, human researchers suffer from fatigue, limited data collection and handling capabilities, and are confined to the visible wavelengths. To add value to clinicopathological research in surgical oncology, research equipment must display a number of capabilities, such as data processing by computer, or light measurement in the infra red and ultraviolet ranges of the spectrum. These include a rapid scanning and data collection capacity, the ability to identify discrete cells or nuclei, the ability to measure meaningful signals from the particles under study, and the ability to collect, store, analyse and present the resulting data.

The quantitative measurement of cell markers in human tissues and solid tumours has been facilitated by a wide range of monoclonal and polyclonal antibodies, and by a series of twentieth century technologies, among the most important of which in applied surgical oncology research have been immunohistochemistry (IHC), image cytometry (ICM), flow cytometry (FCM) and laser scanning cytometry (LSC). These technologies have advanced dramatically our ability to perform quantitative studies on complex biological samples; to store, record large quantities of data in real time; and to undertake studies below human optical resolution or outwith the visible spectrum. Important technical advances include the digital computer, the laser, refined fluorochromatic dyes, monoclonal antibodies and new stain technologies.

There are many conceptual and technical problems which arise from over-ambitious and uncritical interpretation of data generated using modern technologies. Problems and pitfalls with the technologies used in clinicopathological research often go unrecognized. This impairs the testing of the hypothesis.

Immunohistochemistry (IHC) allows the qualitative and semiquantitative study of specifically labelled cells within the preserved two dimensional architecture of the tissue or of cytology smears using light microscopy. It allows cells containing an antigen or epitope to be illuminated within the tissue, such that their presence can be quantified as a proportion of the total number of cells in the field.

Quantitative IHC which depends upon manual counting is limited by observer fatigue and by observer bias. The quality and consistency of IHC is determined by a number of other important factors. These include fixation and staining, and variation in the quality of complex biological reagents such as monoclonal and polyclonal antibodies batches. Thus, a hypothesis in an IHC study should anticipate the conditions to minimize artefact. For example, will the analysis be performed on fresh or archival samples? If archival, what steps will be taken to use fresh samples for comparison? What steps will be taken to control for variation at each stage on the staining process? How will

varying the concentration of applied reagents, in particular the primary antibody, influence the staining? What steps will be taken to ensure that representative samples are selected for counting?

Image cytometry allows quantitative studies of histochemical and fluorochrome stained tissues. Optimism that image processing would allow the substitution of the human observer, for example in the automation of cervical smear cytology, by machine has not yet been borne out. Computerized image processing of cancer cells remains more laborious and inaccurate than the trained human observer.

Flow cytometry (FCM) is a versatile technology for the study of heterogenous suspensions of cells or nuclei such as are obtained from tumour biopsies. Its use has been widely reported in the surgical oncology literature in DNA analysis (diploidy, aneuploidy, aneuploidy and S-Phase Fraction, SPF, measurements) of solid tumours. The flow cytometer recognises cells in suspension during analysis by their light scatter and fluorescence characteristics. The technology is fundamentally the study of uncertainty, in that the morphology of individual cells cannot be correlated with their corresponding light pulses by the nature of the analysis. Where cells have a consistent size, optical morphology and CD antigen characteristics, such as leukocytes, the type of cell can be deduced with reasonable confidence from the machine's data output (dot plots and histograms). However, where the sample is a complex mixture of irregular cells of various lineages, and where it includes complex artefact, such as may be derived from a solid tumour biopsy, the simple assumptions of flow cytomteric analysis are confounded.

The detailed technical limitations of flow cytometry include unrecognized degradation in performance of individual machines, problems of individual observer and interinstitutional variation in the interpretation of data sets, and failure to recognize artefacts.[5 7] Many reports of FCM-based clinicopathological studies fail to address these technical problems, which should be recognized during the formulation of the hypothesis and for which there should be ample compensation in the study design.

Laser scanning cytometry[4] is a newer technology which shares analytical processes with flow cytometry but which may address some of the imitations of FCM in cancer research. It allows direct inspection of cells undergoing laser excitation and constrained on a microscope slide rather than moving past a fixed laser beam. This direct visualization of complex samples, including those derived from smears and fine needle aspirates allows verification and validation in one instrument, and may thus find many applications in cytopathology and tumour research, where it decreases uncertainty about the nature of the cells under study.

Concluding comments

Surgical clinicopathological studies play an important role in translating laboratory research into clinical progress, and in testing biological hypotheses in the rough and ready environment of solid tumours. Many studies in surgical oncology and pathology require greater rigour in the formulation of their underlying hypotheses. We must ensure that the necessary intellectual discipline and care is applied to this important field of human scientific endeavour, if we are to produce cost-effective and clear-sighted research, and to ensure a steady flow of high quality manuscripts into scientific journals.

The checklist

1. The hypothesis and experimental design

What is the principal problem?
What are the subsidiary problems?
Does a hypothesis exist, and what is it?
Is the hypothesis testable using the clinical material, technology, time scale, skills available?
Have the appropriate statistical tests been chosen at the outset?

