use U80. Both doctors and patients would resist doubling the injection volume, which would result if U80 was abandoned and U40 became standard. If we standardised on U80 we would be the only country in the world likely to do so, and old syringes and obsolete instruction leaflets would continue to cause confusion. U100 has already been adopted by the United States, Canada, and Australia, with New Zealand following shortly. After seeing the results from clinical trials at several British centres, considering the experience of other countries, and debating the problems and hazards of change, the British Diabetic Association determined to seek standardisation of insulin strength at U100. Doctors who belong to the BDA voted unanimously for this change. At a meeting in Aberdeen in September 1980 they showed their disquiet that the change was proceeding so slowly. Other professional bodies have also formally welcomed the principle of U100 as the sole insulin strength: the Royal Colleges of Physicians of London, Edinburgh, and Ireland; the Royal College of Physicians and Surgeons of Glasgow; the Royal College of General Practitioners; the British Paediatric Association; and the General Medical Services Committee of the British Medical Association.

With U100 insulin will come a new British Standard U100 insulin syringe, graduated and numbered directly in units of insulin. There will be two sizes—one for injections of up to 50 units, the other up to 100 units and both marked in units. The 50-unit syringe will allow even small doses to be measured and delivered accurately. These syringes have already been shown to be suitable for small children. Plastic syringes for U100 are already available and though in most areas patients still have to

buy them, they are very cheap if used repeatedly, safe,¹ and recommended by most clinics and by the BDA. We hope that the change to U100 will come in 1982. The change-over will probably be largely carried out at diabetic clinics, though general practitioners and all other relevant agencies will be affected. A carefully prepared programme of information and organisation has already been started.

We believe that the great majority of doctors accept that the interest and safety of diabetic patients is best served by this change. Legitimate doubts and hesitations have been expressed—particularly a reluctance to change the practice for a patient already well established for whom there may be no obvious advantage. The whole-hearted support of the medical profession is vital so that the demand for U40 and U80 ends and these strengths are withdrawn. Initiating, organising, and implementing the change to U100 are in the hands of the doctors, for there is no legislative process in Britain by which the withdrawal can be effected. We solicit their support earnestly but confidently.

Reference

- ¹ Greenough A, Cockcroft PM, Bloom A. Disposable syringes for insulin injection. *Br Med J* 1979; *ii*:1467-8.

Requests for reprints should be made to: the British Diabetic Association, 10 Queen Anne Street, London W1M 0BD.

For Debate . . .

What does "response" in cancer chemotherapy really mean?

JAMES V WATSON

Immense difficulty is experienced in assessing the response to chemotherapy of patients with metastases from the more common tumours. Two response categories, "complete" and "partial," are generally recognised; but in a recent review of interferon as a treatment for cancer an appreciable proportion of patients classified as "responders" were included in a new category of "less-than-partial" response.¹

A clinical response of a human tumour can obviously be defined as some measurable reduction in the patient's total tumour burden. A large proportion of reports, however, have failed to give the exact criteria of response, which should state explicitly what has been measured, the percentage reduction implied by "partial response," and the duration of the response. Even when all these data are available the informed reader often cannot decide whether patients should be treated with a particular chemotherapeutic regimen showing a good partial response rate. The reason is quite simple. Apart from certain

rare tumours—for instance, leukaemias, lymphomas, and chorion carcinoma—there is little hard evidence to show that the duration of life of a good quality is usefully extended. Furthermore, it is difficult to document and present reliably the chemotherapeutic side effects that often seem to be relegated to second place in overall assessment. The sceptically non-committed oncologist instinctively feels that the problems associated with chemotherapeutic combinations far outweigh the dubious benefits and that the partial response criterion is far too crude an indicator on which to base the management of patients.

In this article I attempt to delineate some of the known factors related to our current methods of oncological assessment and some of the fallacies inherent in them. I also attempt to assess critically some of the results of interferon treatment in human cancer in the light of those facts and fallacies.

