Quantitative vs. Visual Interpretation of Images

2

Introduction

> Accurate reckoning. The entrance into the knowledge of all existing things and all obscure secrets.
>
> *Rhind Mathematical Papyrus*, ~1650 BCE [1]

> Measure the measurable and try to render measurable what is not yet.
>
> Galileo Galilei (1564–1642)

> When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.
>
> Lord Kelvin, 1893

Members of the public are generally impressed by quantitation and numbers; and scientists, in particular, show great reverence for measurements and quantitative analysis. Relatively simple well-defined parameters of basic qualities such as a person's height and weight can be easily measured in an accurate and reproducible fashion. The person's height and weight can also be estimated visually by a human observer, but the estimate will be neither as accurate nor as reproducible.

At the other end of the spectrum of complexity are images such as the human face and medical images. These images consist of multiple parameters, lack sharp borders or geometric shapes, have variable signal intensity within a single structure, and cannot be easily measured. Yet these images hold very useful information. The question is not a simple one such as height or weight but whether a face is that of specific person, from any projection and in any mood, or whether a medical image is consistent with a certain medical condition or is pathognomonic of a particular disease.

In addition, the details of a pattern for any given disease will vary from patient to patient because of variations in the size, shape, and age of the patient and the

W.C. Klingensmith III, *The Mathematics and Biology of the Biodistribution of Radiopharmaceuticals-A Clinical Perspective*,
DOI 10.1007/978-3-319-26704-3_2

location, extent, and duration of the disease. The possible variation in details of the same disease from patient to patient and from time to time is endless. And, there is essentially no equivalent to a discrete scientific quality that can be measured that would summarize the useful information in the image.

Evolutionary pressures have resulted in a human eye-brain complex which has extraordinary abilities for analyzing images and recognizing patterns (Fig. 2.1). As an example, radiologists refer to a typical pattern of a disease or condition in a medical image that can be immediately recognized once you have seen it before as an "Aunt Minnie." Aunt Minnie patterns have certain features that can be described and listed, but there is no substitute for having seen an image of the disease or condition before.

