

Task-based measures of image quality and their relation to radiation dose and patient risk

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 Phys. Med. Biol. 60 R1

(<http://iopscience.iop.org/0031-9155/60/2/R1>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 146.107.3.4

This content was downloaded on 08/01/2015 at 07:32

Please note that [terms and conditions apply](#).

Topical Review

Task-based measures of image quality and their relation to radiation dose and patient risk

Harrison H Barrett^{1,2}, Kyle J Myers³, Christoph Hoeschen^{4,5},
Matthew A Kupinski^{1,2} and Mark P Little⁶

¹ College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA

² Center for Gamma-Ray Imaging, Department of Medical Imaging, University of Arizona, Tucson, AZ 85724, USA

³ Division of Imaging, Diagnostics, and Software Reliability, Office of Scientific and Engineering Laboratories, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, 20993 MD, USA

⁴ Department of Electrical Engineering and Information Technology, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany

⁵ Research Unit for Medical Radiation Physics and Diagnostics, Helmholtz Zentrum München, Ingolstädter Landstraße 1, 85764 Oberschleissheim, Germany

⁶ Division of Cancer Epidemiology and Genetics, Radiation Epidemiology Branch, National Cancer Institute, 9607 Medical Center Drive, Bethesda, MD 20892-9778, USA

E-mail: barrett@radiology.arizona.edu

Received 4 June 2014, revised 24 September 2014

Accepted for publication 26 September 2014

Published 7 January 2015



Abstract

The theory of task-based assessment of image quality is reviewed in the context of imaging with ionizing radiation, and objective figures of merit (FOMs) for image quality are summarized. The variation of the FOMs with the task, the observer and especially with the mean number of photons recorded in the image is discussed. Then various standard methods for specifying radiation dose are reviewed and related to the mean number of photons in the image and hence to image quality. Current knowledge of the relation between local radiation dose and the risk of various adverse effects is summarized, and some graphical depictions of the tradeoffs between image quality and risk are introduced. Then various dose-reduction strategies are discussed in terms of their effect on task-based measures of image quality.



Content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Keywords: radiation dose, image quality, patient risk, ionizing radiation, tomography

1. Introduction

Like any engineering problem, the design of an imaging system always requires tradeoffs. For medical imaging with ionizing radiation, a major tradeoff is between image quality and risk to the patient from absorbed radiation dose. All else being equal, increasing the radiation dose to the patient will always improve image quality by reducing the Poisson noise in the image, but by how much? At the same time, increasing the radiation dose, above some level at least, will always result in increased risk of adverse effects on the patient, but by how much? The purpose of this paper is to survey our current knowledge of how radiation dose relates to image quality and patient risk, thereby laying the groundwork for a rigorous risk-benefit analysis applicable to such important medical modalities as mammography, nuclear medicine and computed tomography.

The tools required for this risk-benefit analysis are (a) the theory and practice of task-based assessment of image quality; (b) methods for calculating the spatiotemporal distribution of absorbed dose in a patient's body, and (c) knowledge from epidemiological data, mathematical models and laboratory studies of the probability of various adverse effects arising from the absorbed dose distribution.

Task-based assessment of image quality rests on the premise that the only rigorous way of defining image quality for a medical image is in terms of the medical goal of the image, which is often called the *task* of the imaging system (Barrett 1990, Barrett *et al* 1995, Barrett *et al* 1998a, Barrett *et al* 2006, Barrett and Myers 2004). Given an image, the task can be performed by a person such as a radiologist, by a computer algorithm, or by a mathematical construct called an ideal observer. We use the general term *observer* to refer to any of these methods for performing the task. For a given task, observer and imaging system, we can define a suitable figure of merit (FOM) that specifies how well that task can be performed on some ensemble of patients.

If the chosen task is relevant to the diagnosis and/or treatment of the patients in the ensemble, then the FOM quantifies the benefit to patients from the imaging procedure under study. Many other ad hoc figures of merit for image quality have been proposed in the literature (see section 14.1 in Barrett and Myers (2004) for a survey), but if they do not relate to benefits to the patients, i.e. to task performance, then they cannot be used for risk-benefit analysis.

The theory behind task-based assessment of image quality is now well established, and the methodology is used routinely (but not universally) in many investigations of image quality. After some mathematical and statistical preliminaries in section 2, a review of task-based assessment is given in section 3 of this paper.

A task-based FOM depends not only on the task and the chosen ensemble of patients, but also on many details of the imaging system, such as the spatial and temporal resolution of the system and, in the case of imaging with x-rays and gamma rays, the number of high-energy photons recorded in the image. The number of recorded photons, in turn, depends on many factors including the efficiency of the imaging system at collecting and recording the photons and the exposure time over which the image is acquired. Most importantly for this paper, however, all task-based FOMs will necessarily depend on the radiation dose to the patient.

Many physical processes enter into the calculation of the distribution of absorbed dose in a patient's body. A careful calculation requires consideration of how the photons are delivered to the body, the energy spectrum of the photons, and the physical processes of absorption and

scattering in the body. Though complex, these factors are all well understood and precise computational methods have been developed, so there is no difficulty in principle in computing the distribution of the absorbed dose for a given patient and imaging procedure. Unfortunately, it is common practice in the field to reduce the dose distribution, which is really a function of spatial position in the body and time, to a single number which, it is hoped, will capture some indication of the biological hazard from the dose in a way that is easily communicated to the patient or to the scientific and medical community. Some aspects of the dose distribution and various summary measures of it are discussed in section 4.

The biological risk from medical imaging with ionizing radiation is a complicated and controversial subject but, as we shall see in section 5, there are viable approaches to assessing the risks of various adverse effects and incorporating them into an overall framework for assessing the tradeoffs between dose and image quality. It is not the objective of this paper to enter into the debate regarding whether the linear, no-threshold model or some nonlinear model is more nearly correct for some specific radiation hazard, but in section 5 we provide enough information about the models that a designer or user of a medical imaging system can quantify the relationship between image quality and patient risk for any of the extant models and for many different adverse effects of radiation.

Section 6 shows how these major themes—task-based assessment of image quality, spatial and temporal distribution of absorbed dose, and dose-risk models—can be integrated into a general theoretical framework for using and optimizing medical imaging systems. In section 7 we survey current research into methods that might be used for reducing the patient dose in imaging with ionizing radiations, potentially without loss of image quality. In section 8 we try to identify key gaps in our knowledge of image-quality assessment, dose calculation and estimation of biological hazards, and we suggest some research directions that will help fill these gaps.

2. Mathematical basics and notation

2.1. Mathematical descriptions of objects and images

The notation used in this paper closely follows (Barrett and Myers 2004). An object being imaged is a scalar-valued function of space and time, denoted $f(\mathbf{r}, t)$, where $\mathbf{r} = (x, y, z)$ is a three-dimensional (3D) position vector and t is the time. In this paper we consider only radiological imaging with x rays or nuclear-medical imaging with gamma rays.

With x-ray imaging the radiation source is outside the patient's body, and $f(\mathbf{r}, t)$ represents the spatiotemporal distribution of the x-ray attenuation coefficient in the body, which can arise either from the intrinsic properties of the tissue or from an exogenous x-ray contrast medium. The time dependence arises from respiratory or cardiac motion, peristalsis, gross patient motion or redistribution of the contrast medium. The x-ray attenuation coefficient can also be a function of photon energy, and there can be both absorption and scattering components of the attenuation.

In nuclear medicine, $f(\mathbf{r}, t)$ represents the distribution of a radioactive tracer in the body, with the time dependence arising from redistribution of the tracer, radioactive decay or patient motion. For notational ease we denote the object function as $\mathbf{f}$, which can be regarded simply as a shorthand for $f(\mathbf{r}, t)$. Technically, $\mathbf{f}$ denotes a vector in a Hilbert space, but we do not need this level of rigor here.

Although an object being imaged is a function of continuous spatial and temporal variables, a digital image is a discrete sets of numbers, each obtained as a result of a detector making a measurement of the x-ray or gamma-ray flux impinging on it. We denote the set of measurements for one object as $\{g_m, m = 1, \dots, M\}$, where the curly brackets denote a set, M

is the total number of measurements, and index m denotes a particular detector measurement. The measurement consists of integrating the spatially and temporally varying photon flux over some finite detector area and some finite measurement time. Again for notational convenience, we denote the set of measurements as $\mathbf{g}$, which we can think of as an $M \times 1$ column vector.

Because objects are functions of continuous variables and images are discrete vectors, the imaging system is referred to as a continuous-to-discrete (CD) operator (Barrett and Myers 2004), which we denote as $\mathcal{H}$. To be more precise, we denote the mean image as $\bar{\mathbf{g}}$, where the overbar indicates an average over many repeated measurements of exactly the same object function. (Deviations from this average are discussed below.) With this notation, we describe the imaging system abstractly as⁷

$$\bar{\mathbf{g}} = \mathcal{H}\mathbf{f}. \quad (1)$$

For nuclear medicine, the operator $\mathcal{H}$ is linear to a good approximation. That is, if we replace the object $\mathbf{f}$ by the scaled object $\alpha\mathbf{f}$, for example by injecting a different activity of the radio-tracer, then we change each component $\bar{g}_m$ of the mean data vector to $\alpha\bar{g}_m$ (where α is some nonnegative constant). For x-ray imaging, $\mathcal{H}$ is not linear because of the exponential attenuation of x-rays, but we can still write (1).

The general form of a linear CD operator is

$$\bar{g}_m = \int_{\infty} d^3r h_m(\mathbf{r}) f(\mathbf{r}), \quad (m = 1, \dots, M), \quad (2)$$

where $h_m(\mathbf{r})$ is referred to as either the kernel of the operator $\mathcal{H}$ or as the sensitivity function of the system. For single-photon radionuclide imaging, $h_m(\mathbf{r})$ specifies the probability that a gamma-ray photon originating at point $\mathbf{r}$ results in an event in element m of the discrete image; in conventional rotating-camera SPECT, the index m specifies not only the position of the event on the camera face but also the angular position of the camera, so M is the number of angles times the number of pixels in each 2D projection. The subscript ∞ on the integral in (2) indicates that the integral is over the full infinite range of x , y and z , but in practice the range is limited by $h_m(\mathbf{r})$.

All CD operators have *null functions*, functions in object space for which $\mathcal{H}\mathbf{f}_{\text{null}} = 0$. For linear CD operators, the method of *singular-value decomposition* (SVD) can be used to decompose any object uniquely as $\mathbf{f} = \mathbf{f}_{\text{null}} + \mathbf{f}_{\text{meas}}$, where $\mathbf{f}_{\text{meas}}$ is called the *measurement component* of $\mathbf{f}$. By definition, only the measurement component affects the mean data: $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f} = \mathcal{H}\mathbf{f}_{\text{meas}}$.

In tomographic imaging modalities such as PET, SPECT or CT, the raw detector measurements $\mathbf{g}$ do not constitute the final image. Instead, a linear or nonlinear algorithm is used to produce some discrete approximation to the original object function. We express the algorithm with an operator $\mathcal{O}$ and write

$$\hat{\mathbf{f}} = \mathcal{O}\mathbf{g}, \quad (3)$$

where $\hat{\mathbf{f}}$ is a set of N numbers on some reconstruction grid (almost always a voxel grid). Thus $\hat{\mathbf{f}}$ is a vector representation of the set $\{\hat{f}_n, n = 1, \dots, N\}$, and the reconstruction operator $\mathcal{O}$ maps an $M \times 1$ vector to an $N \times 1$ vector; if the algorithm is linear, then $\mathcal{O}$ is an $N \times M$ matrix, which we can denote as $\mathbf{O}$.

In x-ray CT, $\mathcal{O}$ is necessarily nonlinear because of a logarithmic operation that compensates for the exponential attenuation of x rays, and in all modalities the subsequent reconstruction algorithm can be nonlinear as well.

⁷ Typographically, we will always denote vectors of any dimensionality with lower-case bold roman (e.g. $\mathbf{g}$, $\mathbf{f}$, $\hat{\mathbf{f}}$); finite-dimensional matrices with upper-case bold roman (e.g. $\mathbf{O}$, $\mathbf{K}$), and general operators (which may reduce to matrices in special cases) with a bold calligraphic font (e.g. $\mathcal{H}$, $\mathcal{O}$).

2.2. Stochastic descriptions of objects and images

In order to understand and analyze image quality and how it is affected by radiation dose, we need to understand why images are random and how we describe that randomness mathematically.

With ionizing radiation, a primary cause of image randomness is the random interaction of high-energy photons with the detector. Under very general assumptions (Barrett and Myers 2004), virtually always satisfied in medical imaging, the interactions in different detectors, or in the same detector in two different, nonoverlapping time intervals, are statistically independent of each other. From this simple independence assumption, it can be shown (Barrett and Myers 2004) that the random number of detected photons in the m th measurement, denoted N_m , follows the familiar Poisson probability law:

$$\Pr(N_m | \bar{N}_m) = \frac{(\bar{N}_m)^{N_m}}{N_m!} \exp(-\bar{N}_m), \quad (4)$$

where the vertical bar indicates a conditional probability; $\Pr(N_m | \bar{N}_m)$ is to be read: the probability of observing N_m ‘conditional on’ (or ‘given’) its mean $\bar{N}_m$.

For a photon-counting detector, such as an Anger scintillation camera or some of the new cadmium zinc telluride (CZT) detectors being developed for CT, the detector output for each m is the number of photons detected, so that $g_m = N_m$. If this detector is used in an imaging system, then it follows from (1) that $\bar{g}_m = (\mathcal{H}\mathbf{f})_m$, and conditioning on $\bar{g}_m$ is the same as conditioning on $\mathbf{f}$, so

$$\Pr(g_m | \mathbf{f}) = \frac{[(\mathcal{H}\mathbf{f})_m]^{g_m}}{g_m!} \exp[-(\mathcal{H}\mathbf{f})_m]. \quad (5)$$

An important property of Poisson random variables is that the variance equals the mean:

$$\text{Var}(g_m | \mathbf{f}) \equiv \langle (g_m - \bar{g}_m)^2 \rangle_{g_m | \mathbf{f}} = (\mathcal{H}\mathbf{f})_m, \quad (6)$$

where the subscripted angle brackets denote an average using the probability law (5).

Because the photon interactions in different detectors are statistically independent (conditional on the object $\mathbf{f}$), the M -dimensional multivariate probability law for the full data vector $\mathbf{g}$ is given by

$$\Pr(\mathbf{g} | \mathbf{f}) = \prod_{m=1}^M \frac{[(\mathcal{H}\mathbf{f})_m]^{g_m}}{g_m!} \exp[-(\mathcal{H}\mathbf{f})_m]. \quad (7)$$

We can use this multivariate probability to generalize the concept of variance and compute an $M \times M$ conditional covariance matrix, $\mathbf{K}_{\mathbf{g}|\mathbf{f}}$, with elements given by

$$[\mathbf{K}_{\mathbf{g}|\mathbf{f}}]_{mm'} \equiv \langle (g_m - \bar{g}_m)(g_{m'} - \bar{g}_{m'}) \rangle_{\mathbf{g}|\mathbf{f}} = \text{Var}(g_m | \mathbf{f}) \delta_{mm'} = (\mathcal{H}\mathbf{f})_m \delta_{mm'}, \quad (8)$$

where $\delta_{mm'}$ is the Kronecker delta, which equals 1 if $m = m'$ and 0 if $m \neq m'$. Thus, for photon-counting detectors, $\mathbf{K}_{\mathbf{g}|\mathbf{f}}$ is a diagonal matrix, and the diagonal elements, which are the variances of the individual detector outputs, are also the means in this case.

Detectors used in x-ray imaging (Yaffe and Rowlands 1997) are usually not photon-counting, and their outputs are not integers. Instead, g_m is a continuous-valued random variable and hence $\mathbf{g}$ is an $M \times 1$ continuous-valued random vector, with the randomness arising from both the random photon interactions and electronic readout noise. With these detectors, we need to use a conditional probability density function (PDF), denoted $\text{pr}(\mathbf{g} | \mathbf{f})$, rather than the probability $\Pr(\mathbf{g} | \mathbf{f})$ as in (7). The electronic noise has zero mean, so it does not affect the mean

detector output $\bar{\mathbf{g}}$, but it does affect the covariance matrix. If the electronic noise in different detector elements is statistically independent, then we can write the conditional PDF as a product of component PDFs, analogous to (7). The covariance matrix of (8) remains diagonal but the variance is no longer the same as the mean; the covariance matrix has the structure

$$[\mathbf{K}_{\mathbf{g}|\mathbf{f}}]_{mm'} \equiv \langle (g_m - \bar{g}_m) (g_{m'} - \bar{g}_{m'}) \rangle_{\mathbf{g}|\mathbf{f}} = [\alpha(\mathcal{H}\mathbf{f})_m + \sigma_m^2] \delta_{mm'}, \quad (9)$$

where α is a constant related to the random variations in detector output for different x-ray photons and σ_m^2 is the variance of the readout noise in detector m . If we normalize the detector output so that $\bar{g}_m$ is the mean number of detected x-ray photons in some exposure time, then σ_m is in units of the average detector output for one x-ray photon and α , which is the reciprocal of the Swank factor (Swank 1973, Yaffe and Rowlands 1997), is typically in the range 1.2–1.5.

The diagonal form in (9) holds only when fluctuations at different detector elements are statistically independent, but in many x-ray detectors an absorbed photon produces light or charge which spreads out and affects many detector elements, and in those cases the covariance matrix is not diagonal (Kyprianou *et al* 2009).

So far we have discussed probability laws and covariance matrices conditional on $\mathbf{f}$, which means that they apply to repeated measurements on the same object. In imaging, and especially in assessment of image quality, we must also consider many different objects, which we can think of as random functions $\mathbf{f}$. The overall probability for the data vector $\mathbf{g}$ is obtained by averaging the conditional probability over a group of patients or other objects to be imaged⁸:

$$\text{pr}(\mathbf{g}|\mathbf{C}) \equiv \langle \text{pr}(\mathbf{g}|\mathbf{f}) \rangle_{\mathbf{f}|\mathbf{C}}, \quad (10)$$

where $\mathbf{C}$ denotes some class or group of objects (often a theoretical infinite ensemble), such as all patients who could undergo a particular diagnostic examination, and the subscripted angle brackets denote an average over all objects $\mathbf{f}$ in $\mathbf{C}$.

Similarly, the overall covariance matrix is given by

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}} \equiv \langle \mathbf{K}_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}|\mathbf{C}}. \quad (11)$$

Because the conditional covariance matrix is itself an average, given in (8), we can also write (11) in component form as

$$[\mathbf{K}_{\mathbf{g}|\mathbf{C}}]_{mm'} = \left\langle \left\langle (g_m - \bar{g}_m) (g_{m'} - \bar{g}_{m'}) \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}}, \quad (12)$$

where the double overbar indicates the overall average, obtained by averaging first over $\mathbf{g}$ given $\mathbf{f}$, then over $\mathbf{f}$ given that the object was drawn from set $\mathbf{C}$.

A more compact way of writing (12), without explicitly displaying the components, is in terms of outer products; if $\mathbf{x}$ is an $M \times 1$ column vector, then its transpose $\mathbf{x}^t$ is a $1 \times M$ row vector, and $\mathbf{x}\mathbf{x}^t$, called an outer product, is an $M \times M$ matrix with mm' component $x_m x_{m'}$. Thus

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}} = \left\langle \left\langle (\mathbf{g} - \bar{\mathbf{g}}) (\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}}. \quad (13)$$

As we shall see in section 3, this overall covariance plays a key rule in image quality.

Moreover, a simple algebraic manipulation of (13) will prove useful when we discuss the relation between dose and image quality in section 6. We merely add and subtract the conditional mean defined by (1), but written here as $\bar{\mathbf{g}}(\mathbf{f})$ to emphasize that it depends on the random $\mathbf{f}$. Thus

⁸ We use the lower-case pr for PDF here, but the results are general enough to apply to discrete-valued random vectors as well.

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}} = \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f}) + \bar{\mathbf{g}}(\mathbf{f}) - \bar{\bar{\mathbf{g}}}] [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f}) + \bar{\mathbf{g}}(\mathbf{f}) - \bar{\bar{\mathbf{g}}}]^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}}. \quad (14)$$

When the outer product is expanded out and averaged, the cross terms vanish by construction; for example,

$$\left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] [\bar{\bar{\mathbf{g}}}(\mathbf{f}) - \bar{\mathbf{g}}(\mathbf{f})]^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}} = \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] \right\rangle_{\mathbf{g}|\mathbf{f}} [\bar{\bar{\mathbf{g}}}(\mathbf{f}) - \bar{\mathbf{g}}(\mathbf{f})]^t \right\rangle_{\mathbf{f}|\mathbf{C}} = 0, \quad (15)$$

since $\langle \mathbf{g} \rangle_{\mathbf{g}|\mathbf{f}} = \bar{\mathbf{g}}(\mathbf{f})$. Thus we find that

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}} = \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}} + \mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}}, \quad (16)$$

where the first term is the noise covariance matrix averaged over objects and the second term accounts for the covariance induced by the object variability itself. The terms are defined by

$$\bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}} \equiv \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})] [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{f})]^t \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}}; \quad (17)$$

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}} \equiv \left\langle [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\bar{\mathbf{g}}}(\mathbf{f})] [\bar{\mathbf{g}}(\mathbf{f}) - \bar{\bar{\mathbf{g}}}(\mathbf{f})]^t \right\rangle_{\mathbf{f}|\mathbf{C}}. \quad (18)$$

The covariance decomposition in (16) is just a mathematical tautology; it does not depend on statistical independence of the object and image or on any other assumptions about the object or image statistics.

In situations where the conditional noise covariance matrix $\mathbf{K}_{\mathbf{g}|\mathbf{f}}$ is diagonal, such as (9), the averaged noise covariance matrix $\bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}}$ is also diagonal; the average of diagonal matrices is diagonal. The object term $\mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}}$, however, is never diagonal; interesting objects always have point-to-point correlations. Because objects are functions of continuous variables, we must describe these correlations by a covariance function, rather than a covariance matrix; the definition is

$$k_f(\mathbf{r}, t; \mathbf{r}', t') \equiv \langle [f(\mathbf{r}, t) - \bar{f}(\mathbf{r}, t)] [f(\mathbf{r}', t') - \bar{f}(\mathbf{r}', t')] \rangle_{\mathbf{f}|\mathbf{C}}. \quad (19)$$

If we regard $k_f(\mathbf{r}, t; \mathbf{r}', t')$ as the kernel of a continuous-to-continuous operator $\mathcal{K}_{\mathbf{f}|\mathbf{C}}$, and $\mathcal{H}$ is linear, then we can express (18) compactly as

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}} = \mathcal{H} \mathcal{K}_{\mathbf{f}|\mathbf{C}} \mathcal{H}^\dagger, \quad (20)$$

where $\mathcal{H}^\dagger$ is the adjoint (back-projection) operator corresponding to $\mathcal{H}$; for more details, see Barrett and Myers (2004).

For tomography, we can also define a covariance matrix for the reconstructed images rather than raw projection data, just by replacing $\mathbf{g}$ with $\hat{\mathbf{f}}$ in (13). For a linear reconstruction operator and a linear system, it is easy to show that

$$\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{C}} = \mathcal{O} \mathbf{K}_{\mathbf{g}|\mathbf{C}} \mathcal{O}^\dagger = \mathcal{O} \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}} \mathcal{O}^\dagger + \mathcal{O} \mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}} \mathcal{O}^\dagger = \mathcal{O} \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}} \mathcal{O}^\dagger + \mathcal{O} \mathcal{H} \mathcal{K}_{\mathbf{f}|\mathbf{C}} \mathcal{H}^\dagger \mathcal{O}^\dagger. \quad (21)$$

Again we have decomposed the covariance matrix into a term arising from noise (as propagated through the reconstruction algorithm) and a term arising from object variability (as propagated through both the system operator and the algorithm). Now, however, neither term is diagonal; even if the detectors do not introduce correlations, the reconstruction algorithm will.

With nonlinear reconstruction algorithms or systems, decomposition of the covariance matrix into noise and object terms is still valid, but both terms are more complicated than indicated in (21). For a detailed treatment of the noise statistics of images produced by

nonlinear reconstruction algorithms, see chapter 15 in Barrett and Myers (2004), Barrett *et al* (1994), Wilson *et al* (1994), Fessler (1996), Wang and Gindi (1997) and Qi and Leahy (2000), and for a method of estimating the object covariance term for nonlinear reconstructions, see section 6.3 of this paper.

Covariance decompositions are important in this paper because the two terms depend differently on patient dose (see section 6).

3. Principles of task-based assessment of image quality

3.1. Efficacy

In classic papers published in the 1990s, Fryback and Thornbury (1991) and Thornbury (1994) proposed a systematic approach to expressing the efficacy of diagnostic imaging. They identified six distinct stages of efficacy:

- Stage 1. Technical capacity
- Stage 2. Diagnostic accuracy
- Stage 3. Diagnostic impact
- Stage 4. Therapeutic impact
- Stage 5. Patient outcomes
- Stage 6. Societal outcomes

Stage 1 includes all of the things that engineers measure and attempt to optimize in designing medical imaging systems. Functional characterizations such as point spread function (PSF), modulation transfer function (MTF) and noise power spectrum (NPS) and summary measures such as spatial, temporal and energy resolution and count-rate capability all fit under technical capacity, Stage 1 of efficacy. For purposes of this paper, radiation dose is a technical characterization that is also part of Stage 1.

An immediate difficulty in defining efficacy in purely technical terms is that there are many different parameters and functions that one can—and should—determine as part of a comprehensive system characterization, but it is not clear how one can combine them all into a single scalar merit function that can be used for system optimization.

This difficulty is addressed in Stage 2 of efficacy. The most common measures of the accuracy of a diagnostic test are its sensitivity and specificity. As we shall see in section 3.2, these quantities can be combined into receiver operating characteristic (ROC) curves, and scalar metrics such as the area under the curve (AUC) or a detectability index derived from AUC can be used for optimization (Metz 1978, Metz 2006). Again for purposes of this paper, the variation of AUC with some measure of radiation dose is a relevant metric for the effect of dose on Stage 2 efficacy.

Stage 3, diagnostic impact, relates to the ability of a test to improve diagnostic decisions. This stage is the basis for studies of comparative effectiveness, often quantified by pre- and post-test probability of disease or by the expected utility of the diagnosis. Because diagnostic decisions are commonly based on quantitative information derived from images, we also include performance on parameter-estimation tasks as a metric for Stage 3 efficacy. Radiation dose affects the noise in an image and hence the accuracy of any quantitative information extracted from the images.

Stage 4, therapeutic impact, specifies efficacy of a diagnostic test by its ability to improve therapeutic choices, at least in the short term. In radiation therapy, for example, image data are used to plan a treatment that will control tumor growth while minimizing acute radiation damage to normal organs. The success of this plan depends in part on the quality of

the information derived from the images, which in turn, for imaging with ionizing radiation, depends on the radiation dose to the patient in the imaging step.

Stage 5 refers to longer-term outcomes, specified in terms of length or quality of life. A common metric is quality-adjusted life years (QALYs) (Pliskin *et al* 1980, Mortimer and Segal 2007). Much of the literature on Stage 5 efficacy of imaging attempts to relate the benefit of an imaging study, in QALYs, to the monetary cost of the study. In imaging with ionizing radiation, however, improved diagnostic accuracy and impact (Stages 2–4) can increase the length or quality of life, but the concomitant radiation dose can decrease it.

Stage 6, societal outcomes, is the most difficult kind of efficacy to assess. It is here that the contentious issues of the epidemiological effects of low-level radiation arise. Nevertheless the eventual goal of research into the risk-benefit tradeoff of imaging with ionizing radiation is to extend the efficacy analysis to Stages 5 and 6.

3.2. Tasks and figures of merit

This subsection introduces the tasks and the corresponding figures of merit used to quantify efficacy. Most of the emphasis here is on Stage 2 efficacy, but some discussion of the later stages is also included. Mathematical observers for performing the tasks are summarized in sections 3.3 and 3.4.

The task in imaging is always to draw inferences about the object that produced an image. Because the image data are random, as discussed in the previous section, we require statistical inferences, and the underlying theory is referred to as statistical decision theory. Useful general references for this field include (Melsa and Cohn 1978, Kay 1998, Wickens 2002, Schonhoff and Giordano 2006, Van Trees and Bell 2013).

Generically, the task can be classification, estimation or some combination of the two. A classification task is to assign the object that produced the image to one of two or more disjoint classes. In the context of cancer imaging, the classes can be as simple as ‘tumor present’ versus ‘tumor absent’ or ‘benign’ versus ‘malignant’. If there are only two possible classes, as in these examples, the classification task is said to be *binary*, but often there are more than two classes, such as different types of tumor. A binary classification task in which the objective is to determine whether some feature (such as a tumor) is present in the object is called a *detection task*.

An estimation task is to determine numerical values for one or more parameters that describe important characteristics of the object. In cancer imaging, an estimation task might be to estimate the volume of a tumor or its uptake of some tumor-specific radiotracer. If only one number is required, it is a *scalar estimation task*. If the goal is multiple numbers, such as tumor volume, location and uptake, then it is a *vector estimation task*.

Classification and estimation tasks are often combined in one diagnostic study. The clinician may want to know the type and size of a tumor. We refer to this situation as a *joint classification/estimation task*. If the presence of the tumor must also be determined before classification and size estimation, it is a *joint detection/classification/estimation task*.

3.2.1. Binary detection or classification tasks. Consider a binary task in which the goal is to use an image vector $\mathbf{g}$ to classify the object that produced it as belonging to either class $\mathbf{C}_1$ (e.g. signal present) or $\mathbf{C}_0$ (e.g. signal absent). The hypothesis that the object was drawn from class $\mathbf{C}_1$ is denoted H_1 , and the hypothesis that the object was drawn from class $\mathbf{C}_0$ is denoted H_0 .

For algorithmic observers (though not for humans), we can assume that the task is performed without any additional randomness beyond that inherent in the image data (no coin flips or random-number generators can be used) and that no equivocation is permitted (every

Table 1. Possible outcomes for a binary classification task.

1.	True positive (TP):	H_1 is true; observer makes decision $\mathcal{D}_1$.
2.	False positive (FP):	H_0 is true; observer makes decision $\mathcal{D}_1$.
3.	False negative (FN):	H_1 is true; observer makes decision $\mathcal{D}_0$.
4.	True negative (TN):	H_0 is true; observer makes decision $\mathcal{D}_0$.

image must be classified as either $\mathbf{C}_0$ or $\mathbf{C}_1$). Under these two assumptions, it can be shown (see Barrett *et al* (1998) or section 13.2 in Barrett and Myers (2004)) that any decision rule is equivalent to computing a scalar test statistic $t(\mathbf{g})$ and comparing it to a threshold t_c . If $t(\mathbf{g}) > t_c$, then the decision is that the object came from class $\mathbf{C}_1$, or equivalently that hypothesis H_1 is true; we refer to this decision as $\mathcal{D}_1$. Conversely, if $t(\mathbf{g}) < t_c$, then the decision is that the object came from class $\mathbf{C}_0$, so H_0 is true, and in this case we refer to the decision as $\mathcal{D}_0$.

We can express this decision rule succinctly as

$$t(\mathbf{g}) \begin{matrix} \mathcal{D}_1 \\ > \\ < \\ \mathcal{D}_0 \end{matrix} t_c, \quad (22)$$

which is read, ‘choose $\mathcal{D}_1$ if $>$ holds, and vice versa’. Specific choices for $t(\mathbf{g})$, i.e. specific mathematical observers for the binary task, are discussed in section 3.3.

For each image vector $\mathbf{g}$ and any particular choice of t_c , there are four possible outcome scenarios, as summarized in table 1. The true-positive and true-negative outcomes are correct decisions and the false-positive and false-negative decisions are erroneous. In the terminology of hypothesis testing, a false positive is a Type I error and a false negative is Type II.

Associated with each of these outcomes there is a fraction or probability of occurrence. For example, the true-positive fraction, denoted TPF, is the probability that the observer makes the correct decision $\mathcal{D}_1$ when the object that produced $\mathbf{g}$ was actually in class $\mathbf{C}_1$, so that hypothesis H_1 is true. In signal-detection terminology, TPF is the probability of detecting a signal when it is really there. Conversely, the false-positive fraction, FPF, is the probability of deciding that a signal is present when it really is not, so it is the false-alarm probability.

In terms of conditional probabilities, the four fractions are defined as

$$\begin{aligned} \text{TPF} &= \Pr(\mathcal{D}_1|H_1), & \text{TNF} &= \Pr(\mathcal{D}_0|H_0), \\ \text{FPF} &= \Pr(\mathcal{D}_1|H_0) = 1 - \text{TNF}, & \text{FNF} &= \Pr(\mathcal{D}_0|H_1) = 1 - \text{TPF}. \end{aligned} \quad (23)$$

In the medical literature, the TPF is referred to as the *sensitivity*, since it is an indication of the sensitivity of the test to the presence of an abnormality. The TNF is commonly referred to as the *specificity*; it measures the proportion of the tests in which the subject really does not have the abnormality that are correctly identified as normal. Because $\text{FPF} = 1 - \text{TNF}$, a test with low specificity is one where there are many false positive decisions.