Current methods of oncological assessment

DEFINITIONS OF TUMOUR RESPONSE

Two categories of objective tumour response are now generally recognised, in line with the recommendations of the International Union Against Cancer (UICC).²

MRC Unit of Clinical Oncology, The Medical School, Cambridge CB2 2QH

JAMES V WATSON, MSc, FRCP, senior clinical scientist, Medical Research Council, and honorary consultant oncologist, Addenbrooke's Hospital

Complete response

Complete response is "disappearance of all known disease. In the case of lytic bone metastases these must be shown radiographically to have calcified."

Partial response

Partial response is "greater than or equal to a 50% decrease in measurable lesions and objective improvement in evaluable but non-measurable lesions. No new lesions. It is not necessary for every lesion to have regressed to qualify for partial response, but no lesion should have progressed."

Under the section "measurable lesions" the UICC report states, "Ideally, all lesions should be measured at each assessment. When multiple lesions are present this may not be possible and under such circumstances a representative number of eight or more lesions may be selected for measurement." The measurement recommended by the UICC report was "the sum of the products of the diameters (one being the longest and the other the longest perpendicular to it) of each individual lesion, or those selected for study." In those cases where two measurements are not possible the diameter or sums of diameters should be used.

The UICC report did not make any specific recommendations about the duration of response that qualified the patient to be classified in this category. The statement reads, "In a patient who has objective regression, this is to be dated from the start of therapy until either new lesions appear or any existing lesion increases by 25% or more above its smallest size recorded. It is essential to categorise a patient as having a regression at a stated time. It is also essential that all baseline studies should be repeated at this time."

Despite no specific recommendation by the UICC, the minimum response duration is now generally accepted as four weeks. Although this seems pathetically short, let us accept the criterion and examine the remaining elements in the partial response.

TUMOUR COMPOSITION

Before we can attempt to contemplate what a partial response of a tumour might mean in real terms we must consider the composition of that tumour. Human tumours contain normal as well as neoplastic cells. The former include endothelial cells, connective tissue elements—for instance, fibroblasts—and some normal cells of the type from which the tumour arose. Also included are red blood cells, polymorphonuclear leucocytes, lymphocytes, and macrophages, which may be present in considerable numbers in a necrotic neoplasm. The neoplastic tissue may be divided into three "functional" categories: dead cells, differentiating or end cells, and clonogenic cells. The differentiating and clonogenic cell components may both be subdivided into hypoxic and euoxic compartments, and hypoxic cells are less likely to be within the division cycle and undergoing active proliferation. The importance of any clonogenic cells arrested in a "G₀" state owing to hypoxia or other undefined nutrient factors is that they may be relatively insensitive not only to radiation but also to chemotherapeutic attack owing to diffusion problems, and still retain the capacity to enter the division cycle if conditions become more favourable. Actively proliferating cells may be further subdivided into the four stages of the cell cycle—mitosis, G₁, S-phase, and G₂—and each of these phases may have different sensitivities to therapeutic agents. So far a total of no fewer than 24 different "compartments" have been defined that may be present even in the very simplified tumour model described above. A reduction in "size" of the tumour may be due to a decrease in any one, or any combination, of these compartments.

MEASUREMENT PARAMETERS

In some rare tumours biochemical measurements performed on peripheral blood give an indication of the total tumour bulk. Perhaps the best examples are myeloma proteins³ and the human chorionic gonadotrophin of chorion carcinoma.⁴ In the most common tumours—carcinomas of the breast, bronchus, and bowel—no such reliable markers exist, and we have to depend on radiographs, isotope scans, fingers, and callipers to assess the size of the tumour. Measurements so obtained may be expressed as diameters, areas, or volumes. Although the UICC recommendation is equivalent to an area assessment (the product of two diameters), the relationship between the

three parameters will be examined as there are reports that have used each of these for assessment. The volume of a sphere is related to the cube of the radius, the area of a circle is related to the square of the radius, and the diameter is directly related to the radius. If an assessment is being made on *diameter* a 50% reduction will reduce the area to 25% and the volume to 12.5%. When *area* is used as the assessment criterion a 50% reduction will reduce the volume to 36% and the diameter will be reduced to 71%. If *volume* is used a 50% reduction reduces the diameter to 80% and the area to 64%. Clearly a report using volume for assessing partial response, even if it adheres rigorously to the criterion of a 50% reduction for the minimum interval of four weeks, is far less convincing than a similar report using diameter as the assessment criterion. In the latter case the volume is reduced to 12.5%.