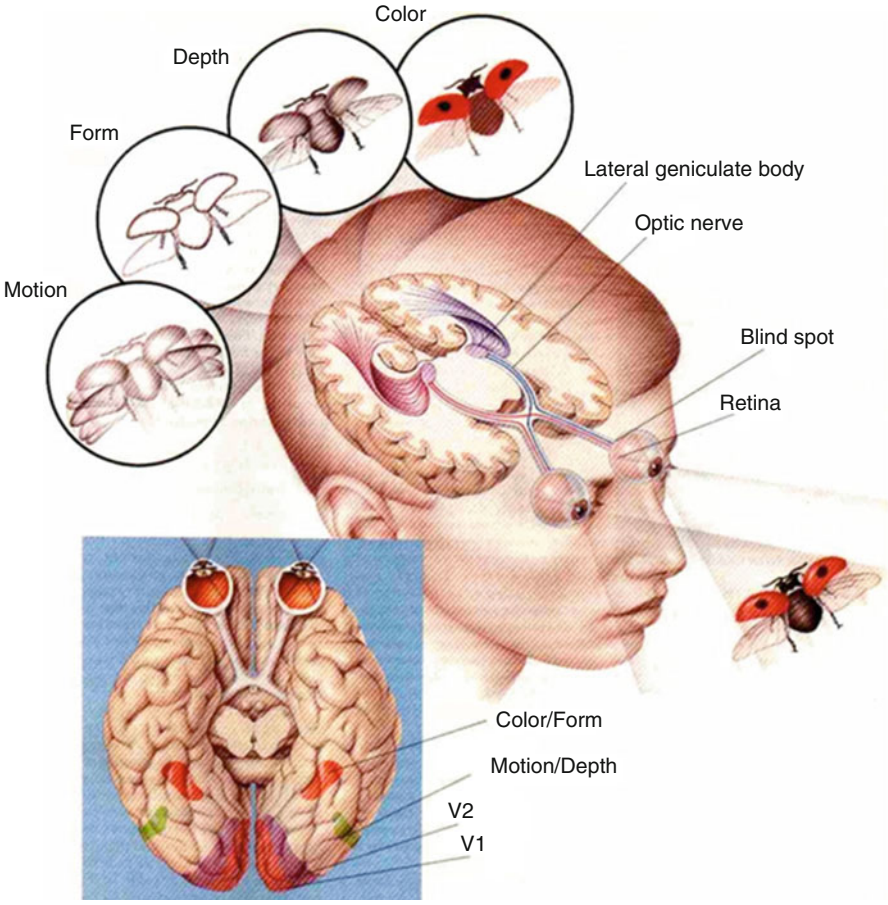

Fig. 2.1 The eye-brain complex. The eye-brain complex, in almost real time, breaks down an image into multiple different parameters, evaluates each one separately, and then synthesizes the data into a coherent result [2]

The same phenomenon applies to faces as the label suggests. Imagine a lineup of your Aunt Minnie and four other older women who were chosen because they look like your Aunt Minnie: same height, weight, hair color, eye color, dress, etc. Since you know your Aunt Minnie, you can visually analyze the faces and immediately pick out your Aunt Minnie. Now imagine that you have to create an algorithm for a colleague, who doesn't know your Aunt Minnie, that will enable your colleague to pick her out of the group. You could specify hair and eye color, distance between eyes and nostrils, a pleasant disposition, etc., but your colleague most likely will do little better than chance. A related dictum in radiology is "A radiologist with a ruler (or electronic calipers) is a radiologist in trouble." The point is that if the radiologist does not recognize the abnormality from previous experience, resorting to a series of measurements is not likely to provide the answer.

> The brain is the last and grandest frontier, the most complex thing we have yet discovered in our universe. It contains hundreds of billions of cells interlinked through trillions of connections. The brain boggles the mind.
>
> James D. Watson

Another critical capability of the eye-brain complex is its ability to recognize a pattern that has been significantly distorted (Fig. 2.2) [3]. This image of Abraham Lincoln from a cover of Science has been digitized at low spatial resolution, but an American eye-brain complex can still recognize who it is almost instantly. The original cover of Science had three additional images of Lincoln with increasing amounts of smoothing, which in turn made the image progressively easier to recognize. Here only the most distorted version is shown.

The ability of the eye-brain complex to recognize distorted images is also demonstrated by the widely used method, CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) (Fig. 2.3). CAPTCHA is a challenge-response test that is used on the internet to determine if the entity on the other end of the connection is a human or a computer software application [4]. Another aspect of the eye-brain complex is demonstrated in Fig. 2.4. Panel a is a matrix of Hounsfield units from a medical image. Every digital image has an underlying matrix of numbers that corresponds to the intensity of the gray scale in each cell of the matrix. Yet the numbers alone, while very quantitative, are essentially uninterpretable to the eye-brain complex. The information must be converted to a gray or color scale before it can be analyzed by the eye-brain complex (panel b).

> Everything that can be counted does not necessarily count;
> everything that counts cannot necessarily be counted.
>
> Albert Einstein

Despite the exquisite sophistication of the eye-brain complex for evaluating patterns, it must be stated that there is no theoretical reason why a computer cannot be programmed to analyze patterns in images, as reflected in a matrix of numbers, as well as, or better than, the eye-brain complex [5]. However, currently the ability of

