PET Quantitation: Blessing and Curse

Positron emission tomography (PET) has been in existence for quite a number of years. Interest in the method increased dramatically 10 years ago when it was applied to the measurement of cerebral glucose utilization, with fluorodeoxyglucose (FDG) as the in vivo tracer.

What distinguished PET from its sister nuclear medicine techniques? Physicists were excited about the tomographic capability, using the annihilation radiation resulting from positron decay as an "electronic" collimator. Chemists were attracted by the ability to tag biologically interesting molecules, such as glucose or deoxyglucose, with positron emitters, and neurophysiologists were stimulated by the prospect that important parameters such as blood flow and glucose utilization could be measured in vivo. Thus PET was born and nurtured in a research environment imbued with the quest for quantitative measurement. The clinician had very little input in the beginning, and the ability of PET to produce clinically useful images was nearly lost sight of. Quantitation, with emphasis on accuracy, became the supreme law.

Obviously one cannot be against accuracy of measurement. Important results have been achieved through careful quantitative analysis, but there are areas in which quantitation is not needed. The day-to-day practice of radiology and nuclear medicine is nonquantitative. Trained specialists routinely read plain x-rays, angiograms, bone scans, etc., with little emphasis on numerical measure. This has remained the case with computed tomography (CT), despite its better capability for accurate measurements of x-ray attenuation, and even appears to be the trend with the younger magnetic resonance imaging (MRI). While the deoxyglucose method was introduced with a degree of quantitative validation seldom found in the medical sciences (*1*), this does not mean that quantitation is necessary to its clinical application, and indeed it may be a burden.

Visual Diagnosis

So our central point is this. It is time for PET to emerge from behind the veil of mathematics and models into the real world of clinical imaging. Our direct experience with brain tumors bears witness to this. We have seen, in over 500 studies, that FDG-PET imaging has a striking superiority over *any other* diagnostic method, including MRI and, in some cases, even histology, in regard to the fundamental issue, i.e., assessment of malignancy. Attempts to quantitate FDG uptake or glucose utilization in brain tumors have not been helpful, and can even lead to diagnostic error. Unsurprisingly, the trained human eye has once again triumphed over computers in the interpretation of anatomic images. Our (GDC) extensive experience with epilepsy and the dementias has also led to a preference for diagnostic conclusions based on image interpretation, rather than quantitation. In other potential diagnostic applications, such as neuro-receptor disease, it remains to be seen which will be the respective roles of visual and quantitative analysis.

It is unfortunate that the quantitative cradle in which PET was born has maintained such a hold that some practitioners cannot recognize when it is no longer needed. We have recently had the "Alice in Wonderland" experience of trying to justify our validated method of diagnosing brain tumors vis a vis a quantitative approach *which admittedly didn't work*, but which was defended, nonetheless, with plausible sounding but ultimately irrelevant mathematic arguments (*2,3*). How sweet is the siren song of mathematics, that it can blind people to what is before their very eyes!

In this respect, the marriage of PET to the deoxyglucose method was a mixed blessing, bringing with it a substantial controversy about the accuracy of the method and pointing a spotlight at quantitation. The many discussions about lumped constants, rate constants, phosphatase correction, etc., that have cast doubt—unjustifiably, in our opinion—on the deoxyglucose method, are irrelevant to the diagnostic usefulness of FDG. While we believe that FDG uptake is closely related to glucose utilization, this belief is not essential to its clinical usefulness. If it were found that shoe polish (to borrow one of Lou Sokoloff's favorite analogies) is an effective tracer for diagnosing tumors, it would not matter whether its mode of operation was understood. Clinical medicine, unlike pure research, can progress solely on the basis of empirical evidence.

Our complaint is not trivial or nit-picking. PET could be a powerful, cost-effective diagnostic method, if freed of its excess baggage. Fluorine-18 compounds, such as fluorodeoxyglucose, have a 2-hr half-life and could be made and distributed regionally at moderate cost, as was done in the days when fluorine was used for bone scanning. Blood sampling and all mathematic manipulations could be eliminated. The PET examination could be carried out by the same technicians who operate gamma cameras, and its interpretation could be delegated to nuclear medicine specialists or radiologists with nuclear medicine expertise. Even the cost of the scanners, if produced in quantity, could at least be brought in line with magnetic resonance imaging scanners.

By urging the use and acceptance of PET as a diagnostic imaging technique, freed from the burden of blood sampling and unneeded quantitation, we are not attempting to deny its obvious application to quantitative problems, such as inter-population comparison. Again referring to our own work, the observation that glucose metabolism per unit volume is inversely proportional to brain size (*4,5*) one of the few original heuristic contributions from PET research, would have been impossible without accurate quantitation.