GROWTH KINETIC CONSIDERATIONS

The net increase in volume of a tumour reflects the excess of cell production over cell loss from the tumour. Steel⁵ has defined a parameter, the cell loss factor (CLF), which is equal to the ratio of the rate of cell loss divided by the rate of cell production.

If a tumour is composed exclusively of tumour cells a CLF of unity represents the condition where cell loss from the tumour exactly equals cell production and the net growth increase is zero. A CLF greater than unity defines the condition where the tumour is shrinking, and a value of less than unity continuing growth. The few experimental data available indicate that the CLF for certain human tumours may be about 90% or greater during unperturbed growth.⁶ To obtain values for the CLF, however, we must know the overall growth rate and the rate of cell production. Breur, in a careful study,⁷ has shown that secondary disease in the lungs grows exponentially in the vast majority of tumours. The doubling time, however, ranged from under 10 days in the fastest-growing tumours to over 160 days in the slowest, and there was considerable variation not only within but also between histological types. These results agreed with those of Collins *et al.*⁸ Determination of growth rate requires at least three observations, and to obtain reasonable accuracy the untreated patient probably needs to be observed for two tumour doublings. This is a four-fold volume increase, and to delay treatment long enough for the information to be acquired may be unethical in many patients. Although the growth rate can be determined, the rate of cell production is notoriously difficult to obtain from human tumours⁹; thus the CLF is not likely to be a practicable aid to assessment within the foreseeable future. Nevertheless, it is useful conceptually. We have already seen that a tumour may be composed of many different compartments and that a reduction in tumour volume could be due to a decrease in any one or any combination of those compartments. The only compartment, however, of any interest to the patient and therapist is the clonogenic fraction. Theoretically, a CLF greater than unity can be obtained, with a subsequent reduction in tumour volume that could be scored as a partial therapeutic response by an increase in the rate of cell loss from non-clonogenic components, including normal cells, within the tumour. Conceivably, the clonogenic fraction could be a minority population occupying under 10% of the total tumour bulk, and a volume reduction may well be just a manifestation of a temporary decrease in production rates within either or both of the clonogenic and non-clonogenic compartments that "unmasks" the normal processes of cell loss. This type of argument, however, may be inverted. Another theoretical possibility is to see an increase in tumour volume due to an influx of inflammatory cells in response to cell death due to treatment. From a cell kinetic standpoint, therefore, any gross measurements of a change in tumour size, particularly over a period as short as four weeks, cannot be regarded as having any meaning since we do not know what we are measuring. This implies not that a change has not occurred but that we cannot know what the change means.

QUALITY OF CLINICAL DATA

The information required to assess response is usually obtained from radiographs, isotope scans, and vernier callipers. Most radiographic data so far have been obtained from chest radiographs and lymphograms but examination by computed tomography is increasing, and this may improve our resolving capacity.

The physical size of a shadow on an x-ray plate is dependent on several factors. These include the geometric set-up, particularly the distance ratio between the lesion, the photographic plate, and the

x-ray source, and the penetration of the beam. Minor rotational differences of the patient from one examination to the next may alter the distance ratios and hence the apparent size of the lesion. Differences in penetration from one examination to the next may also influence the apparent size. Patients with metastatic lung disease where the tumour shadow is roughly circular are rare. This also applies to primary lung tumours, and to make an assessment we often have to measure two or more "representative" diameters and take an average. Even when the utmost care is exercised it is difficult to be sure that we are measuring even reasonably accurately—say, to within 10%—across the same planes from one examination to the next. This is particularly true if the measurements are changing, be the change real or apparent. When magnification, penetration, and orientational factors are all taken into account it is extremely unlikely that consistent measurements to within $\pm 10\%$ can be made in the most favourable cases on sequential examinations. This effectively means that a 50% volume reduction, which occurs from a 20% diameter decrease, cannot be scored. The problems are further compounded when there is collapse or consolidation, or both, in the vicinity of the lesion when we do not know where the lesion "begins" or "ends." In these cases we are probably unrealistic if we expect that we can obtain a measurement to within $\pm 20\%$, which means that area cannot be used for assessment. Whenever the lesion encroaches on the mediastinal shadow we probably cannot make an assessment to within $\pm 50\%$ on diameter, as what we can observe may be only the tip of the iceberg; we cannot be sure, and in these cases an assessment on diameter is effectively eliminated.