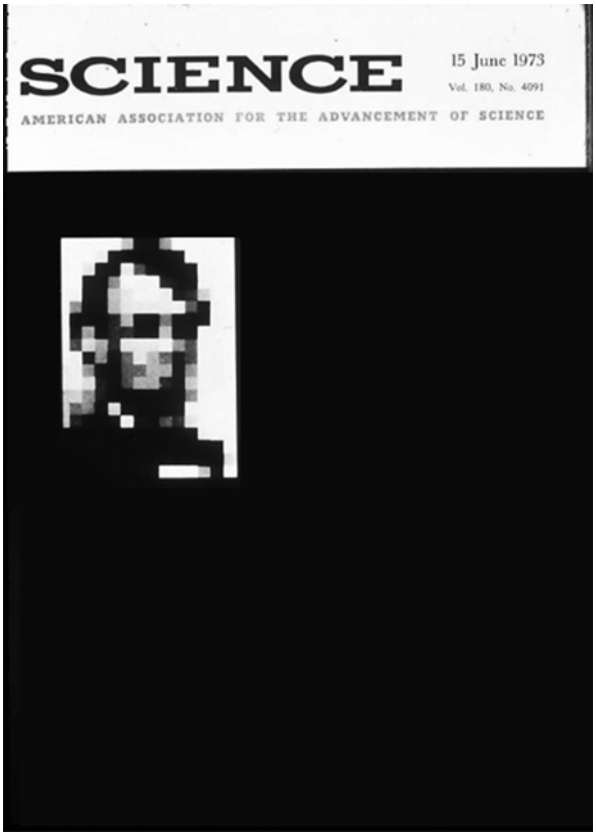

Fig. 2.2 Abraham Lincoln. The original cover of Science showed four digitized images of Abraham Lincoln with increasing degrees of smoothing. Three of them have been removed for clarity. The point is that despite significant distortion of the remaining image, those familiar with Lincoln's face have no trouble recognizing him [3]

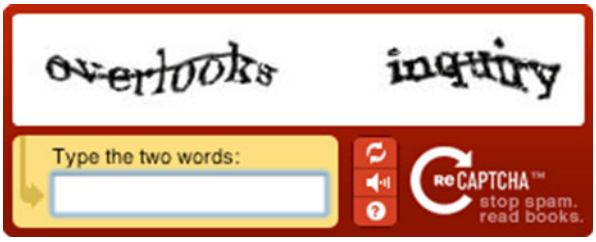

Fig. 2.3 CAPTCHA. CAPTCHA tests consist of distorted letters or figures. Humans can easily determine what the symbols represent, but computer algorithms, in general, are unable to decipher them [4]

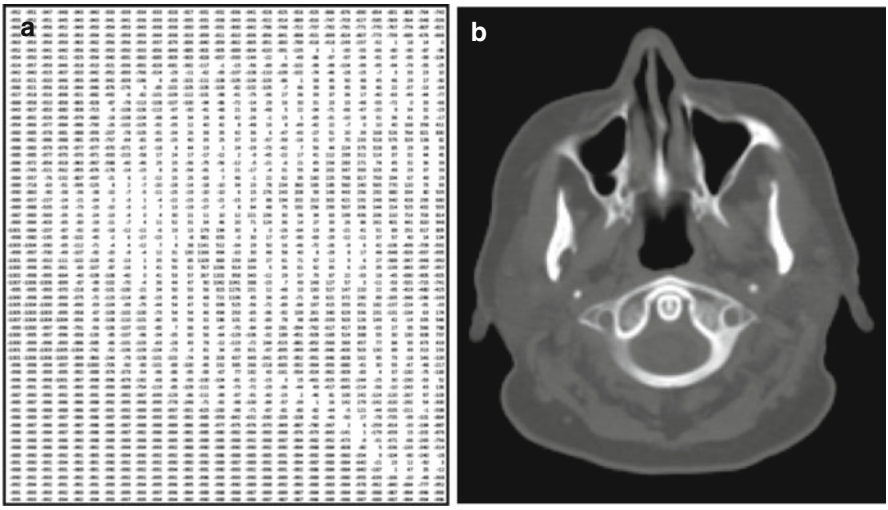

Fig. 2.4 Digital medial image. Panel (**a**) is a matrix of numbers, which in this case are Hounsfield units from a CT image. Although the information is very quantitative, the eye-brain complex is unable to interpret the information until the numbers are converted to a gray scale (Panel **b**)

computers to analyze digital images created by medical imaging machines is quite limited. Because of the inability of computer search engines to make sense of images, beyond their labels, online images have been called the "dark matter of the digital age" [6]. Presumably, it will be a long time before computers or some more advanced analytic device matches the pattern recognition capabilities of the eye-brain complex.