These fractions can be computed if we have knowledge of the probability density functions on the test statistic t . For the generic decision rule of (22) and a given threshold t_c , we can express the four fractions as

$$\text{TPF} = \Pr(t > t_c|H_1) = \int_{t_c}^{\infty} dt \, \text{pr}(t|H_1), \quad (24a)$$

$$\text{FPF} = \Pr(t > t_c|H_0) = \int_{t_c}^{\infty} dt \, \text{pr}(t|H_0), \quad (24b)$$

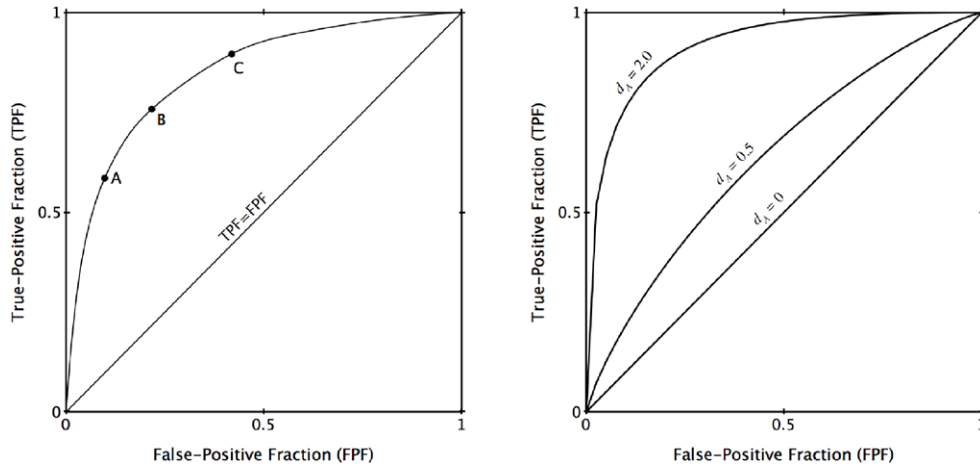


Figure 1. ROC curves. Left: single curve showing three different operating points A, B and C, corresponding to three decision thresholds t_c . Right: three different ROC curves, corresponding to different values of the detectability index d_A (hence different AUCs).

$$\text{TNF} = 1 - \text{FPF} = \int_{-\infty}^{t_c} dt \, \text{pr}(t|H_0), \quad (24c)$$

$$\text{FNF} = 1 - \text{TPF} = \int_{-\infty}^{t_c} dt \, \text{pr}(t|H_1). \quad (24d)$$

Recall that Pr denotes a probability and pr denotes a probability density function; thus $0 \leq \text{Pr}(\cdot) \leq 1$ and $\int_{-\infty}^{\infty} dt \, \text{pr}(t|\cdot) = 1$.

3.2.2. The ROC curve and figures of merit. The integrals in (24) show that all four fractions depend on the chosen decision threshold t_c , and in particular that decreasing t_c increases both TPF and FPF. We can achieve any desired probability of detection, but only at the expense of an increased false-alarm probability. This trade-off is illustrated in a Receiver Operating Characteristic (ROC) curve (see figure 1), which is a plot of TPF versus FPF (or equivalently, sensitivity versus $1 - \text{specificity}$) as t_c is varied.

As seen from (24a) and (24b), an ROC curve is determined by the two PDFs on the test statistic, $\text{pr}(t|H_1)$ and $\text{pr}(t|H_0)$. If these PDFs are identical, then there is no way to discriminate between the two hypotheses, and the ROC curve is the diagonal line $\text{TPF} = \text{FPF}$ as shown in figure 1. Conversely, if there is no overlap of $\text{pr}(t|H_1)$ and $\text{pr}(t|H_0)$, then perfect discrimination is possible and the ROC curve is a line vertically from the origin to the upper left corner, then horizontally to the upper right corner. Perfect separation of the two PDFs (hence perfect performance on this binary classification task) corresponds to an area under the curve (AUC) of 1.0. The line $\text{TPF} = \text{FPF}$, indicating no ability to perform the task, has $\text{AUC} = 0.5$. Real ROC curves will have AUCs intermediate between these extremes, as shown on the right in figure 1, and AUC is a commonly used figure of merit for binary tasks.

In an imaging context, AUC depends on the classes of objects considered (C_0 and C_1), on the form of the test statistic $t(\mathbf{g})$, and on the statistical properties of the image vectors $\mathbf{g}$. For imaging with ionizing radiation, the statistics of the images, and hence the AUC, depend on radiation dose.

An alternative FOM for separability of $\text{pr}(t|H_1)$ and $\text{pr}(t|H_0)$ is the *signal-to-noise ratio* of the test statistic, denoted SNR_t . With the notation developed in section 2.2, we can express the means and variances of the test statistic under the two hypotheses as

$$\langle t \rangle_j \equiv \langle t(\mathbf{g}) \rangle_{\mathbf{g}|\mathbf{C}_j}, \quad \sigma_j^2 \equiv \langle [t(\mathbf{g}) - \langle t \rangle_j]^2 \rangle_{\mathbf{g}|\mathbf{C}_j}, \quad (j = 0, 1). \quad (25)$$

Then SNR_t is defined as

$$\text{SNR}_t = \frac{\langle t \rangle_1 - \langle t \rangle_0}{\sqrt{\frac{1}{2}\sigma_0^2 + \frac{1}{2}\sigma_1^2}}. \quad (26)$$

When $\text{pr}(t|H_1)$ and $\text{pr}(t|H_0)$ are both univariate normal PDFs, the area under the ROC curve is given by Barrett *et al* (1998) as

$$\text{AUC} = \frac{1}{2} + \frac{1}{2} \text{erf} \left(\frac{1}{2} \text{SNR}_t \right), \quad (27)$$

where $\text{erf}(\cdot)$ is the error function, defined as $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z dt \exp(-t^2)$.

Even when $\text{pr}(t|H_1)$ and $\text{pr}(t|H_0)$ are not normals, AUC can be used to compute an effective signal-to-noise ratio simply by inverting (27). The resulting figure of merit is denoted d_A and defined as (ICRU 1997)

$$d_A = 2 \text{erf}^{-1} [2 (\text{AUC}) - 1]. \quad (28)$$

This definition shows that d_A is a nonlinear but monotonic transformation of AUC, so these two figures of merit are equivalent.

3.2.3. Multiclass tasks. There has been some effort to extend the ROC paradigm to classification problems with three or more classes, but the results are at best unwieldy. With N classes, $N - 1$ scalar variables are needed to make a decision and the ROC curve becomes an $(N - 1)$ -dimensional surface.

It is not clear how to relate these generalized ROC concepts to Stage 2 efficacy, but as discussed below it is relatively straightforward in principle to use Stage 3 efficacy for any number of classes.

3.2.4. Stage 3 efficacy in classification tasks. The FOMs used for binary classification tasks—AUC, d_A and SNR_t —are properties of the complete ROC curve, so they are independent of the operating point t_c . In practice, however, the choice of operating point itself depends on the clinical task and on the consequences of the outcome. In screening mammography, for example, a clinician might choose a low t_c (high TPF and FPF) in order to avoid missing any true malignancies, especially if the consequence of a false positive is just that a second test will be performed to confirm the result from mammography. If, on the other hand, the consequence of a false positive is that the patient will undergo an unnecessary surgery, then a higher t_c (lower TPF and FPF) is more appropriate.

In order to account for the consequences of clinical decisions, we must use Stage 3 efficacy. The key feature that sets Stage 3 apart from Stage 2 is the introduction of some measure of the clinical usefulness of the diagnostic information. One way to do this is to introduce the concept of *utility*.

The utility of a classification outcome has been defined, somewhat subjectively, as a measure of the desirability of the outcome relative to other outcomes (Patton and Woolfenden

1989). The utility can also be viewed as the negative of the cost of the outcome, especially if ‘cost’ is interpreted more broadly than simply monetary cost. The methodology of Fryback and Thornbury is to elicit estimates of utility from questionnaires submitted to expert clinicians.

The overall diagnostic utility of an imaging test can be evaluated as an expectation of the utility of the test over the population of patients and outcomes. For classification tasks with N classes and imaging system S_k , the overall utility U_k is defined as

$$U_k = \sum_{i=1}^N \sum_{j=1}^N u_{ij} \Pr(\mathcal{D}_i | H_j, S_k) \Pr(H_j), \quad (29)$$

where u_{ij} is the utility assigned *a priori* to decision $\mathcal{D}_i$ when hypothesis H_j is true, and $\Pr(H_j)$ is the probability that H_j is true (i.e. the object being imaged came from class C_j) in the population under study. The conditional probability $\Pr(\mathcal{D}_i | H_j, S_k)$ is the generalization of TPF, FPF, etc to more than 2 classes; it depends on the decision rule and on the statistical properties of the image vectors $\mathbf{g}$, hence on the imaging system and the radiation dose. Note that the other factors in the summand of (29), u_{ij} and $\Pr(H_j)$, are independent of the imaging system and the radiation dose.

Wunderlich and Abbey (2013) have shown that utility can also be used for choosing the performance-assessment paradigm itself.

The concept of expected utility, and the assessment of diagnostic image quality by Stage 3 efficacy in general, has been criticized on the grounds that it is difficult to know what utilities to assign to different outcomes and what probabilities of disease to use for computing the expected utility (Metz 2006). The increasing availability of Big Data should mitigate the latter problem, and the dependence of overall utility on the choice of the component utilities u_{ij} can be regarded as an opportunity for further research rather than a problem. We can in principle study the variation of $\Pr(\mathcal{D}_i | H_j, S_k)$ with radiation dose for different true hypotheses and decision outcomes and determine not only the overall effect of radiation dose but also its effect on various possible decision outcomes; if the outcomes for which large values of $\Pr(\mathcal{D}_i | H_j, S_k)$ require large dose are associated with relatively small utilities, then the dose can be reduced with little effect on overall utility.

Moreover, expected utility can be used to compare different imaging systems or to study the effect of radiation dose with a given system just by selecting reasonable values for u_{ij} and $\Pr(H_j)$ and holding them constant across the comparisons. The robustness of the conclusions can then be checked by varying u_{ij} and $\Pr(H_j)$ after the fact. In many cases it will suffice if the rank order of the comparisons is maintained as utilities and probabilities are varied over reasonable ranges.

On the other hand, defensible utilities are essential if we want to evaluate image quality not only for Stage 3 efficacy but also for Stages 5 and 6, and in particular if we want to study the impact of radiation dose on long-term patient outcomes and on society.

3.2.5. Estimation tasks. In image-quality analysis, the goal of an estimation task is to use image data to estimate one or more numerical parameters *defined in terms of the object*. This goal is distinct from the goals in other areas of imaging, such as data mining, pattern recognition, texture analysis and computer-aided diagnosis (CAD), which often use parameters *of an image*.

To illustrate this distinction, there is considerable current interest in the heterogeneity of the uptake of fluorodeoxyglucose (FDG) in tumors, as observed in PET scans, because heterogeneity can have prognostic value for radionuclide therapy. Various metrics from the literature on texture analysis (see, for example, Tixier *et al* (2012)) have been proposed for this purpose,

but they describe the heterogeneity of the image, not the object. Noise in the image can introduce heterogeneity in the image where none exists in the object, and limited resolution of the imaging system can hide fine structures that are really present. The identification of image texture features with physical properties of the object, much less with biological significance, has been criticized by Brooks (2013). An approach to avoiding this problem and directly estimating object texture parameters is given in Kupinski *et al* (2003b).

A scalar parameter $\theta(\mathbf{f})$, which is defined in terms of the object $\mathbf{f}$, is said to be linear if it can be written as a scalar product. If $\mathbf{f}$ represents a time-independent spatial function $f(\mathbf{r})$, the scalar product has the form (see (23)):

$$\theta(\mathbf{f}) = \int_{\infty} d^3r \chi(\mathbf{r}) f(\mathbf{r}) \equiv \chi^\dagger \mathbf{f}, \quad (30)$$

where χ represents the spatial function $\chi(\mathbf{r})$ (strictly, as a vector in a Hilbert space), and $\chi^\dagger \mathbf{f}$ is a notation for the scalar product of functions. As an example, if $f(\mathbf{r})$ is a radionuclide distribution (activity per unit volume) and $\chi(\mathbf{r})$ is an indicator function equal to 1 for $\mathbf{r}$ inside the boundary of some region of interest (ROI) and 0 outside, then the integral in (30) is the total amount of activity in the ROI.

If we have a raw image data vector $\mathbf{g}$, we can use one of the estimation rules to be discussed in section 3.4 to form an estimate $\hat{\theta}(\mathbf{g})$. On the other hand, if the data vector is a reconstructed image $\hat{\mathbf{f}}$, then the estimate is denoted $\hat{\theta}(\hat{\mathbf{f}})$. For simplicity, we consider only $\hat{\theta}(\mathbf{g})$ here.

Because $\mathbf{g}$ is random, $\hat{\theta}(\mathbf{g})$ is also random. That is, if you repeatedly image the same object, hence keeping $\theta(\mathbf{f})$ constant, you will get a distribution of random images $\mathbf{g}$ and hence a distribution of values of $\hat{\theta}(\mathbf{g})$.

The full statistics of $\hat{\theta}(\mathbf{g})$ can be difficult to derive. For this reason, it is common practice to provide the bias and variance only. These quantities are defined, respectively, as

$$b(\mathbf{f}) \equiv \langle \hat{\theta}(\mathbf{g}) \rangle_{\mathbf{g}|\mathbf{f}} - \theta(\mathbf{f}) = \bar{\hat{\theta}}(\mathbf{f}) - \theta(\mathbf{f}), \quad \text{Var}\{\hat{\theta}|\mathbf{f}\} \equiv \left\langle [\hat{\theta}(\mathbf{g}) - \bar{\hat{\theta}}(\mathbf{f})]^2 \right\rangle_{\mathbf{g}|\mathbf{f}}. \quad (31)$$

Thus the bias is the average deviation of the estimate from the true $\theta(\mathbf{f})$, and the variance is the mean-square deviation of the estimate from the average of the estimates themselves. Bias is the systematic error and variance is the random error in the measurement of θ .

Bias and variance can be combined into a single FOM, the mean-square error (MSE), which is defined as

$$\text{MSE}(\mathbf{f}) \equiv \left\langle [\hat{\theta}(\mathbf{g}) - \theta(\mathbf{f})]^2 \right\rangle_{\mathbf{g}|\mathbf{f}} = [b(\mathbf{f})]^2 + \text{Var}\{\hat{\theta}|\mathbf{f}\}. \quad (32)$$

Thus MSE accounts for the total deviation (systematic and random) between the estimate and the true θ .

As the notation implies, bias, variance and MSE all depend on the object $\mathbf{f}$; it is not correct, however, to say that these metrics depend on the true value of the parameter. The problem is that a linear CD imaging operator $\mathcal{H}$ necessarily has null functions, for which $\mathcal{H}\mathbf{f}_{\text{null}} = 0$. Moreover, any object vector can be decomposed uniquely into a null component and an orthogonal measurement component, $\mathbf{f} = \mathbf{f}_{\text{meas}} + \mathbf{f}_{\text{null}}$. Objects that differ by a null function produce the same mean data $\bar{\mathbf{g}}$, and the same conditional PDF $\text{pr}(\mathbf{g}|\mathbf{f})$, so they must produce the same conditional mean estimate, $\bar{\hat{\theta}}(\mathbf{f})$, and the same conditional variance, $\text{Var}\{\hat{\theta}|\mathbf{f}\}$. Objects that differ by a null function do not, however correspond to the same true value of the parameter unless $\theta(\mathbf{f})$ is independent of the null functions. This condition can be stated as $\theta(\mathbf{f}) = \theta(\mathbf{f}_{\text{meas}})$. A parameter that satisfies this condition is said to be *estimable*, which means that there exists an unbiased estimate of it for all true values.

For linear parameters as defined in (30), we can state the estimability condition more directly by writing

$$\theta(\mathbf{f}) = (\chi, \mathbf{f}) = (\chi_{\text{meas}}, \mathbf{f}_{\text{meas}}) + (\chi_{\text{null}}, \mathbf{f}_{\text{null}}). \quad (33)$$

There are no cross terms $(\chi_{\text{meas}}, \mathbf{f}_{\text{null}})$ or $(\chi_{\text{null}}, \mathbf{f}_{\text{meas}})$ because null space and measurement space are orthogonal subspaces of object space (Barrett and Myers 2004). Thus, for a general object, a linear parameter is estimable only if the function that defines it, $\chi(\mathbf{r})$, contains no null functions, as determined, for example, by an SVD analysis. In particular, the integral of an object over a voxel is virtually never estimable. Intuitively, the sharp edges of the voxel function contain high spatial frequencies that are not captured by the imaging system; for more detail, see Whitaker *et al* (2008) and Kupinski *et al* (2013).

One way to account for null functions in the figure of merit for a scalar estimation task is to average the MSE over some class of objects, called an *ensemble* in this context. The ensemble mean-square error (EMSE) for object class $\mathbf{C}$ is defined as

$$\text{EMSE} \equiv \langle \text{MSE}(\mathbf{f}) \rangle_{\mathbf{f}|\mathbf{C}} = \langle [b(\mathbf{f})]^2 + \text{Var}\{\hat{\theta}|\mathbf{f}\} \rangle_{\mathbf{f}|\mathbf{C}} = \left\langle \left\langle [\hat{\theta}(\mathbf{g}) - \theta(\mathbf{f})]^2 \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}}. \quad (34)$$

3.2.6. Vector parameters. The discussion above on estimation of a single scalar parameter is readily extended to a set of P parameters. The parameters $\{\theta_p, p = 1, \dots, P\}$ can be organized into a $P \times 1$ parameter vector $\boldsymbol{\theta}$, and the estimates can be similarly organized as a $P \times 1$ vector $\hat{\boldsymbol{\theta}}$. Likewise the biases and variances are components of $P \times 1$ vectors. A vector parameter $\boldsymbol{\theta}(\mathbf{f})$ is estimable if and only if every component is estimable: $\theta_p(\mathbf{f}) = \theta_p(\mathbf{f}_{\text{meas}})$.

For P parameters, the EMSE is given by

$$\text{EMSE} = \sum_{p=1}^P \left\langle [b_p(\mathbf{f})]^2 + \text{Var}\{\hat{\theta}_p|\mathbf{f}\} \right\rangle_{\mathbf{f}|\mathbf{C}} = \left\langle \|\hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta}(\mathbf{f})\|^2 \right\rangle_{\mathbf{g}|\mathbf{f}} \Big|_{\mathbf{f}|\mathbf{C}}, \quad (35)$$

where $\|\cdot\|$ denotes the norm of the vector.

3.2.7. Efficacy in estimation tasks. If the parameter values obtained in an estimation task are used to make diagnostic decisions, then Stage 2 efficacy (diagnostic accuracy) becomes relevant, and if we want to discuss the diagnostic impact of an estimation task, thereby moving on to Stage 3 efficacy, we must assign utilities to the outcomes. For classification tasks, we defined u_{ij} as the utility of making decision $\mathcal{D}_i$ when hypothesis H_j is true. A natural analog of this definition for a vector estimation task is the utility of obtaining estimate $\hat{\boldsymbol{\theta}}$ when the true parameter is $\boldsymbol{\theta}$, denoted $u(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta})$.

It is common in the estimation literature to use a quadratic cost function $c(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$. Since a cost can be viewed as a constant minus a utility, the negative of the EMSE in (35) is a possible expected utility associated with this quadratic cost.

EMSE is not directly related to diagnosis or Stage 3 efficacy, in part because it treats all components of the vector $\boldsymbol{\theta}$ as equally important to the figure of merit. Some of the components may be *nuisance parameters*, defined as ones that influence the image data $\mathbf{g}$ but are of no direct clinical interest. It may be necessary to estimate the nuisance parameters in order to get good estimates of the parameters of interest, but they should be omitted from the EMSE if diagnostic relevance is to be ascribed to that metric. Conversely, high utilities should be assigned to accurate estimates of clinically relevant parameters.

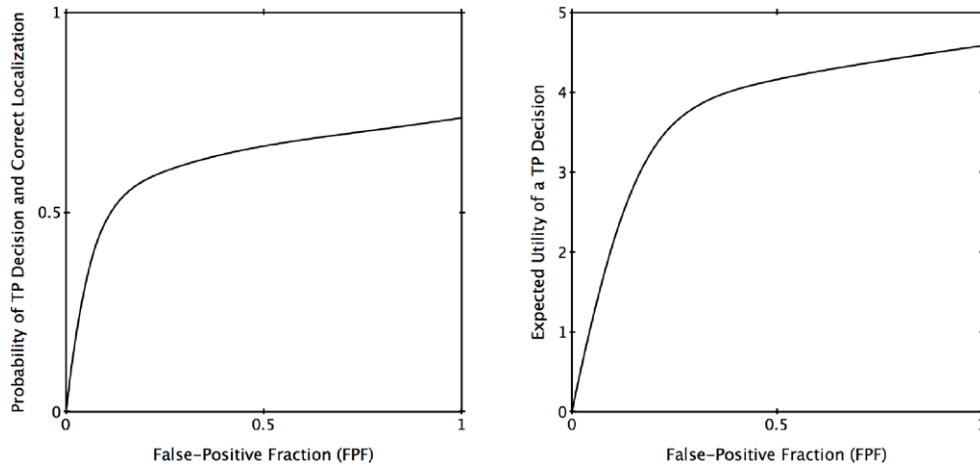


Figure 2. Left: schematic localization ROC (LROC) curve. Right: schematic estimation ROC (EROC) curve.

One way to bring these judgments about clinical impact into the estimation figure of merit is to define a weighted EMSE by

$$\text{WEMSE} = \sum_{p=1}^P w_p \left\langle [b_p(\mathbf{f})]^2 + \text{Var} \{ \hat{\theta}_p | \mathbf{f} \} \right\rangle_{\mathbf{f} | \mathbf{C}}, \quad (36)$$

where all of the weights w_p are nonnegative. A component considered to be a nuisance parameter is assigned the weight $w_p = 0$, and components of high clinical relevance are assigned high weights. As a side benefit, if different components of $\theta(\mathbf{f})$ have different units, the weights can be used to normalize each component by either the variance or the square of the mean of that component, thereby making WEMSE dimensionless.

Even with these modifications, however, there can be objections to relating EMSE to efficacy. There is no reason other than mathematical convenience to assume that the cost increases quadratically with the discrepancy between an estimate of a parameter and its true value. We know that the bias term in MSE or EMSE is influenced by null functions of the imaging system, and this term can be arbitrarily large because a null function remains a null function when it is multiplied by a constant. Rather than accepting a large MSE as indicative of low clinical utility, it is better to construct utility functions more directly related to the clinical tasks.

Diagnosis is ultimately about classification. If one or more of the estimated parameters is used for classification, then the expected utility as defined in (29) is a relevant FOM for an estimation task.

3.2.8. Joint detection and estimation. In some clinical scenarios, one wants to detect a tumor or other lesion and, if it is present, estimate some numerical parameters such as the size or location of the lesion. These situations lead to several new kinds of ROC curve, including the localization ROC (LROC) and estimation ROC (EROC), both of which are illustrated in figure 2.

The LROC curve plots the joint probability of detection and correct localization (within some tolerance) versus the false-positive fraction as the decision threshold for detection is varied. The EROC curve (Clarkson 2007), a generalization of LROC, plots the expected utility

of a true positive decision against the false-positive fraction, again as the decision threshold is varied. An LROC curve is an EROC where the parameter being estimated is the lesion location and the utility function is unity if the estimate is within the set tolerance of the true location, and zero otherwise.

3.2.9. Phantoms for search tasks. Popescu and Myers (2013) recently published a paper describing a signal-search paradigm, along with a phantom design, that enables the evaluation and comparison of iterative reconstruction algorithms in terms of signal detectability versus dose performance using a few tens of images. Standard commercial phantoms for CT quality assessment have limited utility for quantitative evaluation of image quality, because of the limited area available for deriving the statistics of a model observer in the signal-absent condition. Moreover, such phantoms yield only one signal size/contrast combination per image. Phantoms that allow for multiple realizations of background-only and signal-present regions in each image, along with the addition of search as a component of the detection task, enable the investigator to customize the signal detectability level through adjustment of the search region area so that meaningful image quality comparisons can be made across algorithms or imaging systems with a fairly small number of images.

3.2.10. Tasks related to radiation therapy and Stage 4 efficacy. Imaging with ionizing radiation is often used in the planning and monitoring of radiation therapy. For a discussion of anatomical imaging for radiotherapy applications, see Evans (2008), and for a discussion of the role of PET in radiotherapy, see Nestle *et al* (2009).

A recent paper in this journal (Barrett *et al* 2013) discussed how radiobiological models could be used to assess image quality in terms of therapeutic efficacy, defined by the tradeoff between the probability of tumor control and the probability of some critical adverse reaction in normal tissues. In essence, this paper defines image quality in terms of Stage 4 efficacy, and if we assign utilities, the theory is applicable to Stage 5.

3.3. Observers for classification and detection tasks

3.3.1. Ideal observer for binary classification tasks. The ideal observer (IO) for a binary classification task can be defined in several equivalent ways: It maximizes the TPF for any given FPF; it maximizes the area under the ROC curve (AUC); it maximizes the detectability index d_A , and it maximizes the overall utility. A well-known result from statistical decision theory is that the ideal observer implements the generic decision strategy (22) with the test statistic $t(\mathbf{g})$ being the *likelihood ratio* (LR), defined as

$$\Lambda(\mathbf{g}) \equiv \frac{\text{pr}(\mathbf{g}|H_1)}{\text{pr}(\mathbf{g}|H_0)}, \quad (37)$$

where $\text{pr}(\mathbf{g}|H_j)$ is the multivariate PDF for the data given that hypothesis H_j ($j = 0, 1$) is true.

An important, though little recognized, point is that the LR is invariant to any invertible transformation of the data (see Barrett and Myers (2004), pp 829–30). For example, if we base the classification decision on a reconstructed image $\hat{\mathbf{f}}$ instead of raw projection data $\mathbf{g}$, and it is possible to go backwards and compute $\mathbf{g}$ from $\hat{\mathbf{f}}$, then $\Lambda(\hat{\mathbf{f}}) = \Lambda(\mathbf{g})$, and hence the ROC and any FOM derived from it will be the same for the two data sets. If the reconstruction algorithm is not invertible in this sense, then the IO performance on the reconstruction will be less than or equal to its performance on the projection data; if this were not the case, the ideal observer, being ideal, would include the reconstruction algorithm in her decision process when given $\mathbf{g}$.

An important example of an invertible transformation is the logarithm. The logarithm of the likelihood ratio (LLR),

$$\lambda(\mathbf{g}) \equiv \ln \Lambda(\mathbf{g}) = \ln \text{pr}(\mathbf{g}|H_1) - \ln \text{pr}(\mathbf{g}|H_0), \quad (38)$$

produces exactly the same ROC curve as the LR itself, hence the same value for any figure of merit derived from the ROC curve. Thus LR and LLR are equivalent test statistics for the ideal observer.

Analytical forms for the LR and LLR are available for several different families of data PDFs, $\text{pr}(\mathbf{g}|H_j)$, including multivariate normal (Gaussian), log-normal, exponential and Poisson distributions (see sections 13.2.8 and 13.2.9 in Barrett and Myers (2004)).

Here we focus on multivariate normal data, for which the PDF is given by

$$\text{pr}(\mathbf{g}|H_j) = \left[(2\pi)^M \det(\mathbf{K}_j) \right]^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}_j)^T \mathbf{K}_j^{-1} (\mathbf{g} - \bar{\mathbf{g}}_j) \right], \quad (39)$$

where $\det(\cdot)$ denotes determinant, and $\bar{\mathbf{g}}_j$ and $\mathbf{K}_j$, ($j = 0, 1$) are shorthand forms for the mean data vector and covariance matrix, respectively, when the object class is C_j . We remind the reader that the means and covariances include averages over both the conditional noise for a given object, described by the PDF $\text{pr}(\mathbf{g}|\mathbf{f})$, and the randomness of the object itself; this double averaging is indicated by the double overbar on $\bar{\mathbf{g}}_j$.

With (38) and (39), we see that the LLR is given by

$$\lambda(\mathbf{g}) = -\frac{1}{2} \ln \left[\frac{\det(\mathbf{K}_1)}{\det(\mathbf{K}_0)} \right] - \frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}_1)^T \mathbf{K}_1^{-1} (\mathbf{g} - \bar{\mathbf{g}}_1) + \frac{1}{2} (\mathbf{g} - \bar{\mathbf{g}}_0)^T \mathbf{K}_0^{-1} (\mathbf{g} - \bar{\mathbf{g}}_0). \quad (40)$$

Thus the test statistic for a binary classification task with multivariate-normal data is, in general, a quadratic function of the data vector, referred to as a *quadratic discriminant*.

In x-ray and gamma-ray imaging, however, it is frequently an excellent approximation to say that the two covariance matrices, $\mathbf{K}_0$ and $\mathbf{K}_1$, are nearly equal. In transmission x-ray imaging, the signal to be detected results from a localized change in the x-ray attenuation coefficient, but it is observed by its effect on a line integral of the attenuation coefficient through the patient's body. Similarly, in radionuclide imaging a localized change in the activity per unit volume of the tracer is observed by its effect on a line integral through all of the background activity within the body. In neither case is a significant change in the value of a single line-integral projection to be expected; a large change in one projection would be too easily detected from the full set of projections and would be of little use in comparing imaging systems or studying the effect of increased radiation dose. Thus we are interested in detecting small signals in this sense, and the signal is in the mean data vector, so we can usually neglect its effect on the covariances.

If we assume $\mathbf{K}_0 = \mathbf{K}_1 \equiv \mathbf{K}_g$, and take advantage of the fact that covariance matrices are symmetric, then (40) becomes

$$\lambda(\mathbf{g}) = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \mathbf{g} - \frac{1}{2} \bar{\mathbf{g}}_1^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_1 + \frac{1}{2} \bar{\mathbf{g}}_0^T \mathbf{K}_g^{-1} \bar{\mathbf{g}}_0, \quad \text{where } \Delta \bar{\mathbf{g}} \equiv \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0. \quad (41)$$

The last two terms in (41) are constants, independent of the data, so they can be lumped into the decision threshold. Thus the ideal observer for a binary classification task with multivariate normal data and equal covariance matrices is a linear discriminant, with test statistic given by

$$t_{10}(\mathbf{g}) = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \mathbf{g}. \quad (42)$$

Because a covariance matrix is symmetric and positive definite, we can define its square root $\mathbf{K}_g^{1/2}$ and then rewrite (42) as

$$t_{IO}(\mathbf{g}) = [\mathbf{K}_g^{-1/2} \Delta \bar{\mathbf{g}}]^T [\mathbf{K}_g^{-1/2} \mathbf{g}]. \quad (43)$$

The $M \times M$ matrix $\mathbf{K}_g^{-1/2}$ is called a *prewhitening* operator because the transformed data, $\mathbf{K}_g^{-1/2} \mathbf{g}$, exhibit uncorrelated or ‘white’ noise. More precisely, the covariance matrix of $\mathbf{K}_g^{-1/2} \mathbf{g}$ is the $M \times M$ unit matrix. Thus the IO test statistic for the present problem is just the scalar product of the prewhitened data and the prewhitened mean difference signal.

Because any linear transformation of a normal random vector or random variable is also normal, it follows that $\text{pr}(t_{IO}|H_1)$ and $\text{pr}(t_{IO}|H_0)$ are both univariate normal PDFs. Thus AUC_{IO} is given by (27), with the SNR as defined in (26); it can be shown that

$$\text{SNR}_{IO}^2 = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \Delta \bar{\mathbf{g}}. \quad (44)$$

Note that calculation of both the test statistic in (43) and its SNR in (44) requires prior knowledge of the mean data vector for both classes and their common covariance matrix.

Some alternative ways of writing this SNR are:

$$\text{SNR}_{IO}^2 = [\mathbf{K}_g^{-1/2} \Delta \bar{\mathbf{g}}]^T [\mathbf{K}_g^{-1/2} \Delta \bar{\mathbf{g}}] = \|\mathbf{K}_g^{-1/2} \Delta \bar{\mathbf{g}}\|^2 = \text{tr} \{ \mathbf{K}_g^{-1} \Delta \bar{\mathbf{g}} \Delta \bar{\mathbf{g}}^T \}, \quad (45)$$

where $\text{tr}(\cdot)$ denotes the trace (sum of the diagonal elements) of a matrix.

3.3.2. The Hotelling observer. In 1931, the American statistician Harold Hotelling devised a multivariate generalization of Student’s t test for determining the separability of two groups of N -dimensional random vectors (Hotelling 1931). In essence, Hotelling’s separability measure, which he called T^2 , was a linear discriminant function (Fisher’s linear discriminant (Fisher 1936), in fact, though Hotelling’s work was five years earlier) in the MD space.

The concept we now refer to as the *Hotelling observer* was introduced into the image-quality literature by Smith and Barrett (1986) and Fiete *et al* (1987). Because T^2 was in the form of a trace of a matrix, Smith and Fiete referred to this proposed FOM for classification tasks as the Hotelling trace criterion. A key difference, however, is that T^2 is a random variable (a test statistic) but the Hotelling trace, as Smith and Fiete used the term, is an ensemble average separability (a figure of merit).