All these arguments apply to lymphograms but here we have added complications. The image we are measuring is a filling defect often on the "edge" of lymph nodes, and we cannot know how far the tumour extends beyond the limits of the contrast medium. Isotope scanning is similarly fraught with hazards so far as measurements are concerned, and no one in their right mind would expect to be able to palpate an abdominal mass (including liver) and measure its linear dimensions with an accuracy of $\pm 50\%$. In a fairly recent review¹⁰ secondary colorectal disease in the lungs was found to "respond" to fluorouracil in only 6% of cases. This contrasted sharply with "response rates" of 20%-30% for abdominal disease. Is this a real site difference or is it a manifestation of the much greater difficulty that we experience in making a truly objective assessment of intra-abdominal masses? Superficial lumps—for instance, lymph nodes containing tumour—are obviously easier to measure but the apparent size will depend on the thickness of overlying tissue, and the true size will depend on the proportion of the mass that is tumour tissue.

I know of no work that has related the true volume of human tumour deposits to estimates made with callipers. Although this is probably impracticable in man it has been carried out for transplantable tumours in animals¹¹ (J V Watson, PhD thesis submitted to London University). Some surprising, and in this context disturbing, observations have been made (J V Watson, submitted thesis). A total of 263 mouse tumours with a volume spread of nearly three orders of magnitude were measured to the nearest 0.5 mm across three mutually perpendicular diameters. The tumours were then excised and the volumes were measured by Archimedes's principle on a chemical balance to give the true volume. The product of the three diameters was then plotted against the true volume to give a calibration curve that was found to be a "log-log" relationship. Regression analysis of the data showed that for a given product of the three diameters there was a spread of a factor of 3.35 in the true volume at the 95% confidence interval (the "magic" statistical limit of $p < 0.05$). This 3.35 factor corresponds to an "error" of 50% on mean diameter measurements at a given volume. Data on rodent tumours in many respects, of course, have little relevance to man. From simple "size" measurements, however, some thing can be extrapolated to man. The large spread in the true volume at the 95% confidence limit for a given product of three diameters was due to different overlying skin thicknesses, different tumour "textures," and irregularly shaped tumours. In my experience human tumour deposits are considerably more irregular than the mouse tumours I measured. Can we realistically expect, then, to measure the size of human tumours more "accurately" than we can measure tumours in mice?

Moertel and Hanley¹² have attempted to quantify measurement errors by simulating tumour deposits. Twelve solid spheres measuring from 1.8 to 14.5 cm in diameter were placed on a soft mattress and covered with foam rubber that was 1.3 cm thick for the six smallest and 3.8 cm thick for the six largest sizes. A careful statistical analysis of the measurements of 16 experienced oncologists was carried out, and some of the conclusions read as follows. "In this study the 'placebo response rate' (that is, the false-positive response rate) in unchanged masses as measured by the same investigator was 25% if a 25%

regression criterion (product of two diameters) was employed and 7% if a 50% criterion was employed. When, however, the oncologist is measuring a deep-seated irregular mass through the muscle guarding of an uncomfortable patient, the errors of his measurement will undoubtedly be far greater than that recorded under the structure of this study. Certainly under real life circumstances, a 25% regression must be considered meaningless, and even a 50% regression is questionable."

