In vivo measurement of cerebral blood flow: a review of methods and applications

Paul D Griffiths a, N Hoggard a, Wayne R Dannels b and Iain D Wilkinson a

Abstract: This article reviews the concepts of cerebral blood flow for the clinician involved in the management of patients with carotid stenosis and/or ischaemic stroke. Methods of assessing cerebral blood flow in vivo using nuclear medicine, magnetic resonance and X-ray computed tomography are described. Applications of magnetic resonance and X-ray computed tomographic methods are reviewed and illustrated by examples from the authors’ radiological practice.

Key words: carotid stenosis; cerebral blood flow; computed tomography; magnetic resonance; stroke

Introduction

There have been major advances in diagnostic imaging over the last two decades of the twentieth century with structural imaging delineating most body structures at sub-millimeter resolution. The new frontier for imagers is functional imaging, which refers to a wide range of physiological activities. Potentially one of the most valuable techniques has been the development of methods to measure cerebral blood flow (CBF). The high prevalence of craniocervical atherosclerosis in the western world results in high rates of ischaemic stroke. Intuitively it would seem that methods of quantifying CBF would be useful to investigate patients with stroke and carotid stenosis, and an evolving literature supports this thesis. In this review we discuss what is meant by CBF, outline some of the methods that can be used to measure flow and introduce some clinical situations where CBF measurements are proving to be useful.

What is CBF?

Stewart produced the indicator-dilution methods for blood flow when studying the heart in 1898, which led to the fundamental definition of flow:

\[ F = \frac{V}{T} \]

where \( F \) is tissue blood flow, \( V \) is volume of distribution and \( T \) is transit time.

For the brain the transit time is the mean transit time (MTT) and the volume of distribution is the cerebral blood volume (CBV); so, by rearranging:

\[ MTT = \frac{CBV}{bF} \]

where \( bF \) is brain tissue flow.

With realistic flow in tissue, transit time is not a simple constant. In particular, the time for blood to traverse tissue will entail a temporal distribution, and its measurement will likely yield a curve. With appropriate definitions of a mean transit time (MTT) for a curve, the relationship above will still hold. We must specify the amount of tissue associated with the reported blood flow. In some cases it may be natural to report the total flow of an entire organ or anatomical region. When investigating regionally varying flow, however, the natural measurement becomes fluid per unit time per mass of tissue. Physiological literature traditionally reports blood flow in units of ml of blood/100 g of tissue per minute. For volumetric imaging techniques, ml of blood/100 ml per minute describes the actual measurement better, but since the density of brain tissue is close to 1 g/ml, these two values are used interchangeably.

Another distinction arises when we consider localized imaging results instead of organs or whole vascular regions. We would wish to ascertain the effective physiological flow occurring at the level of microvasculature or capillaries. When we desire metabolic supply or tissue perfusion in small regions of an image, we are not interested in flow simply passing through major vessels on its way to supply other areas. In some methods, the physics of the imaging mechanism and the choice of acquisition parameters may lead to a tissue weighting which naturally emphasizes capillary or parenchymal flow. In other methods, if the acquisition does not naturally exhibit such selectivity, contribution from major vessels may be suppressed as part of the data analysis process.

Principles of CBF estimation

Most methods of measuring blood flow involve an indicator-dilution method. Reference 2 is a review of those methods, including the detailed mathematical approach, whereas here we attempt to provide a background to the physical principles without recourse to detailed equations.
An indicator may be introduced into the vascular compartment in a number of ways and must be identifiable by some means at the target organ with respect to position and change in concentration with time. There are two methods of applying the indicator: sudden (bolus) introduction and by constant level introduction in an attempt to produce a steady-state system.