Current terminology in the image-quality literature is that the Hotelling observer is a linear discriminant in the form of a scalar product, $t(\mathbf{g}) = \mathbf{w}^T \mathbf{g} = \sum_{m=1}^M w_m g_m$, where $\mathbf{w}$ is a vector of the same dimension as the data $\mathbf{g}$. We refer to $\mathbf{w}$ as a *template* because the scalar product can be visualized by thinking of $\mathbf{g}$ as a transparency on a light box and $\mathbf{w}$ as a second transparency laid over the first one. The total light that emerges, in this metaphor, is proportional to the product of the two transmittances, integrated over position on the light box.

Specifically, the Hotelling observer uses the linear test statistic that maximizes SNR^2 . It can be shown by a straightforward calculation (Barrett and Myers 2004) that the Hotelling test statistic is given by

$$t_{\text{Hot}}(\mathbf{g}) = \mathbf{w}_{\text{Hot}}^T \mathbf{g} = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1} \mathbf{g}, \quad (46)$$

where $\mathbf{w}_{\text{Hot}}^T = \Delta \bar{\mathbf{g}}^T \mathbf{K}_g^{-1}$, so that $\mathbf{w}_{\text{Hot}} = \mathbf{K}_g^{-1} \Delta \bar{\mathbf{g}}$ (because the inverse of a covariance matrix is symmetric), and now

$$\mathbf{K}_g \equiv \frac{1}{2} \mathbf{K}_0 + \frac{1}{2} \mathbf{K}_1. \quad (47)$$

The corresponding SNR^2 is found to be

$$\text{SNR}_{\text{Hot}}^2 = \Delta \bar{\mathbf{g}}^t \mathbf{K}_{\mathbf{g}}^{-1} \Delta \bar{\mathbf{g}}. \quad (48)$$

Comparing these results to (42) and (44), respectively, we see that the Hotelling observer is also the ideal observer, in the sense of maximizing AUC, when the data PDFs are multivariate normal with equal covariances under the two hypotheses. Even when these conditions are not satisfied, however, the Hotelling observer is still optimal among all linear observers in the sense of maximizing SNR^2 .

On the other hand, no linear observer performs well on a signal detection task when the signal is at a random location in the image. Recall that the double overbar on $\Delta \bar{\mathbf{g}}$ indicates averages over both image noise (Poisson or electronic) and object randomness, and Δ indicates the difference in these averages between the signal-absent and signal-present classes. If the background (non-signal) part of the data has the same statistics in the two classes, then $\Delta \bar{\mathbf{g}}$ reflects only the randomness of the signal. If the signal can appear randomly anywhere in the image, the components of the Hotelling template vector, $[\mathbf{w}_{\text{Hot}}]_m$, will have very little variation with the pixel index m , and the test statistic $t_{\text{Hot}}(\mathbf{g})$ is little more than a sum over the components of $\mathbf{g}$, which conveys very little information about the presence or absence of the signal.

For this reason, the use of the Hotelling observer, as we have described it so far, is restricted to situations where the location of a potential signal is known, and the task is to determine whether or not it is present at that location. If we further restrict the signal to be completely nonrandom, then the task is said to be SKE (signal known exactly). These restrictions will be lifted below when we discuss *scanning* observers that can perform well even when the signal location and other parameters are random variables. The converse of SKE is SKS (signal known statistically).

We might also consider the case of BKE (background known exactly), which for example would describe the task of detecting a lesion in a water tank or other spatially uniform phantom. The much more realistic case is BKS (background known statistically), where there can be organs of unknown size, shape, location and radiographic appearance in the phantom or object being imaged, and there can be unknown fine structure within the organs. The SKE/BKS combination is often useful in image-quality studies, but SKE/BKE should be avoided because the BKE assumption eliminates the object-variability term in the covariance matrix. In particular, as we shall see in section 6, the noise and object-covariance terms depend differently on radiation dose, so conclusions about the effect of dose in a BKE task are not applicable in the real world of random, unknown anatomical backgrounds.

An important early example of a mathematically tractable random background is the *lumpy background* developed by Rolland and Barrett (1992) and used by Myers *et al* (1990) to study the optimization of pinhole apertures in nuclear medicine. A later variant called the *clustered lumpy background* provides an excellent model for the random background inhomogeneity seen in mammography (Bochud *et al* 1998).

3.3.3. Approximating the covariance matrix and performing its inverse. An immediate difficulty with implementing the Hotelling observer is that the covariance matrix is enormous. As a simple example, a rotating-camera SPECT system with 64 2D projection images, each consisting of a grid of 128×128 pixels, acquires a data vector with approximately 10^6 elements, so the covariance matrix of the data is $10^6 \times 10^6$, or 10^{12} elements. Modern CT systems can deliver even larger data vectors and covariance matrices; a typical single-organ volumetric scan might consist of 2000 projection images, each 2500×320 pixels.

The daunting prospect of computing and storing such large matrices, much less inverting them, has led to a malady termed megalopinakophobia (fear of large matrices) (Barrett *et al* 2001). The key to the treatment of this syndrome is the covariance decomposition of (16)

or (21), where the covariance matrix is split into two terms, one arising from Poisson and other noise sources, and one arising because the objects being imaged are unknown and unpredictable, hence best regarded as random processes.

Many studies of task-based image quality these days are performed with simulated objects. Simulations can be very realistic and flexible, and they avoid all issues related to knowing the true class of the object or the true values of parameters being estimated. Most importantly for the present paper, they permit arbitrary variation of the radiation dose delivered to the ‘patient’ and hence allow control of the relative importance of the two terms in the covariance decomposition. In particular, we can simulate images with no noise, and with modern computers we can simulate as many of them as we wish. Therefore we can generate image data or reconstructions related just to the second term in the decomposition, with no noise term. Analytic models such as (9) then permit construction of the first term at any desired level of radiation dose.

If we have simulated a set of N_s noise-free sample images, and we denote the k th such image as $\mathbf{g}_k$, then we can form the $M \times N_s$ matrix $\mathbf{R}$ given by Smith and Barrett (1986) and Fiete *et al* (1987),

$$\mathbf{R} = \frac{1}{\sqrt{N_s - 1}} [\delta \mathbf{g}_1, \delta \mathbf{g}_2, \dots, \delta \mathbf{g}_{N_s}], \quad (49)$$

where

$$\delta \mathbf{g}_k \equiv \mathbf{g}_k - \frac{1}{N_s} \sum_{j=1}^{N_s} \mathbf{g}_j. \quad (50)$$

Thus each column of $\mathbf{R}$ is $1 / \sqrt{N_s - 1}$ times the difference between a sample image and the arithmetic average of all of the sample images.

With these definitions, we can estimate the object covariance by a sample covariance matrix of the form,

$$\hat{\mathbf{K}}_{\mathbf{g}}^{\text{obj}} = \mathbf{R} \mathbf{R}^t, \quad (51)$$

where the hat indicates an estimate. The overall covariance matrix can then be estimated via

$$\hat{\mathbf{K}}_{\mathbf{g}} = \bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}} + \mathbf{R} \mathbf{R}^t. \quad (52)$$

The noise covariance matrix $\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}$ will be diagonal if we use noise models (8) or (9) but it will still depend on the grand mean of the data, which for a linear system is $\bar{\mathbf{g}} = \mathcal{H} \bar{\mathbf{f}}$, where $\bar{\mathbf{f}}$ can be estimated as the average of the noise-free images used to construct $\mathbf{R}$. Logically, therefore, we should include a hat on $\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}$, but in practice the error in the estimate of this term is negligible because we are making use of the analytic noise models and we can treat $\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}$ as a known, full-rank, diagonal matrix.

An exact inverse of the estimated overall covariance matrix $\hat{\mathbf{K}}_{\mathbf{g}}$ can be obtained without onerous calculations or impossibly large matrix inversions by use of the Woodbury matrix inversion lemma (Tylavsky and Sohie 1986, Barrett and Myers 2004). From (52), the lemma allows us to write

$$[\hat{\mathbf{K}}_{\mathbf{g}}]^{-1} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} - [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} \mathbf{R} \left\{ \mathbf{I} + \mathbf{R}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} \mathbf{R} \right\}^{-1} \mathbf{R}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1}, \quad (53)$$

where $\mathbf{I}$ is the $N_s \times N_s$ identity matrix.

The advantage of this form is that $\mathbf{I} + \mathbf{R}^t [\mathbf{K}_g^{\text{noise}}]^{-1} \mathbf{R}$ is an $N_s \times N_s$ matrix, readily inverted numerically for N_s up to 10^5 or so. The matrix $\mathbf{K}_g^{\text{noise}}$ is still $M \times M$, but it is full rank and in fact diagonal if we work with the noise model of (9). Hence all of the inverses indicated in (53), including $[\hat{\mathbf{K}}_g]^{-1}$ itself, will exist.

3.3.4. Estimating the Hotelling template. An alternative to inverting $\hat{\mathbf{K}}_g$ is to use it in an iterative algorithm to find the Hotelling template. A simple manipulation of (46) shows that this template satisfies $\mathbf{K}_g \mathbf{w}_{\text{Hot}} = \Delta \bar{\mathbf{g}}$, so an estimate of the template can be found by solving (see (52))

$$[\bar{\mathbf{K}}_g^{\text{noise}} + \mathbf{R}\mathbf{R}^t] \hat{\mathbf{w}}_{\text{Hot}} = \Delta \bar{\mathbf{g}}. \quad (54)$$

A variant of the iterative Landweber algorithm for solving this equation is given in Barrett and Myers (2004); the iteration rule is

$$\hat{\mathbf{w}}_{\text{Hot}}^{(n+1)} = \hat{\mathbf{w}}_{\text{Hot}}^{(n)} + \alpha [\bar{\mathbf{K}}_g^{\text{noise}}]^{-1} [\Delta \bar{\mathbf{g}} - \hat{\mathbf{K}}_g \hat{\mathbf{w}}_{\text{Hot}}^{(n)}], \quad (55)$$

where $\hat{\mathbf{w}}_{\text{Hot}}^{(n)}$ is the estimated template at the n th iteration and α is a positive constant that controls the convergence rate.

Once the template has been estimated, the SNR can be found by applying the template to a set of sample images, determining the mean and variance of the resulting scalar test statistic under each hypothesis, and estimating the observer's performance via (26). Alternatively, we can directly estimate $\text{SNR}_{\text{Hot}}^2$ by (48) as $\Delta \bar{\mathbf{g}}^t \hat{\mathbf{w}}_{\text{Hot}}$. For a discussion of the errors in the estimate, see Kupinski *et al* (2007).

3.3.5. Channelized Hotelling observers. A channel vector is a relatively low-dimensional vector used in place of the original large data vector to perform the desired detection and/or estimation task. In almost all of the copious literature on this topic (which began with Myers and Barrett (1987)), the j th component of the channel vector is the scalar product of a channel template vector with a data vector of the same dimensionality:

$$v_j \equiv \mathbf{u}_j^t \mathbf{g} = \sum_m u_{jm} g_m, \quad (j = 1, \dots, N_c), \quad (56)$$

where N_c is the number of channels and u_{jm} is the m th component of template vector $\mathbf{u}_j$. Equation (56) can also be written in matrix-vector form as

$$\mathbf{v} \equiv \mathbf{U}^t \mathbf{g}, \quad (57)$$

where the j th column of the $M \times N_c$ matrix $\mathbf{U}$ is the template vector for the j th channel.

By analogy to (46), the weak-signal form of the Hotelling discriminant for operating on channelized data is

$$t_{\text{CHO}}(\mathbf{v}) = \Delta \bar{\mathbf{v}}^t \mathbf{K}_v^{-1} \mathbf{v}, \quad (58)$$

where

$$\Delta \bar{\mathbf{v}} \equiv \mathbf{U}^t [\bar{\mathbf{g}}_{H_1} - \bar{\mathbf{g}}_{H_0}], \quad (59)$$

$$\mathbf{K}_v = \mathbf{U}^t [\bar{\mathbf{K}}_g^{\text{noise}} + \mathbf{K}_g^{\text{obj}}] \mathbf{U}. \quad (60)$$

Since $\mathbf{K}_v$ is an $N_c \times N_c$ matrix, computation of the CHO test statistic is much less onerous than for the full Hotelling observer.

The signal-to-noise ratio for the channelized Hotelling observer is given by (see (48))

$$\text{SNR}_{\text{CHO}}^2 = \Delta \bar{\mathbf{v}}' \mathbf{K}_{\bar{\mathbf{v}}}^{-1} \Delta \bar{\mathbf{v}}. \quad (61)$$

One possible objective in designing a set of channels is to make $\text{SNR}_{\text{CHO}}^2$ as close to $\text{SNR}_{\text{Hot}}^2$ as possible. Channels that achieve this goal to a good approximation are said to be *efficient*, meaning that there is little loss of information (defined by the performance of a Hotelling observer) as a result of the dimensionality reduction inherent in projecting the full data vector onto the channels. The condition for approximate efficiency in this sense is

$$[\mathbf{U}' \Delta \bar{\mathbf{g}}]' [\mathbf{U}' \mathbf{K}_{\mathbf{g}} \mathbf{U}]^{-1} \mathbf{U}' \Delta \bar{\mathbf{g}} \approx [\Delta \bar{\mathbf{g}}]' \mathbf{K}_{\bar{\mathbf{g}}}^{-1} \Delta \bar{\mathbf{g}}, \quad (62)$$

for all signals of interest. Channel vectors that satisfy this condition can be regarded as approximate sufficient statistics for the detection problem.

One way to construct efficient channels is to take advantage of the symmetries of the signal and noise in the images under study. For example, if we are dealing with 2D images and we know that the signal is rotationally symmetric and that the noise covariance is isotropic (independent of direction in the plane), then we can expand the channel template in rotationally symmetric functions such as Laguerre–Gauss polynomials (Barrett *et al* 1998b). For a detailed study of Laguerre–Gauss (LG) channels and the errors in estimating the CHO SNR from sample images, see Gallas and Barrett (2003).

3.3.6. Human observers and anthropomorphic models. Much of the early development of statistical decision theory was aimed at analyzing the results of psychophysical studies with human observers (Tanner *et al* 1954, Peterson *et al* 1954, Swets 1964, Green and Swets 1966, Chesters 1992), and some of the early papers on the Hotelling observer showed that $\text{SNR}_{\text{Hot}}^2$ was a good predictor of human performance on SKE/BKS tasks with uncorrelated noise (Fiete *et al* 1987, Rolland and Barrett 1992). The work of Myers on channels (Myers and Barrett 1987), however, was motivated by the observation that the Hotelling observer was *not* a good predictor of human performance when the noise was correlated as a result of the high-pass filter required in tomographic reconstruction. At the time, available psychophysical data on perception in tomographic images (e.g. Judy and Swensson 1981, Burgess 1984, Myers *et al* 1985) were generally summarized by saying that humans were unable to prewhiten the noise. Myers, however, was able to build on many previous experiments in human vision (e.g. Campbell and Robson 1968, Sachs *et al* 1971, Graham and Nachmias 1971, Strohmeyer and Julesz 1972, Mostafavi and Sakrison 1976) which showed that humans process visual information through spatial-frequency-selective channels. Typically, the channels had about one octave of bandwidth in the magnitude of the spatial-frequency vector and $\sim \pm 45^\circ$ selectivity in the orientation of the spatial frequency vector. By incorporating such channels in a channelized Hotelling model, Myers was able to account for the extant data on human detection performance in tomographic images.

Many subsequent studies (Barrett *et al* 1992, Barrett *et al* 1993, Abbey and Barrett 2001) confirmed the broad applicability of the CHO as a predictor of human task performance and refined the model by including an additional source of noise internal to the visual system; no adjustable parameters were needed in the resultant model (Burgess and Colborne 1988, Eckstein *et al* 1997, Abbey and Barrett 2001).

Of particular importance to the present paper, Abbey and Barrett (2001) showed that the CHO correctly predicted the minimum contrast required for a human observer to detect a weak signal as a function of radiation dose, over a range of several decades in dose.

This same paper (Abbey and Barrett 2001) showed that the CHO performance and its correlation with human performance were very insensitive to the spatial-frequency profiles of the

channels, so long as all channels had a bandpass structure with zero response at zero spatial frequency. Interestingly, the response of Laguerre–Gauss channels is maximum, not zero, at zero frequency, so LG channels should *not* be used as models for human observers.

For a comprehensive review of research through 2000 on visual detection tasks in correlated image noise with linear model observers, see Abbey and Bochud (2000), and for a practical guide to model observers for visual detection in synthetic and natural noisy images, see Eckstein *et al* (2000). A more recent review of these topics is given in Burgess (2011).

Yu *et al* (2013) investigated how well a channelized Hotelling observer predicts human observer performance to validate their use for the evaluation of CT reconstruction algorithms. Their work made use of a 2-alternative forced choice (2AFC) lesion-detection paradigm. They considered two different reconstruction algorithms: a filtered-backprojection (FBP) and an iterative reconstruction (IR) method. Images of a torso-shaped phantom containing three rods of different diameters that simulated small, medium and large lesions were reconstructed. Because the goal of the work was to show the predictive ability of the model observer as a surrogate for human observers, the observer model incorporated Gabor channels intended to model the human visual system. Sixty channels in all were utilized, including six frequency-selective passbands, five orientations, and two phases. With this model the investigators found that the performance of the human and model observers were highly correlated. Investigations of this type provide strong evidence that the channelized Hotelling observer with anthropomorphic channels can be used to predict human performance for the evaluation of CT image reconstruction algorithms.

The use of a CHO for the assessment of volumetric imaging data sets is presented in Platisa *et al* (2011).

3.3.7. Ideal observers for detection in non-Gaussian random backgrounds. Linear observers are computationally convenient, and in section 6.3 they will prove to be useful vehicles for the study of task performance as a function of radiation dose. If the background statistics are not Gaussian, however, the performance of the ideal observer can significantly exceed that of the ideal linear observer, especially at high dose where object variability can dominate over measurement noise.

One fruitful approach to ideal-observer computation in medical imaging is the use of Markov-chain Monte Carlo (MCMC) techniques (Kupinski *et al* 2003c, Barrett and Myers 2004, He *et al* 2008, Park and Clarkson 2009), in which one generates a sequence of likelihood ratios drawn from the posterior density of the background given a measured image $\mathbf{g}$ or $\hat{\mathbf{f}}$. The average of this sequence is then an estimate of the LR for that image, and repeating the estimation for many images leads to an ROC curve for the ideal observer. This method can then be used for optimizing imaging hardware (Gross *et al* 2003) or for determining the efficiency of a CHO (Park and Clarkson 2009) or a human observer (Park *et al* 2005) relative to the ideal observer.

Another approach is to use surrogate figures of merit based on Fisher information (see section 3.4), which accurately approximate the performance of the ideal observer (Shen and Clarkson 2006, Clarkson and Shen 2010, Clarkson 2012).

3.3.8. Stationary noise and Fourier methods for detection tasks. Much of the early research on task-based assessment of image quality in radiology used continuous Fourier transforms to define quantities such as TF (transfer function), MTF (modulation transfer function), NPS (noise power spectrum), NEQ (noise equivalent quanta) and DQE (detective quantum efficiency), all of which are functions of the spatial frequency vector $\boldsymbol{\rho}$. One of the first papers to make use of these quantities was the treatment of image quality in CT presented by Hanson

(1979). In that paper Hanson lays out the various contributors to image quality in CT, including the noise power spectrum, the MTF, and the NEQ. Further, Hanson presents signal detection theory and the concept of an optimum receiver that makes use of knowledge of the noise correlations in the image as part of its test statistic, first in tutorial/general form and then as it applies to the particular case of CT, including the concept of dose efficiency in the comparison of systems. For a history of the use of Fourier methods in the evaluation of image quality, see Barrett (2009), and for a detailed critique of its applicability to task-based assessment, see section 16.1.6 of Barrett and Myers (2004).

Continuous Fourier transforms are useful when the image is a function of continuous spatial variables instead of discrete pixel indices; when the imaging system is linear and shift invariant, so that the operator $\mathcal{H}$ can be modeled as a convolution; and when both the noise and object randomness are statistically stationary continuous random processes, so that the autocovariance function of the continuous data depends only on the vector distance between two points of interest. When all of these conditions are satisfied for 2D continuous data, the Hotelling test statistic and SNR in the spatial-frequency domain are given, respectively, by (see (46) and (48))

$$t_{\text{Hot}}(\mathbf{G}) = \int_{\infty} d^2\rho \frac{[\Delta\bar{G}(\rho)]^*}{\text{NPS}(\rho)} G(\rho), \quad \text{SNR}_{\text{Hot}}^2 = \int_{\infty} d^2\rho \frac{|\Delta\bar{G}(\rho)|^2}{\text{NPS}(\rho)}, \quad (63)$$

where ρ is a 2D spatial-frequency vector, $G(\rho)$ is the 2D transform of the continuous data $g(\mathbf{r})$, the asterisk denotes complex conjugate, and $\text{NPS}(\rho)$ is the noise power spectrum, defined as the 2D Fourier transform of the autocovariance function of the data.

One thing to note about (63), in contrast to all of the other Hotelling expressions in this section, is that there is no inverse covariance present in either t_{Hot} or $\text{SNR}_{\text{Hot}}^2$. In essence, under the stated assumptions, the Fourier transform simultaneously diagonalizes both the system operator $\mathcal{H}$ and the covariance operator $\mathcal{K}$. The system operator (now a convolution) is realizable as a product in the Fourier domain and the inverse of the covariance becomes a simple division by the NPS.

As in the covariance decomposition of (16), the NPS can be expressed as a sum of two terms, one arising from the noise in the data conditional on the object, and one arising from object randomness, provided that both of these effects are described by stationary random processes. Even with the simplest of noise models, (8) or (9), however, the noise in imaging with ionizing radiation is not stationary because it depends on the mean image, which must necessarily vary from point to point in any interesting image. Similarly, the object variability, described in continuous terms by (19), cannot be expected to depend on only $\mathbf{r} - \mathbf{r}'$.

Another conclusion to be drawn from (63) is that this form of $\text{SNR}_{\text{Hot}}^2$ is independent of signal location; a shift in position of a signal in the space domain corresponds to a phase change in its Fourier transform, so the square modulus of $\Delta\bar{G}(\rho)$ is invariant to signal location.

Further problems arise when one replaces the Fourier integrals required in (63) with discrete Fourier transforms (DFTs) for purposes of computer implementation. The correct discretization of a linear, shift-invariant operator or a stationary autocovariance function is a *Toeplitz* matrix, but DFTs diagonalize only *circulant* matrices; for a graphical depiction of the difference between a Toeplitz matrix and its circulant approximation, see figure 7.15 in Barrett and Myers (2004). A DFT will actually *undiscretize* a noise covariance described by (8) or (9).

In spite of these difficulties, there is considerable current interest in Fourier-based approaches to task-based assessment, with their simple, intuitive interpretations (Siewerdsen *et al* 2002, Richard *et al* 2005, Tward *et al* 2007, Richard and Siewerdsen 2008, Pineda *et al*

2008, Pineda *et al* 2012, Tward and Siewerdsen 2008, Tward and Siewerdsen 2009, Gang *et al* 2010). In many cases deviations from strict shift-invariance or strict stationarity can be accommodated by considering detection of a signal at a specific location $\mathbf{r}_0$ (in 2D or 3D as appropriate), then defining a local point spread function and a local autocovariance function, from which we can use a Fourier transform to get a local modulation transfer function $\text{LMTF}(\boldsymbol{\rho}; \mathbf{r}_0)$, and a local noise power spectrum $\text{LNPS}(\boldsymbol{\rho}; \mathbf{r}_0)$, respectively. There are many options as to how the intermediate functions are defined and how one handles the discretization, but the end result will be a Fourier-based formula for a Hotelling-like figure of merit appropriate to detection of a signal localized at $\mathbf{r}_0$.

For more on the mathematics of quasistationary noise and local Fourier metrics for image quality, see sections 13.2.3 and 13.2.4 in Barrett and Myers (2004). Fourier-based approaches to task-based assessment will be discussed further in section 7 of this paper.

3.4. Observers for estimation and joint classification/estimation tasks

For an estimation task in imaging, the term *observer* refers to an algorithm for assigning numerical values to all components of a parameter vector on the basis of some data vector. In most of this subsection we denote the data available for performing the task as $\mathbf{g}$ and refer to the estimate as $\hat{\boldsymbol{\theta}}(\mathbf{g})$, but the results still apply if we substitute a reconstructed image $\hat{\mathbf{f}}$ for $\mathbf{g}$.

We begin with a discussion of general estimation principles and then specialize them to the kinds of data and parameters of interest in imaging with ionizing radiation. The dose dependence of the resulting figures of merit will be discussed in section 6.4.

3.4.1. Likelihood and Fisher information. The likelihood function for estimating a $P \times 1$ parameter vector $\boldsymbol{\theta}$ from an $M \times 1$ data vector $\mathbf{g}$ is defined as the conditional probability density function (or probability, for discrete-valued data), $\text{pr}(\mathbf{g}|\boldsymbol{\theta})$. For a particular data vector, $\text{pr}(\mathbf{g}|\boldsymbol{\theta})$ can be regarded as a function in a P -dimensional parameter space.

The *score* $\mathbf{s}(\mathbf{g})$ is a $P \times 1$ vector of partial derivatives of the log of the likelihood function, with components given by

$$s_p(\mathbf{g}) \equiv \frac{\partial}{\partial \theta_p} \log \text{pr}(\mathbf{g}|\boldsymbol{\theta}), \quad (p = 1, \dots, P). \quad (64)$$

In terms of repeated data acquisitions with $\boldsymbol{\theta}$ fixed, the score is a random vector, and its mean is readily shown to be zero, i.e. $\langle \mathbf{s}(\mathbf{g}) \rangle_{\mathbf{g}|\boldsymbol{\theta}} = 0$. The $P \times P$ covariance matrix of the score, which is called the *Fisher Information Matrix* (FIM) and denoted $\mathbf{F}(\boldsymbol{\theta})$, is defined in component form by

$$F_{pp'}(\boldsymbol{\theta}) \equiv \left\langle \frac{\partial}{\partial \theta_p} \log \text{pr}(\mathbf{g}) \cdot \frac{\partial}{\partial \theta_{p'}} \log \text{pr}(\mathbf{g}) \right\rangle_{\mathbf{g}|\boldsymbol{\theta}}. \quad (65)$$

The FIM sets a fundamental lower bound, called the *Cramér-Rao Bound* (CRB), on the conditional variance of any unbiased estimator. If the bias of $\hat{\boldsymbol{\theta}}(\mathbf{g})$ is zero, then it can be shown that the variance of the p th component must satisfy

$$\text{Var} \left\{ \hat{\theta}_p(\mathbf{g}) | \boldsymbol{\theta} \right\} \geq [\mathbf{F}^{-1}(\boldsymbol{\theta})]_{pp}. \quad (66)$$

An unbiased estimator for which this inequality becomes an equality is said to be *efficient*, which means that it achieves the best possible variance among all unbiased estimators for a specific estimation task.

It is always possible to find an estimator with zero bias just by ignoring the data and setting $\hat{\theta}$ to a constant. More generally, variance can always be made arbitrarily small just by accepting as much bias as needed. The tradeoff between variance and bias in an estimation task is analogous to the tradeoff between sensitivity and specificity in a detection task; in both cases, the tradeoff should be made in terms of clinical efficacy.

The form of the CRB in (66) immediately suggests that one or more of the diagonal elements of the inverse of the FIM can be used as FOMs for estimation tasks (Kupinski *et al* 2003a).

3.4.2. Maximum-likelihood estimation. The maximum-likelihood estimate (MLE) is defined as the point in the P -dimensional parameter space where the likelihood function, $\text{pr}(\mathbf{g}|\boldsymbol{\theta})$ for a given, observed $\mathbf{g}$, is maximal. This condition is expressed mathematically as

$$\hat{\boldsymbol{\theta}}_{\text{ML}}(\mathbf{g}) \equiv \underset{\boldsymbol{\theta}}{\text{argmax}} \{ \text{pr}(\mathbf{g}|\boldsymbol{\theta}) \} = \underset{\boldsymbol{\theta}}{\text{argmax}} \{ \ln [\text{pr}(\mathbf{g}|\boldsymbol{\theta})] \}, \quad (67)$$

where the second form yields the same MLE as the first because the maximum of the logarithm of a function occurs at the same point as the maximum of the function itself.

The MLE is efficient (unbiased and best possible variance) in any problem where an efficient estimator exists. Further properties of the MLE are discussed in Barrett and Myers (2004), section 13.3.6.

3.4.3. Bayesian estimates. In Bayesian estimation, the parameter of interest is treated as a random variable, described *a priori* (i.e. before any data are acquired) by a PDF $\text{pr}(\boldsymbol{\theta})$. The estimation rule is then based on the posterior density, $\text{pr}(\boldsymbol{\theta}|\mathbf{g})$, which is interpreted as the probability of the parameter after a data set is acquired. By Bayes' rule, this posterior can be written as $\text{pr}(\mathbf{g}|\boldsymbol{\theta}) \text{pr}(\boldsymbol{\theta}) / \text{pr}(\mathbf{g})$, and because $\text{pr}(\mathbf{g})$ is independent of $\boldsymbol{\theta}$, the *maximum a posteriori* (MAP) estimate is given by

$$\hat{\boldsymbol{\theta}}_{\text{MAP}}(\mathbf{g}) \equiv \underset{\boldsymbol{\theta}}{\text{argmax}} \{ \text{pr}(\mathbf{g}|\boldsymbol{\theta}) \text{pr}(\boldsymbol{\theta}) \} = \underset{\boldsymbol{\theta}}{\text{argmax}} \{ \ln[\text{pr}(\mathbf{g}|\boldsymbol{\theta})] - \ln[\text{pr}(\boldsymbol{\theta})] \}. \quad (68)$$

Thus the MAP estimator maximizes the product of the likelihood and the prior. Equivalently, it is the *mode* of the posterior density.

Another important Bayesian estimator is the *posterior mean*,

$$\hat{\boldsymbol{\theta}}_{\text{PM}}(\mathbf{g}) \equiv \int_{\infty} d^P \boldsymbol{\theta} \boldsymbol{\theta} \text{pr}(\boldsymbol{\theta}|\mathbf{g}), \quad (69)$$

which can be shown to minimize the EMSE defined in (35).

3.4.4. Linear estimators for linear parameters. As we saw in (30), a linear, scalar parameter has the form of a scalar product, $\theta(\mathbf{f}) = \boldsymbol{\chi}^\dagger \mathbf{f}$. The general linear estimator of this parameter has the same structure as a general linear discriminant, $\hat{\theta} = \mathbf{w}'\mathbf{g}$, and design of a linear estimator is equivalent to choosing the template $\mathbf{w}$. The template can range from a simple indicator function that defines a region of interest to ones that account for the statistics of the object and image noise; whatever template is chosen, however, the task performance will depend on these statistics.

The *Gauss–Markov Estimator* (GME) accounts for the noise statistics but not the object variability. It applies to data obtained with a linear system operator $\mathcal{H}$, and it minimizes the conditional MSE for a given object $\mathbf{f}$ by choosing (Barrett 1990)

$$\hat{\boldsymbol{\theta}}_{\text{GM}}(\mathbf{g}) \equiv \boldsymbol{\chi}^\dagger [\mathbf{K}_{\mathbf{g}|\mathbf{f}}^{-1/2} \mathcal{H}]^+ \mathbf{K}_{\mathbf{g}|\mathbf{f}}^{-1/2} \mathbf{g}, \quad (70)$$

where $\mathbf{K}_{\mathbf{g}|\mathbf{f}}^{-1/2}$ is recognized from section 3.3 (see (43)) as a prewhitening operator and $[\cdot]^+$ denotes a *Moore–Penrose pseudoinverse*; for many properties of pseudoinverses and algorithms for computing them, see chapter 1 in Barrett and Myers (2004).

For completeness, the GME for estimation from a reconstructed image with a linear reconstruction operator $\mathcal{O}$ is given by

$$\hat{\boldsymbol{\theta}}_{\text{GM}}(\hat{\mathbf{f}}) \equiv \chi^{\dagger} \left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{-1/2} \mathcal{O} \mathcal{H} \right]^+ \mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{-1/2} \hat{\mathbf{f}}. \quad (71)$$

Even though the GME was derived by minimizing MSE for a chosen object, we can still get the EMSE by averaging over an ensemble of objects. An expression for the resulting EMSE is given in Barrett (1990). This averaging step does not make the GME a Bayesian estimator because the statistics of the object ensemble are not used in defining the estimator.

3.4.5. Optimal Bayesian linear estimator. We have noted above that the Bayesian estimator that minimizes EMSE among all estimators is the posterior mean estimator defined in (69), but the integral in that expression usually requires Markov-chain Monte Carlo techniques as in Kupinski *et al* (2003c). A more tractable approach is to use the *Wiener Estimator* (WE) which minimizes the EMSE among all linear estimators.

The most compact way of writing the WE is to assume that the prior means of the parameter and the data are zero, in which case the best linear estimate of a $P \times 1$ parameter vector is given by

$$\hat{\boldsymbol{\theta}}_{\text{W}}(\mathbf{g}) \equiv \mathbf{K}_{\boldsymbol{\theta},\mathbf{g}} \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{g}, \quad (72)$$

where $\mathbf{K}_{\boldsymbol{\theta},\mathbf{g}}$ is the $P \times M$ cross-covariance matrix of the random parameter and the random data. When the prior means of the parameter and the data are not zero but known, the WE becomes

$$\hat{\boldsymbol{\theta}}_{\text{W}}(\mathbf{g}) = \bar{\boldsymbol{\theta}} + \mathbf{K}_{\boldsymbol{\theta},\mathbf{g}} \mathbf{K}_{\mathbf{g}}^{-1} [\mathbf{g} - \bar{\mathbf{g}}]. \quad (73)$$

In this form, the WE is an affine transformation of the data, not a strictly linear one.