2. Biological variables

What is the biological function of the study moiety?
What is its dynamic status, half life and stability in the tissue under the experimental conditions?
What is the impact of tissue collection, processing and storage?
Will the appropriate comparisons be made between fresh and archival material?
How representative of the whole tumour is the sample analysed?
What steps will be taken to assess all forms of intra- and inter-tumour heterogeneity within the sample population?
Will sufficient samples be studied to address the hypothesis competently?

3. Technological variables

Will the appropriate technical issues be addressed for the measurement technique employed?
What are the causes of artefact? Have these been controlled?
What are the problems in data interpretation and analysis? Have these been declared and addressed?

4. The finished paper

Does the title reflect the original hypothesis and aims?
Is the hypothesis clearly expounded?
Is the methods section designed to address all foreseen biological and technical constraints on the study?

Is the results section an accurate report of the findings? Is undue weight afforded to spurious or coincidental findings?
Is the discussion an honest and realistic appraisal of the constraints and the results?
Is a fair and balanced review of the pre-existing literature acknowledged in the references?

References

1. Greenhalgh T. How to read a paper: getting your bearings. *BMJ* 1997; **315**: 243–6.
2. Potten CS, Kellett M, Roberts S, Rew DA, Wilson GD. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut* 1992; **33**: 71–8.
3. Wyllie AH. Apoptosis and the regulation of cell numbers in normal and neoplastic tissues: an overview [Review] *Cancer & Metastasis Rev* 1992; **11**: 95–103.
4. Reeve L, Rew DA. New technology in cell analysis: the Laser Scanning Cytometer. *Eur J Surg Oncol* 1997; **23**: 445–50.
5. Wheeless LL, Coon JS, *et al.* Precision of DNA flow cytometry in inter institutional analyses. *Cytometry* 1991; **12**: 405–12.
6. Wersto RP, Liblit RL, Koss LG. Flow cytometric analysis of human solid tumours: a review of the interpretation of DNA histograms. *Hum Pathol* 1991; **22**: 1085–98.
7. Shankey TV, Rabinovitch PS, Bagwell B, *et al.* Guidelines for implementation of clinical DNA cytometry. *Cytometry* 1993; **14**: 472–7.

Answers

(to self-assessment exercise on p. 59)

1. Although the incidence of cutaneous melanoma has risen more rapidly than any other malignancy, the evidence from Europe suggests a plateauing of incidence. In some parts of Europe (e.g. Switzerland), however, the rate has continued to double over the past decade.

2. Although the aetiology of melanoma is poorly defined, a number of risk factors have been established (e.g. giant congenital naevus, naevi with architectural disorder, large number of naevi, celtic complexion, excessive sunburn, family history of melanoma).

3. This is not correct. Growth patterns define four major types (superficial spreading (65%), nodular (25%), lentigo and acral lentiginous). Clinical characteristics are important in establishing a diagnosis and are based on the ABCDE concept (A = asymmetry; B = border irregularity; C = colour variegation; D = diameter >6 mm; E = enlargement).

4. That is correct. Nodal status affects significant primary management and long-term outcome. In superficial lesions (<1.5 mm), radiographic and laboratory screening tests are of little practical value.

5. That is not correct. An excisional biopsy is the most appropriate diagnostic procedure for a skin lesion thought to be a melanoma, provided it is anatomically and cosmetically feasible. Incisional biopsy can be done for very large or advanced lesions and/or where complete excision can lead to a disfiguring outcome (e.g. on the face).

6. That is correct; two recognized methods (Breslow thickness, Clarke levels) are used. Breslow micro-staging is accurate and the most important prognostic parameter (melanomas of $\leq$0.75 mm show a 10-year survival of 95%; lesions between 1.51 and 3.00 mm about 60%; melanomas >3.00 mm have a 40% 5-year survival).

7. These are indeed key factors in assessing prognosis. For clinically localized lesions various other parameters have also been shown to be important—gender (females better than males); anatomic site (limbs better than axis); tumour characteristics (ulceration, regression, lymphocytic reaction, angio-invasion, etc).

8. Prospective randomized trials have shown that the margin of skin excised should be based on the Breslow thickness of the tumour (e.g. 1-cm margin for tumours $\leq$1 mm thick, possibly for 1–2 mm lesions as well; 2-cm margins for lesions 2–4 mm thick). The incision should extend down to the deep fascia. Undermining of skin edges is acceptable to achieve primary closure.

9. Although amputation is the best method of treatment for such lesions the level can vary. In the case of toes it is at the metatarso-phalangeal joints, whereas for the hand digits it is at the proximal interphalangeal joint.

10. The role of regional lymph node dissection is an area of continuing debate. Patients with thin (<1 mm) or thick (>4 mm) melanomas will not benefit from elective surgical removal of nodes. For lesions of between 1 and 4 mm, various studies have produced conflicting results. The operative staging is unreliable. Sentinel node biopsy may be an important advance in management. When the nodes are known to be involved (e.g. FNAC established) regional resection and *en bloc* removal is carried out to achieve control of loco-regional disease.