Generating Numbers: General

There are a total of 39 diagnostic nuclear medicine studies included in this book (Appendix A). Sixteen of them routinely involve the explicit generation of one or more numbers, which are integrated with the visual evaluation of the images to form the overall interpretation of the study. In addition, quantification is performed in a number of other studies as an optional maneuver. In the 16 studies, in which a number is generated routinely, visual evaluation is directly involved in the quantitative process either by determining when activity reaches a certain anatomic structure or by determining the appropriate anatomic boundaries of one of more regions of interest (ROIs). This concept even applies to the thyroid uptake measurement with I-123 in that correct placement of the thyroid probe depends on visual assessment of the patient's neck and thigh although here the visual evaluation is done by the nuclear medicine technologist. In addition, visual interpretation should always be used to evaluate all of the images in a study whether there is any quantitation involved or not.

As will be mentioned below and discussed in detail in Part II, Mathematics of the Biodistribution of Radiopharmaceuticals, and Part III, Quantitative Evaluation of Nuclear Medicine Studies, the numbers that are generated by the various quantitative processes are subject to a number of errors [7–9]. Whenever the visual appearance of the structures involved in generating numbers does not agree with the number, the various steps in the quantitative process, especially the placement of ROI(s), should be reviewed. In general, the nuclear medicine physician should not let a number override his/her visual interpretation.

We will now look at three different types of quantitation as examples of how quantitation in nuclear medicine studies can be conceptually categorized. These three categories are (1) measurement of transit time by visual evaluation of a sequence of images; (2) measurement of relative function, in terms of either time or clearance; and (3) measurement of absolute function in terms of clearance. Each of these categories, as well as others, will be discussed more extensively in Parts II and III of this book.

Generating Numbers: Transit Times

Table 2.1 lists the five nuclear medicine studies in which visual evaluation is used to directly determine a number, the leading edge transit time, through a flowing system. This works because the beginning and end points are operationally well defined and/or easily identified in the images. Because of laminar flow, the leading edge transit time will always be faster than the mean transit time (see Chap. 5, Mean Transit Time: Central Volume Principle).

In the case of the Cisternography Study, the start time is the time of injection of the tracer into the lumbar subarachnoid space and the end time is the time at which the tracer reaches the region of the superior sagittal sinus. Normally, the tracer reaches the convexity of the brain by 24 h after intrathecal injection in the lumbar region. Blockage of the subarachnoid space and enlargement of the cerebrospinal fluid space are two causes of a prolonged leading edge transit time.

Table 2.1 Studies that measure leading edge transit times visually

Study	Radiopharmaceutical	Parameter	Upper limit of normal
Cisternography	In-111-DTPA	Subarachnoid space: lumbar to superior sagittal sinus	24 h
Cardiac Gated Blood Pool Study	Tc-99m-red blood cells	Vascular space: right ventricle to left ventricle	6 s
Hepatobiliary Study	Tc-99m-trimethylbromo-IDA	Biliary tract: hepatocytes to extrahepatic bile ducts	10 min
Renal Glomerular Filtration Study	Tc-99m-DTPA	Tubular lumens: glomeruli to calyces	5 min
Renal Tubular Secretion Study	Tc-99m-MAG3	Tubular lumens: tubular cells to calyces	5 min

In the case of the Cardiac Gated Blood Pool Study with Tc-99m-red blood cells, the leading edge transit time through the pulmonary vasculature is defined as the time it takes for the leading edge of the tracer to pass from the right ventricle to the left ventricle. Visually this is the time difference between when the tracer is first seen in the right ventricle and when the tracer is first seen in the left ventricle. The serial images during the first circulation of the Tc-99m-red blood cells are acquired in the left anterior oblique projection in order to provide good separation between the right and left ventricles.