3.4.6. Signal parameters. To apply the estimators discussed above to imaging, it is useful to decompose the object being imaged into signal and background components. The distinction is that the signal contributes to task performance and hence to efficacy and the background does not; therefore we treat the background as a nuisance parameter, albeit a very high-dimensional one.

For imaging with ionizing radiation, where no part of the object is opaque to the radiation, we can write the signal/background decomposition as a sum, and for estimation tasks we can assume that the parameters of interest are associated only with the signal. Thus we can write the object function as

$$f(\mathbf{r}; \boldsymbol{\theta}) = f_{\text{bg}}(\mathbf{r}) + f_{\text{sig}}(\mathbf{r}; \boldsymbol{\theta}), \quad (74)$$

or, in our usual shorthand, $\mathbf{f} = \mathbf{f}_{\text{bg}} + \mathbf{f}_{\text{sig}}(\boldsymbol{\theta})$. Common choices for the components of $\boldsymbol{\theta}$ might include its spatial coordinates, some measures of its shape and size, and a measure of the strength of the signal, such as its mean x-ray attenuation coefficient or the uptake of a radiotracer.

3.4.7. Gaussian likelihood models. It may be a useful approximation to treat the likelihood for the signal parameters as a multivariate normal of the form

$$\text{pr}(\mathbf{g}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^M \det(\mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}})}} \exp \left[-\frac{1}{2} [\mathbf{g} - \bar{\mathbf{g}}(\boldsymbol{\theta})]^t \mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}}^{-1} [\mathbf{g} - \bar{\mathbf{g}}(\boldsymbol{\theta})] \right]. \quad (75)$$

This PDF is a complicated function of the parameter vector through the covariance matrix, $\mathbf{K}_{\mathbf{g}|\theta}$, but there are many cases where this dependence can be neglected. The easiest approach is to consider signals that are very weak compared to the background. If the signal is not weak, we may still be able to approximate $\mathbf{K}_{\mathbf{g}|\theta}$ by either $\mathbf{K}_{\mathbf{g}|\bar{\theta}}$ or $\langle \mathbf{K}_{\mathbf{g}|\theta} \rangle_{\theta}$ (Whitaker *et al* 2008, Kupinski *et al* 2013).

If the θ dependence of the covariance matrix can be neglected or approximated away, then a Gaussian log-likelihood can be written as

$$\ln p(\mathbf{g}|\theta) = [\bar{\mathbf{g}}(\theta)]^t \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{g} - \frac{1}{2} [\bar{\mathbf{g}}(\theta)]^t \mathbf{K}_{\mathbf{g}}^{-1} [\bar{\mathbf{g}}(\theta)] + \text{terms independent of } \theta, \quad (76)$$

which is a linear function of $\mathbf{g}$. Under the same assumptions, the components of the Fisher information matrix can be shown to be:

$$F_{pp'} = \left[\frac{\partial \bar{\mathbf{g}}(\theta)}{\partial \theta_p} \right]^t \mathbf{K}_{\mathbf{g}}^{-1} \left[\frac{\partial \bar{\mathbf{g}}(\theta)}{\partial \theta_{p'}} \right]. \quad (77)$$

3.4.8. Linear observers for joint detection/estimation tasks. As we mentioned in section 3.3, linear discriminant functions do not perform well on detection tasks when the signal, if present, can be at a random location in the image. Similarly when the task is to estimate the location, size or other parameters of a lesion at a random location, then linear estimators do not perform well either (Whitaker *et al* 2008).

A *Scanning Linear Estimator* (SLE) is an estimator that computes a scalar product of the data vector with a template $\mathbf{w}(\theta)$ and then searches for the point in parameter space where the scalar product is maximum. Formally,

$$\hat{\theta}_{\text{SL}}(\mathbf{g}) \equiv \underset{\theta}{\operatorname{argmax}} [\mathbf{w}(\theta)]^t \mathbf{g}. \quad (78)$$

If the parameter vector consists only of spatial coordinates of a tumor or other signal, then $\hat{\theta}_{\text{SL}}(\mathbf{g})$ has a natural interpretation as spatial scanning of a template; more generally, (75) describes scanning in parameter space, not just spatially. The argmax operator is a nonlinear selection of the point in parameter space where the maximum occurs, so the SLE is nonlinear even though it performs only linear operations on the data.

Moreover, if the task requires detection of a signal before estimating its parameters, the same search in parameter space yields the test statistic for detection:

$$t(\mathbf{g}) = \max_{\theta} [\mathbf{w}(\theta)]^t \mathbf{g}. \quad (79)$$

As in any detection problem, we make the decision that the signal is present if $t(\mathbf{g})$ exceeds the decision threshold (see (22)).

Conversely, any of the linear observers for detection discussed in section 3.3 can be converted to an SLE simply by showing its dependence on the signal parameters explicitly. For example, the Hotelling observer of (46) leads to

$$[\mathbf{w}(\theta)]^t = [\Delta \bar{\mathbf{g}}(\theta)]^t \mathbf{K}_{\mathbf{g}}^{-1}. \quad (80)$$

If the data covariance arises entirely from the background, as in the case of a weak signal, then $\mathbf{K}_{\mathbf{g}}$ is independent of θ and $\mathbf{w}(\theta)$ can be computed with methods from section 3.3 without performing a matrix inverse at each point in parameter space.

The performance of an SLE on a joint detection/estimation task can be specified by the area under an LROC or EROC curve (see figure 2). If detection is not at issue, the SLE performance can be specified by some variant of EMSE or by an expected utility. All of these FOMs will vary with radiation dose (see section 6.4).

Khurd and Gindi presented ideal-observer decision strategies that maximize the area under the LROC curve (Khurd and Gindi 2005), and Clarkson extended that work to the area under the EROC curve (Clarkson 2007). For a general treatment of methods for performing, analyzing and validating observer studies involving detection and localization, see Chakraborty and Berbaum (2004).

An interesting practical application is the work of Yendiki and Fessler (2007), who considered model observers for which the decision variable was the maximum value of a local test statistic within a search area. Their goal was to optimize regularized image reconstruction for emission tomography with respect to lesion detectability. They made use of approximations of tail probabilities for the maximum of correlated Gaussian random fields to facilitate the analytical evaluation of detection performance. They illustrated the reasonable accuracy of these approximations in the regime of low probability of false alarms and their potential use in the optimization of regularization functions.

4. Radiation dose

4.1. Basic definitions

A confusing mélange of terminologies and definitions related to radiation dose has appeared in the literature, but in this paper we adhere to the 2007 recommendations of the International Commission on Radiological Protection (ICRP 2007, Wrixon 2008).

The basic quantity on which these recommendations are based is the *absorbed dose*, which is a measure of the energy per unit mass deposited in a material by ionizing radiation. Absorbed dose can depend on the location in the material, denoted by a 3D vector $\mathbf{r}$, and it can also vary with time. Often we are interested in just the spatial distribution, so we integrate the dose per unit time over some relevant time, such as the duration of an x-ray imaging procedure or over $0 < t < \infty$ for radionuclide imaging, taking into account the physical half-life and the time-dependent biodistribution of the radiopharmaceutical.

The spatial distribution of absorbed radiation dose for a single study on a given patient is denoted $d(\mathbf{r})$ and measured in grays (Gy), where $1 \text{ Gy} = 1 \text{ J kg}^{-1}$. Because dose is defined as absorbed energy per unit mass, the absorbed energy per unit volume is given by $d(\mathbf{r})\rho(\mathbf{r})$, where $\rho(\mathbf{r})$ is the tissue density at point $\mathbf{r}$. Thus the basic definition of $d(\mathbf{r})$ is that the differential energy dE absorbed in the differential volume element d^3r is given by $d(\mathbf{r})\rho(\mathbf{r})d^3r$.

The dose distribution can be integrated spatially to get the dose to any organ. If K different organs are considered, the dose to organ k , ($k = 1, \dots, K$), is given by

$$D_k = \frac{\int_{V_k} d(\mathbf{r}) \rho(\mathbf{r}) d^3r}{\int_{V_k} \rho(\mathbf{r}) d^3r}, \quad (81)$$

where V_k is the volume of organ k and the denominator is the mass of the organ. An expression similar to (81) applies if we consider a particular tissue type t confined to a volume V_t , and in that case we refer to the tissue dose (in Gy) as D_t .

The *equivalent dose* to tissue t , denoted as H_t , is defined by multiplying D_t by a factor Q_R for the relative biological effectiveness of different types of radiation, indicated by the

subscript R. By international convention, this factor is taken to be 1.0 for x-ray photons and electrons of all energies. In this paper, we do not consider neutrons or alpha particles, which would have a factor Q_R different from 1, so for our purposes the numerical values for the absorbed tissue dose and the equivalent tissue dose are the same. However, according to the conventions of the ICRP, the equivalent dose is expressed in a unit with a different name, the sievert (Sv), where $1 \text{ Sv} = 1 \text{ J kg}^{-1}$. Therefore, we can define the equivalent tissue dose, in the context of medical imaging, as the absorbed dose in the tissue multiplied by 1 Sv Gy⁻¹. The same conclusion applies to organ doses.

The *effective dose*, denoted E and also measured in sieverts, is defined by summing up all equivalent doses to all irradiated organs k , each weighted by a numerical factor w_k defined by ICRP:

$$E = \sum_{k=1}^K w_k D_k, \text{ where } \sum_{k=1}^K w_k = 1. \quad (82)$$

The weighting factor w_k is explained by ICRP (2007) as the

‘... factor by which the equivalent dose in a tissue or organ ... is weighted to represent the relative contribution of that tissue or organ to the *total health detriment* resulting from *uniform irradiation of the body*.’ [Emphasis added.]

Radiation detriment is defined by ICRP as a

‘... function of several factors, including incidence of radiation-related cancer or heritable effects, lethality of these conditions, quality of life, and years of life lost owing to these conditions.’

In the ICRP view, effective dose has a very limited range of applications. According to a summary by Wrixon (2008),

‘ICRP emphasises that the effective dose provides a measure of radiation detriment for protection purposes only. It does not provide an individual-specific dose and should not be used for epidemiological evaluations. Furthermore, the collective effective dose, the main use of which is in the optimisation of radiological protection, should not be used in epidemiological studies and in assessing the hypothetical number of cases of cancer or heritable disease in an exposed population.’

Because the weighting factors are defined so that they sum to unity, the ratio of the effective dose (in Sv) to the absorbed dose (in Gy) depends strongly on the volume of tissue irradiated. As an example, suppose that only the liver receives any absorbed dose in a CT scan, and that $d(\mathbf{r})$ is constant over the liver and zero everywhere else. Moreover, take the density of the liver as 1 g cm^{-3} at all points and assume that the average liver has a mass of 1 kg. Then, if the absorbed energy is 0.2 J, the absorbed dose to the liver is 0.2 Gy or 200 mGy, but because the weighting factor for liver is 0.04, the reported effective dose would be 8 mSv. This reported value would cause little concern compared to other sources of whole-body radiation, but in fact the local absorbed dose for the liver would still be 200 mGy. The same 200 mGy dose delivered uniformly to the whole body, corresponding to about a 70-fold increase in energy absorbed in the body for an average adult male, would yield an effective dose of 200 mSv.

4.2. Other dose measures

4.2.1. Dose distribution concepts from radiation oncology. Radiation treatment planning almost always starts by segmenting a CT image to identify the boundaries of the tumors and surrounding normal tissues. It is usually assumed for purposes of treatment planning that the tissue distribution within an organ is homogeneous and that the boundaries returned by the segmentation algorithm are exact. In some case substructures within an organ, such as cortical bone and marrow, are considered separately, and there are some current investigations into how much the substructure in organs like the lung would change the results of the dose calculations if they were not considered separately.

Even with the assumption that the tissues in an organ are homogeneous, the simulation tools used in therapy planning are able to provide some information about the distribution of doses within an organ. However, at the current state, this information is not used; the absorbed doses predicted by the plan are averaged over the organs, and often even these average values are then summarized by a single value, typically the effective dose.

If one wants to use the information available from the planning algorithm for more in-depth evaluation one could use the so-called dose–volume histograms (DVHs), preferably for each organ. This concept was introduced in 1979 in the context of radiation oncology (Shipley *et al* 1979, Drzymala *et al* 1991), where it is used to predict probability of tumor control and/or probability of healthy tissue detriment.

One might expect the DVH for an organ to be a histogram of the dose values in the volume elements (voxels) in that organ, hence a discretized approximation of the probability density function for occurrence of different values of the absorbed dose in a voxel, D_v . Mathematically, this PDF could be denoted $pr(D_v)$. In the radiation-oncology literature, however, the term DVH corresponds to a cumulative probability distribution function rather than a PDF. Specifically, the DVH is defined as the fraction of voxels that receive a dose greater than some amount D_0 , plotted as a function of D_0 . Thus it is a discretized approximation of the probability (not PDF) $Pr(D_v > D_0)$, which is expressed mathematically as $1 - \int_0^{D_0} pr(D_v) dD_v$.

The equivalent uniform dose (EUD), a concept similar to that of DVH but more related to possible biological effects, was established by Niemierko (1997) in the context of radiation oncology. This concept can be used for any organ specifically or for the whole body. It is designed in a way that the current dose distribution is compared to another one which is assumed to have the same radiobiological effect. If the biological effect follows the linear-no-threshold theory, and if only photons and electrons are involved, the EUD is equivalent to average of the dose distribution. If so-called deterministic effects (see section 5) might occur, for example due to large dose variations inside an organ like the skin, the EUD can be much higher than the average dose.

4.2.2. Mean glandular dose. The breast is not a homogeneous tissue. Rather, it consists of glandular tissue, fibrous tissue, fatty tissue and the skin. Mainly depending on the size of the breast, the fractional content of the various tissues is different and hence the optimum imaging conditions are different (Thilander-Klang *et al* 1997). Besides the different contrast appearances of the different tissues, it is also important to note that the glandular (and the fibrous) tissues are much more sensitive to ionizing radiation than the fatty tissue and the skin. To compare mammographic imaging systems on the basis of radiation risk, it is therefore more useful to evaluate the average dose to the glandular and fibrous tissue. This can be done by measuring the air kerma at the entrance of a breast phantom, multiplying this value by a conversion coefficient to get the entrance dose for the real breast and then multiplying by a factor relating this

entrance dose to the mean glandular dose. These factors have been determined by Dance *et al*, first for 50% glandularity and then for various different glandularities (Dance 1990, Dance *et al* 2000). While Dance was assuming a homogeneous distribution in the main area of the breast between fatty and glandular tissue, this is obviously not the case. Therefore there are approaches that use more realistic phantoms of the breast either by evaluating large number of mammograms and finding mathematical representations resulting in 3D tissue distributions (Lau *et al* 2012, Bakic *et al* 2011, Bakic *et al* 2002) or by high-resolution 3D-scanning of specially fixed breast specimens from corpses (Hoeschen *et al* 2005). The segmented high-resolution phantoms need to be extended to whole-body phantoms in order to gain additional dosimetric information for comparison of different imaging modalities.

4.2.3. DLP and CTDI. The effective dose is intended to be a way to compare systems regarding their radiation burden for a large group of patients undergoing similar examinations, but it is not a directly measurable unit. In addition, for CT applications, which generate the highest per capita radiation burden in most of the developed countries today, there is no available measurement system for really measuring radiation dose to the patient directly. Instead, various dose indicators have been proposed. They are typically based on measuring absorbed dose within polymethyl methacrylate (PMMA) phantoms during standard dose protocols. With these values and the known system parameters for a particular patient examination, dose indicators such as the so-called dose-length product (DLP) and the computed tomography dose index (CTDI) can be calculated and displayed. The DLP in general is the dose in a certain area multiplied by the length of the investigated area; in a CT scanner this would mean the scan length (typically calculated as the slice thickness multiplied by the number of slices). However, there is not a measured or calculated value for the absorbed dose within the body of the patient. Therefore the CTDI_{vol} is used as the multiplier.

The CTDI as it was originally defined by the Food and Drug Administration (FDA) in 1981 was introduced for axial scanners (Shope *et al* 1981). It was meant to characterize the dose from the primary beam and that resulting from scattered radiation emitted from surrounding slices in such a scanning geometry and for standard conditions. The CTDI is thus defined as the equivalent value of absorbed dose in a central slice assuming that the total radiation is concentrated in that specific slice. Normalization of this value by the current output of the tube gives a hint for the dose profile which can be reached by a scanner but not about the dose of the patient or the dose efficiency of the scanner. This basic concept has been modified in various approaches to make CTDI meaningful for spiral and multi-detector-row CT. CTDI₁₀₀ is assumed to represent the dose contribution from a 100 mm measurement length along the axis with 100 mm dosimetric chambers measured in standard conditions for a single collimation scan still in a phantom. Doing such a measurement in the center of the acrylic phantom typically 32 cm or 16 cm diameter and on various positions at the periphery permits calculation of the weighted CTDI (CTDI_w). For this purpose, one weights the central measurement by 1/3 and the average of the peripheral measurements by 2/3 as expressed in the following equation:

$$\text{CTDI}_w = \frac{1}{3}\text{CTDI}_c + \frac{2}{3} \frac{\sum_{n=1}^N \text{CTDI}_{\text{per},n}}{N}, \quad (83)$$

From CTDI_w one can calculate the volume CTDI (CTDI_{vol}), which is the most commonly used CT dose index today, by dividing the weighted CTDI by the pitch factor. This dose index is somehow related to an absorbed dose of the patient in a specific volume, but only indirectly because it does not refer to either a phantom measurement or to measurements in the center and the periphery with a fixed weighting. Instead, CTDI refers to a standard operation protocol

for the scanner. Using the CTDI_{vol} , the scan length, the body part investigated and conversion coefficients resulting from calculation programs based on results from Monte Carlo simulations seems to give results consistent with thermoluminescent dosimeter (TLD) measurements in anthropomorphic phantoms (Lechel *et al* 2007) for calculated effective doses on standard patients as indicators for dose comparisons of systems. Using these values for risk estimates on single patients is probably not useful for various reasons.

4.2.4. Approaches to patient-specific dose. One approach to getting a better patient-specific dose estimate is the size-specific dose estimate as introduced by the AAPM (AAPM 2011), which tries to find a way to calculate dose values dependent on the size of the patient. However, again this value does not take into account the patient-specific radiation sensitivity and it therefore does not give a true patient-specific risk value, but only an indicator.

Li *et al* (2011b) developed a method to estimate patient-specific dose and cancer risk from CT examinations by combining a validated Monte Carlo program with patient-specific anatomical models. The organ dose, effective dose, and risk index (a surrogate of cancer risk) was estimated for clinical chest and abdominopelvic protocols. In Li *et al* (2011a), this same group evaluated the dependence of dose and risk on patient size and scanning parameters. The reported relationships can be used to estimate patient-specific dose and risk in clinical practice for a given pediatric patient.

Schmidt and Kalender (2002) describe a voxel-based Monte Carlo method for scanner- and patient-specific dose calculations in computed tomography.

4.3. Computation of dose

4.3.1. X-ray imaging. Dose values cannot usually be measured inside a patient. It is possible to measure the skin dose directly on the patient or to calculate it from dose measurements (mostly air kerma measurements) in the x-ray beam, but this measurement does not yield the dose distributions in specific organs throughout the body. Therefore, ICRP and ICRU as well as many research groups have been using simulation tools in combination with models of the human body and biokinetic models to determine absorbed doses to organs. The simulation tools that are most commonly used today are Monte Carlo simulations (e.g. EGS, MCNP, Fluka, Penelope and GEANT (Bielajew *et al* 1994, Baró *et al* 1995, Agostinelli *et al* 2003, Ferrari *et al* 2005, Allison *et al* 2006, Reed 2007, Ljungberg and King 2012, Pozuelo *et al* 2012)). The models range from pure mathematical models (ICRP 1979, Zankl *et al* 1988)) to various anthropomorphic voxel phantoms (Zubal *et al* 1994, Xu *et al* 2000, Petoussi-Henss *et al* 2002, Kramer *et al* 2004, Lee *et al* 2006, Menzel *et al* 2009, Xu and Eckerman 2010).

The voxel models for average adults in ICRP 110 (Menzel *et al* 2009) are based on real patient images that were scanned and segmented. These models were then modified to fit the conditions of the organs of average human beings. ICRP and ICRU are currently working on models for children of different ages. There are also other phantoms available (Zhang *et al* 2007, Becker *et al* 2007, Segars *et al* 2010). For all these phantoms and all these kinds of simulation tools, there are many publications investigating the variation of dose with different imaging parameters e.g. for angiography or CT scanning in Schlattl *et al* (2007, 2012), Li *et al* (2012).

Many details on computational phantoms were presented at the 4th International Workshop on Computational Phantoms for Radiation Protection, Imaging, and Radiotherapy in 2013, and selected papers from this workshop were published in a special section of *Physics in Medicine and Biology* in September, 2014; included is a topical review on computational phantoms (Xu 2014).

4.3.2. Radionuclide imaging. Two separate steps are required for rigorous dose calculation in emission imaging and radionuclide therapy. First, the spatiotemporal radiopharmaceutical distribution should be determined. Then, in principle, the distribution of absorbed dose, $d(\mathbf{r}, t)$, should be calculated, but in practice usually only time-integrated organ doses are considered.

Time-dependent radiopharmaceutical distributions for nuclear medical imaging have been investigated in Zankl *et al* (2012) and Giussani *et al* (2012). A detailed review of dosimetry in nuclear medicine is given by Stabin (2006).

5. Relating dose to risk

5.1. Types and measures of risk

Risks associated with ionizing radiation have been known for almost as long as ionizing radiation itself (Doll 1995). Within a year of the discovery of x-rays by Röntgen skin burns had been reported (Gilchrist 1897, Stevens 1896), and within 7 years a case of skin cancer was observed (Friebe 1902), in all cases associated with high-dose x-ray exposure. In general, risks associated with ionizing radiation can be divided into the so-called stochastic effects (genetic risks in offspring, somatic effects (cancer) in directly exposed population), and so-called tissue-reaction effects (previously termed deterministic effects) (ICRP 2007, 2012). We briefly review the stochastic and tissue-reaction risks associated with exposure to radiation.

There are a number of commonly used measures of risk employed in epidemiological studies. The *odds ratio* (OR) is the ratio of the odds of disease occurrence in a group with exposure to a factor to that in an unexposed group: within each group, the odds are the ratio of the numbers of diseased and non-diseased individuals.

The *relative risk* (RR) is the ratio of the disease rate in a group under study to that in a comparison group, with adjustment for confounding factors such as age, if necessary. The *excess relative risk* (ERR) is $RR - 1$. The *excess absolute risk* (EAR) is the difference in absolute risk between exposed and control populations.

5.2. Tissue-reaction (deterministic) effects

Tissue-reaction (deterministic) effects generally occur only after high-dose acute exposure (mostly > 0.1 Gy, see table 2) and are characterized by nonlinear dose responses, with a threshold dose below which the effect is not observed. Because of these features, tissue-reaction effects are of most relevance in radiotherapy; normal tissue therapy doses are limited to avoid these effects. Tissue-reaction effects are thought to arise from the killing of large groups of cells in the tissues concerned, leading to functional deterioration in the organs affected. Tissue-reaction effects generally arise within days (e.g. gastrointestinal syndrome or central nervous system syndrome) or weeks (e.g. haematopoietic syndrome, pulmonary syndrome) of exposure; however, certain tissue-reaction effects (e.g. cataracts, hypothyroidism) are manifest only over periods of years or more. Most of the information on tissue-reaction effects of radiation comes from (a) medically exposed groups, (b) the survivors of the atomic bombings of Hiroshima and Nagasaki, (c) radiation accidents and (d) animal experiments (Edwards and Lloyd 1998, ICRP 2012).

Edwards and Lloyd (1998) propose that the probability, P , of most tissue-reaction effects following an acute dose, D , is given by a modified Weibull distribution:

$$P = \left\{ 1 - \exp \left[- \ln 2 \cdot \left(\frac{D}{D_{50}} \right)^V \right] \right\} \cdot 1_{D > T}, \quad (84)$$

Table 2. Estimates of the threshold doses for approximately 1% incidence of morbidity in tissues and organs in adults exposed to acute, fractionated or protracted, and chronic irradiation (taken from ICRP (2012)).

Effect	Organ/tissue	Time to develop effect	Acute exposure (Gy)	Highly fractionated (2 Gy fractions) or equivalent protracted exposures (Gy)	Annual (chronic) dose rate for many years (Gy/year)
Temporary sterility	Testes	3–9 weeks	~ 0.1	NA	0.4
Permanent sterility	Testes	3 weeks	~ 6	< 6	2.0
Permanent sterility	Ovaries	< 1 week	~ 3	6.0	>0.2
Depression of haematopoiesis	Bone marrow	3.7 d	~ 0.5	10–14	>0.4
Xerostomia	Salivary glands	1 week	NA	< 20	NA
Dysphagia, stricture	Oesophagus	3–8 months	NA	55	NA
Dysphagia, ulceration	Stomach	2 years	NA	50	NA
Stricture	Small intestine	1.5 years	NA	45	NA
Stricture	Colon	2 years	NA	45	NA
Anorectal dysfunction	Rectum	1 year	NA	60	NA
Hepatomegaly, ascites	Liver	2 weeks–3 months	NA	< 30–32	NA
Main phase of skin reddening	Skin (large areas)	1–4 weeks	< 3–6	30	NA
Skin burns	Skin (large areas)	2–3 weeks	5–10	35	NA
Temporary hair loss	Skin	2–3 weeks	~ 4	NA	NA
Late atrophy	Skin (large areas)	> 1 year	10	40	NA
Telangiectasia at 5 years	Skin (large areas)	> 1 year	10	40	NA
Cataract (visual impairment)	Eye	> 20 years	~ 0.5	~ 0.5	~ 0.5 divided by years duration
Acute pneumonitis	Lung	1–3 months	6–7	18	NA
Oedema	Larynx	4–5 months	NA	70	NA
Renal failure	Kidney	> 1 year	7–8	18	NA
Fibrosis/necrosis	Bladder	> 6 months	15	55	NA
Stricture	Ureters	> 6 months	NA	55–60	NA
Fracture	Adult bone	> 1 year	NA	50	NA
Fracture	Growing bone	< 1 year	NA	25	NA
	Muscle	Several years	NA	55	NA
Endocrine dysfunction	Thyroid	> 10 years	NA	> 18	NA
Endocrine dysfunction	Pituitary	> 10 years	NA	≤ 10	NA
Paralysis	Spinal cord	> 6 months	NA	55	NA
Necrosis	Brain	> 1 year	NA	55–60	NA
Cognitive defects	Brain	Several years	1–2	< 20	NA
Cognitive defects infants <18 months	Brain	Several years	0.1–0.2	NA	NA

where V is the shape factor, determining the steepness of the risk function, T is the threshold dose below which no effect is observed, and $1_{D>T}$ is an indicator function which equals 1 if $D > T$ and 0 otherwise. The quantity D_{50} , defined as the dose at which the effect is expected to be observed in half the population, is a function of the radiation dose rate DR (in Gy h^{-1}):

$$D_{50}(DR) = \theta_{\infty} + \frac{\theta_1}{DR}. \quad (85)$$

ICRP (2012) recommends use of a practical threshold dose for various tissue-reaction effects, below which morbidity would occur at <1%, as indicated in table 2.

Although (84) is thought to describe most tissue-reaction effects, there are certain effects for which it probably does not apply. Particularly problematic in this respect are severe mental retardation and reduction in IQ following irradiation of the fetus. Otake *et al* (1996) and Otake and Schull (1998) observed a dose-related increase in severe mental retardation in those exposed *in utero* as a result of the atomic-bomb explosions in Hiroshima and Nagasaki. This effect was particularly marked for those exposed in the period 8–15 weeks post-conception. A threshold in the region of 0.1–0.3 Gy was indicated. The authors found a linear no-threshold reduction in IQ with increasing uterine dose in the atomic-bomb survivors, although thresholds of at least 0.1 Gy are consistent with the data (Otake and Schull 1998).

It is also arguable that a threshold may not apply for radiation-induced circulatory disease. There have been a number of reviews of circulatory disease in people exposed to low and moderate doses of radiation (Little *et al* 2008b, McGale and Darby 2005, McGale and Darby 2008, UNSCEAR 2008a, McMillan *et al* 2010, Little *et al* 2010, Little *et al* 2012). The systematic review and meta-analysis of Little *et al* (2012) documented statistically significant excess relative risk coefficients for three of the four major subtypes of circulatory disease (IHD, heart disease apart from IHD, stroke, all other circulatory disease) in people exposed to low and moderate doses (mean < 0.5 Sv) of radiation. For ischemic and non-ischemic heart disease, there was no significant heterogeneity in risks between the various studies; however, this was not the case for stroke and other circulatory diseases (Little *et al* 2012). There is also accumulating evidence for radiation-associated cataracts following moderate dose and low dose-rate radiation exposure (Little 2013).

5.3. Stochastic effects

Stochastic effects are the main late health effects that are expected to occur in populations exposed to ionizing radiation; somatic risks dominate the overall estimate of health detriment (see section 4.1). For both somatic and genetic effects the probability of their occurrence, but not their severity, is taken to depend on the radiation dose. The dose response may be non-linear, as for tissue-reaction effects. However, in contrast to the situation for tissue-reaction effects, for most stochastic effects it is generally accepted that at sufficiently low doses there is a non-zero linear component to the dose response, i.e. there is no threshold. There is little biological or epidemiological evidence for thresholds for stochastic effects (Little and Muirhead 1996, Little and Muirhead 1998, Pierce and Preston 2000, Little *et al* 2009).

5.3.1. Heritable genetic effects. The heritable genetic risks associated with radiation exposure are estimated directly from animal studies in combination with data on baseline incidence of disease in human populations (UNSCEAR 2001). There are no usable human data on radiation-induced germ cell mutations, let alone induced genetic diseases, and the results of the largest and most comprehensive of human epidemiological studies, namely that carried out on the children of the Japanese atomic bomb survivors, are negative; there are no statistically significant radiation-associated adverse hereditary effects in this cohort (Neel *et al* 1990). The

Table 3. Cancer and heritable genetic percentage risk estimates associated with low-dose irradiation (taken from UNSCEAR (2001, 2008b)).

Somatic (cancer) mortality risks ^a		% Lifetime risk/Gy
Generalized relative risk models for solid cancer and for leukemia	Leukemia	0.50 (0.06, 1.09) ^b
	Solid cancer	5.45 (2.67, 8.51) ^b
Generalized relative risk models for solid cancer and for leukemia	Leukemia	0.52 (0.08, 0.99) ^b
	Solid cancer	4.48 (1.90, 7.33) ^b
Heritable genetic risk ^c		
Autosomal dominant, X-linked	Risk to first generation	0.075–0.15
	Risk to second generation	0.05–0.10
Autosomal recessive	Risk to first generation	≈ 0
	Risk to second generation	≈ 0
Chronic multifactorial disease	Risk to first generation	0.025–0.12
	Risk to second generation	0.025–0.12
Developmental abnormalities	Risk to first generation	0.20
	Risk to second generation	0.04–0.10
<i>Total</i>	Risk to first generation	0.30–0.47
	Risk to second generation	0.12–0.32

^a Risk calculated for a UK population assuming a test dose of 0.1 Gy administered.

^b Mean and 95% BCI.

^c Risk calculated for radiation exposure in one generation only, with a dose and dose-rate effectiveness factor of 3 applied.

data on the induction of germline mutations at human minisatellite loci (Dubrova *et al* 1996), although of importance from the standpoint of direct demonstration of radiation-induced heritable genetic changes in humans, are probably not suitable for risk estimation; they occur in non-coding DNA and are thought not to be associated with heritable disease (UNSCEAR 2001, Little *et al* 2013).

Simple linear models of dose are generally employed to model genetic effects. For example, the preferred risk model used in the 2001 report of the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR 2001) assumes that the excess heritable genetic risk for disease class i associated with a dose D of radiation to the parental gonads is given by:

$$R_i = P_i \cdot \left[\frac{1}{DD_i} \right] \cdot MC_i \cdot PCRF_i \cdot D, \quad (86)$$

where P_i is the baseline incidence of that disease class, MC_i is the mutational component of the disease class (defined as the relative increase in disease frequency (relative to the baseline) per unit relative increase in mutation rate (relative to the spontaneous rate)), $PCRF_i$ is the potential recoverability correction factor for the disease class (the fraction of induced mutations compatible with live births) and DD_i is the mutational doubling dose (i.e. the dose required to double the mutational load associated with the disease). Table 3 summarizes estimates of risk for the first- and second-generation progeny of an irradiated population that has sustained radiation exposure in the parental generation and no radiation subsequently. These risk estimates assume a mutational doubling dose, DD_i , of 1 Gy, that is based on human data on spontaneous mutation rates and mouse data on induced mutation rates (UNSCEAR 2001). Neel *et al* (1990) derived an estimate of the doubling dose in the atomic bomb survivors of 3.4–4.5 Sv, based on examination of a combination of five endpoints. However, as discussed in UNSCEAR (2001), in view of the differences between the endpoints considered by Neel *et al*

(1990) and those employed by UNSCEAR, and the uncertainties in the estimates of Neel *et al*, their estimate of the doubling dose is entirely compatible with that used by UNSCEAR.