An indicator must have certain characteristics to be used for flow calculation and some assumptions must be made about the system under study. The indicator should mix completely with blood at the point of entry. Although not all the indicator molecules will travel through the system under study at the same rate, the distribution of transit times of the indicator must be the same as blood. The mean transit time should not change during the course of the measurement. Recirculation of the indicator should not occur or must be corrected for by some method. The indicator should not perturb the physiology we wish to measure. Another distinction among tracer techniques is that of freely diffusible versus blood pool tracers. Some tracers, generally smaller molecules, exchange freely from capillaries into brain parenchyma. In this case, analytical approaches must deal with a few distinct ‘pools’ of contrast, and distinct wash-in versus wash-out dynamics. Freely diffusible tracer goes from arterial blood supply to the tissue parenchyma, with net transfer rates depending upon accumulated tracer concentrations besides the flow rate. The tissue parenchyma exchanges tracer into the venous spaces at a rate virtually independent of the instantaneous arterial concentrations. Non-diffusing or blood pool agents present a different natural description, passing from arteries through capillaries and out via veins, with essentially no interaction occurring with the non-vascular spaces. The entire subject is sometimes denoted as ‘tracer kinetics’.

In neuroradiology, several common contrast agents can normally be considered to be non-diffusible tracers. This assumption may become invalid when the blood–brain barrier is disrupted. Where brain tissue exhibits contrast uptake, more sophisticated analyses are required. The enhancement time curves in the presence of uptake may still provide useful information, for example as angiogenic markers, but analysis strategies must be adapted accordingly.

A single input and single output is assumed for the mathematical derivation and all of the indicator entering the system must leave the system in the same timescale as blood.
The value of the single input/single output assumption lies in the fact that it allows a mathematically rigorous definition of MTT, which will work correctly in the fundamental flow definitions. For data collected with radiological acquisition and tracers, the directly observable signal is usually from tissue and/or microvasculature, not from major arteries and veins. This implies analysis may no longer follow the steps of: (1) determine CBV; (2) use arterial and venous measurements to give MTT; (3) divide to get CBF. Regardless of the analysis steps used, decomposing flow effects into two contributions, volume changes and transit time changes, is often a useful concept.

Transit times for blood flow in brain through capillary beds are in the order of a few seconds. With venous injection of contrast, the bolus will be broadened and dispersed as it passes through the cardiopulmonary system before reaching the brain. Thus, rigorous quantification attempts in the brain are likely to depend upon measuring the time course of contrast agent concentration in the arterial supply to the tissue. This ‘arterial input function’ comes into play both for CBV and for MTT determinations. Intuitively, we may expect that the observed concentration in tissue is going to be a function of the CBV times and the concentration in the pure arterial supply. Similarly, observed time courses in tissue should be some combination of the time course of the actual bolus in the arterial supply, plus the transit time characteristics through the tissue for an ideal (short) bolus.

Bolus methods with blood pool agents in human brain require temporal resolution given by a measurement every few seconds. Slower data acquisition results in erroneous shapes and widths of the tissue intensity curves or arterial curves. The total width of the broadened bolus, plus uncertainty of when the bolus will reach the brain, plus the practical need for baseline images before enhancement, add up to the requirement to acquire data for about a minute when using blood pool agents.

Clinical methods of measuring CBF

Nuclear medicine methods
Nuclear medicine imaging can be divided into two main techniques: single photon and coincidence detection methods. By far the commonest in clinical practice is the single photon method using gamma cameras to detect the single photon product of a number of radioactive products such as iodine-123 and metastable technetium-99. However, the reference standard for in vivo brain flow imaging according to many authorities are techniques using coincidence detection of the two gamma-rays produced by positron decay. Positrons are produced by the decay of some low atomic number isotopes that have too few neutrons, such as carbon-11, oxygen-15 and fluorine-18. These isotopes are produced in cyclotrons which are exceptionally expensive and need to be on-site because of the short half-lives of the isotopes (carbon-11 20 min, oxygen-15 2 min, fluorine-18 110 min).