The normal leading edge pulmonary transit time is 6 s or less. Any disease or condition that decreases cardiac output (flow) or increases the blood volume of the lungs will increase the leading edge pulmonary transit time (see Chap. 5, Mean Transit Time: Central Volume Principle). Left ventricular failure both decreases cardiac output and increases pulmonary blood volume and is the most common cause of a prolonged pulmonary leading edge transit time.

In the case of the Hepatobiliary Study with Tc-99m-trimethylbromo-IDA, Renal Glomerular Filtration Study with Tc-99m-DTPA, and the Renal Tubular Secretion Study with Tc-99m-MAG3, the start time is the time of injection even though the parameter being measured starts with extraction of the tracer from blood into the liver or kidneys. However, the time from site of injection to the liver or kidneys is short compared to the parenchymal transit time, the injection time is convenient, and the start of the clearance is not easy to identify. The end times are the times at which the tracers leave the organ in question and appear in the intrahepatic ducts for the Hepatobiliary Study and in the calyces for the two renal studies. The upper limit of normal is 10 min for the Hepatobiliary Study and 5 min for the renal studies. The upper limit of normal is longer for the Hepatobiliary Study than for the kidneys presumably because the liver is larger than the kidneys and has longer pathways. The most common cause of a prolonged leading edge transit time in both organs is acute disease of any kind, possibly because the acute process causes edema and constriction of the intrahepatic bile ducts or renal tubular lumens.

Generating Numbers: Relative Measurements

Table 2.2 lists the 14 studies that routinely include generation of a number in relative terms. Quantitative values in relative terms are measurements that compare two values at two different locations in one or more images measured at the same point in time or two values at the same location in more than one image measured at the different times. For any given comparison, the attenuation of activity between the images is usually similar and in some cases can be ignored. However, background correction is always performed unless the images are tomographic (SPECT or PET) or there is no activity outside of the organ of interest. The amount of activity introduced to perform the study, i.e., the dose, is not part of the calculation.

The visual component in generating these numbers is the placement of the ROIs. The position, size, and shape of each ROI are important because misplacement can result in large numerical errors. The ROI should always be recorded on the pertinent

Table 2.2 Studies with relative quantitation

Procedure	Radiopharmaceutical	Functional parameter	Image type	Attenuation correction	Background subtraction
Cardiovascular system					
Cardiac Gated Blood Pool Study	Tc-99m-red blood cells	Ejection fraction	Planar	No	Background ROI
Myocardial Perfusion Study	N-13-ammonia	Myocardial perfusion	PET-CT	CT density map	Tomography
Myocardial Perfusion Study	Rb-82 as rubidium chloride	Myocardial perfusion	PET-CT	CT density map	Tomography
Myocardial Perfusion Study	Tc-99m-sestamibi	Myocardial perfusion	SPECT	Comparison to normal range	Tomography
Myocardial Perfusion & Viability Study	Tl-201 as thallous chloride	Myocardial perfusion and viability	SPECT	Comparison to normal range	Tomography
Myocardial Viability Study	F-18-FDG	Myocardial viability	PET-CT	CT density map	Tomography
Central nervous system					
Ventricular Shunt Study	Tc-99m-DTPA	Washout time	Planar	No	No background activity
Gastrointestinal system					
Esophageal Motility Study	Tc-99m-sulfur colloid	Transit time	Planar	No	No background activity
Gastric Emptying Study	Tc-99m-sulfur colloid	Transit time	Planar	Geometric mean of opposing images	No background activity
Hepatic Artery Perfusion Study	Tc-99m-MAA	Hepatic artery perfusion	Planar	Geometric mean of opposing images	No background activity
Hepatobiliary Study	Tc-99m-trimethylbromo-IDA	Gallbladder ejection fraction	Planar	No	Background ROI
Genitourinary system					
Renal Glomerular Filtration Study	Tc-99m-DTPA	Half-time of peak renal activity	Planar	No	Background ROI
Renal Tubular Excretion Study	Tc-99m-MAG3	Half-time of peak renal activity	Planar	No	Background ROI
Renal Tubular Function Study	Tc-99m-DMSA	Renal tubular clearance	Planar	No	Background ROI

image, submitted for review at the time of interpretation, and saved permanently as part of the study record.