5.3.2. Somatic effects (cancer). Most of the information on radiation-induced cancer risk comes from (a) the Japanese atomic bomb survivors, (b) medically exposed populations and (c) occupationally exposed groups (UNSCEAR 2008b). The Life Span Study (LSS) cohort of Japanese atomic bomb survivors is unusual among exposed populations in that both genders and a wide range of ages were exposed, comparable with those of a general population (Ozasa *et al* 2012, Preston *et al* 2007). Most medically treated groups are more restricted in the age and gender mix. For example, the International Radiation Study of Cervical Cancer patients (IRSCC), a cohort of women followed up after treatment for cancer of the cervix, were all treated as adults, most above the age of 40 (Boice *et al* 1987, Boice *et al* 1988). Organ doses among those treated with radiotherapy tend to be higher than those received by the Japanese atomic bomb survivors, although there are some exceptions, e.g. breast doses in the IRSCC patients (Boice *et al* 1987, Boice *et al* 1988).

Occupationally exposed groups are also more restricted in their age and gender mix. For example, the cohorts of workers exposed in the nuclear industry (Cardis *et al* 2007, Muirhead *et al* 2009) are overwhelmingly male and exposed in adulthood, as are the groups of underground miners (National Research Council 1999). For these reasons, most standard-setting bodies (National Research Council 1999, National Research Council 2006, ICRP 2007) use the LSS as the basis for estimates of population cancer risk associated with exposure to low-LET radiation. For certain cancer sites and types of radiation exposures, other groups are occasionally used; in particular for high-LET (α -particle) exposure to the lung, underground miners are used (National Research Council 1999). However, lung-cancer risks estimated for the miners by applying those estimated in the LSS in combination with the current ICRP dosimetric model (ICRP 1994) are close to those that can be estimated directly (Little 2002).

5.3.3. Temporal patterns of risk for radiation-induced cancer. One of the principal uncertainties that surround the calculation of population cancer risks from epidemiological data results from the fact that few radiation-exposed cohorts have been followed up to extinction. For example, 58 years after the atomic bombings of Hiroshima and Nagasaki, about 42% of the survivors were still alive (Ozasa *et al* 2012). In attempting to calculate lifetime population cancer risks, it is therefore important to predict how risks might vary as a function of time after radiation exposure, in particular for that group for whom the uncertainties in projection of risk to the end of life are most uncertain, namely those who were exposed in childhood.

Analyses of solid cancers in the LSS and other exposed groups have found that the radiation-induced excess risk can be approximately described by a constant-relative-risk model (UNSCEAR 2008b). The time-constant excess relative risk (ERR) model assumes that if a dose of radiation is administered to a population, then, after some latent period, there is an increase in the cancer rate, the excess rate being proportional to the underlying cancer rate in an unirradiated population.

It is well known that for all cancer subtypes (including leukemia) the ERR diminishes with increasing age at exposure (UNSCEAR 2008b). For those irradiated in childhood there is some evidence of a reduction in the ERR of solid cancer 25 or more years after exposure. Therefore, even for solid cancers various factors have to be employed to modify the ERR. For many solid cancers a generalized relative risk model is commonly used, in which the cancer rate t years after exposure for gender s following exposure at age e to a dose D of radiation is given by

$$R(a, s, e, D) = r_0(a, s) [1 + F(D) \phi(t, e, s)], \quad (87)$$

where $r_0(a, s)$ is the cancer rate in the absence of irradiation, i.e. the baseline cancer rate, $a = t + e$ is the age at observation (attained age) of the person and $F(D)$ is the function determining the dose dependency of the cancer risk, to be discussed below. The expression $\phi(t, e, s)$ describes the modification to the ERR, $F(D)$, as a function of time since exposure t , age at exposure e and gender s . Similar types of model have been fitted to leukemia (Hsu *et al* 2013). However, generalized absolute risk models can also be used to model both solid cancers and leukemia, in which the cancer rate t years after exposure for gender s following exposure at age e to a dose D of radiation is given by:

$$R(a, s, e, D) = R_0(a, s) + F(D)\psi(t, e, s). \quad (88)$$

The expression $\psi(t, e, s)$ describes the modification to the EAR, $F(D)$, as a function of time since exposure t , age at exposure e and gender s .

Given appropriate forms of the modifying functions $\phi(t, e, s)$ and $\psi(t, e, s)$ of the relative and absolute risk respectively, equivalently good fits to the solid cancer and leukemia mortality dataset were achieved using both generalized ERR and generalized EAR models (Little *et al* 2008a). It is to some extent arbitrary which of these two models one uses. However, as can be seen from table 3, models with equivalent fit can yield slightly different lifetime population risks. The reason for this is that, as noted above, about 42% of the LSS cohort are still alive (Ozasa *et al* 2012), so that population risk calculations based on this dataset (used by many scientific committees (National Research Council 2006, ICRP 2007)) crucially depend on extrapolating the current mortality and incidence follow-up of this group to the end of life.

Uncertainties due to risk projection are greatest for solid cancers, because the radiation-associated excess risk in the LSS is still increasing (Ozasa *et al* 2012, Preston *et al* 2007). For leukemia the excess risk is reducing over time (Hsu *et al* 2013), and most models used predict very few radiation-associated leukemia deaths or cases from the current follow-up point in the LSS to extinction.

Use can also be made of so-called biologically based cancer risk models, as summarized by Little (2010). However, these would not be expected to yield materially different population cancer risks to those given by purely empirical models of the form discussed above, although they may be parametrically more parsimonious than purely empirical models.

5.3.4. Forms of cancer dose response. It has been customary to model the dose-response function $F(D)$ that appears in (87) and (88) in fits to biological (UNSCEAR 1993) and epidemiological data (UNSCEAR 2008b) by the linear-quadratic expression:

$$F(D) = \alpha D + \beta D^2. \quad (89)$$

There is significant curvilinearity in the dose response for leukemia in the LSS (Hsu *et al* 2013), although for solid cancers, apart from non-melanoma skin cancer (Ron *et al* 1998, Preston *et al* 2007), there is little evidence for anything other than a linear dose response in the Japanese cohort or in any other group (UNSCEAR 2008b). It should be noted that as well as modifications in effectiveness (per unit dose) relating to alterations in the total dose there are also possible variations of effectiveness as a result of dose fractionation (the process of splitting a given dose into a number of smaller doses suitably separated in time) and dose rate (UNSCEAR 1993). This is not surprising radiobiologically; by administering a given dose at progressively lower dose rates (i.e. giving the same total dose over longer periods of time), or by splitting it into many fractions, the biological system has time to repair the damage, so that the total damage induced will be less (UNSCEAR 1993). Therefore, although for cancers other than leukemia there is generally little justification for assuming anything other than a linear dose response, i.e. $\beta = 0$, it may nevertheless be justifiable to employ a dose and

dose-rate effectiveness factor (DDREF) other than 1. The DDREF is the factor by which one divides risks for high-dose and high-dose-rate exposure to obtain risks for low doses and low dose rates. The ICRP (2007) recommended that a DDREF of 2 be used together with models linear in dose for all cancer sites, on the basis largely of the observations in various epidemiological datasets. UNSCEAR (2008b) used linear-quadratic models to derive cancer risks (as shown in table 3), so that a DDREF should not be applied.

Another form of dose response, perhaps less commonly used, slightly generalizes (89):

$$F(D) = (\alpha D + \beta D^2) \exp(\gamma D), \quad (90)$$

and this form has been employed in fits to biological data (UNSCEAR 1993) and epidemiological data (Boice *et al* 1987, Thomas *et al* 1992, Little *et al* 1999). The $\alpha D + \beta D^2$ component represents the effect of (carcinogenic) mutation induction, while the $\exp(\gamma D)$ term represents the effect of cell sterilization or killing. In general the cell sterilization coefficient γ is negative. Essentially this implies that there is a competing effect due to cell killing which is greater at higher radiation doses. A dead cell cannot proliferate and become the focus of a malignant clone. Variant forms of the cell-sterilization term $\exp(\gamma D)$, incorporating higher powers of dose D , i.e. $\exp(\gamma D^k)$ for $k > 1$, are sometimes employed (Little and Charles 1997).

6. Relating task-based image quality to dose and risk

6.1. General considerations

The most basic relationship between image quality and radiation dose in x-ray and gamma-ray imaging is that they both depend, albeit in complicated ways, on the number of high-energy photons that interact with the patient's body. How one defines this number is different for external-source x-ray imaging (e.g. planar radiography, CT, digital breast tomosynthesis) and internal-source radionuclide imaging (e.g. planar nuclear medicine, SPECT, PET).

In external-source imaging, the dose distribution $d_j(\mathbf{r})$ for patient j is proportional to the mean number of x-ray photons (averaged over Poisson fluctuations) incident on the patient, denoted $\bar{N}_j^{\text{inc}}$. Similarly, in radionuclide imaging, the dose distribution is proportional to the mean number of gamma rays emitted by the radiotracer before it decays fully or is eliminated from the body. We shall denote this number also as $\bar{N}_j^{\text{inc}}$. In external-source imaging, $\bar{N}_j^{\text{inc}}$ can be controlled while holding all other factors constant by varying the current in the x-ray tube, and in radionuclide imaging it is controlled by varying the injected activity of the radiopharmaceutical.

For both external-source imaging and radionuclide imaging, we can write

$$d_j(\mathbf{r}) = \bar{N}_j^{\text{inc}} d_{1j}(\mathbf{r}), \quad (91)$$

where $d_{1j}(\mathbf{r})$ is interpreted as the mean dose at point $\mathbf{r}$ in patient j per incident photon. Of course, $d_{1j}(\mathbf{r})$ is a complicated function of the parameters of the imaging system and the anatomy of patient j .

Because the dose to organ k in patient j is a linear functional of $d_j(\mathbf{r})$ (see (81)), we also have

$$D_{kj} = \bar{N}_j^{\text{inc}} D_{1kj}, \quad (92)$$

where D_{1kj} is the dose to organ k of patient j per incident photon. Averaging both sides of (92) over an ensemble of patients, we can write the average organ dose as

$$\bar{D}_k = \bar{\bar{N}}^{\text{inc}} \bar{D}_{1k}, \quad (93)$$

where the overbar on D_k indicates an average over patients; the second overbar is added to $\bar{N}^{\text{inc}}$ to account for any patient-specific variations in imaging protocol.

In contrast to the linear relations between dose and number of incident photons in (92) and (93), image quality depends nonlinearly on the number of detected photons and hence on the number of incident photons and the dose.

Let Q be some task-based figure of merit for an ensemble of patients undergoing a specific diagnostic examination with a specific instrument and imaging protocol. The figure of merit depends in a complicated way on the task, the patient ensemble and various system and algorithmic parameters, but if we hold these factors constant and focus on the effect of the number of detected photons in the image, it always turns out that more photons are better and that image quality can be increased by increasing the radiation dose to the patient.

To be more precise, suppose an imaging examination for patient j is based on detection of N_j^{det} photons, where N_j^{det} is a Poisson random variable with mean $\bar{N}_j^{\text{det}}$. If we average this number over the ensemble of patients undergoing the given examination, we can denote the resulting mean total number of detected photons per patient as $\bar{\bar{N}}^{\text{det}}$, and with all other factors held constant, it will be found that Q is a nondecreasing function of $\bar{\bar{N}}^{\text{det}}$. This statement is true even if there are other sources of noise besides Poisson statistics and even if the imaging instrument does not operate in a photon-counting mode.

To make the connection between image quality and radiation dose, we define an average detection efficiency (averaged over patients in the ensemble) as

$$\bar{\eta} \equiv \frac{\bar{\bar{N}}^{\text{det}}}{\bar{\bar{N}}^{\text{inc}}}. \quad (94)$$

Thus $\bar{\eta}$ is the average fraction of photons incident on the patient that result in measurable signals in the detector system in the imaging instrument. For external-source x-ray imaging, $\bar{\eta}$ is determined by the transmission through the patient and the geometry and efficiency of the detectors. For radionuclide imaging, there is an additional factor of the image-acquisition time relative to the decay and washout of the radiotracer; all gamma-ray emissions in the patient's body contribute to the dose, but only those during the acquisition time contribute to the image.

We can now find an expression for the derivative of the image-quality metric Q with respect to the average dose $\bar{D}_k$ to any organ of interest:

$$\frac{dQ}{d\bar{D}_k} = \frac{dQ/d\bar{\bar{N}}^{\text{inc}}}{d\bar{D}_k/d\bar{\bar{N}}^{\text{inc}}} = \bar{\eta} \frac{dQ/d\bar{\bar{N}}^{\text{det}}}{d\bar{D}_k/d\bar{\bar{N}}^{\text{inc}}} = \frac{\bar{\eta}}{\bar{D}_{k1}} \frac{dQ}{d\bar{\bar{N}}^{\text{det}}}, \quad (95)$$

where $\bar{D}_{k1} \equiv d\bar{D}_k/d\bar{\bar{N}}^{\text{inc}}$ is the mean dose to organ k for one incident photon. Because dose depends linearly on the number of incident photons, this quantity can be estimated by any Monte Carlo dose-calculation program just by dividing the computed dose by the number of incident photons used. Similarly, the factor $dQ/d\bar{\bar{N}}^{\text{det}}$ is the increment in Q for one additional detected photon. This derivative can be determined by analytic methods surveyed in sections 6.3 and 6.4 or numerically by use of image-quality programs available on the internet (Ljungberg and King 2012), and we can then use (95) to relate the result to the derivative of the task-based image quality with respect to the mean organ dose. The conversion factor is simply the mean detection efficiency divided by the mean organ dose produced by one incident photon.

In some imaging studies, it will be possible to identify one or two critical organs that are at risk of damage. In these cases, (95) can be used to determine when an increased organ dose will result in little or no increase in image quality.

The relations in (95) will still hold if we perform a weighted average of the organ doses to get the effective dose E , as defined in (82), but (95) has the huge advantage that organ doses can be related directly to adverse effects of the radiation on particular critical organs, which is no longer possible if we try to summarize the organ doses with an effective dose.

6.2. Dose dependence of mean vectors and covariance matrices

In many cases, especially for complicated tasks and realistic object models, there is no substitute for simulation and numerical evaluation of figures of merit if one wants to study the relationship of image quality to radiation dose and patient risk. Nevertheless, some general conclusions and useful scaling laws can be derived from the analytical expressions for image-quality metrics developed in section 3.

In order to discuss how the various expressions for task performance developed in section 3 vary with absorbed dose, we must first decide how quantities such as the object $\mathbf{f}$, system operator $\mathcal{H}$, the mean and covariance of the data and various probability density functions relate to the mean number of detected photons. The discussion is different for linear operators as in SPECT or PET and nonlinear operators as in CT and mammography, and it is different for photon-counting and integrating detectors. It is also different for linear and nonlinear observers.

6.2.1. Emission computed tomography. In SPECT and PET, the object of interest is the spatiotemporal distribution of a radiopharmaceutical, described in detail by the function $f(\mathbf{r}, t)$, which describes the mean number of emitted gamma-ray photons per second per unit volume in the patient's body. The mean number of photons incident (from within) on the patient's body, following an injection at $t = 0$ is given by

$$\bar{N}^{\text{inc}}(\mathbf{f}) = \int_0^\infty dt \int_{\text{body}} d^3r f(\mathbf{r}, t), \quad (96)$$

where the single overbar indicates that $\bar{N}^{\text{inc}}(\mathbf{f})$ is averaged only over Poisson statistics, not over object variability, so it remains a function of $\mathbf{f}$. We can define a normalized object function $f_0(\mathbf{r}, t)$ by

$$f_0(\mathbf{r}, t) = \frac{f(\mathbf{r}, t)}{\bar{N}^{\text{inc}}(\mathbf{f})}, \quad (97)$$

so that $f(\mathbf{r}, t) = \bar{N}^{\text{inc}}(\mathbf{f}) f_0(\mathbf{r}, t)$.

Because SPECT and PET systems are linear continuous-to-discrete mappings, the mean detector output (averaged only over Poisson statistics) is given by (see (2))

$$\bar{g}_m(\mathbf{f}) = \int_{t_1}^{t_2} dt \int_{\text{body}} d^3r h_m(\mathbf{r}, t) f(\mathbf{r}, t) = \bar{N}^{\text{inc}}(\mathbf{f}) \int_{t_1}^{t_2} dt \int_{\text{body}} d^3r h_m(\mathbf{r}, t) f_0(\mathbf{r}, t), \quad (98)$$

where the image acquisition takes place over $t_1 < t < t_2$ and $h_m(\mathbf{r}, t)$ is the probability (not PDF) that a gamma-ray photon emitted at time t and location $\mathbf{r}$ will be detected in detector bin m . The mean number of detected photons, in all bins, for a given object is

$$\bar{N}^{\text{det}}(\mathbf{f}) = \sum_{m=1}^M \bar{g}_m(\mathbf{f}) = \int_{t_1}^{t_2} dt \int_{\text{body}} d^3r s(\mathbf{r}, t) f(\mathbf{r}, t) = \bar{N}^{\text{inc}}(\mathbf{f}) \int_{t_1}^{t_2} dt \int_{\text{body}} d^3r s(\mathbf{r}, t) f_0(\mathbf{r}, t), \quad (99)$$

where $s(\mathbf{r}, t)$, defined as $\sum_{m=1}^M h_m(\mathbf{r}, t)$ is called the system sensitivity function; in SPECT it can vary with time as the detector rotates around the patient's body.

The mean number of incident photons, $\bar{N}^{\text{inc}}(\mathbf{f})$, varies from patient to patient because the total injected activity is not always the same and because different patients may have different biological elimination rates. It is reasonable to assume that these effects are statistically independent of the more complicated patient-to-patient variations captured in the normalized distribution $f_0(\mathbf{r}, t)$. Thus the efficiency defined in (94) is given, for patients in group $\mathbf{C}$, by

$$\bar{\eta} = \frac{\langle \bar{N}^{\text{det}}(\mathbf{f}) \rangle_{\mathbf{f}|\mathbf{C}}}{\langle \bar{N}^{\text{inc}}(\mathbf{f}) \rangle_{\mathbf{f}|\mathbf{C}}} = \int_{t_1}^{t_2} dt \int_{\text{body}} d^3r \langle s(\mathbf{r}, t) f_0(\mathbf{r}, t) \rangle_{\mathbf{f}|\mathbf{C}}, \quad (100)$$

where the expectation within the integral is the average of the normalized object distribution weighted by the system sensitivity. We see immediately that the efficiency can be increased just by increasing the acquisition time $t_2 - t_1$, though practical issues of patient motion and patient throughput may intrude.

As we saw in section 3, many figures of merit for image quality depend on some form of data covariance matrix. By (16) or (21), these matrices can always be decomposed into two terms, one arising from measurement noise and one from the randomness in the object being imaged. As noted in section 2.2, these terms have different dependences on $\bar{N}^{\text{det}}$ and hence on any measure of the radiation dose administered during the imaging procedure.

For SPECT and PET, the noise term in the covariance expansion is obtained from (8) by averaging over object variability, which requires only adding an overbar to $\mathbf{f}$. From this result and (97) we obtain

$$[\bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}}]_{mm'} = (\mathcal{H}\bar{\mathbf{f}})_m \delta_{mm'} = \bar{N}^{\text{inc}} (\mathcal{H}\bar{\mathbf{f}}_0)_m \delta_{mm'}, \quad (101)$$

where the last step has used the fact that $\mathcal{H}$ is linear for SPECT and PET. We can express (101) in vector form, without explicit components, as

$$\bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}} = \text{Diag}(\mathcal{H}\bar{\mathbf{f}}) = \bar{N}^{\text{inc}} \text{Diag}(\mathcal{H}\bar{\mathbf{f}}_0) \equiv \bar{N}^{\text{inc}} \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{C}}^{\text{noise}}, \quad (102)$$

where Diag indicates a diagonal matrix and the final covariance on the right is the noise covariance for the normalized object $\mathbf{f}_0$. Similar manipulations show that

$$\mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}} = [\bar{N}^{\text{inc}}]^2 \mathbf{K}_{\mathbf{g}|\mathbf{C}}^{\text{obj}}. \quad (103)$$

Thus, for the linear system operators in SPECT and PET, the noise covariance term for the raw projection data varies linearly with the mean number of incident photons, and hence linearly with any measure of radiation dose, and the object-variability term varies quadratically. The same conclusions hold if we consider SPECT and PET images reconstructed with linear or nonlinear algorithms.

6.2.2. Transmission computed tomography. The situation is more complicated for x-ray modalities where the system operator is nonlinear and integrating detectors are usually used rather than photon-counting detectors as in SPECT and PET. We begin by examining the statistics of the raw detector outputs, and how they vary with dose, and then we derive expressions for the mean vectors and covariance matrices after logarithmic mapping and also after image reconstruction.

The object of interest in CT is the 3D distribution of x-ray attenuation coefficient in some portion of the patient's body. Though the attenuation coefficient is conventionally called μ , we refer to it as $\mathbf{f}$ in order to maintain a common notation across modalities. As noted in section 2, we can continue to write $\bar{\mathbf{g}}(\mathbf{f}) = \mathcal{H}\mathbf{f}$, but the system operator is now nonlinear because

$\mathbf{f}$ appears in an exponent. Moreover, in a complete treatment, the energy spectrum of the source, the energy dependence of the attenuation coefficient and the variation of the detector efficiency with photon energy would all have to be taken into account, yielding a very complicated nonlinear expression for the mean data (see section 16.1.4 in Barrett and Myers (2004), especially equation (16.109)).

For the purposes of this paper, however, we approximate the mean number of primary (unscattered) x-ray photons reaching the m th detector element in some exposure time by

$$\bar{g}_m^{\text{pri}}(\mathbf{f}) = \bar{N}_m^{\text{inc}} \exp [-(\mathcal{R}\mathbf{f})_m], \quad (104)$$

where $\mathcal{R}$ is a linear operator that integrates the object over a thin tube of directions from the x-ray source to the detector element and $\bar{N}_m^{\text{inc}}$ is the mean number of photons that would be incident on the detector element in the absence of the object, and hence it is the number incident on the object in the tube of integration when the object is present. The steps involved in getting to (104) are detailed in section 16.1.7 in Barrett and Myers (2004).

In addition to the primary photons, there are also some photons scattered in the patient's body that reach the detector element, so we can write

$$\bar{g}_m(\mathbf{f}) = \bar{g}_m^{\text{pri}}(\mathbf{f}) + \bar{g}_m^{\text{sc}}(\mathbf{f}). \quad (105)$$

Detected scattered photons convey very little information about the task, but it is important that we retain them in the noise analysis.

An initial step in the processing of CT data is usually to normalize each detector output to the number of incident photons (presumably known by careful scanner calibration) and take the negative logarithm of the ratio:

$$y_m \equiv -\ln \left(\frac{g_m}{\bar{N}_m^{\text{inc}}} \right). \quad (106)$$

The conditional mean of y_m is

$$\bar{y}_m(\mathbf{f}) = \ln(\bar{N}_m^{\text{inc}}) - \langle \ln(g_m) \rangle_{\mathbf{g}|\mathbf{f}} = \ln(\bar{N}_m^{\text{inc}}) - \langle \ln[\bar{g}_m(\mathbf{f}) + g_m - \bar{g}_m(\mathbf{f})] \rangle_{\mathbf{g}|\mathbf{f}} \quad (107)$$

$$\approx -\ln \left(\frac{\bar{g}_m(\mathbf{f})}{\bar{N}_m^{\text{inc}}} \right) + \frac{1}{2} \frac{\alpha \bar{g}_m(\mathbf{f}) + \sigma_m^2}{[\bar{g}_m(\mathbf{f})]^2}, \quad (108)$$

where we have used a truncated Taylor expansion, $\ln(1+x) \approx x - \frac{1}{2}x^2$, and the last step has used the noise model of (9). Notice that the variance of g_m affects the mean of y_m because of the logarithmic nonlinearity.

The usual assumption in CT is that $\bar{y}_m(\mathbf{f})$ is equal to the tube integral $(\mathcal{R}\mathbf{f})_m$, but it isn't because of the last term in (108) and the scattered radiation. Instead we get

$$\bar{y}_m(\mathbf{f}) = (\mathcal{R}\mathbf{f})_m + b_m(\mathbf{f}), \quad (109)$$

where

$$b_m(\mathbf{f}) = \frac{1}{2} \frac{\alpha \bar{g}_m(\mathbf{f}) + \sigma_m^2}{[\bar{g}_m(\mathbf{f})]^2} - \ln \left[1 + \frac{\bar{g}_m^{\text{sc}}(\mathbf{f})}{\bar{g}_m^{\text{pri}}(\mathbf{f})} \right]. \quad (110)$$

The second term in b_m is negligible if the scatter-to-primary ratio is small, and the first term is small if the mean number of detected photons for one detector element is large; neither assumption is guaranteed to be valid in practice.

Calculation of the conditional covariance matrix, $\mathbf{K}_{\mathbf{y}|\mathbf{f}}$, follows the same pattern that led to (108), but with a different truncated Taylor series: $[\ln(1+x)]^2 \approx x^2$. The result is

$$[\mathbf{K}_{\mathbf{y}|\mathbf{f}}]_{mm'} = \frac{\alpha \bar{g}_m(\mathbf{f}) + \sigma_m^2}{[\bar{g}_m(\mathbf{f})]^2} \delta_{mm'}. \quad (111)$$

Averaging over all objects $\mathbf{f}$ in set $\mathbf{C}$ yields the average noise covariance matrix for data vector $\mathbf{y}$:

$$[\bar{\mathbf{K}}_{\mathbf{y}|\mathbf{C}}^{\text{noise}}]_{mm'} = \left\langle \frac{\alpha \bar{g}_m(\mathbf{f}) + \sigma_m^2}{[\bar{g}_m(\mathbf{f})]^2} \right\rangle_{\mathbf{f} \in \mathbf{C}} \delta_{mm'}. \quad (112)$$

In principle, this average can be estimated by collecting raw projection data for a large number of real or accurately simulated patients in some defined set $\mathbf{C}$ and estimating the ensemble average with an arithmetic average, separately for each index m .

The object term in the covariance matrix for $\mathbf{y}$ is complicated by the object-dependent bias b_m . We should write (see (18))

$$\mathbf{K}_{\mathbf{y}|\mathbf{C}}^{\text{obj}} \equiv \langle [\bar{\mathbf{y}}(\mathbf{f}) - \bar{\mathbf{y}}][\bar{\mathbf{y}}(\mathbf{f}) - \bar{\mathbf{y}}]^t \rangle_{\mathbf{f} \in \mathbf{C}} = \langle [\mathcal{R}(\mathbf{f} - \bar{\mathbf{f}}) + (\mathbf{b} - \bar{\mathbf{b}})][\mathcal{R}(\mathbf{f} - \bar{\mathbf{f}}) + (\mathbf{b} - \bar{\mathbf{b}})]^t \rangle_{\mathbf{f} \in \mathbf{C}}, \quad (113)$$

but for simplicity we assume that $b_m - \bar{b}_m$ is small, and we find

$$\mathbf{K}_{\mathbf{y}|\mathbf{C}}^{\text{obj}} \approx \langle [\mathcal{R}(\mathbf{f} - \bar{\mathbf{f}})][\mathcal{R}(\mathbf{f} - \bar{\mathbf{f}})]^t \rangle_{\mathbf{f} \in \mathbf{C}} = \mathcal{R} \mathcal{K}_{\mathbf{f}|\mathbf{C}} \mathcal{R}^t, \quad (114)$$

where $\mathcal{K}_{\mathbf{f}|\mathbf{C}}$ was introduced in (19) and (20).

By analogy to (21), we can express the covariance matrix for a CT image reconstructed by a linear algorithm $\mathcal{O}$, such that $\hat{\mathbf{f}} = \mathcal{O}\mathbf{y}$, as

$$\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{C}} = \mathcal{O} \bar{\mathbf{K}}_{\mathbf{y}|\mathbf{C}}^{\text{noise}} \mathcal{O}^t + \mathcal{O} \mathbf{K}_{\mathbf{y}|\mathbf{C}}^{\text{obj}} \mathcal{O}^t = \mathcal{O} \bar{\mathbf{K}}_{\mathbf{y}|\mathbf{C}}^{\text{noise}} \mathcal{O}^t + \mathcal{O} \mathcal{H} \mathcal{K}_{\mathbf{f}} \mathcal{H}^t \mathcal{O}^t. \quad (115)$$

6.2.3. Comparison of ECT and TCT. We are now able to discuss the relationship of these various mean vectors and covariance matrices to radiation dose and to compare the results for transmission computed tomography (TCT, or CT for short) to what we found earlier for ECT (SPECT and PET). One obvious difference is that the object in ECT is the source of the ionizing radiation, so we immediately have that $\mathbf{f}_{\text{ECT}}$ is linearly related to the absorbed dose at any point or to any organ dose. In CT, on the other hand, $\mathbf{f}_{\text{CT}}$ is an x-ray attenuation coefficient, hence unrelated to dose. In both cases, however, $\mathcal{H}\mathbf{f}$ is linearly related to dose; the proportionality to dose comes from $\mathbf{f}$ in ECT and from $\mathcal{H}$ in TCT.

These and other dose dependences are summarized in table 4, and their relation to figures of merit for image quality are discussed in sections 6.3 and 6.4.

6.3. Dose dependence of figures of merit for classification tasks

The various task-based figures of merit introduced in section 3 can be classified according to the task (classification, estimation or both) and how much knowledge of the underlying object and data statistics is used by the observer. Even if the observers do not make use of certain statistical properties, the performance of the observers must necessarily depend on these statistics and hence on the radiation dose. In this section we illustrate these points for classification tasks with various examples.

6.3.1. Signal detection with linear discriminants and dose-independent templates. Consider first a simple linear discriminant for the task of detecting the presence of a lesion at a single,

Table 4. Dose dependences of mean vectors and covariance matrices for emission computed tomography (ECT) and transmission computed tomography (TCT). Photon-counting detectors are assumed for ECT and integrating detectors are assumed for TCT. Symbol D indicates a generic dose quantity (such as absorbed dose at a point, organ dose or effective dose), and D^2 , for example, indicates that the quantity in the left column varies quadratically with dose; D^0 is used for a quantity independent of dose. The notation $\approx D^0$ in the TCT column means that the quantity in the first column would be independent of dose if we neglected the noise-induced bias resulting from the logarithmic step in CT reconstruction. Angle brackets, included only when they affect the dose dependence, denote an average over object variability. Constant c relates to the dose-independent electronic noise in integrating detectors, and constant a serves to scale the dose-dependent part of the noise covariance to the dose-independent part. Double right arrow indicates asymptotic dependence for large dose. All other symbols are defined in the text.

Quantity	ECT	TCT
$\mathbf{f}, \bar{\mathbf{f}}$	D	D^0
$\mathcal{H}$	D^0	D
$\bar{\mathbf{g}}, \bar{\bar{\mathbf{g}}}$	D	D
$\bar{\mathbf{y}}, \bar{\bar{\mathbf{y}}}$	NA	$\approx D^0$
$\bar{\mathbf{f}}, \bar{\bar{\mathbf{f}}}$	D	$\approx D^0$
$\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}$	D	$\langle bD + c \rangle \Rightarrow D$
$\mathbf{K}_{\mathbf{g}}^{\text{obj}}$	D^2	D^2
$\bar{\mathbf{K}}_{\mathbf{y}}^{\text{noise}}$	NA	$\left\langle \frac{bD + c}{D^2} \right\rangle \Rightarrow D^{-1}$
$\mathbf{K}_{\bar{\mathbf{y}}}^{\text{obj}}$	NA	D^0
$\bar{\mathbf{K}}_{\bar{\mathbf{f}}}^{\text{noise}}$	D	$\left\langle \frac{bD + c}{D^2} \right\rangle \Rightarrow D^{-1}$
$\mathbf{K}_{\bar{\bar{\mathbf{f}}}}^{\text{obj}}$	D^2	D^0

specified location. If the task is performed with raw projection data $\mathbf{g}$ (as opposed to a reconstructed image $\bar{\mathbf{f}}$), then the linear discriminant is defined by a template $\mathbf{w}_{ld}$, and the test statistic is a simple scalar product:

$$t_{ld}(\mathbf{g}) = \mathbf{w}_{ld}^t \mathbf{g}. \quad (116)$$

We saw in (46) that the Hotelling observer uses a test statistic of this form, but there the template depends on the data mean vectors and covariance matrices, hence on the radiation dose. Here we consider templates that do not depend on the dose.

The template might have been designed for a certain signal $\mathbf{s}_0$ and without regard to the background in which it is immersed, but if it is applied in a real clinical setting (as opposed to a simulation study or a simplified psychophysical study with synthetic data), then the inevitable random variations in signal and background determine the statistics of $\mathbf{g}$ and hence of $t_{ld}(\mathbf{g})$.