Pitfalls in Quantitation

This brings us to our second axe to grind. Even where quantitation is necessary and appropriate, there are those who seem to use it as a smoke screen, deluging us with numbers and numerical arguments, to the point where we (and one suspects, they) can't see the forest for the trees. The reverse side of Lord Kelvin's oft-quoted dictum, "When you cannot express it in numbers, your knowledge is meager and unsatisfactory", might be "A meager and unsatisfactory knowledge can be concealed by excessive use of numbers."

We have seen sophisticated statistical methods used by investigator, the material being generated by statisticians who are often blind to the physiological or pathological implications. We believe that a potful of statistics should at least be accompanied by a teaspoon of intuition, if not the other way around. Any valid quantitative conclusion is usually apparent, or at least supported, by a visual examination of the images or (in the case of population comparisons) graphs. A good rule of thumb is that, if no difference is seen visually or graphically, it either does not exist, or is too small compared to methodologic error to have great significance.

One of us (RAB), who spent his early career in pure physics, was surprised by the over-reliance on statistical calculations upon entering the medical sciences. Physicists present their results with error bars (including systematic error, which is often ignored in medicine), and let them speak for themselves. Some medical researchers can't seem to trust their own judgment. There is something magical about a t-test; when the distribution percentile surpasses some arbitrary limit, say 0.05, those values may be declared "significant" and earn a star (*) while other values are declared "normal", even though their deviations from strict normalcy may not be much different than those of the "significant" data. Overlooked is the fact that given, say, 20 regions of interest, it is probable that one or two of them will show a significance level of 0.05 by chance alone.

The significance problem becomes worse if there happens to be a small systematic shift in the data (whether real or artifactual is irrelevant) that, by itself, is below the arbitrary significance level, but which pushes more regional values into the magic region of "significance". Blind reliance on t-tests can even be exploited for less than exemplary motives, as in the case of an investigator who, to justify attendance at an international meeting, analyzed enough regional ratios of metabolic rates so that he could find some that were "significant".

We believe there have been many errors and wrong conclusions from misinterpretation of numbers and statistics in the quantitation of PET images. The PET methodology would benefit by thinning out the mighty legions of experts in the Bonferroni, Monte Carlo, and Cox's nonparametric regression methods, and by beefing up the tiny maniples of PET practitioners who have a modicum of understanding of neuroanatomy. We would not have to witness again presentations at prestigious meetings where, as happened at one PET symposium, "occipital lobes" were analyzed to statistical consummation, with the minor problem that the chosen regions of interest had nothing to do with the occipital lobes, but encompassed only the cerebellum. We would, once and for all, agree on the respective borders of the imaged parietal and temporal lobes, and we would be spared the investigator who discusses the relative degrees of involvement of the cuneus or supramarginal gyrus while using a scanner with a 2-cm spatial resolution. Nor would we see claims of differences in global metabolism, made on the basis of statistical manipulation without regard to the fact that the tasks performed or pathologies evidenced by the subject population have no conceivable connection with global metabolism.

Obviously, the matters discussed above are controversial. While many investigators use quantitation and statistics appropriately, there have been misuses, and there will always be those who bury themselves in numbers to the exclusion of common sense. Therefore we appeal to the nuclear medicine and radiology audiences: Be wary of articles that rely overly on numerical analysis. Look for visual documentation and, above all, common sense reinforcement. As for clinical diagnosis, surely Lord Kelvin himself, a practical man who was knighted for introducing the telephone into Great Britain, would see the need to abandon quantitation when it is not appropriate to the task at hand.

Giovanni Di Chiro
Rodney A. Brooks
National Institutes of Health
Bethesda, Maryland

REFERENCES

1. Sokoloff L, Reivich M, Kennedy C, et al. The [^{14}C]deoxyglucose method for the measurement of local cerebral glucose utilization: theory, procedure, and normal values in the conscious and anesthetized albino rat. *J Neurochem* 1977; 28:897–916.

2. Tyler JL, Diksic M, Villemure J-G, et al. Metabolic and hemodynamic evaluation of gliomas using positron emission tomography. *J Nucl Med* 1987; 28:1123–1133.

3. Di Chiro G, Brooks RA. PET-FDG of untreated and treated cerebral gliomas [Letter]. *J Nucl Med* 1988; 29:421–422.

4. Hatazawa J, Brooks RA. Di Chiro G, Bacharach SL. Glucose utilization rate vs. brain size in humans. *Neurology* 1987; 37:583–588.

5. Hatazawa J, Brooks RA, Di Chiro G, Campbell G. Global cerebral glucose utilization is independent of brain size: a PET study. *J Comput Assist Tomogr* 1987; 11:571–576.