Because of the expense of this method it is limited to larger centres and is unlikely to be widely available in its current form. One major advantage of using these isotopes for positron emission tomography (PET) is that isotopes retain their usual chemical characteristics so that they can be combined into many pharmaceutical agents. With respect to brain flow imaging the most useful compound is water, using the positron-emitting oxygen-15 to produce $\text{H}_2^{15}\text{O}$. From the previous discussion, water satisfies the criteria for a CBF indicator and direct, repeated measurements of arterial and venous levels of oxygen-15 during the experiment allow formal quantification. Quantifiable measures of cerebral blood volume can be obtained using carbon-11, which binds to haemoglobin in red blood cells.

It is not possible to label water with any compound that emits single photons. Therefore, single photon emission computed tomography (SPECT) must use indirect measures of brain blood flow. The compounds that have shown most promise are hexamethylpropyleneamineoxime (HMPAO).
Figure 3  Investigation of stroke using MR. The fluid-attenuated inversion recovery (FLAIR) image (a) shows a small volume of abnormality in the distribution of the left middle cerebral artery which matches the hyperintense abnormality on diffusion-weighted imaging (b). Two regions of interest (ROI) are depicted on an image selected from the base perfusion data set (c): ROI#1 is placed in the normal right hemisphere and ROI#2 in the area of abnormality depicted in (a) and (b). The resultant concentration curves show longer time-to-peak in ROI#2 (d), having a skewed envelope consistent with a major reduction in CBF. The calculated time-to-peak map (e) indicates that most of the hemisphere has reduced CBF.

labelled with technetium-99m and amphetamine derivatives labelled with iodine-123. It is not possible to obtain quantifiable blood flow data from these methods. Regional CBV can be quantified using technetium-99m-labelled red cells.

X-ray computed tomography methods
X-ray computed tomography (CT) has been the mainstay of cross-sectional imaging since the 1970s. The physical principle that underlies the technique is a highly collimated X-ray beam directed at the area of interest. A proportion of the beam is stopped (attenuated) by the patient whilst the remainder of the beam is detected and quantified. Contrast in the image is produced by variations in attenuation of adjacent structures, which in CT is primarily related to tissue density. For example, bone has high density, and therefore high attenuation when compared with the surrounding soft tissues or air. This produces high contrast, making CT a superb technique for studying bone pathology. In the brain, grey matter and white matter structures have similar densities (not identical because of the higher lipid content in white matter); therefore, CT is not a good method to assess subtle abnormalities such as cortical dysplasias. The same argument is applied for the early detection of ischaemic stroke; CT can only detect stroke when there is sufficient cytotoxic oedema to significantly reduce density of the affected brain.

For many years external agents have been introduced into the bodies of patients undergoing CT examinations in order to increase tissue contrast. These are usually positive contrast agents (higher attenuation values than tissue) and include iodine and barium compounds. Iodine is particularly suited to the vascular compartment because of its
low toxicity when combined to organic molecules. The halide, iodine, acts as contrast agent because of its high density (atomic number 53, atomic weight 127) and favourable k-edge with the energy of X-rays used in diagnostic imaging. The inert gas, xenon (atomic number 54, atomic weight 131), is adjacent to iodine in the periodic table and has similar attenuation characteristics. Both iodine and xenon have been used to study CBF.

A rapidly injected bolus of iodinated contrast agent satisfies many of the criteria for the indicator-dilution principle. However, it is difficult to measure arterial input and venous output which is necessary for assessment of the correct mean transit time. This means that it is difficult to quantify cerebral blood flow by this method, although many groups are pursuing this goal. Using iodinated contrast in CT, typical analysis presumes that a bolus injection can be measured in an arterial sample such as a portion of the middle cerebral artery, and that the contrast agent remains wholly in the vascular space. Spiral (or helical) scanning is the main enabling technology which can be programmed to not move the couch while scanning in order to collect a time series of data at a fixed slice position. Trade-offs of dose versus temporal resolution versus signal-to-noise are critical, since the slice being studied will in effect be scanned perhaps 30 times in a minute. Typically, the data analysis includes a step of searching for pixels with large contrast changes (say 50 Hounsfield units), and those pixels are excluded on the grounds of being contaminated by major vessels. The remaining pixels can then be analysed as time curves, and, as it turns out, the maximum upslope before peak enhancement divided by the peak of each enhancement curve yields a blood flow value.