These statistics can be different for images that do not contain the signal and those that do, so we define two subsets of objects $\mathbf{C}_0$ and $\mathbf{C}_1$, respectively, and we denote averages over data vectors in the subsets with the corresponding subscripts. Thus

$$\bar{t}_0 \equiv \left\langle \left\langle \mathbf{w}_{ld}^t \mathbf{g} \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}_0} = \mathbf{w}_{ld}^t \bar{\mathbf{g}}_0, \quad \bar{t}_1 \equiv \left\langle \left\langle \mathbf{w}_{ld}^t \mathbf{g} \right\rangle_{\mathbf{g}|\mathbf{f}} \right\rangle_{\mathbf{f}|\mathbf{C}_1} = \mathbf{w}_{ld}^t \bar{\mathbf{g}}_1. \quad (117)$$

Because $t_{ld}(\mathbf{g})$ is a scalar, it does not have a covariance matrix, but its variance can still be decomposed into terms arising from noise and object variability. From (16), it is easy to show that

$$\text{Var} \{t_{ld}(\mathbf{g}) | \mathbf{C}_j\} = \mathbf{w}_{ld}^t \mathbf{K}_{\mathbf{g}|\mathbf{C}_j} \mathbf{w}_{ld} = \mathbf{w}_{ld}^t \bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}} \mathbf{w}_{ld} + \mathbf{w}_{ld}^t \mathbf{K}_{\bar{\mathbf{g}}|\mathbf{C}_j}^{\text{obj}} \mathbf{w}_{ld}, \quad (j = 0, 1). \quad (118)$$

The signal-to-noise ratio associated with $t_{ld}(\mathbf{g})$ is defined by (see (26))

$$\text{SNR}_{ld}^2 = \frac{[\bar{t}_1 - \bar{t}_0]^2}{\frac{1}{2}[\text{Var} \{t_{ld}(\mathbf{g}) | \mathbf{C}_0\} + \text{Var} \{t_{ld}(\mathbf{g}) | \mathbf{C}_1\}]}. \quad (119)$$

For a weak signal, it will usually be a valid approximation to replace the average variance in the denominator here with the signal-absent variance (see section 3.3).

The template $\mathbf{w}_{ld}$ is independent of dose, and the dose dependence of the terms in (117) and (118) can be found in table 4. For ECT, we find that

$$\text{SNR}_{ld}^2 \propto \frac{D^2}{aD^2 + bD}, \quad (120)$$

where a and b are constants that depend not only on the task, system operator and object and image statistics, but also on the particular measure of dose used. If D is an organ dose, D_k , the constants depend on which organ is chosen.

For TCT with integrating detectors and the task performed on raw data, we find a different dependence,

$$\text{SNR}_{ld}^2 \propto \frac{D^2}{aD^2 + bD + c}, \quad (121)$$

where c arises from the electronic readout noise, introduced in (9). The discussion below that equation shows that the readout noise is negligible if $\bar{g}_m(\mathbf{f}) \gg \sigma_m^2$, where σ_m is the standard deviation of the noise in detector m in units of the mean output for one x-ray photon, but this condition might be difficult to achieve for small detector pixels and low-dose imaging.

In TCT, the situation is more complicated when the detection task is performed on reconstructed images rather than raw projections, as it almost always will be in clinical practice. In that case $\mathbf{w}$ is a vector with the same dimension as $\hat{\mathbf{f}}$, and $t_{ld}(\hat{\mathbf{f}}) = \mathbf{w}_{ld}^t \hat{\mathbf{f}}$. Equations (117)–(119) still hold if we replace $\mathbf{g}$ with $\hat{\mathbf{f}}$ everywhere, but when we consult the TCT column of table 4 we find

$$\text{SNR}_{ld}^2 \propto \frac{1}{a + \left\langle \frac{bD + c}{D^2} \right\rangle_{\mathbf{f}|\mathbf{C}_0}}, \quad (122)$$

where a , b and c depend on the object statistics but are independent of D . The necessity for averaging the noise term over objects in this manner arises from the logarithmic nonlinearity used in CT reconstruction.

It follows from (121) and (122) that the SNR^2 for any linear discriminant with a dose-independent template must asymptotically approach an upper limit as dose is increased.

6.3.2. Signal detection from raw projection data with Hotelling observers. As we saw in section 3.3, the Hotelling observer is also a linear discriminant of the form (116), but now the

observer has full knowledge of the mean and covariance matrix of the data and uses it optimally to construct the template $\mathbf{w}$; as a result the template depends on the dose.

A convenient way to study the dose-dependence of the performance of the Hotelling observer is to use the Karhunen-Loève (KL) expansion of the estimated covariance matrix $\hat{\mathbf{K}}_{\mathbf{g}}$, defined in (52). In brief, KL is an expansion of a covariance matrix in terms of its eigenvectors, with its eigenvalues as coefficients.

Following Kupinski *et al* (2007), we rewrite (52) as

$$\hat{\mathbf{K}}_{\mathbf{g}} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2} \left\{ \mathbf{I} + [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \mathbf{R} \mathbf{R}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \right\} [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2}, \quad (123)$$

where $\mathbf{I}$ is the $M \times M$ unit matrix and $\mathbf{R}$ is the $M \times N_s$ matrix of noise-free sample images as defined by (49) and (50). We recognize that $[\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2}$ has the same form as the prewhitening operator, introduced in (43), but it prewhitens with respect to the average noise covariance only. Because $\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}$ is often a diagonal matrix, as in (9), it can be very simple in practice to perform this kind of prewhitening operation.

After prewhitening in this sense, the data vector $\mathbf{g}$ and the matrix $\mathbf{R}$ become, respectively,

$$\tilde{\mathbf{g}} \equiv [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \mathbf{g}, \quad \tilde{\mathbf{R}} \equiv [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \mathbf{R}. \quad (124)$$

Thus (123) becomes

$$\hat{\mathbf{K}}_{\mathbf{g}} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2} [\mathbf{I} + \tilde{\mathbf{R}} \tilde{\mathbf{R}}^t] [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2}. \quad (125)$$

The next step is to solve the eigenvalue problem for the $M \times M$ prewhitened sample covariance matrix $\tilde{\mathbf{R}} \tilde{\mathbf{R}}^t$:

$$\tilde{\mathbf{R}} \tilde{\mathbf{R}}^t \boldsymbol{\phi}_j = \mu_j \boldsymbol{\phi}_j. \quad (126)$$

If the complete set of $M \times 1$ eigenvectors, $\{\boldsymbol{\phi}_j, j = 1, \dots, M\}$, is found, it will form a complete orthonormal basis for the M -dimensional data space.

However, because a sample covariance matrix with $N_s < M$ (where N_s is the number of samples) has rank $N_s - 1$, only $N_s - 1$ of the eigenvalues will be nonzero. We can find these nonzero eigenvalues and the corresponding eigenvectors by solving the eigenvalue problem for the much smaller $N_s \times N_s$ matrix $\tilde{\mathbf{R}}^t \tilde{\mathbf{R}}$:

$$\tilde{\mathbf{R}}^t \tilde{\mathbf{R}} \boldsymbol{\psi}_j = \mu_j \boldsymbol{\psi}_j. \quad (127)$$

It can be shown that all of these eigenvalues are nonnegative, and it is convenient to order them from largest to smallest, i.e. $\mu_1 \geq \mu_2 \geq \mu_3 \geq \dots \geq \mu_{N_s-1} > 0$.

Equation (127) can be solved by standard linear-algebra programs such as Matlab or Lapack for N_s up to 10^4 or so, and once the $N_s \times 1$ eigenvectors $\boldsymbol{\psi}_j$ and the corresponding eigenvalues μ_j are determined, the requisite $M \times 1$ eigenvectors can be found from (Barrett and Myers 2004)

$$\boldsymbol{\phi}_j = \frac{1}{\sqrt{\mu_j}} \tilde{\mathbf{R}} \boldsymbol{\psi}_j, \quad \text{for } j < N_s \text{ (i.e. } \mu_j \neq 0 \text{)}. \quad (128)$$

With these eigenvectors and eigenvalues, we can express $\tilde{\mathbf{R}} \tilde{\mathbf{R}}^t$ exactly by its spectral representation,

$$\tilde{\mathbf{R}} \tilde{\mathbf{R}}^t = \sum_{j=1}^{N_s-1} \mu_j \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t. \quad (129)$$

If we choose $M - N_s + 1$ additional vectors that form an orthonormal set and satisfy $\bar{\mathbf{R}}\bar{\mathbf{R}}^t\boldsymbol{\phi}_j = 0, N_s < j \leq M$, then we can express the $M \times M$ unit matrix as

$$\mathbf{I} = \sum_{j=1}^M \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t. \quad (130)$$

It turns out that we will not need any particular eigenvectors for $j \geq N_s$; it will suffice to know that they exist.

The estimated covariance matrix in the KL domain is now

$$\hat{\mathbf{K}}_{\mathbf{g}} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2} \left[\sum_{j=1}^M (1 + \mu_j) \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t \right] [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{1/2}, \quad (131)$$

and by the orthonormality of the eigenfunctions, the inverse of the estimated covariance is

$$\hat{\mathbf{K}}_{\mathbf{g}}^{-1} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \left[\sum_{j=1}^M \frac{1}{1 + \mu_j} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t \right] [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2}. \quad (132)$$

If we use $\frac{1}{1 + \mu_j} = 1 - \frac{\mu_j}{1 + \mu_j}$, we find

$$\hat{\mathbf{K}}_{\mathbf{g}}^{-1} = [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} - [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \left[\sum_{j=1}^{N_s-1} \frac{\mu_j}{1 + \mu_j} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t \right] [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2}. \quad (133)$$

This form of the inverse is equivalent to (53) but now in terms of KL eigenfunctions.

The estimated Hotelling FOM, $\widehat{\text{SNR}}_{\text{Hot}}^2 \equiv \Delta \bar{\mathbf{g}}^t \hat{\mathbf{K}}_{\mathbf{g}}^{-1} \Delta \bar{\mathbf{g}}$, is given in terms of these eigenfunctions by

$$\widehat{\text{SNR}}_{\text{Hot}}^2 = \bar{\mathbf{g}}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} \Delta \bar{\mathbf{g}} - \Delta \bar{\mathbf{g}}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \left[\sum_{j=1}^{N_s-1} \frac{\mu_j}{1 + \mu_j} \boldsymbol{\phi}_j \boldsymbol{\phi}_j^t \right] [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \Delta \bar{\mathbf{g}}. \quad (134)$$

The first term in this expression is exactly what we would have for an SKE/BKE problem, and the second term (with its minus sign) is the estimated *reduction* in the FOM resulting from the object randomness. Because $\frac{\mu_j}{1 + \mu_j} < 1$, the overall estimated SNR^2 cannot go negative.

From table 4, we see that $\bar{\mathbf{g}}, \bar{\mathbf{g}}$ and hence $\Delta \bar{\mathbf{g}}$ are $\propto D$ and that $[\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \propto D^{-1/2}$ in ECT. It follows that $\bar{\mathbf{R}}\bar{\mathbf{R}}^t \propto D$. Moreover, because the eigenvectors $\boldsymbol{\phi}_j$ always have unit norm, $\mu_j \propto D$. Thus we can write $\mu_j = \mu_{j1} D$, where μ_{j1} is the eigenvalue found by solving (127) for $D = 1$ in whatever units we wish (e.g. Gy or mGY) and for whatever particular dose quantity we are discussing (e.g. organ dose, ROI dose or even effective dose.)

Assembling these results, we find that the estimated FOM scales with dose as

$$\widehat{\text{SNR}}_{\text{Hot}}^2 = aD - \sum_{j=1}^{N_s-1} |\gamma_j|^2 \frac{\mu_{j1} D^2}{1 + \mu_{j1} D}, \quad (135)$$

where a is $\Delta \bar{\mathbf{g}}^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1} \Delta \bar{\mathbf{g}}$ as computed for $D = 1$ and γ_j is the difference signal after prewhitening for the noise covariance and projecting it onto the j th basis vector in the KL expansion of the object covariance:

$$\gamma_j = \boldsymbol{\phi}_j^t [\bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}}]^{-1/2} \Delta \bar{\mathbf{g}}. \quad (136)$$

The result in (134) holds for TCT as well as ECT, in spite of the fact that the system operator $\mathcal{H}$ is nonlinear for TCT; the exponential nonlinearity is included in the simulation code that generates noise-free sample images. From results in table 4, we can also show that (135) holds for TCT provided the excess detector noise (c in the table) can be neglected.

6.3.3. Hotelling observer applied to reconstructed images. So far we have discussed application of the Hotelling observer to raw projection data $\mathbf{g}$, but the expression in (134) still applies formally if we substitute the reconstruction $\hat{\mathbf{f}}$ for $\mathbf{g}$ everywhere. The practical difference is that the prewhitening operator is more complicated with reconstructions.

For linear algorithms in ECT, we know from (21) that $\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}} = \mathbf{O} \bar{\mathbf{K}}_{\mathbf{g}}^{\text{noise}} \mathbf{O}^t$, where the reconstruction operator $\mathbf{O}$ is an $N \times M$ matrix with elements O_{nm} , with index n specifying a voxel located at position $\mathbf{r}_n$ in the reconstructed image. With the pure Poisson noise model of (8), we find (see Barrett and Myers (2004), section 15.2.6)

$$\left[\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}} \right]_{nn'} = \sum_{m=1}^M O_{nm} O_{n'm} [\mathcal{H}\mathbf{f}]_m. \quad (137)$$

To understand this expression and its implications for prewhitening, it is useful to look at the elements of the conditional covariance matrix for a single object, $\left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{\text{noise}} \right]_{nn'}$, obtained from (137) just by omitting the overbars on $\mathbf{K}$ and $\mathbf{f}$. With (2), we find

$$\left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{\text{noise}} \right]_{nn'} = \sum_{m=1}^M O_{nm} O_{n'm} \int_{\infty} d^3r h_m(\mathbf{r}) f(\mathbf{r}). \quad (138)$$

The conditional variance of the reconstruction at voxel n is the nn diagonal element of this covariance matrix, which can be written as

$$\text{Var}(\hat{f}_n) = \left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{\text{noise}} \right]_{nn} = \int_{\infty} d^3r \mathfrak{N}_n(\mathbf{r}) f(\mathbf{r}), \quad (139)$$

where $\mathfrak{N}_n(\mathbf{r})$, called the *noise kernel* (see Barrett and Myers (2004), section 15.2.6), is defined by

$$\mathfrak{N}_n(\mathbf{r}) \equiv \sum_{m=1}^M [O_{nm}]^2 h_m(\mathbf{r}). \quad (140)$$

In 2D filtered backprojection (FBP), $\mathfrak{N}_n(\mathbf{r})$ varies approximately as $|\mathbf{r} - \mathbf{r}_n|^{-1}$, with the disconcerting consequence that the variance at the center of the image of a disk object increases linearly, without bound, as the diameter of the disk increases (see Barrett and Myers (2004), section 17.3.2). A variance map (image of $\text{Var}(\hat{f}_n)$ versus $\mathbf{r}_n$) looks like a highly blurred version of the object $f(\mathbf{r})$. Moreover, the correlations predicted by (138) have very long range, especially if a line drawn through voxels n and n' passes through the center of rotation in a tomographic imaging system; low-dose FBP images exhibit radial noise pattern extending across the whole reconstruction.

In short, the noise in tomographic reconstructions with linear algorithms is very nonlocal, with strong, long-range correlations. For the present discussion, this means that $\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}}$, though in principle full rank, is nowhere close to being diagonal, and computation of the prewhitening operator $\left[\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}} \right]^{-1/2}$ is far from trivial.

The noise covariance matrix behaves quite differently when nonlinear reconstruction algorithms such as MLEM (maximum-likelihood expectation maximization) are used. The noise

is much more local, the variance map looks very much like the object, and $\left[\mathbf{K}_{\mathbf{f}|\mathbf{f}}^{\text{noise}}\right]_{nn'}$ falls off rapidly as $|\mathbf{r}_n - \mathbf{r}_{n'}|$ increases (Barrett *et al* 1994, Wilson *et al* 1994). The average noise covariance matrix $\bar{\mathbf{K}}_{\mathbf{f}}^{\text{noise}}$ is sparse, and there are efficient algorithms for prewhitening.

6.3.4. Channelized Hotelling observers. All of the computational difficulties associated with estimating SNR^2 for the Hotelling observer, and its dependence on radiation dose, disappear if we use the channelized Hotelling observer. If the number of sample images is much greater than the number of channels ($N_s \gg N_c$), then the sample covariance matrix for the channel outputs, $\hat{\mathbf{K}}_v$, will be invertible and well conditioned, for both raw data and for reconstructed images. There is no need to use the covariance decomposition, and there may not be much advantage to simulating noise-free images.

6.4. Dose dependence of figures of merit for estimation tasks

6.4.1. Linear estimation of tracer uptake in a defined ROI. The estimators used for estimating radiotracer uptake in SPECT or PET are usually simple templates applied to the reconstructed images. The templates do not account for the spatial resolution of the imaging system, the background statistics, the noise correlations introduced by the reconstruction algorithm or the effects of radiation dose. They do not acknowledge the presence of null functions in the system operator or the effect of nuisance parameters. As a result, the estimates can be highly biased and suboptimal in terms of variance. Various correction methods referred to as contrast recovery or partial-volume correction have been invented, but they usually assume that the background is spatially uniform and nonrandom. For a recent review of partial-volume-correction techniques for ECT, see Erlandsson *et al* (2012).

The activity of a SPECT or PET tracer inside a specified ROI is a scalar defined as $\theta \equiv \chi^\dagger \mathbf{f}$, where χ represents an indicator function $\chi(\mathbf{r})$ which equals 1 inside the ROI and 0 outside. The expression $\chi^\dagger \mathbf{f}$ denotes a scalar product in object space as defined by the integral in (30).

A linear estimate of the uptake from a reconstructed image is $\hat{\theta}(\hat{\mathbf{f}}) \equiv \mathbf{w}^\dagger \hat{\mathbf{f}}$, where $\mathbf{w}$ is some template in reconstruction space, perhaps just a voxelized version of χ . For simplicity we assume that the reconstruction is linear so the operator $\mathcal{O}$ is the $N \times M$ matrix $\mathbf{O}$.

It is straightforward to show that the bias and variance of this estimate, conditioned on a particular object, are given, respectively, by

$$b_{\hat{\theta}}(\mathbf{f}) = \mathbf{w}^\dagger \mathbf{O} \mathcal{H} \mathbf{f} - \chi^\dagger \mathbf{f}; \quad \text{Var}(\hat{\theta}|\mathbf{f}) = \mathbf{w}^\dagger \mathbf{K}_{\mathbf{f}|\mathbf{f}} \mathbf{w} = \mathbf{w}^\dagger \mathbf{O} \mathbf{K}_g \mathbf{O}^\dagger \mathbf{w}. \quad (141)$$

The conditional mean-square error, $\text{MSE}(\mathbf{f})$, is related to conditional bias and variance by (32), and a further average over objects, followed by some algebra, yields

$$\text{EMSE} = \mathbf{w}^\dagger \mathbf{O} \bar{\mathbf{K}}_g^{\text{noise}} \mathbf{O}^\dagger \mathbf{w} + [\mathbf{w}^\dagger \mathbf{O} \mathcal{H} - \chi^\dagger] \mathcal{K}_f [\mathbf{w}^\dagger \mathbf{O} \mathcal{H} - \chi^\dagger]^\dagger + [\mathbf{w}^\dagger \mathbf{O} \mathcal{H} \mathbf{f} - \chi^\dagger \mathbf{f}]^2. \quad (142)$$

The first term in this equation is the average variance in the estimate arising from noise in the image; the second term is the variance arising from object randomness, and the third term is the square of the ensemble-average bias. Both the second and third terms would be zero if we could choose the template so that $\mathbf{w}^\dagger \mathbf{O} \mathcal{H} = \chi^\dagger$, or equivalently $(\mathbf{O} \mathcal{H})^\dagger \mathbf{w} = \chi$, but this is not possible if the indicator function has null components, where $\mathbf{O} \mathcal{H} \chi_{\text{null}} = 0$. In essence, the last two terms in (142) arise because we are trying to estimate a parameter that is not estimable.

With the dose dependences from table 4, we find that the first term in (142) varies linearly with any measure of dose, D , and the second and third terms go as D^2 . Of course, the parameter being estimated—activity in an ROI—is also linearly related to dose, and it is reasonable to

normalize the estimate to the total injected activity. The normalized EMSE (denoted NEMSE) then has a noise term that varies as $1/D$ plus two dose-independent terms that come from object randomness and from the null functions. Smaller NEMSE is better, so its reciprocal can be used as a figure of merit, and we see that

$$\frac{1}{\text{NEMSE}} \propto \frac{D^2}{aD^2 + bD}, \quad (143)$$

which is the same dependence as seen in (120) for a detection task with a dose-independent template.

6.4.2. More sophisticated estimators of tracer uptake in an ROI. We cannot eliminate the bias term in (142), but we can minimize it by finding a least-squares solution to $(\mathbf{O}\mathbf{H})^\dagger \mathbf{w} = \boldsymbol{\chi}$; that is, we seek a template whose backprojection is, as closely as possible, the indicator function that defines the parameter being estimated. From Barrett and Myers (2004), section 1.7.4, the solution is

$$\mathbf{w}_{\text{LS}} = (\mathbf{O}\mathbf{H}\mathbf{H}^\dagger\mathbf{O}^\dagger)^\dagger \mathbf{O}\mathbf{H}\boldsymbol{\chi}, \quad (144)$$

where superscript $+$ denotes a Moore–Penrose pseudoinverse. The operator $\mathbf{O}\mathbf{H}\mathbf{H}^\dagger\mathbf{O}^\dagger$ is an $N \times N$ matrix, and in many cases the pseudoinverse becomes an ordinary matrix inverse.

As we saw in section 3.4, the Gauss–Markov estimator (GME) minimizes the first term in (143). From (70), as transcribed to apply to reconstructed images, its template satisfies

$$\mathbf{w}_{\text{GM}}^\dagger = \boldsymbol{\chi}^\dagger \left[\mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{-1/2} \mathbf{O}\mathbf{H} \right]^\dagger \mathbf{K}_{\hat{\mathbf{f}}|\mathbf{f}}^{-1/2}. \quad (145)$$

The dose dependence of the FOM given in (143) still holds for the LS and GM templates, though of course the constants are different.

If we wish to minimize the sum of the three terms in (142), we should use the Wiener estimator from (73). For the special case of estimating a scalar linear parameter such as uptake in an ROI from a reconstructed image, the Wiener estimator becomes (Whitaker *et al* 2008)

$$\hat{\theta}_W(\hat{\mathbf{f}}) = \bar{\theta} - \boldsymbol{\chi}^\dagger \mathbf{K}_{\mathbf{f}} \mathbf{H}^\dagger \mathbf{O}^\dagger \mathbf{K}_{\hat{\mathbf{f}}}^{-1} (\hat{\mathbf{f}} - \bar{\bar{\mathbf{f}}}), \quad (146)$$

where $\bar{\theta}$ is the prior mean of θ . The corresponding EMSE is

$$\text{EMSE} = \sigma_\theta^2 - \boldsymbol{\chi}^\dagger \mathbf{K}_{\mathbf{f}} \mathbf{H}^\dagger \mathbf{O}^\dagger \mathbf{K}_{\hat{\mathbf{f}}}^{-1} \mathbf{O}\mathbf{H}\mathbf{K}_{\mathbf{f}} \boldsymbol{\chi} = \sigma_\theta^2 - \boldsymbol{\chi}^\dagger \mathbf{K}_{\mathbf{f}} \mathbf{H}^\dagger \mathbf{O}^\dagger \left[\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}} + \mathbf{K}_{\hat{\mathbf{f}}}^{\text{obj}} \right]^{-1} \mathbf{O}\mathbf{H}\mathbf{K}_{\mathbf{f}} \boldsymbol{\chi}, \quad (147)$$

where σ_θ^2 is the prior variance of θ . From the ECT column of table 4, $\sigma_\theta^2 \sim D^2$, $\mathbf{K}_{\mathbf{f}} \sim D^2$, $\bar{\mathbf{K}}_{\hat{\mathbf{f}}}^{\text{noise}} \sim D$ and $\mathbf{K}_{\hat{\mathbf{f}}}^{\text{obj}} \sim D^2$. To get the overall dose dependence, we can approximate $\mathbf{K}_{\hat{\mathbf{f}}}^{-1}$ as in (133), but with eigenvalues and eigenvectors determined by use of simulated noise-free reconstructions $\hat{\mathbf{f}}$ rather than simulated, noise-free $\mathbf{g}$.

6.4.3. Estimation of signal parameters. Now suppose we separate the object into background and signal parts as in (74), $\mathbf{f} = \mathbf{f}_{bg} + \mathbf{f}_{\text{sig}}(\boldsymbol{\theta})$, and our task is to estimate the $P \times 1$ signal parameter vector $\boldsymbol{\theta}$. For example, the components of $\boldsymbol{\theta}$ might be the 3D coordinates, diameter and tracer uptake for a tumor modeled as a sphere, or they might include other parameters to incorporate departures from sphericity or spatial variations of uptake within the boundaries. Some of these components might be null functions (e.g. fine structures in the signal beyond the resolution limit of the system) and some may be nuisance parameters (e.g. precise location of a tumor when the task is to estimate the total tracer uptake).

The use of models specified by a small number of parameters usually means that all of the parameters are estimable—provided the model is correct. On the other hand, if the model is wrong (as all models are), then there is a new form of bias arising from modeling error. There is no reason to think that bias from modeling error will depend on dose in the same way as bias arising from properties of the estimator or bias stemming from null functions in the template that defines the parameter, but this point does not seem to have been studied in the imaging literature.

We will proceed as if the model used in defining $\mathbf{f}_{\text{sig}}(\boldsymbol{\theta})$ is an accurate description of the real object and $\boldsymbol{\theta}$ is estimable. We assume also that the set of P parameters (components of $\boldsymbol{\theta}$) is adequate to describe the signal completely, with no remaining randomness.

6.4.4. Dose dependence of ML and MAP estimates of signal parameters. In section 3.4 we introduced a multivariate normal data model and some reasonable approximations under which the log-likelihood, $\ln \text{pr}(\mathbf{g}|\boldsymbol{\theta})$, reduces to a linear function of $\mathbf{g}$, albeit a nonlinear function of $\boldsymbol{\theta}$ (see (76)). The corresponding $P \times P$ Fisher information matrix is given in (77).

This data model was also considered by Abbey *et al* (1998), who used earlier work by Barndorff-Nielsen (1983) and Fessler (1996) to develop methods for approximating the PDF of maximum-likelihood and maximum a posteriori estimates. For low noise and ML estimation, Abbey *et al* showed that

$$\text{pr}(\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}) \approx \left[\frac{\det [\mathbf{F}(\boldsymbol{\theta})]}{(2\pi)^P} \right]^{1/2} \exp \left[-\frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^t [\mathbf{F}(\boldsymbol{\theta})] (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \right], \quad (148)$$

which has the same structure as the generic multivariate normal PDF (see (39)) except that the Fisher information matrix $\mathbf{F}(\boldsymbol{\theta})$ has replaced the inverse of the covariance matrix, $[\mathbf{K}_{\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}}]^{-1}$, and the true parameter vector $\boldsymbol{\theta}$ has replaced the mean of the estimate.

The PDF in (148) could have been obtained by citing a theorem that says that in the asymptotic (low-noise) limit, a maximum-likelihood estimate is unbiased, efficient and normally distributed, but Abbey *et al* derived it from a different viewpoint that facilitated its extension to higher noise where the PDF is not multivariate normal and to MAP estimates which are necessarily biased. The approximations derived by Abbey *et al* were in excellent agreement with Monte Carlo simulations.

The FIM that appears in (148), defined in (77), involves $\mathbf{K}_{\mathbf{g}}^{-1}$, which can be approximated by (133) if we can simulate realistic noise-free images. If all components of $\boldsymbol{\theta}$ are defined to be independent of dose, then the derivatives in the FIM, $\partial \bar{g}_m / \partial \theta_p$, have the same dependence on dose as $\bar{g}_m$ itself, and the dose dependence of any component of the FIM has the same form as the detection SNR^2 in (135). The variance of any component of an ML estimate as a function of dose can then easily be found numerically by performing a P -dimensional inverse of the FIM, and from there we can get the dose dependence of the EMSE (35), the WEMSE (36), or any other desired measure of efficacy for the estimation task.

6.5. Graphical depictions of tradeoffs

The results in sections 6.3 and 6.4 enable us to plot task-based measures of image quality as a function of any desired measure of dose. Examples of such plots are given in figure 3, where the performance of a linear discriminant with a dose-independent template, referred to here as a non-prewhitening matched filter (NPWMF), is compared to the performance of a Hotelling observer, which is a linear discriminant where the template varies optimally with dose. The SNR^2 was calculated from (120) for the NPWMF and from (135) for the Hotelling observer; in both cases, SNR^2 was converted to AUC by use of (28). The abscissa in this figure, D , can be any convenient measure of dose, including absorbed dose at a point, dose in a region of interest, organ dose, or effective dose.

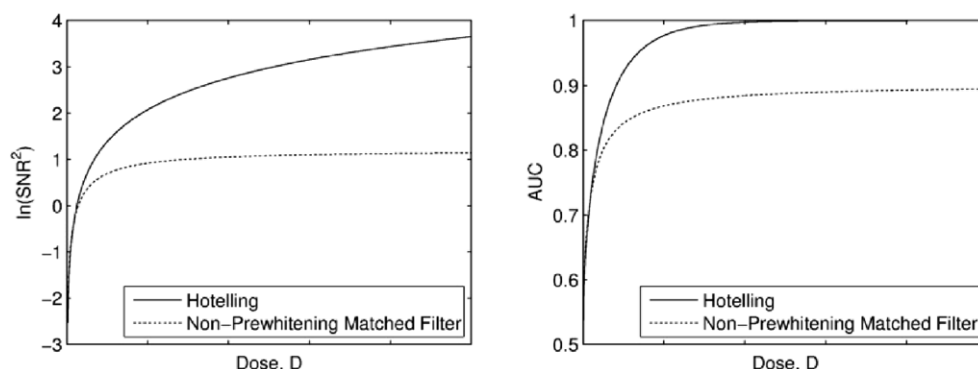


Figure 3. Schematic quality-dose characteristic (QDC) curves for a binary classification task. Left: quality specified by SNR^2 . Right: quality specified by area under the ROC curve (AUC). For both panels D can be any suitable dose measure (see text).

An interesting observation from figure 3 is that AUC as a function of D saturates at a level less than $\text{AUC} = 1$ for the NPWMF, so perfect performance cannot be obtained with this observer even at infinite dose. For the Hotelling observer, on the other hand, AUC is asymptotic to 1 as the dose increases (provided the covariance matrix is known accurately).

Note that the logarithm of SNR^2 is plotted in figure 3. This makes it easy to see how the curves vary with signal contrast C . For both observers, $\text{SNR}^2 \propto C^2$, so the graphs of $\ln[\text{SNR}^2]$ simply shift vertically as the contrast is changed. Even though the AUC saturates at 1.0 at high dose for some specific contrast, it does not follow that more dose is not useful; it could be useful for detecting a signal with smaller contrast. For this reason, it can be very helpful to plot detection performance in terms of SNR rather than AUC.

The QDC curves of figure 3 can be converted to plots of task performance versus some measure of risk from the ionizing radiation. For this purpose we consider three response models introduced in section 5 and plotted in figure 4. The resulting quality-risk characteristic (QRC) curves derived by combining figures 3 and 4 are shown in figure 5. The qualitative features of the QDC are relatively insensitive to the response model chosen; in particular, the performance of the NPWMF saturates at $\text{AUC} < 1$ as a function of risk for all models.

None of the three RDCs illustrated in figure 4 and used to construct figure 5 explicitly includes a threshold. The only threshold-based model introduced in section 5, and commonly used in the community, is the Weibull model of (84). This model for tissue-reaction effects has three free parameters: threshold T , shape factor V and dose for 50% response, D_{50} . As discussed in section 5, there is no strong evidence for a non-zero threshold for certain of the tissue-reaction effects (in particular circulatory disease, cataracts and CNS damage); nevertheless, ICRP (2012) recommends use of a practical threshold dose, for various tissue-reaction effects, below which morbidity would occur at $<1\%$, as indicated in table 2.

With this definition of the threshold, the Weibull distribution for various shape factors is shown in figure 6. The interesting range of doses on this plot is $D \ll D_{50}$, and the corresponding QRCs are very similar to those for the no-threshold models.

It is important to note that the ICRP-recommended thresholds for tissue-reaction effects are on the response axis, not the dose axis. In essence, ICRP merely says that if you keep the dose below the threshold listed in table 2, the incidence of the effect will be less than 1%.

Moreover, the dose must be in Gy, not Sv. A threshold, if there is one, must be local to where the adverse effect occurs in the body, not somehow averaged over the body or even necessarily over an organ. For example, radiation-induced cataractogenesis depends on dose

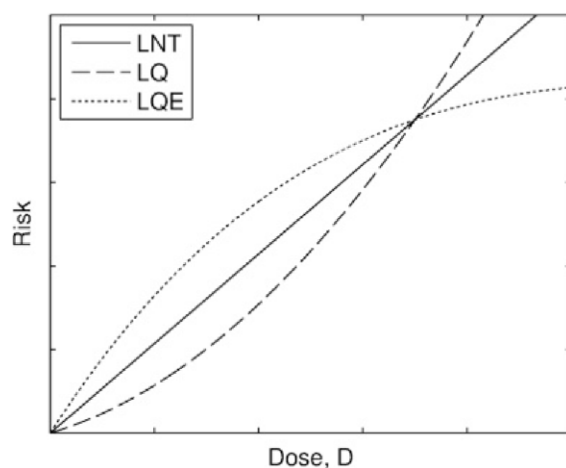


Figure 4. Schematic curves of risk versus dose for three response models. The term ‘risk’ is used broadly here and in subsequent graphics to include any of the absolute or relative risks defined in section 5 or the rates of cancer incidence given by (87) or (88). The solid curve is the conventional linear response model with no threshold (LNT); the dashed curve is the linear-quadratic (LQ) model of (89), and the dotted line is a linear-quadratic model modified with an exponential factor (LQE) as in (90).