In the section above, Generating Numbers: Transit Times, the clinical aspects of the studies were discussed as examples of putting everything together and because there are only five studies. However, in this section, because there are so many studies with relative quantitative measurements, the clinical details will be deferred until Part III, Quantitative Evaluation in Nuclear Medicine Studies.

Generating Numbers: Absolute Measurements

Table 2.3 lists four studies that routinely include generation of a number in absolute terms. The list has been arbitrarily limited to quantitative processes that measure a function rather than just an amount. All four studies achieve absolute quantification of a function by forming a ratio of the activity in the organ or lesion of interest in the numerator divided by the amount of activity that was administered to the patient in the denominator. In addition, the issues of attenuation, background activity, and field of view of the imaging or measuring device must be normalized for the numerator and denominator.

There is one other study that involves a routine absolute measurement, but not of a bodily function: the Cystogram with Tc-99m-DTPA quantitates the post-void residual volume of urine in the bladder in absolute terms. In addition, there is a study that involves an absolute measurement, but not on a routine basis: the Cisternogram with In-111-DTPA occasionally quantitates the amount of leakage of tracer into the nasal cavity in absolute terms. These two studies are discussed in Chap. 10, Other Quantitative Techniques, at the end of Part II.

The visual component in generating the absolute quantitation of function is the placement of the ROIs. As stated above for relative measurements, the position, size, and shape of each ROI are important because misplacement can result in large numerical errors. The ROI should always be recorded on the pertinent image, submitted for review at the time of interpretation, and saved permanently as part of the study.

In absolute measurements, in which activity in an organ or lesion in the body is compared to the administered dose, there are even more possibilities for error than in relative measurements. As an example, a partial extravasation of the injected dose of a radiopharmaceutical will decrease the amount of tracer in the blood that is available for clearance into the organ or lesion and will result in a spuriously low measurement from the organ or lesion. But in a relative measurement, the effect of the extravasation has a proportional effect in the two or more ROIs that are involved and the errors cancel.

An important step in the interpretation of all studies is to evaluate the images for artifacts and in the case of studies that have quantitative measurements to look for findings that might render the measurements invalid. Table 2.4 (modified from reference [8]) is a partial summary of the large number of potential sources of error in the commonly used standard uptake value (SUV) in the Tumor Glucose Metabolism Study with F-18-fluorodeoxyglucose [8]. These errors are discussed in greater detail in Chap. 20, Tumor Imaging, in Part III.

Table 2.3 Studies with absolute quantitation

Procedure	Radiopharmaceutical	Functional parameter	Image type	Attenuation correction	Background subtraction	Absolute reference measurement
Endocrine system						
Thyroid Uptake Measurement	!-123	Thyroid uptake (clearance)	Probe	Neck phantom	Similar shaped structure without thyroid: Thigh	Measure dose in phantom with probe
Genitourinary system						
Renal Glomerular Filtration Study	Tc-99m-DTPA	Glomerular filtration (clearance)	Planar	Estimated from height, weight, and age	Background ROI	Measure dose with gamma camera and collimator
Renal Tubular Excretion Study	Tc-99m-MAG3	Tubular excretion (clearance)	Planar	Estimated from height, weight, and age	Background ROI	Measure dose with gamma camera and collimator
Tumor imaging						
Tumor Glucose Metabolism Study	F-18-FDG	Glucose metabolism (clearance)	PET-CT	CT density map	Tomography	Measure dose in dosimeter

Table 2.4 Causes of inaccurate SUV measurements (partial list)