The main limitation to date of this CT blood flow method is that in practice it is restricted to one slice level of anatomical coverage. New commercially available multislice CTs may collect four to eight slices over a 2-cm range, but this still falls short of covering the whole brain, so it is not quite appropriate for use when the area of flow abnormality cannot be specified in advance.

If native (non-radioactive) xenon is inhaled it enters the circulation and hence the brain in proportion to cerebral blood. CT can detect the change in concentration of xenon as increased attenuation and the high anatomical resolution capacity of CT allows regional measurements. This has given rise to a large amount of work using xenon-computed tomography (Xe-CT) to measure regional CBF in vivo. Xenon acts as a freely diffusible tracer, and is especially soluble in lipids. Xe-CT flow measurements are predominantly acquired in the wash-out phase.

**Magnetic resonance (MR) methods**

MR can be used to study cerebral blood flow by several techniques but by far the most frequently used in clinical practice is dynamic susceptibility perfusion imaging using chelates of gadolinium (e.g., Gd-DTPA). Gd-DTPA is used primarily in neuroimaging (and in other areas) because of its effect on the longitudinal relaxation-enhancing effect seen on T1-weighted sequences. However, as well as altering the T1 signal, the perturbing influence of the gadolinium ion also shortens the T2* relaxation time of the neighbouring protons. The T2* effect produces a decrease of signal on T2*-weighted sequences such as gradient echo or gradient-recalled echoplanar sequences. The most commonly used technique is a gradient-recalled echoplanar sequence run during the injection of a high dose of Gd-DTPA into a large peripheral vein using an MR-compatible pump injector. The pump injector will minimize experimental error due to inter- and intra-individual variation in injection technique.

The brain is imaged using a T2*-weighted sequence which has the capacity to image the same slice of the brain each second or less. In most situations it is important to image all of the brain with multiple slices of 10 mm or thinner. This requires very rapid imaging techniques such as echoplanar imaging, which to date is mainly available as an option on high performance scanners. The primary enabling technology here is fast and powerful magnetic gradient subsystems. As the gadolinium bolus passes through the brain capillaries a measurable loss of signal occurs. The signal reductions can be mathematically...

---

*Vascular Medicine* 2001; 6: 51–60
Figure 4  MR shows an ischaemic stroke with anatomically matched diffusion and perfusion abnormalities. A large lesion can be seen in the right middle cerebral artery territory using the fluid-attenuated inversion recovery technique (a). This abnormality is matched with the diffusion-weighted abnormality (b), area of increased time-to-peak (c) and reduced regional blood volume (d).

inverted into a quantity proportional to the concentration of the tracer.8,9

Two mathematical techniques are among the analysis methods commonly applied to MR tissue perfusion data. In the first method, the data is fitted to a curve with the expected shape, the gamma variate function.10 Only the portion of the data obtained before the onset of a recirculating second pass is used in the fit. Then, extrapolating the fitted curve, areas can be calculated which include estimations of the correct tail of the curve. In the second method, the time course of an arterial sample is mathematically stripped away from the observed tissue curves, leaving ideal flow curves. This process is known as deconvolution, and can yield notoriously noisy results if not done carefully with algorithms such as singular value decomposition. Several parameters can be measured. Time to peak (TTP) is the time at which the signal loss produced by the Gd-DTPA-bolus is maximal, widths such as the full-width-at-half-max are indicators of the time course of the passage of Gd-DTPA and the area under the curve is proportional to the CBV. A subtle detail is that the widths of tissue enhancement curves need not give rigorously correct values for MTT.11 These values are frequently displayed as maps of the slice after the computer software has calculated the values on a pixel-by-pixel basis.