Conclusions

The title of this article poses the question of what "response" in cancer chemotherapy really means. If we exclude the minority groups in whom an increased span of useful life has been shown to occur and turn our attention to the problem of metastatic cancer of the breast, bowel, and bronchus the answer must be "Not very much." What is perhaps even worse is that most of the time we really do not know what is worth pursuing and what is not, and the largest single reason for this sorry state of affairs is our inadequate capacity for assessment. A complete response as currently defined leaves little to argue about except the duration, which is just as important as the "quantity" of response. We cannot, however, be so charitable about some of the criteria of partial response that have been used. Quite clearly, the minimum possible partial response criterion, a 50% volume reduction for four weeks, should be completely abandoned. A partial response assessed as a 50% area reduction for four weeks is of dubious merit, and even a 50% diameter reduction for four weeks must often be a less-than-adequate criterion, not because of the size change but owing to the pitifully short duration. This period is simply not long enough to gain any meaningful data on human tumours, particularly as the doubling times may vary from 10 to more than 160 days,⁷ and we have virtually no reliable data on regression rates during or after treatment.

Recently, a less-than-partial response category has been introduced to describe the responses to interferon of patients with myeloma, breast cancer, and lymphoma.^{13,14} In these studies the responses of the patients with breast cancer and lymphoma were assessed as the product of two diameters, which complies admirably with the UICC recommendations. This is synonymous with an area assessment, and in this context the less-than-partial response criterion is met by an area reduction of between 25-50%, which is equivalent to a mean diameter decrease of only 13-29%. As the previous section shows, there must be few patients in whom such small mean decrease in tumour diameter can have any importance. But, even worse, we cannot identify those patients in whom such a small change might be important. Moreover, in one report¹⁴ the response duration was "from 14 to 121 days." In the other study the minimum duration of response for patients with breast cancer was eight weeks.¹³ In the overall analysis of results¹³ a total of 19 out of 38 patients were classified as responders, which confers equal "status" or "quality" to less-than-partial and complete responders alike. This way of presenting results also gives the same status to partial responses for 14 days and for 121 days. A significant correlation of response to interferon and previous responsiveness to chemotherapy was found,¹³ but this required that the less-than-partial responders should be included in the responders group. The UICC recommendation does not include less-than-partial response within the "objective regression" category, and when these contentious cases are excluded from the response group there is no significant correlation of interferon response and previous responsiveness to chemotherapy ($p > 0.05$, χ^2 test).

The real problem in cancer treatment at present is that we do not have anything specific to measure in most tumours, and without such measurements we are severely limited. After all, physics did not begin until the standard length and time interval were invented, and only with the introduction of the chemical balance did chemistry emerge as a discipline, relegating alchemy and the concept of phlogiston to the dustbin of human endeavour. The phlogiston theory, however, may be said to

have been based on scientific observation—the burning piece of paper obviously “gives something off” and the weight of the ash is less than that of the paper before it was burnt. But the wrong conclusions were drawn, mainly owing to inability to measure the weight of the effluent gases. The analogy between the concepts of phlogiston and partial response seems to be uncomfortably close at times, and we are in real danger of drawing incorrect conclusions from the way the latter now tends to be abused. Hence the concept it should be dispatched along with that of phlogiston.

Although we do not have anything specific to measure, we could use what we have to better advantage. Firstly, the application of plastic strips with radio-opaque divisions set at one centimetre intervals placed on the front and back of every patient having a chest x-ray examination to assess tumour “size” would enable us to be sure that any changes occurring were real and not artifactual, as magnification factors could be taken into account on a routine basis. Secondly, it would be an advantage, and would probably not be too difficult to arrange, to have each patient’s radiographs taken on the same machine with the same settings at each visit. Finally, and most importantly, time could be incorporated into the response criterion just as size is at present. At each visit the patient would be assessed according to the UICC recommendations, the product of two diameters being recorded. As an addition, however, the area under the response curve would be recorded to give the “cumulative response index” as the assessment criterion. This is illustrated in the figure, which shows such response curves of two hypothetical patients. In the top panel we see the percentage

Hypothetical response curves of two patients during treatment. Top panel—response at weekly intervals; bottom panel—cumulative response index plotted against time.