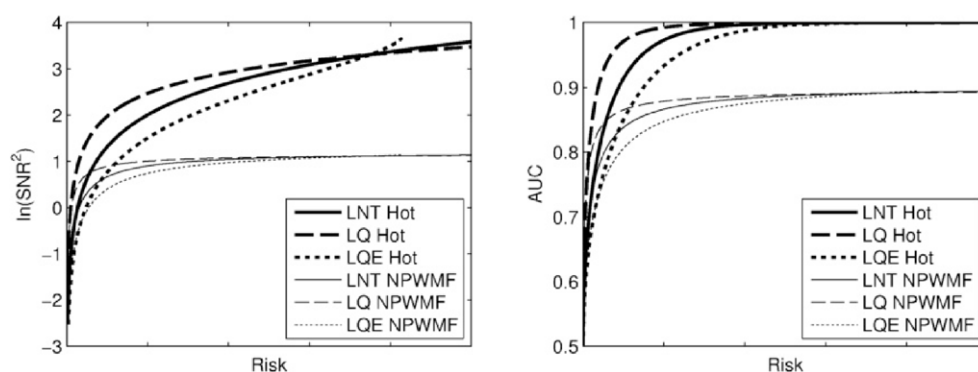


Figure 5. Schematic quality-risk characteristic (QRC) curves. Left: SNR^2 on a binary detection task for each of the three risk models shown in figure 4 and for NPWMF and Hotelling observers. Right: AUC for the same observers and risk models.

to the lens, not the whole eye and certainly not the whole body. The energy absorbed in the lens divided by the mass of the lens is the only quantity that can matter.

7. Methods of dose reduction and their effect on task performance

7.1. General considerations

In SPECT, gamma-ray detectors are used with image-forming elements such as collimators, pinholes or coded apertures to form an image of the distribution of a gamma-ray-emitting radiopharmaceutical. Factors that influence task performance include: the properties of the image-forming elements; the characteristics of the image detector; the kind of information

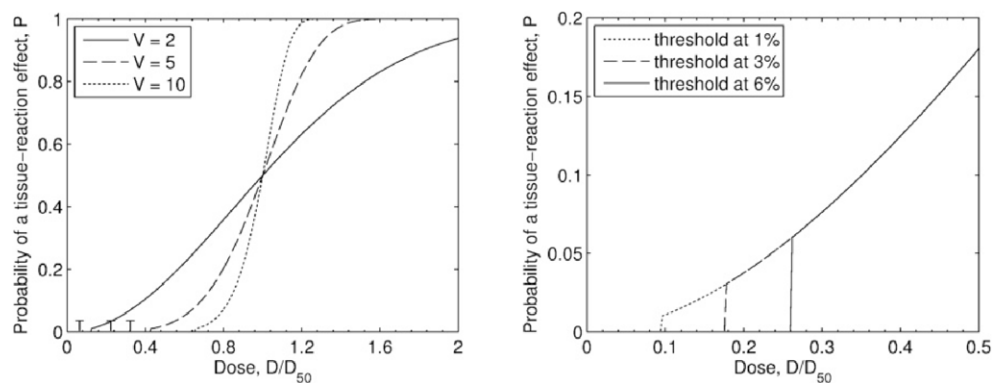


Figure 6. Plots of the Weibull probability of a tissue-reaction effect. Left: threshold set at 1% incidence as recommended in ICRP (2012) and various shape factors. Right: fixed shape factor and various thresholds. Note that the graph on the right covers a small area in the lower left corner of the graph at the left.

acquired and stored by the image detector; the reconstruction algorithm and, most importantly, the radiopharmaceutical itself.

PET imaging usually uses arrays of discrete detector elements operated in coincidence, so the image formation is implicit in the detector configuration itself. Other factors that can influence task performance include the coincidence timing resolution, the reconstruction algorithm and, again, the radiopharmaceutical.

In these emission modalities, the absorbed dose distribution in the patient's body, and hence any potential risk to the patient, is determined entirely by the radiopharmaceutical chosen and the amount of activity administered. Progress in detectors, image formation and reconstruction algorithms can improve task performance and diagnostic efficacy, but they will result in lower dose or risk only if the administered activity of the radiotracer is reduced.

Similarly, in transmission x-ray imaging modalities such as mammography and CT, the absorbed dose distribution and potential risk are determined entirely by the radiation incident on the patient's body. Image quality can be increased by improvements in the image detector, reconstruction algorithm and other imaging factors, but these measures result in dose reduction only if the incident radiation is reduced in some way. For a survey of strategies for reducing radiation dose in CT, see McCollough *et al* (2009) or Mattsson (2011).

Moreover, as we saw in sections 6.4 and 6.5, the relationship between radiation dose and objective measures of image quality can be very nonlinear, depending on the task, the observer and the class of objects being imaged. In particular, it is very important to consider both image noise, as expressed by $\mathbf{K}^{\text{noise}}$, and object randomness, as given by $\mathbf{K}^{\text{obj}}$. These covariance terms depend differently on dose (see table 4), and both must be considered in meaningful assessment of image quality and strategies for dose reduction. Because of the object randomness, there are many situations where the task performance saturates as a function of dose or risk (see figures 3 and 5), and in these cases dose reduction by reducing the injected activity in ECT or the tube current in TCT will not degrade image quality at all.

Conversely, any study of image quality or dose reduction that considers only a single patient or phantom, and states the result as contrast-to-noise ratio, subjective appearance or any other metric that does not involve an average over objects, will have little to do with clinical imaging.

In the subsections that follow, we present the major categories of approaches to dose reduction, provide some references to research on their relation to objective measures of image quality, and make general comments on the state of the art in each category.

7.2. *DRAPE: Dose Reduction by Aperture Enhancement*

It has long been recognized that pinholes and collimators are very inefficient in collecting gamma-ray photons from a radionuclide source. More photons can be collected by increasing the diameter of a pinhole or the bore of a collimator, but the concomitant loss in spatial resolution can degrade image quality as measured by task performance. One of the earliest papers on task-based assessment was by Tsui, Metz and others on optimizing collimators by using statistical decision theory and observer performance (Tsui *et al* 1978).

An alternative to pinholes and collimators is coded-aperture imaging, introduced into nuclear medicine in the early 1970s (Barrett 1972a, 1972b, Rogers *et al* 1972). Coded apertures can collect far more photons than pinholes or collimators for the same exposure time and spatial resolution, but the overlap of projection images, called *multiplexing*, can reduce task performance. A proper analysis of the effect of multiplexing must account for both noise and object randomness. Myers *et al* (1990) evaluated the impact of spatially varying, or lumpy, backgrounds on the optimum aperture size for both detection and discrimination tasks. They found that, while detectability of a signal in a uniform background is maximized by enlarging the aperture size, a lumpy background results in an optimum aperture that is matched naturally to the size of the signal. For the task of discriminating a single signal from a double signal, a so-called Rayleigh task, the optimum aperture size is smaller in the presence of a lumpy background than the optimum aperture size for the flat-background case. These evaluations were among the first to demonstrate the use of statistical backgrounds in the objective evaluation of image quality using model observers in general, and the particular finding that there is a greater need for resolution in the presence of a non-uniform background than is needed when the background is known and uniform. Hesterman *et al* (2007) performed a detailed experimental and simulation-based study of multiple-pinhole apertures for lesion detection in a lumpy background. The experimental lumpy-background phantom consisted of plastic beads immersed in a radioactive solution. Measured covariance functions exhibited a power-law dependence previously seen in simulation studies, helping to validate the use of simulated lumpy backgrounds in observer studies. Detection performance was quantified by use of channelized Hotelling observer is used with Laguerre–Gauss channels as a function of radiation dose, signal size, lump size and aperture configuration. These studies reveal an improvement in AUC for certain ranges of signal and lump combinations through the use of multiplexed, multiple-pinhole apertures, indicating a need for task-specific aperture optimization.

There are many other papers in the literature that use task-based methods to optimize collimators for SPECT; examples include (Kijewski *et al* 2001, Zhou *et al* 2008, Zhou and Gindi 2009). For a review of assessment methodology based on Monte Carlo simulation of SPECT systems, see Ljungberg and King (2012).

7.3. *DRUID: Dose Reduction with Upgraded Image Detectors*

This subsection deals with SPECT and PET detectors; x-ray detectors are discussed in section 7.4.

7.3.1. Detectors for SPECT. The detectors commonly used in SPECT and planar nuclear medicine are scintillation cameras, modern variants of the venerable Anger camera (Peterson

and Furenlid 2011), but there is also considerable current interest in semiconductor detector arrays. The important characteristics of these detectors that can affect image quality, and eventually radiation dose, are detection efficiency, energy resolution, spatial resolution and overall detector area. The detector area divided by the area of one resolution cell is referred to as the *space-bandwidth product*.

Of these detector properties, perhaps the most important in terms of dose and image quality is the space-bandwidth product, which is basically the dimension M of the data vector $\mathbf{g}$ in the preceding sections. All else being equal, increasing M will always increase the task performance at any given dose, or reduce the dose required for a specified task performance for any task.

Moreover, increasing the space-bandwidth product of the detector also opens up the opportunity for new imaging configurations that improve task performance. For example, in a simulation of new pinhole imaging systems for brain SPECT, Rogulski *et al* (1999) demonstrate that use of a detector with a very large number of resolvable elements permits the use of a large number of pinholes, but without overlap of the images on the detector. The result is that the system can collect more photons without multiplexing or loss of spatial resolution than would be possible with a detector having smaller space-bandwidth product.

Improvements in energy resolution in SPECT detectors result immediately in improvements in task performance by reducing the number of scattered photons that contribute to the data. In addition, better energy resolution facilitates imaging studies with multiple radioisotopes or multiple emission lines from a single isotope. Improved energy resolution is the main motivation for developing semiconductor detector arrays.

One way to improve the spatial and energy resolution of either scintillation cameras or semiconductor detector arrays is by using maximum-likelihood estimation to determine the position and energy of the gamma-ray interaction event. For a review of ML estimation applied to scintillation cameras, see Barrett *et al* (2009), and for a discussion of real-time implementations, see Hesterman *et al* (2010). For more on this topic, see section 7.4.

7.3.2. Detectors for PET. For a review of the basic principles of PET imaging, see Muehllehner and Karp (2006), and for a tutorial on recent developments in PET detector technology, see Lewellen (2008).

Current PET systems usually use discrete arrays of scintillation detectors. Space-bandwidth product in these detectors can be interpreted simply as the number of individual scintillation crystals. In an attempt to get high absorption efficiency for 511 keV gamma rays and good spatial resolution, these crystals are usually several cm long and a few mm in cross section.

Early PET systems used a single ring of scintillation detector to image a single 2D slice of the patient at a time, but recent trends are in the direction of multiple contiguous rings for imaging a significant volume of the patient with one position of the detector array. This benefits the photon collection efficiency, and it facilitates dynamic studies which are an important application of PET.

Advances in scintillation crystals, photodetector technology and readout electronics have led to improvements in coincidence timing resolution, thereby improving the rejection of accidental coincidences and allowing some degree of localization of a positron annihilation event in three dimensions by estimating the difference in time of flight (TOF) of the two gamma rays to their respective detectors.

These technological advances should lead directly to improvements in image quality or allow a reduction in the administered activity (hence reducing the absorbed dose) at the same image quality. A few groups are now using task-based assessment methods to investigate these gains quantitatively.

El Fakhri, Surti and coworkers have used model and human observers to demonstrate improvements in lesion detection in oncological PET both by volumetric imaging and with TOF-PET (El Fakhri *et al* 2007, El Fakhri *et al* 2011, Surti *et al* 2011). Lartizien *et al* (2004) compared the impact of 2-dimensional (2D) and fully 3-dimensional (3D) acquisition modes on the performance of human observers in detecting and localizing tumors in whole-body 18F-FDG images.

7.4. DRACO: Dose Reduction with the Advent of Counting Options

Detectors for high-resolution digital x-ray imaging usually consist of either a monolithic (non-segmented) slab of scintillator material with an array of photodetectors or a monolithic slab of a semiconductor and an array of electrodes (Yaffe and Rowlands 1997).

These detectors are used in an integrating mode where the accumulated signal after x-ray exposure for some time is read out and stored as the vector $\mathbf{g}$ of previous sections. An immediate difficulty is that the charge or light produced by a single x-ray photon usually spreads out over several contiguous readout elements (electrodes or photodetectors), making the noise covariance matrix $\mathbf{K}_{\mathbf{g}}^{\text{noise}}$ nondiagonal and reducing task performance for all tasks. Further performance reduction arises from the inevitable noise contributions of dark current and various noise sources in the electronics. There is also the so-called Swank noise (Swank 1973) from the random amount of light or charge produced by different x-ray photons.

All of these deleterious effects can be ameliorated by reading out the charge or light signals produced from individual x-ray photons. The term *photon counting* is often used for this approach, but it is very misleading. The simplest, and most common, way of ‘counting photons’ is to determine the sensor element that receives the maximum signal for a particular x-ray interaction and assign the event to a data bin determined by the coordinates of that element. Moreover, the amount of signal in this chosen readout element is interpreted as the energy of the photon. At best it is an estimate of the energy deposited by the photon (which may not be all of its energy), and in practice it is a very poor estimate of the deposited energy because much of the deposited energy produces signals in adjacent elements, which are not even read out, much less used in the estimation of position and energy.

A much better approach, first applied to integrating semiconductor detectors by Marks *et al* (1999), is to read out the cluster of array elements that receive charge or light from a single x-ray photon and use all of that data to do ML estimation of the position and energy of the interaction. The rate at which these operations can be performed with modern readout technologies and computers is increasing rapidly, and there is no other approach that will produce more dose reduction in x-ray imaging.

To be specific, a digital x-ray detector fast enough to respond to individual x-ray interactions, to read out all sensor elements that respond to each event and to do ML estimation of position and energy would accomplish the following goals (Barrett 2007):

- Eliminate Swank noise
- Be immune to readout noise (thresholded away)
- Have pure Poisson image statistics (conditional on $\mathbf{f}$)
- Have subpixel spatial resolution
- Provide an estimate of depth of interaction
- Provide accurate energy information.

The best current approach to meeting these goals might be the use of crystalline silicon detectors where the charge spreading is small and the output of a single electrode is a good estimate of energy.

In 1985 Tapiovaara and Wagner (1985) showed that a moderate increase in detection performance could be achieved by using a linear discriminant function that assigned different weights to photons of different energy. In a recent paper Kalluri *et al* (2013) investigated the improvements in task performance with energy weighting for photon-counting breast CT.

7.5. DROLL: Dose Reduction by Obtaining Likelihood Lists

A common misconception in imaging is that it is necessary to detect a large number of photons on each detector pixel in order to achieve a large detector signal-to-noise ratio (DSNR). This quantity, not to be confused with task-based observer SNR, is defined as the mean signal from a detector element divided by its standard deviation for a constant x-ray flux; for a pure photon-counting detector, DSNR is just the square root of the mean number of detected photons per detector pixel. Some CT manufacturers go so far as to bin together the outputs of adjacent detector elements, thereby deliberately degrading spatial resolution in a quest for increased DSNR. A recent paper by Baek *et al* (2013) shows that such binning also degrades performance on detection tasks, or equivalently requires more radiation dose for the same performance. In fact, one can show quite generally that best ideal-observer task performance for any photon-counting detector is obtained by making the detector elements very small, even to the point where the average number of detected photons per element is $\ll 1$.

Rather than trying to construct detectors with ever smaller pixels, it is preferable to use a continuous, monolithic detector material and use ML estimation as discussed in section 7.4 to estimate the position and energy of each interaction. Once this is accomplished, the complete information about each photon interaction is preserved if we store the estimated position, energy and possibly other attributes in a list, what is commonly called *list-mode storage*.

The relation of list-mode data to task performance has been studied in Barrett *et al* (1997), Parra and Barrett (1998) and Caucci and Barrett (2012). A concept that emerges from these studies is that of the *Ideal Dose Utilizer*. Specifically, an Ideal Dose Utilizer is an x-ray or gamma-ray imaging system that maximizes the area under the quality-dose characteristic curve (AUQDC) by:

- Responding to individual photons;
- Collecting all sensor signals for each photon;
- Using the sensor signals to compute ML estimates of photon attributes;
- Recording the estimated attributes in list mode;
- Implementing the list-mode ideal observer.

The list-mode ideal observer, is discussed in detail in Caucci and Barrett (2012) for detection tasks, and a similar development for estimation tasks is possible.

7.6. DRAMA: Dose Reduction by Accurate Modeling Algorithms

Iterative image-reconstruction algorithms are used routinely in PET and SPECT (Qi and Leahy 2006), and rapid advances in computer power are making them increasingly feasible for the large data sets in CT as well (Nuyts *et al* 2013). Iterative algorithms are not expected to be useful in dose reduction just because they are iterative, but rather because they reduce the level of inaccuracy in the modeling of the object and imaging system. Currently, all CT manufacturers implement some kind of iterative algorithm involving a system model, but details are usually not available in the literature.

Many iterative algorithms are variants on the well-known maximum-likelihood expectation-maximization (MLEM) algorithm (Dempster *et al* 1977, Shepp and Vardi 1982), so they

require knowledge of the likelihood function, $\text{pr}(\mathbf{g}|\mathbf{f})$. The models used in practice for the likelihood functions are inaccurate for several reasons: (a) the object is represented by a finite set of numbers (usually voxel values) instead of the true continuous function; (b) the system is represented by a matrix rather than a CD operator; (c) various aspects of the physics of the radiation transport and detection process are omitted or oversimplified; (d) prior knowledge of the object support may not be used, and (e) the fact that the object cannot be negative may not be incorporated into the algorithm.

One of the earliest papers to make use of an objective assessment framework for the evaluation of image reconstruction algorithms, with attention to accurate modeling of the image-formation process, was the 1990 publication by Hanson (1990). In that work Hanson compared unconstrained ART (Algebraic Reconstruction Technique) to ART with a non-negativity constraint, and considered both noise-free and noisy data scenarios. Of most relevance to our emphasis in this section is that he explicitly modeled the continuous-to-discrete nature of the image formation process and incorporated object support into the reconstruction algorithm. The figure of merit was the detectability of low-contrast signals, where the signals had known, but random, locations. The task was further complicated in that the scenes contained significant artifacts generated by randomly-located high contrast disks, allowing for the evaluation of image quality in the presence of such artifacts particularly in the case of a limited number of projection views.

In spite of the high degree of interest in iterative algorithms for CT, the great majority of papers in this field do not make use of objective, task-based assessment of image quality. Instead they use subjective ratings of anecdotal images; opinion polls, such as preference studies or conspicuity and sharpness ratings with experienced radiologists; or various noise measures such as variances and contrast-to noise ratio. Recent examples of this kind of non-task-based assessment of iterative methods for CT include (May *et al* 2011, Hou *et al* 2012, Husarik *et al* 2012, Kalra *et al* 2012, Katsura *et al* 2012, Sato *et al* 2012, Singh *et al* 2012, Qi *et al* 2012, Ebersberger *et al* 2013).

Willemink *et al* (2013) performed a systematic review of the literature on iterative reconstruction in CT, covering the period January, 2006 through January 2012. They found that there were many indications that iterative reconstruction preserved image quality (compared to FBP) and allowed some dose reduction, but they concluded that ‘IR has not yet been investigated with clinical diagnosis and accuracy as endpoints’.

A few recent papers have begun to fill this gap in the literature, using human observers, model observers or both for objective assessment of reconstruction algorithms.

Miéville *et al* (2013) evaluated human observer performance as well as a number of Stage 1 efficacy metrics (CT number accuracy, pixel standard deviation, noise power spectrum and modulation transfer function) for three commercial CT iterative reconstruction methods to assess their image quality and dose reduction performance. The human observers performed four-alternative forced-choice (4AFC) experiments to assess the detectability of low- and high-contrast objects embedded in two pediatric phantoms. These investigators found, not surprisingly, that the largest potential improvements in dose efficiency are to be found by making use of model-based iterative reconstruction algorithms that utilize a full volumetric model of the image acquisition process.

Xu and Tsui (2014) used approximations to the likelihood function and the ideal observer to investigate the effect of inaccurate statistical modeling on a detection task in CT.

An often overlooked form of modeling error is the use of an $M \times N$ matrix $\mathbf{H}$ to represent a CD operator $\mathcal{H}$. Of course the use of a matrix to approximate an operator is inevitable in computer implementations, but inaccuracies in the representations can impair task performance. In general, the number of reconstruction voxels N should be large, and the elements of the matrix should be as nearly as possible $H_{mn} = h_m(\mathbf{r}_n)$, where $\mathbf{r}_n$ is the 3D location of voxel n .

Finally, accurate modeling of the physics of the photon transport, image formation and photon detection is essential in order to get the best possible task performance. Many groups are addressing this issue in SPECT (Ljungberg and King 2012) and PET (Nuyts *et al* 2013).

7.7. DRESS: Dose Reduction Enabled by Sparse Sampling

Sparse sampling, also called compressive sensing, refers to the use of ‘incomplete data’ to form an image. In a way, this term is misleading because all imaging systems that map functions to discrete data sets have null functions, hence there are no complete data sets. In practice, sparse sampling in tomography refers to using fewer projections than would be indicated by sampling theory if the object is both space-limited and band-limited (which is not possible).

The number of projections can be restricted by collecting over a limited range of projection angles, collecting at larger angular increments, or both. In transmission tomography, restricting the number of projections results in lower dose if the dose per projection is held constant.

Hanson presented a framework for the objective evaluation of sparse data-sampling methods for CT in 1990 (Hanson 1990). In that work he compared variations of the ART algorithm in the presence of a limited number of projection views, where the task was the detection of low-contrast disk signals. The task was confounded by the presence of artifacts caused by high-contrast disks at random locations in the object. It is important to evaluate sparsely sampled systems with a framework of this kind, where the artifacts are randomized, to fully and realistically account for their associated degradation in task performance.

There is currently a high level of interest in applying sophisticated task-based assessment methodology to digital breast tomosynthesis (DBT) and related limited-angle CT systems. For reviews of DBT, see Dobbins and Godfrey (2003), Sechopoulos (2013a) and Sechopoulos (2013b), and for a comprehensive review of breast CT, see Glick (2007).

A recent paper by Young *et al* (2013) presented a virtual-trial framework for the assessment of DBT using digital phantoms, open-source x-ray transport codes, and a projection-space model observer (the channelized Hotelling observer). Using this framework, the authors compared various DBT system geometries (in terms of numbers of projection images and overall scan angle range, for example) in terms of the resulting detectability of a small signal embedded in a large sample of digital breast phantoms with randomly-varying anatomical backgrounds. Experiments were repeated for three test cases, where the detectability-limiting factor was dominated by anatomical variability, quantum noise, or electronic noise. It is notable that the authors made use of 1000 digital breast phantoms in an effort to more completely sample the range of patient variability in clinical imaging and thereby derive meaningful estimates of both the mean system performance and its variation resulting from random patient anatomical structure.

Young *et al* (2012) used the same virtual-trial framework to evaluate various exposure delivery schemes for DBT, including the standard approach that exposes the patient uniformly over projection angles, as well as exposure delivery schemes that emphasize either the central projections or the most oblique projections. This work further illustrates the power of virtual clinical trials using large samples of virtual patients and simulated image acquisition systems for the objective evaluation of image quality and patient exposure settings.

Rupcich *et al* (2013) evaluated various methods for reducing dose to the breast during CT coronary angiography. In a simulation study, they compare breast shielding, angular tube current modulation, reduced kV, and partial angle protocols. They make use of both a detection and a localization task in this work. They plot a variety of figures of merit versus dose,

including contrast-to-noise ratio (CNR) and detectability. They state that the system rankings are pretty similar for these figures of merit, but in their discussion and conclusion they point out the concern that CNR does not account for noise correlations or artifacts.

Makeev and Glick (2013) have investigated a penalized maximum likelihood (PML) algorithm (using Huber and Total Variation priors) for breast CT, and have reported its performance with varying dose levels using a Fourier domain ideal observer model. This same group also studied the PML reconstruction algorithm for breast tomosynthesis, and showed improved performance at lower dose than with filtered backprojection (Das *et al* 2011).

Das *et al* (2009) present the results of human observer studies (no model observers) using an LROC study design to compare a tomosynthesis system with uniform dose distribution to an acquisition with variable dose distribution. This work thus considers the question of how best to allocate the dose to be incurred for a given imaging exam. They investigated both mass and microcalcification detection tasks; the images were simulated (without scatter or breast motion) and reconstructed using FBP.

Other recent papers on task-based assessment of DBT and related systems include (Das *et al* 2009, Park *et al* 2010, Reiser and Nishikawa 2010, Van de Sompel *et al* 2011, Warren *et al* 2012).

8. Conclusions and recommendations for future research

This topical review has attempted to synthesize current knowledge on image-quality assessment, calculation and specification of radiation dose, and risk of various adverse effects from ionizing radiation, thereby providing the tools for a rigorous risk-benefit analysis of medical imaging systems based on gamma rays and x-rays.

In section 2 we introduced the mathematical framework for this synthesis, emphasizing two key points. First, objects to be imaged are correctly described as functions of spatial variables and possibly time, so the imaging system is a linear or nonlinear operator that, on average, maps a function of continuous variables to a discrete data vector. The basic equation $\bar{\mathbf{g}} = \mathcal{H}\mathbf{f}$ runs throughout this paper. Second, statistical deviations from this average arise from two fundamentally different sources: noise in the measurement process and randomness in the objects themselves. Another recurring basic equation is that the covariance matrix for the data can be rigorously decomposed in two terms, $\mathbf{K}_{\mathbf{g}} = \mathbf{K}_{\mathbf{g}}^{\text{noise}} + \mathbf{K}_{\mathbf{g}}^{\text{obj}}$.

Section 3 is a survey of the principles of objective or task-based assessment of image quality. We consider only image-quality metrics that are defined in terms of patient benefit (efficacy), leaving aside metrics like contrast-to-noise ratio or subjective preference that are only indirectly related to task performance or efficacy. Even with this restriction, however, there is a wide variety of tasks and observers from which to choose. Section 3 starts with a review of the Fryback–Thornbury classification of stages of efficacy in medical imaging. Almost all of the past research on task-based assessment has focused on Stage 2, diagnostic accuracy, and in fact the same is true of this paper, but the long-term goal of image-quality research must be quantitative evaluation of the higher stages, especially Stage 5, patient outcomes, and Stage 6, societal outcomes. The remainder of section 3 surveys observers and figures of merit for classification tasks, parameter-estimation tasks and joint classification/estimation tasks.

Section 4 is a summary of definitions and computational methods for absorbed dose. The important point is that absorbed dose is a function of spatial position and time, $d(\mathbf{r}, t)$. It is futile to reduce this function to a single global number, effective dose in sieverts, if one wants to discuss the localized risks that can arise in medical imaging. As quoted in section 4, ICRP makes it clear that effective dose is not intended for many of the uses to which it is commonly

put, such as projecting the number of cancers in a population undergoing medical examinations with ionizing radiation.

Section 5 surveys our current understanding of these risks and the mathematical models for them. A useful advance in the terminology used in this field is that the category of *deterministic effects* is now called *tissue-reaction effects*; nothing is deterministic in medicine. Moreover, it is important that all of the models express risk in terms of absorbed dose in grays, localized at least to the organ level.

Section 6 brings the three threads—image quality, dose and risk—together by developing mathematical expressions and graphical depictions showing at least schematically how image quality varies with measures of dose and risk. Key to this discussion is table 4, showing how the various mean vectors and covariance matrices that are needed for calculating task performance depend on radiation dose. This table shows that the noise and object terms in the covariance decomposition depend differently on radiation dose, and these dependences are built into figures of merit for task performance.

Section 7 considers nine categories of methods that have been proposed for dose reduction in CT and other modalities. These categories differ widely in how far the concepts of task-based assessment have penetrated into them. For some categories, there are many groups working on rigorous assessment of image quality and, at least indirectly, the relation between quality and dose. In these categories we can cite many references, but in other categories there is almost no use of quality metrics related to patient benefit, hence very little to cite in this survey of risk-benefit analysis.

One goal for future research in this area is to begin to convert the schematic quality-dose and quality-risk characteristic curves introduced in section 6 into real curves with real numbers applicable to specific medical examinations and specific risks.

As noted in sections 3 and 6, task-based image-quality analysis requires rigor and realism in modeling imaging systems and the objects that will be imaged with them. The current rapid advances in using realistic object models, accurate simulated images and validated mathematical model observers to evaluate and optimize imaging systems bode very well for the future of this effort. The term *virtual clinical trials* has been coined for this direction of research, and an AAPM Task Group on Virtual Tools for the Evaluation of New 3D/4D Breast Imaging Systems is currently formulating methodologies and standards for it. Examples of virtual clinical trials relevant to analysis of the tradeoff between radiation dose and image quality include (Pineda *et al* 2006, Gifford *et al* 2008, Chawla *et al* 2009, Lu and Chan 2011, Van de Sompel *et al* 2011, Young *et al* 2012, Young *et al* 2013).

These advances also open up the prospect of extending image-quality studies to higher levels of efficacy. This step will require quantitative risks and utilities, especially for Stages 5 and 6, but these assignments seem to be within reach. For a survey of approaches to this problem, see Mortimer and Segal (2007).

A long-term goal for this field would be the ability to create societal efficacy characteristic (SEC) curves for different radiological imaging examinations. An SEC would be a plot of the change in quality-adjusted life years (QALYs) against the radiation dose delivered in an imaging study. For imaging with ionizing radiation, this curve would start at zero change for zero dose, because the study would not be performed, then it would increase with dose and reach a plateau where an increase in dose has little or no effect. Finally, as dose increases still further, the change in QALYs would eventually come down and maybe even go negative, because at some level ionizing radiation always does damage. The methods described in this paper may assist in locating the plateau rigorously and determining how much latitude clinicians have in choosing the operating point on an SEC curve.

Acknowledgments

This research was supported by the National Institutes of Health under grants R01 EB000803 and P41 EB002035 at the University of Arizona. This work was also supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, Division of Cancer Epidemiology and Genetics.

We wish to thank Luca Caucci for careful reading of the manuscript and thoughtful comments and Helen Fan for preparing the figures.