The CBF may be derived from calculations a little more complicated than CBV/MTT, but it should be noted that at present MR bolus methods cannot give absolute flow quantification – the measurements lack a known internal standard. Most of the maps generated by these analyses give results which are described as ‘relative’. Comparisons across similar tissue types in the same patient in the same

Vascular Medicine 2001; 6: 51–60
Figure 5 An atypical stroke case investigated by MR and Xe-CT. The MR examination was performed 15 h after the ictus and shows a large area of ischaemic stroke as high signal on T2-weighted fast spin-echo and diffusion-weighted images (a, b). The time-to-peak map shows more rapid arrival of blood to the affected territory (c) and a focal increase in CBV (d). The CT examination (e, f) demonstrates increased blood flow to the region (~80 ml/mg per min). No abnormality was shown on MR angiography (not shown). It was assumed that this resulted from an embolic stroke with rapid breakdown of the clot and reactive hyperaemia.

scan can yield valid ratios of flows. But any single image of CBV or CBF may have unknown or dubious constants associated with the reported results. In some cases, the results are clearly uncalibrated, in that arterial input functions are not used to generate the maps at all. In other cases, the limitations of the technique are made evident by the empirical observation that choosing different arterial samples from within the images can yield different maps. There is no general agreed-upon solution to date regarding which arterial input functions can be faithfully used for which tissue locations.

The major drawback to produce quantifiable data is the lack of robust arterial input and venous output functions. The difficulties with obtaining robust measurements of arterial or venous curves in perfusion MR imaging (MRI) relate back to the ubiquitous problems of signal-to-noise, image contrast, spatial resolution, and the partial volume effect. As an example of this problem, imagine that 20% signal decreases are induced by a certain dose of contrast in tissue parenchyma. But arterial concentrations can be expected to be 20–50 higher than tissue concentrations, and so the signal in a voxel containing only arterial blood will be extremely darkened, often to the level of noise. Choosing bigger voxels may overcome noise problems, but then it can become impossible to find voxels which consist of only arterial blood. For the present, it is probably prudent to treat any MRI blood flow maps as ‘relative’, i.e. in some sense or another, either uncalibrated or lacking certain cor-

Vascular Medicine 2001; 6: 51–60
Figure 5  Continued.

Figure 6  A MR time-to-peak map depicts hemispheric asymmetry in a patient with a severe, symptomatic stenosis (90%) of the right internal carotid artery (a). A follow-up MR was performed 2 h after the insertion of an intravascular stent. The resultant time-to-peak map (b) shows resolution of the interhemispheric flow asymmetry.

rection steps. But the situation is not quite so bad. Going back to the original time course data can yield highly diagnostic information in many forms. For example, we have not yet considered the arrival time of the contrast bolus. While this is not a pure descriptor of steady-state perfusion through capillaries, haemodynamic arrival time delays are often indicative of arterial concerns.

Numerous parameters can be reported from tissue intensity curves. Each is likely to be correlated, either directly or inversely, to CBF. Examining a few maps is likely to present better insight to the flow characteristics than relying on a single parameter. Low regional blood flow, for example, can be suggested from low peak signal change, low areas (time integrals of change), reduced maximum slope, increased half-widths (i.e. increased MTT), or delayed times at which peak enhancement occurs. Keeping in mind the relationship

\[ F = \frac{CBV}{MTT} \]

it should be straightforward to decide whether a particular measure of a concentration curve relates to the total concentration (like CBV), or how rapidly the tracer passes (like MTT). In turn, it should then be clear whether that measurement or map correlates directly or inversely with flow.

Clinical applications

There is much interest in the possible clinical application of measuring CBF in the clinical environment, particularly in the assessment of tumours in many parts of the body (notably brain, breast and uterus). However, in this commentary we concentrate on the two major clinical conditions affecting the cerebral vasculature: ischaemic stroke and carotid stenosis.