response at each weekly visit (hypothetical interval). Each of these patients would qualify as showing a partial response on the present criteria. The bottom panel, however, which shows the cumulative response index (obtained by summing the past and present response levels) shows clearly that the quality of response in patient A (so far as the tumour is concerned) is much better than in patient B. As long as the cumulative response index is increasing a therapeutic “gain” has been achieved. An increase in slope indicates an increase in response and a constant slope indicates that a given percentage response is being maintained. A decrease in slope while this remains positive indicates a decrease in response. No change in the cumulative response index shows that the tumour has regrown to its pre-

treatment size. This is illustrated between 10 and 11 weeks in patient B. A negative slope of the cumulative response index curve shows that the tumour has increased in size beyond its pretreatment value. If this system were to be adopted, and cumulative response index curves were to be published, the reader could much more easily assess the quality of response. Furthermore, frequency distributions of the cumulative response index at given times after treatment started could also be constructed. This would be achieved by “slicing” the cumulative response index data set vertically, which would give frequency versus cumulative response level at that time. We could then discard the arbitrary criteria for complete, partial, and less-than-partial response and submit the frequency distributions to statistical analysis and follow changes in the distributions with time. We might then get nearer to discovering the truth about what we are doing or trying to do.

The use of chemotherapy in the more common tumours will gain general acceptance among the more traditional oncologists only when better and more stringent, not worse and less stringent, criteria for assessment are used by the devotees of the art. Until then non-committed surgical and radiotherapeutic oncologists will remain cynically uncommitted.

References

- Sikora K. Does interferon cure cancer? *Br Med J* 1980;**281**:855-8.
- Hayward JL, Carbone PP, Heuson J-C, Kumaoka S, Segaloff A, Rubens RD. Assessment of response to therapy in advanced breast cancer. *Cancer* 1977;**39**:1289-94.
- Sullivan PW, Salmon SE. Kinetics of tumour growth and regression in IgG myeloma. *J Clin Invest* 1972;**51**:1697-1708.
- Bagshaw KD, Searle F. Tumour markers. *Essays in Medical Biochemistry* 1977;**3**:25-73.
- Steel GG. Cell loss from experimental tumours. *Cell Tissue Kinet* 1968;**1**:193-207.
- Steel GG. Cell population kinetics of human tumours. In: *Growth kinetics of tumours*. Oxford: Clarendon Press, 1977:185-216.
- Breuer K. Growth rate and radiosensitivity of human tumours. I Growth rate of human tumours. *Eur J Cancer* 1966;**2**:157-71.
- Collins VP, Loeffler RK, Tivey H. Observations on growth rates of human tumours. *AJR* 1956;**76**:988-1000.
- Steel GG, Lamerton LF. The growth rate of human tumours. *Br J Cancer* 1966;**20**:74-86.
- Watson JV. Colorectal tumours. *Hospital Update* 1978;**1**:149-60.
- Steel GG, Adams K, Barrett JC. Analysis of cell population kinetics of transplanted tumours of widely differing growth rates. *Br J Cancer* 1966;**20**:784-800.
- Moertel CG, Hanley JA. The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer* 1976;**38**:388-94.
- Gutterman JU, Blumenschein GR, Alexanian R, et al. Leukocyte interferon induced tumour regression in human metastatic breast cancer, multiple myeloma and malignant lymphoma. *Ann Intern Med* 1980;**93**:399-406.
- Borden E, Dao T, Holland J, Gutterman JU, Merrigan T. Interferon in recurrent breast cancer: preliminary report to the American Cancer Society Clinical Cancer Trials Program. *Proceedings of the American Association for Cancer Research* 1980: Abstract 750.

(Accepted 12 May 1981)

My practice contains 10 epileptic children, all of whom have at least one first-degree relative who reacts to some allergen. Is there any evidence about the neurological manifestations of atopy?

Epilepsy is a common problem in childhood, affecting between six and eight per thousand children of school age (5 to 16 years) in Britain. Allergic disorders, such as eczema and asthma, are also common, and the fact that the 10 children with epilepsy all have at least one close relative who reacts to an allergen is unlikely to have any significance. Probably the same might be true of 10 other children chosen at random. I know of no convincing evidence for a link between atopy and epilepsy (which is a symptom of very varied causation rather than a disease entity). With regard to other possible neurological manifestations of atopy, a “neuroallergic” mechanism has been postulated for certain postinfectious disorders of the nervous system, such as the Guillain-Barré syndrome and postinfectious encephalomyelitis, but the patients affected by these conditions do not seem to show an abnormally high incidence of the commoner manifestations of atopy.