Category/problem	Explanation	Effect	Usual range (maximum effect)
Technical errors			
Relative calibration between PET scanner and dose calibrator	Can create an incorrect calibration factor	Increases or decreases SUV	−10 to 10 % (±50 %)
Residual activity in syringe or administration system	Decreases administered dose	Decreases SUV	0–5 % (typically <15 %)
Dose calibrator and PET scanner clocks not synchronized	–	Increases or decreases SUV	−10 to 10 % (21 %)
Dose calibrator inaccurate	Affects denominator of SUV equation	Increases or decreases SUV	Unknown
Injection vs. calibration time	Inaccurate decay correction	Increases or decreases SUV	−10 to 10 % (unknown)
Extravasation of dose	Decreases blood tracer concentration	Decreases SUV	0–50% (or more)
Biologic factors			
Elevated blood glucose level	Increases competition for clearance of FDG	Decreases SUV	0–15 % (−75 %)
Prolonged FDG uptake time	Increases FDG available for clearance	Increases SUV	0–15 % (30 %)
Patient motion or breathing	Spreads FDG activity into greater number of pixels	Decreases SUV	0–30 % (60 %)
Nontumorous tissue competing for blood FDG	FDG taken out of blood by normal structures, e.g., myocardium, brown fat, muscle	Decreases SUV	Unknown
Decreased extraction efficiency	Down regulation of glucose receptors secondary to elevated blood glucose level	Decrease SUV	Large
Obesity	Attenuation not fully corrected	Decreases SUV	Unknown
Inflammation	Tumors frequently contain white blood cells	Increases SUV	Unknown
Physical factors			
Scan acquisition parameters	Lower SNR causes increase in SUV	Increases SUV	0–15 % (15 %)
Image reconstruction parameters	Insufficient convergence lowers SUV	Decreases SUV	−30 to 0 % (−30 %)
ROI	SUV strongly affected by ROI	Increases or decreases SUV	−30 to 30 % (−50 %)

(continued)

Table 2.4 (continued)

Category/problem	Explanation	Effect	Usual range (maximum effect)
Normalization method	SUV varies with method: weight, surface area, lean mass	Increases or decreases SUV	Trivial
Changing blood CT contrast levels during PET scanning	Inaccurate attenuation of PET data	Increases SUV	0–15 % (50 %)
Spatial resolution of PET scanner	Partial volume effect when lesion under 1.5 cm	Decreases SUV	−50 to 0 % (−100 %)

Conclusion

The approach to any imaging study is to first determine whether the protocol was followed and whether there are any artifacts. Then the images should be evaluated visually. In any study that includes quantitation with ROIs, the ROIs should be evaluated for correct placement. Then the quantitative numbers are assessed. If the numbers do not agree with the visual interpretation, the quantitative process should be scrutinized for errors.

Quantitation can improve the precision and accuracy of the final interpretation and convey useful information to the referring clinician, but the quantitative results should be used as a "consultant" and should not override the visual appearance [10]. For the foreseeable future, quantitation should inform the eye-brain complex, but not replace it.

References

1. Chace AB. Rhind mathematical papyrus; 1927. p. 27. National Council of Teachers of Mathematics, 1979; ISBN 0-87353-133-7.
2. Internet: Extraordinary design in the eye, 2016. http://www.designanduniverse.com/articles/design_in_the_eye.php
3. Harmon LD, Julesz B. Masking in visual recognition: effects of two-dimensional filtered noise. Science. 1973;180:1194.
4. von Ahn L, Maurer B, McMillen C, et al. reCAPTCHA: human-based character recognition via web based security measures. Science. 2008;321:1465–8.
5. Bohannon J. Unmasked: facial recognition software could soon ID you in any photo. Science. 2015;347:492–4.
6. Quote from Fei-Fei Li. In: Beyond the turing test. Science. 2015;347:116.
7. Keyes JW. SUV: standard uptake or silly useless value? J Nucl Med. 1995;36:1836–9.
8. Boellaard R. Standards for PET image acquisition and quantitative data analysis. J Nucl Med. 2009;50:11S–20.
9. Kumar V, Nath K, Berman CG, et al. Variance of SUVs for FDG-PET-CT is greater in clinical practice than under ideal study settings. Clin Nucl Med. 2013;38:175–82.
10. Coleman RE, Graham MM. Is quantitation necessary for oncological PET studies? Eur J Nucl Med. 2002;29:133–5.