Ischaemic stroke

Dynamic, contrast-enhanced MR perfusion, Xe-CT and CT perfusion imaging using iodinated compounds are all currently under review in patients with stroke. Routine imaging fulfills the important role of excluding other pathology such as tumour or haemorrhage and it is hoped perfusion techniques may give extra information on the suitability of therapeutic intervention. It is generally accepted that these decisions need to be made soon after the ictus, probably within 6 h. This will place significant demands on the paramedical and imaging services if the acute treatment of stroke becomes mainstream. The rationale for treatment is based around the idea of an ischaemic penumbra. When ischaemic stroke occurs from whatever reason, brain tissue

Vascular Medicine 2001; 6: 51–60
at the centre (or umbra) of the affected region will die and nothing can be done to save that tissue. However, on the periphery, neurons may be metabolically inactive but alive. They may be damaged further by persisting or evolving reduced blood flow or by substances released from dying cells which are toxic (e.g. glutamate, calcium, free radicals). The major class of drugs used in the treatment of acute ischaemic stroke are thrombolytics, which are designed to break down intravascular clot and restore blood flow. These drugs are not without risk, the major complication being a relatively high risk of intracranial haemorrhage. Because of the risk of serious complication, patient selection needs to be thoughtful because there is no point in giving thrombolytics to patients where the CBF has already been restored or if there is no ischaemic penumbra – all of the affected tissue is already dead. Therefore, all methods are directed at resolving salvageable from non-salvageable tissue. Groups researching this area with Xe-CT believe that the detection of regional flow below a certain level indicates cell death whereas flow above indicates that it is potentially salvageable. For Imaizumi and colleagues the value is 7 ml/mg per min.12 Groups working with iodinated CT perfusion imaging believe that their methods are also quantifiable and can also be used for this application. At present dynamic contrast-enhanced MR perfusion imaging does not allow formal quantification, only comparison of indirect measures of flow among different regions is possible. However, MR perfusion imaging is used in conjunction with MR diffusion imaging, which gives different information. We will not describe the method or applications in detail here (see ref. 13) but present a simplified overview. Diffusion-weighted (DW) imaging detects the ease with which water can diffuse within the brain. Normally the extracellular water diffuses freely and therefore produces no signal on DW images. However, in ischaemic stroke, failure of the Na+/K+ pump causes movement of water into the cell which causes swelling and restricted diffusion of the remaining extracellular water. Therefore, acute stroke appears as an area of high signal on DWI and is currently the most sensitive and specific imaging method of detecting early stroke. It has been proposed that the volume of DW abnormality correlates exceptionally well with the final volume of infarcted tissue and therefore DW abnormality locates tissue that is already dead. This contention is the subject of much current research. We show examples of CT over MR methods used in our clinical practice in Figures 1–5.

Carotid stenosis

The role of CBF measurement is not as clear as in ischaemic stroke. Patients are selected for surgical/endovascular treatment or medical treatment on the basis of clinical assessment and a measurement of the degree of narrowing of the internal carotid artery. Intervention is usually reserved for symptomatic patients with 70–99% stenosis.14,15 The stenosis is usually assessed by screening ultrasound and catheter angiography, although CT and MR angiographic methods are playing an increasing role. Intuitively, assessment of CBF would be useful in symptomatic patients and many groups do use nuclear medicine and CT methods to assess resting flow and flow after a challenge with acetazolamide in patients with symptomatic carotid stenosis.

Our group has used dynamic contrast-enhanced MR perfusion to assess CBF before and after both carotid endarterectomy16 and stenting.17,18 The main value of this in the acute setting is to define those patients with post-procedure hyperperfusion that are at risk of haemorrhagic complications. Examples of this are shown in Figure 6.

Conclusion

In this review we have introduced the basic principles of measuring cerebral blood flow in vivo and illustrated the potential clinical applications in stroke and atherosclerotic carotid stenosis. This is an evolving practice with relative advantages and disadvantages between the competing methodologies, but more importantly the techniques still have to prove their worth in routine clinical management.

References

1 Stewart GN. Researches on the circulation time and on the influences which affect it. IV. The output of the heart. J Physiol 1898; 22: 159–83.
15 ECST collaborative group. MRC European carotid surgery trial: interim results for symptomatic patients with severe or with mild carotid stenosis. Lancet 1991; 337: 1235–44.