References

- AAPM 2011 Position statement on radiation risks from medical imaging procedures (pp 25-A) (www.aapm.org/policies/details.asp?id=318&type=pp)
- Abbey C K and Barrett H H 2001 Human and model-observer performance in ramp-spectrum noise: effects of regularization and object variability *J. Opt. Soc. Am. A* **18** 473–88
- Abbey C K and Bochud F 2000 Modeling visual detection tasks in correlated image noise with linear model observers *Handbook of Medical Imaging (SPIE Press vol 1)* (Bellingham WA: SPIE) pp 629–54
- Abbey C K, Clarkson E, Barrett H H, Müller S P and Rybicki F J 1998 A method for approximating the density of maximum-likelihood and maximum a posteriori estimates under a gaussian noise model *Med. Image Anal.* **2** 395–403
- Agostinelli S *et al* 2003 GEANT4—a simulation toolkit *Nucl. Instrum. Methods Phys. Res. A* **506** 250–303
- Allison J *et al* 2006 Geant4 developments and applications *IEEE Trans. Nucl. Sci.* **53** 270–8
- Baek J, Pineda A R and Pelc N J 2013 To bin or not to bin? The effect of CT system limiting resolution on noise and detectability *Phys. Med. Biol.* **58** 1433
- Bakic P R, Albert M, Brzakovic D and Maidment A D 2002 Mammogram synthesis using a 3D simulation. I. Breast tissue model and image acquisition simulation *Med. Phys.* **29** 2131–9
- Bakic P R, Zhang C and Maidment A D 2011 Development and characterization of an anthropomorphic breast software phantom based upon region-growing algorithm *Med. Phys.* **38** 3165–76
- Barndorff-Nielsen O E 1983 On a formula for the distribution of the maximum likelihood estimator *Biometrika* **70** 343–65
- Barrett H H 1972a Pulse compression techniques in nuclear medicine *Proc. IEEE* **60** 723–4
- Barrett H H 1972b Fresnel zone plate imaging in nuclear medicine *J. Nucl. Med.* **13** 382–5
- Barrett H H 1990 Objective assessment of image quality: effects of quantum noise and object variability *J. Opt. Soc. Am. A* **7** 1266–78
- Barrett H H 2007 Statistical characterization of radiological images: basic principles and recent progress *Proc. SPIE* **6510** 651002
- Barrett H H 2009 NEQ: its progenitors and progeny *Proc. SPIE* **7263** 72630F
- Barrett H H, Abbey C K and Clarkson E 1998a Objective assessment of image quality: III. ROC metrics, ideal observers and likelihood-generating functions *J. Opt. Soc. Am. A* **15** 1520–35
- Barrett H H, Abbey C K, Gallas B and Eckstein M 1998b Stabilized estimates of Hotelling-observer detection performance in patient-structured noise *Proc. SPIE* **3340** 27–43
- Barrett H H, Denny J L, Wagner R F and Myers K J 1995 Objective assessment of image quality: II. Fisher information, Fourier crosstalk, and figures of merit for task performance *J. Opt. Soc. Am. A* **12** 834–52
- Barrett H H, Gooley T A, Girodias K A, Rolland J P, White T A and Yao J 1992 Linear discriminants and image quality *Image Vision Comput.* **10** 451–60
- Barrett H H, Hunter W C J, Miller B W, Moore S K, Chen Y and Furenlid L R 2009 Maximum-likelihood methods for processing signals from gamma-ray detectors *IEEE Trans. Nucl. Sci.* **56** 725–35
- Barrett H H, Kupinski M A, Mueller S, Halpern H J, Morris J C and Dwyer R 2013 Objective assessment of image quality VI: Imaging in radiation therapy *Phys. Med. Biol.* **58** 8197–213
- Barrett H H and Myers K J 2004 *Foundations of Image Science* (Hoboken, NJ: Wiley)
- Barrett H H, Myers K J, Devaney N and Dainty J C 2006 Objective assessment of image quality: IV. Application to adaptive optics *J. Opt. Soc. Am. A* **23** 3080–105
- Barrett H H, Myers K J, Gallas B D, Clarkson E and Zhang H 2001 Megalopinakophobia: its symptoms and cures *Proc. SPIE* **4320** 299–307
- Barrett H H, Parra L and White T A 1997 List-mode likelihood *J. Opt. Soc. Am. A* **14** 2914–23

- Barrett H H, Wilson D W and Tsui B M W 1994 Noise properties of the EM algorithm: I. Theory *Phys. Med. Biol.* **39** 833–46
- Barrett H H, Yao J, Rolland J P and Myers K J 1993 Model observers for the assessment of image quality *Proc. Natl Acad. Sci.* **90** 9758–65
- Baró J, Sempau J, Fernández-Varea J M and Salvat F 1995 PENELOPE: an algorithm for Monte Carlo simulation of the penetration and energy loss of electrons and positrons in matter *Nucl. Instrum. Methods Phys. Res. B* **100** 31–46
- Becker J, Zankl M, Fill U and Hoeschen C 2007 About Katja, a virtual human phantom of a 24 week pregnant woman *Sci. Rep.: J. Univ. Appl. Sci. Mittweide* **3** 5–7
- Bielajew A F, Hirayama H, Nelson W R and Rogers D W O 1994 History, overview, and recent improvements of EGS4 *KEK Internal* **94** 4
- Bochud F O, Abbey C K and Eckstein M P 1998 Statistical texture synthesis of mammographic images with clustered lumpy backgrounds *Opt. Express* **4** 33–43
- Boice J D Jr *et al* 1987 Radiation dose and leukemia risk in patients treated for cancer of the cervix *J. Natl Cancer Inst.* **79** 1295–311
- Boice J D Jr *et al* 1988 Radiation dose and second cancer risk in patients treated for cancer of the cervix *Radiat. Res.* **116** 3–55
- Brooks F J 2013 On some misconceptions about tumor heterogeneity quantification *Eur. J. Nucl. Med. Mol. Imaging* **40** 1292–4
- Burgess A E 1984 Statistical efficiency of perceptual decisions *Proc. SPIE* **454** 18–26
- Burgess A E 2011 Visual perception studies and observer models in medical imaging *Semin. Nucl. Med.* **41** 419–36
- Burgess A E and Colborne B 1988 Visual signal detection IV. Observer inconsistency *J. Opt. Soc. Am. A* **5** 617–27
- Campbell F W and Robson J G 1968 Application of Fourier analysis to the visibility of gratings *J. Physiol.* **197** 551–66
- Cardis E *et al* 2007 The 15-country collaborative study of cancer risk among radiation workers in the nuclear industry: estimates of radiation-related cancer risks *Radiat. Res.* **167** 396–416
- Caucci L and Barrett H H 2012 Objective assessment of image quality: V. Photon-counting detectors and list-mode data *J. Opt. Soc. Am. A* **29** 1003–16
- Chakraborty D P and Berbaum K S 2004 Observer studies involving detection and localization: modeling, analysis, and validation *Med. Phys.* **31** 2313–30
- Chawla A, Lo J, Bake J and Samei E 2009 Optimized image acquisition for breast tomosynthesis in projection and reconstruction space *Med. Phys.* **36** 4859
- Chesters M S 1992 Human visual perception and ROC methods in medical imaging *Phys. Med. Biol.* **37** 1433–76
- Clarkson E 2012 Asymptotic ideal observers and surrogate figures of merit for signal detection with list-mode data *J. Opt. Soc. Am. A* **29** 2204–16
- Clarkson E 2007 Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks *J. Opt. Soc. Am. A* **24** B91–8
- Clarkson E and Shen F 2010 Fisher information and surrogate figures of merit for the task-based assessment of image quality *J. Opt. Soc. Am. A* **27** 2313–26
- Dance D R 1990 Monte Carlo calculation of conversion factors for the estimation of mean glandular breast dose *Phys. Med. Biol.* **35** 1211–9
- Dance D R, Skinner C L, Young K C, Beckett J R and Kotre J 2000 Additional factors for the estimation of mean glandular breast dose using the UK mammography dosimetry protocol *Phys. Med. Biol.* **45** 3225
- Das M, Gifford H C, O'Connor M and Glick S J 2011 Penalized maximum likelihood reconstruction for improved microcalcification detection in breast tomosynthesis *IEEE Trans. Med. Imaging* **30** 904–14
- Das M, Gifford H C, O'Connor J M and Glick S J 2009 Evaluation of a variable dose acquisition technique for microcalcification and mass detection in digital breast tomosynthesis *Med. Phys.* **36** 1976–84
- Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm *J. R. Stat. Soc. B* **39** 1–38
- Dobbins J T and Godfrey D J 2003 Digital x-ray tomosynthesis: current state of the art and clinical potential *Phys. Med. Biol.* **48** R65–106
- Doll R 1995 Hazards of ionising radiation: 100 years of observations on man *Br. J. Cancer* **72** 1339–49
- Dubrova Y E, Nesterov V N, Krouchinsky N G, Ostapenko V A, Neumann R, Neil D L and Jeffreys A J 1996 Human minisatellite mutation rate after the Chernobyl accident *Nature* **380** 683–6

- Drzymala R E, Mohan R, Brewster L, Chu J, Goitein M, Harms W and Urie M 1991 Dose–volume histograms *Int. J. Radiat. Oncol. Biol. Phys.* **21** 71–8
- Ebersberger U *et al* 2013 CT evaluation of coronary artery stents with iterative image reconstruction: improvements in image quality and potential for radiation dose reduction *Eur. Radiol.* **23** 125–32
- Eckstein M P, Abbey C K and Bochud F O 2000 Practical guide to model observers for visual detection in synthetic and natural noisy images *Handbook of Medical Imaging* vol 1 (Bellingham WA: SPIE) pp 629–54
- Eckstein M P, Ahumada A J and Watson A B 1997 Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise *J. Opt. Soc. Am. A* **13** 1777–87
- Edwards A A and Lloyd D C 1998 Risk from ionising radiation: deterministic effects *J. Radiol. Prot.* **18** 175–83
- El Fakhri G, Santos P, Badawi R D, Holdsworth C H, Van den Abbeele A D and Kijewski M F 2007 Impact of acquisition geometry, image processing, and patient habitus on tumor detection in whole-body FDG-PET *J. Nucl. Med.* **58** 1951–60
- El Fakhri G, Surti S, Trott C M, Scheuermann J and Karp J S 2011 Improvement in lesion detection with whole-body oncologic TOF-PET *J. Nucl. Med.* **52** 347–53
- Erlandsson K, Buvat I, Pretorius P H, Thomas B A and Hutton B F 2012 A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology *Phys. Med. Biol.* **57** R119–59
- Evans P M 2008 Anatomical imaging for radiotherapy *Phys. Med. Biol.* **53** R151
- Ferrari A, Fasso A, Ranft J and Sala P R 2005 FLUKA: a multi-particle transport code *CERN Technical Report* CERN-2005-010
- Fessler J A 1996 Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): application to tomography *IEEE Trans. Image Proc.* **5** 493–506
- Fisher R 1936 On the mathematical foundations of theoretical statistics *Trans. R. Soc. Lond.* **222** 309–68
- Fiete R D, Barrett H H, Smith W E and Myers K J 1987 The Hotelling trace criterion and its correlation with human observer performance *J. Opt. Soc. Am. A* **4** 945–53
- Friebe A 1902 Demonstration eines Cancroids des rechten Handrückens, das sich nach langdauernder Einwirkung von Röntgenstrahlen bei einem 33 jährigen Mann entwickelt hatte *Fortsch. Röntgenstrahlen* **6** 106
- Fryback D and Thornbury J 1991 The efficacy of diagnostic imaging *Med. Decis. Making* **11** 88–94
- Gallas B D and Barrett H H 2003 Validating the use of channels to estimate the ideal linear observer *J. Opt. Soc. Am. A* **20** 1725–39
- Gang G J, Tward D J, Lee J and Siewerdsen J H 2010 Anatomical background and generalized detectability in tomosynthesis and cone-beam CT *Med. Phys.* **37** 1948–65
- Gifford H C, Didier C S, Das M and Glick S J 2008 Optimizing breast-tomosynthesis acquisition parameters with scanning model observers *Proc. SPIE* **6917** 69170S
- Gilchrist T C 1897 A case of dermatitis due to the x-rays *Bull. Johns Hopkins Hospital* **8** 17–22
- Giussani A *et al* 2012 A compartmental model for biokinetics and dosimetry of 18F-choline in prostate cancer patients *J. Nucl. Med.* **53** 985–93
- Glick S J 2007 Breast CT *Annu. Rev. Biomed. Eng.* **9** 501–26
- Graham N and Nachmias J 1971 Detection of grating patterns containing two spatial frequencies: a comparison of single channel and multichannels models *Vision Res.* **11** 251–9
- Green D M and Swets J A 1966 *Signal Detection Theory and Psychophysics* (New York: Wiley)
- Gross K A, Kupinski M A, Peterson T E and Clarkson E 2003 Optimizing a multiple-pinhole SPECT system using the ideal observer *Proc. SPIE* **5034** 314–22
- Hanson K M 1979 Detectability in computed tomographic images *Med. Phys.* **6** 441–51
- Hanson K M 1990 Method to evaluate image-recovery algorithms based on task performance *J. Opt. Soc. Am. A* **7** 1294–304
- He X, Caffo B S and Frey E C 2008 Toward realistic and practical ideal observer (IO) estimation for the optimization of medical imaging systems *IEEE Trans. Med. Imaging* **27** 1535–43
- Hesterman J Y, Caucchi L, Kupinski M A, Barrett H H and Furenlid L R 2010 Maximum-likelihood estimation with a contracting-grid search algorithm *IEEE Trans. Nucl. Sci.* **57** 1077–84
- Hesterman J Y, Kupinski M A, Clarkson E and Barrett H H 2007 Hardware assessment using the multi-module, multi-resolution system (M3R)—a signal-detection study *Med. Phys.* **34** 2007
- Hoeschen C, Fill U, Zankl M, Panzer W, Requella D and Dohring W 2005 A high-resolution voxel phantom of the breast for dose calculations in mammography *Med. Decis. Making* **114** 406–9
- Hotelling H 1931 The generalization of Student's ratio *Ann. Math. Stat.* **2** 360–78

- Hou Y, Liu X, Shu X, Guo W and Guo Q 2012 Comparisons of image quality and radiation dose between iterative reconstruction and filtered back projection reconstruction algorithms in 256-MDCT coronary angiography *Am. J. Roentgenol.* **199** 588–94
- Hsu W-L *et al* 2013 The incidence of leukemia, lymphoma and multiple myeloma among atomic bomb survivors: 1950–2001 *Radiat. Res.* **179** 361–82
- Husarik D B, Marin D, Samei E, Richard S, Chen B, Jaffe T A, Bashir M R and Nelson R C 2012 Radiation dose reduction in abdominal computed tomography during the late hepatic arterial phase using a model-based iterative reconstruction algorithm: How low can we go? *Invest. Radiol.* **47** 468–74
- ICRP 1979 Report of the task group on reference man *Ann. ICRP* **3** iii
- ICRP 1994 Publication 66: Human respiratory tract model for radiological protection: a report of a task group of the International Commission on Radiological Protection *Ann. ICRP* **24** 1–482
- ICRP 2007 Publication 103 The 2007 recommendations of the International Commission on Radiological Protection *Ann. ICRP* **37** 1–332
- ICRP 2012 Publication 118: ICRP statement on tissue reactions and early and late effects of radiation in normal tissues and organs—threshold doses for tissue reactions in a radiation protection context *Ann. ICRP* **41** 1–322
- ICRU 1997 Report 54: Medical imaging: the assessment of image quality *Radiography* **3** 243–4
- Judy P F and Swensson R G 1981 Lesion detection and signal-to-noise ratio in CT images *Med. Phys.* **8** 13–23
- Kalluri K S, Mahd M and Glick S J 2013 Investigation of energy weighting using an energy discriminating photon counting detector for breast CT *Med. Phys.* **40** 081923
- Kalra M K *et al* 2012 Radiation dose reduction with sinogram affirmed iterative reconstruction technique for abdominal computed tomography *J. Comput. Assisted Tomography* **36** 339–46
- Katsura M, Matsuda I, Akahane M, Sato J, Akai H, Yasaka K, Kunimatsu A and Ohtomo K 2012 Model-based iterative reconstruction technique for radiation dose reduction in chest CT: comparison with the adaptive statistical iterative reconstruction technique *Eur. Radiol.* **22** 1613–23
- Kay S M 1998 *Fundamentals of Statistical Signal Processing: Detection Theory* (Englewood Cliffs, NJ: Prentice Hall)
- Khurd P and Gindi G 2005 Decision strategies that maximize the area under the LROC curve *IEEE Trans. Med. Imaging* **24** 1626–36
- Kijewski M F, Moore S C, Jadvar H, Zimmerman R E and Müller S P 2001 Effects of SPECT collimation and system geometry on classification tasks related to Parkinson's disease *IEEE Trans. Nucl. Sci.* **48** 734–8
- Kramer R, Khoury H J, Vieira J W, Loureiro E C, Lima V J, Lima F R and Hoff G 2004 All about FAX: a female adult voxel phantom for Monte Carlo calculation in radiation protection dosimetry *Phys. Med. Biol.* **49** 5203–16
- Kupinski M K, Barrett H H and Clarkson E 2013 Scanning linear estimation: improvements over region of interest (ROI) methods *Phys. Med. Biol.* **58** 1283–301
- Kupinski M A, Clarkson E, Gross K and Hoppin J W 2003a Optimizing imaging hardware for estimation tasks *Proc. SPIE* **5034** 309–13
- Kupinski M A, Clarkson E and Hesterman J Y 2007 Bias in Hotelling observer performance computed from finite data *Proc. SPIE* **6515** 65150
- Kupinski M A, Clarkson E, Hoppin J, Chen L and Barrett H H 2003b Experimental determination of object statistics from noisy images *J. Opt. Soc. Am. A* **20** 421–9
- Kupinski M A, Hoppin J, Clarkson E and Barrett H H 2003c Ideal-observer computation using Markov-chain Monte Carlo *J. Opt. Soc. Am. A* **20** 430–8
- Kyprianou I S, Badano A, Park S, Liu H and Myers K J 2009 Noise and signal detection in digital x-ray detectors using the spatial definition of SNR *Proc. SPIE* **7258** 725819
- Lartizen C, Kinahan P E and Comtat C 2004 A lesion detection observer study comparing 2-dimensional versus fully 3-dimensional whole-body PET imaging protocols *J. Nucl. Med.* **45** 714–23
- Lau B A, Reiser I, Nishikawa R M and Bakic P R 2012 A statistically defined anthropomorphic software breast phantom *Med. Phys.* **39** 3375–85
- Lechel U, Becker C R, Langenfeld-Jäger G and Brix G 2007 Dosisreduktion durch strommodulierte Dosisautomatik bei der MSCT: Vergleich von Messung und Rechnung *Fortschr. Röntgenstrahlen* **179** WI_PO_100
- Lee C, Williams J L and Bolch W E 2006 Whole-body voxel phantoms of paediatric patients—UF series B *Phys. Med. Biol.* **51** 4649–61

- Lewellen T K 2008 Recent developments in PET detector technology *Phys. Med. Biol.* **53** R287–317
- Li X, Samei E, Segars W P, Sturgeon G M, Colsher J G and Frush D P 2011a Patient-specific radiation dose and cancer risk for pediatric chest CT *Radiology* **259** 862–74
- Li X, Samei E, Segars W P, Sturgeon G M, Colsher J G, Tonchey G, Yoshizumi T T and Frush D P 2011b Patient-specific radiation dose and cancer risk estimation in CT: part II. Application to patients *Med. Phys.* **38** 408–19
- Li X, Samei E, Williams C H, Segars W P, Tward D J, Miller M I, Ratnanather J T, Paulson E K and Frush D P 2012 Effects of protocol and obesity on dose conversion factors in adult body CT *Med. Phys.* **39** 6550–71
- Little M P and Muirhead C R 1998 Curvature in the cancer mortality dose response in Japanese atomic bomb survivors: absence of evidence of threshold *Int. J. Radiat. Biol.* **74** 471–80
- Little M P 2002 Comparisons of lung tumour mortality risk in the Japanese A-bomb survivors and in the Colorado Plateau uranium miners: support for the ICRP lung model *Int. J. Radiat. Biol.* **78** 145–63
- Little M P 2010 Cancer models, genomic instability and somatic cellular Darwinian evolution *Biol. Direct* **5** 19
- Little M P 2013 A review of non-cancer effects, especially circulatory and ocular diseases *Radiat. Environ. Biophys.* **52** 435–49
- Little M P, Tawn E J, Tzoulaki I, Wakeford R, Hildebrandt G, Paris F, Tapio S and Elliott P 2010 Review and meta-analysis of epidemiological associations between low/moderate doses of ionizing radiation and circulatory disease risks, and their possible mechanisms *Radiat. Environ. Biophys.* **49** 139–53
- Little M P and Charles M W 1997 The risk of non-melanoma skin cancer incidence in the Japanese atomic bomb survivors *Int. J. Radiat. Biol.* **71** 589–602
- Little M P and Muirhead C R 1996 Evidence for curvilinearity in the cancer incidence dose-response in the Japanese atomic bomb survivors *Int. J. Radiat. Biol.* **70** 83–94
- Little M P *et al* 2012 Systematic review and meta-analysis of circulatory disease from exposure to low-level ionizing radiation and estimates of potential population mortality risks *Environ. Health Perspectives* **120** 1503–11
- Little M P, Goodhead D T, Bridges B A and Bouffler S D 2013 Evidence relevant to untargeted and transgenerational effects in the offspring of irradiated parents *Mutation Res. Rev.* **753** 50–67
- Little M P, Hoel D G, Molitor J, Boice J D Jr, Wakeford R and Muirhead C R 2008a New models for evaluation of radiation-induced lifetime cancer risk and its uncertainty employed in the UNSCEAR 2006 report *Radiat. Res.* **169** 660–76
- Little M P, Tawn E J, Tzoulaki I, Wakeford R, Hildebrandt G, Paris F, Tapio S and Elliott P 2008b A systematic review of epidemiological associations between low and moderate doses of ionizing radiation and late cardiovascular effects, and their possible mechanisms *Radiat. Res. Rev.* **169** 99–109
- Little M P, Wakeford R, Tawn E J, Bouffler S D and Berrington G A 2009 Risks associated with low doses and low dose rates of ionizing radiation: Why linearity may be (almost) the best we can do *Radiology* **251** 6–12
- Little M P, Weiss H A, Boice J D Jr, Darby S C, Day N E and Muirhead C R 1999 Risks of leukemia in Japanese atomic bomb survivors, in women treated for cervical cancer, and in patients treated for ankylosing spondylitis *Radiat. Res.* **152** 280–92
- Ljungberg M and King M A 2012 Introduction to task-based assessment of image quality for investigators employing Monte Carlo simulation of imaging systems *Monte Carlo Calculations in Nuclear Medicine* 2nd edn Ljungberg M *et al* (Boca Raton, FL: CRC Press) pp 287–312
- Lu Y and Chan H P 2011 Image quality of microcalcifications in digital breast tomosynthesis: effects of projection-view distributions *Med. Phys.* **38** 5703
- Makeev A and Glick S J 2013 Investigation of statistical iterative reconstruction for dedicated breast CT *Med. Phys.* **40** 081904
- Marks D G, Barber H B, Barrett H H, Tueller J and Woolfenden J M 1999 Improving performance of a CdZnTe imaging array by mapping the detector with gamma rays *Nucl. Instrum. Methods Phys. Res. A* **428** 102–12
- Mattsson S 2011 Radiation dose management in CT, SPECT/CT and PET/CT techniques *Radiat. Prot. Dosim.* **147** 13–21
- May M S, Wüst W, Brand M, Stahl C, Allmendinger T, Schmidt B, Uder M and Lell M M 2011 Dose reduction in abdominal computed tomography: Intraindividual comparison of image quality of

- full-dose standard and half-dose iterative reconstructions with dual-source computed tomography *Invest. Radiol.* **46** 465–70
- McCollough C H, Primak A N, Braun N, Kofler J, Yu L and Christner J 2009 Strategies for reducing radiation dose in CT *Radiol. Clinics North Am.* **47** 27–40
- McGale P and Darby S C 2005 Low doses of ionizing radiation and circulatory diseases: a systematic review of the published epidemiological evidence *Radiat. Res.* **163** 247–57
- McGale P and Darby S C 2008 Commentary: a dose-response relationship for radiation-induced heart disease—current issues and future prospects *Int. J. Epidemiol.* **37** 518–23
- McMillan T J *et al* 2010 Circulatory disease risk: Report of the independent Advisory Group on Ionising Radiation Documents of the HPA Report RCE-16 pp 1–116
- Melsa J L and Cohn D L 1978 *Decision and Estimation Theory* (New York: McGraw Hill)
- Menzel H G, Clement C and DeLuca P 2009 ICRP publication 110. Realistic reference phantoms: an ICRP/ICRU joint effort: a report of adult reference computational phantoms *Ann. ICRP* **39** 1–6
- Metz C E 1978 Basic principles of ROC analysis *Semin. Nucl. Med.* **8** 283–98
- Metz C E 2006 Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems *J. Am. College Radiol.* **3** 413–22
- Miéville F A, Gudinchet F, Brunelle F, Bochud F O and Verdun F R 2013 Iterative reconstruction methods in two different MDCT scanners: physical metrics and 4-alternative forced-choice detectability experiments—a phantom approach *Phys. Med.* **29** 99–110
- Mortimer D and Segal L 2007 Comparing the incomparable? A systematic review of competing techniques for converting descriptive measures of health status into QALY-weights *Med. Decis. Making* **28** 66–89
- Mostafavi H and Sakrison D 1976 Structure and properties of a single channel in the human visual system *Vision Res.* **16** 957–68
- Muehllehner G and Karp J S 2006 Positron emission tomography *Phys. Med. Biol.* **51** R117
- Muirhead C R, O'Hagan J A, Haylock R G E, Phillipson M A, Willcock T, Berridge G L C and Zhang W 2009 Mortality and cancer incidence following occupational radiation exposure: third analysis of the National Registry for Radiation Workers *Br. J. Cancer* **100** 206–12
- Myers K J and Barrett H H 1987 Addition of a channel mechanism to the ideal-observer model *J. Opt. Soc. Am. A* **4** 2447–57
- Myers K J, Barrett H H, Borgstrom M C, Patton D D and Seeley G W 1985 Effect of noise correlation on detectability of disk signals in medical imaging *J. Opt. Soc. Am. A* **2** 1752–9
- Myers K J, Rolland J P, Barrett H H and Wagner R F 1990 Aperture optimization for emission imaging: effect of a spatially varying background *J. Opt. Soc. Am. A* **7** 1279–93
- National Research Council 1999 *Committee on the Health Effects of Exposure to Radon (Beir IV)* (Washington, DC: National Academies Press) pp 1–500
- National Research Council 2006 *Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation (BEIR VII Phase 2)* (Washington, DC: National Academies Press) pp 1–406
- Neel J V, Schull W J, Awa A A, Satoh C, Kato H, Otake M and Yoshimoto Y 1990 The children of parents exposed to atomic bombs: estimates of the genetic doubling dose of radiation for humans *Am. J. Human Genet.* **46** 1053–72
- Nestle U, Weber W, Hentschel M and Grosu A L 2009 Biological imaging in radiation therapy: role of positron emission tomography *Phys. Med. Biol.* **54** R1–25
- Niemierko A 1997 Reporting and analyzing dose distributions: A concept of equivalent uniform dose *Med. Phys.* **24** 103–10
- Nuyts J, De Man B, Fessler J A, Zbijewski W and Beekman F J 2013 Modelling the physics in the iterative reconstruction for transmission computed tomography *Phys. Med. Biol.* **58** R63–96
- Otake M and Schull W J 1998 Radiation-related brain damage and growth retardation among the prenatally exposed atomic bomb survivors *Int. J. Radiat. Biol.* **74** 159–71
- Otake M, Schull W J and Lee S 1996 Threshold for radiation-related severe mental retardation in prenatally exposed A-bomb survivors: a re-analysis *Int. J. Radiat. Biol.* **70** 755–63
- Ozasa Y, Shimizu K, Suyama A, Kasagi F, Soda M, Grant E J, Sakata R, Sugiyama H and Kodama K 2012 Studies of the mortality of atomic bomb survivors, report 14, 1950–2003: an overview of cancer and noncancer diseases *Radiat. Res.* **177** 229–43
- Park S and Clarkson E 2009 Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds *J. Opt. Soc. Am. A* **26** B59–71
- Park S, Clarkson E, Kupinski M A and Barrett H H 2005 Efficiency of the human observer detecting random signals in random backgrounds *J. Opt. Soc. Am. A* **22** 3–16

- Park S, Jennings R, Liu H, Badano A and Myers K J 2010 A statistical, task-based evaluation method for three-dimensional x-ray breast imaging systems using variable-background phantoms *Med. Phys.* **37** 6253–70
- Parra L and Barrett H H 1998 List-mode likelihood: EM algorithm and noise estimation demonstrated on 2D-PET *IEEE Trans. Med. Imaging* **MI-17** 228–35
- Patton D D and Woolfenden J M 1989 A utility-based model for comparing the cost-effectiveness of diagnostic studies *Invest. Radiol.* **24** 263–71
- Peterson W W, Birdsall T G and Fox W C 1954 The theory of signal detectability *Proc. IRE Prof. Group Inform. Theory* **4** 171–212
- Peterson T E and Furenlid L R 2011 SPECT detectors: the Anger camera and beyond *Phys. Med. Biol.* **56** R145–82
- Petoussi-Henss N, Zanki M, Fill U and Regulla D 2002 The GSF family of voxel phantoms *Phys. Med. Biol.* **47** 89–106
- Pierce D A and Preston D L 2000 Radiation-related cancer risks at low doses among atomic bomb survivors *Radiat. Res.* **154** 178–86
- Pineda A, Siewerdsen J H and Tward D J 2008 Analysis of image noise in 3D cone-beam CT: spatial and Fourier domain approaches under conditions of varying stationarity *Proc. SPIE* **6913** 69131Q
- Pineda A R, Tward D J, Gonzalez A and Siewerdsen J H 2012 Beyond noise power in 3D computed tomography: the local NPS and off-diagonal elements of the Fourier domain covariance matrix *Med. Phys.* **39** 3240–52
- Pineda A R, Yoon S, Paik D S and Fahrig R 2006 Optimization of a tomosynthesis system for the detection of lung nodules *Med. Phys.* **33** 1372–9
- Platisa L, Goossens B, Vansteenkiste E, Park S, Gallas B D, Badano A and Philips W 2011 Channelized Hotelling observers for the assessment of volumetric imaging data sets *J. Opt. Soc. Am. A* **28** 1145–63
- Pliskin J S, Shepard D S and Weinstein M C 1980 Utility functions for life years and health status *Operat. Res.* **28** 206–24
- Popescu L M and Myers K J 2013 CT image assessment by low contrast signal detectability evaluation with unknown signal location *Med. Phys.* **40** 111908
- Pozuelo F, Gallardo S, Querol A, Verdu G and Rodenas J 2012 X-ray simulation with the Monte Carlo code PENELOPE: application to quality control *Conf. Proc. IEEE Engineering in Medicine and Biology Society (San Diego, CA, 28 August–1 September 2012)* vol 2012, pp 5777–80
- Preston D L, Ron E, Tokuoka S, Funamoto S, Nishi N, Soda M, Mabuchi K and Kodama K 2007 Solid cancer incidence in atomic bomb survivors: 1958–1998 *Radiat. Res.* **168** 1–64
- Qi J and Leahy R M 2000 Resolution and noise properties of MAP reconstruction for fully 3-D PET *IEEE Trans. Med. Imaging* **19** 493–506
- Qi J and Leahy R M 2006 Iterative reconstruction techniques *Phys. Med. Biol.* **51** R541–78
- Qi L-P, Li Y, Tang L, Li Y-L, Li X-T, Cui Y, Sun Y-S and Zhang X-P 2012 Evaluation of dose reduction and image quality in chest ct using adaptive statistical iterative reconstruction with the same group of patients *Br. J. Radiol.* **85** 906–11
- Reed A L 2007 Medical physics calculations with MCNP: a primer LANL *Technical Report* LA-UR-07-413
- Reiser I and Nishikawa R 2010 Task-based assessment of breast tomosynthesis: effect of acquisition parameters and quantum noise *Med. Phys.* **37** 1591–600
- Richard S and Siewerdsen J H 2008 Comparison of model and human observer performance for detection and discrimination tasks using dual-energy x-ray images *Med. Phys.* **35** 5043–53
- Richard S, Siewerdsen J H, Moseley D J, Jaffray D A and Bakhtiar B 2005 A generalized DQE description of dual-energy imaging with flat-panel detectors *Med. Phys.* **32** 1397–413
- Rogers W L, Han K S, Jones L W and Beierwaltes W H 1972 Application of a Fresnel zone plate to gamma-ray imaging *J. Nucl. Med.* **13** 612
- Rogulski M M, Barber H B, Barrett H H, Shoemaker R L and Woolfenden J M 1999 Ultra-high-resolution brain SPECT: simulation results *IEEE Trans. Nucl. Sci.* **40** 1123–9
- Rolland J P and Barrett H H 1992 Effect of random background inhomogeneity on observer detection performance *J. Opt. Soc. Am. A* **9** 649–58
- Ron E, Preston D L, Kishikawa M, Kobuke T, Iseki M, Tokuoka S, Tokunaga M and Mabuchi K 1998 Skin tumor risk among atomic-bomb survivors in Japan *Cancer Causes Control* **9** 393–401
- Rupcich F, Badal A, Popescu L M, Kyprianou I and Gilat-Schmidt T 2013 Reducing radiation dose to the female breast during CT coronary angiography: a simulation study comparing breast shielding, angular tube current modulation, reduced kV, and partial angle protocols using an unknown-location signal-detectability metric *Med. Phys.* **40** 081921

- Sachs M, Nachmias J and Robson J 1971 Spatial-frequency channels in human vision *J. Opt. Soc. Am.* **61** 1176–86
- Sato J, Akahane M, Inano S, Terasaki M, Akai H, Katsura M, Matsuda I, Kunimatsu A and Ohtomo K 2012 Effect of radiation dose and adaptive statistical iterative reconstruction on image quality of pulmonary computed tomography *Japan. J. Radiol.* **30** 146–53
- Schlattl H, Zankl M, Becker J and Hoeschen C 2012 Dose conversion coefficients for paediatric CT examinations with automatic tube current modulation *Phys. Med. Biol.* **57** 6309–26
- Schlattl H, Zankl M, Hausleiter J and Hoeschen C 2007 Local organ dose conversion coefficients for angiographic examinations of coronary arteries *Phys. Med. Biol.* **52** 4393–408
- Schmidt B and Kalender W A 2002 A fast voxel-based Monte Carlo method for scanner- and patient-specific dose calculations in computed tomography *Eur. J. Med. Phys.* **18** 43–53
- Schonhoff T A and Giordano A A 2006 *Detection and Estimation Theory and Its Applications* (Upper Saddle River, NJ: Pearson Education)
- Sechopoulos I 2013a A review of breast tomosynthesis. Part I. The image acquisition process *Med. Phys.* **40** 014301
- Sechopoulos I 2013b A review of breast tomosynthesis. Part II. Image reconstruction, processing and analysis, and advanced applications *Med. Phys.* **40** 014302
- Segars W P, Sturgeon G, Mendonca S, Grimes J and Tsui B M 2010 4D XCAT phantom for multimodality imaging research *Med. Phys.* **37** 4902–15
- Shen F and Clarkson E 2006 Using Fisher information to approximate ideal-observer performance on detection tasks for lumpy-background images *J. Opt. Soc. Am. A* **23** 2406–14
- Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imaging* **1** 113–22
- Shipley W U, Tepper J E, Prout G R, Verhey L J, Mendiondo O A, Goitein M, Koehler A M and Suit H D 1979 Proton radiation as boost therapy for localized prostatic carcinoma *J. Am. Med. Assoc.* **241** 1912–5
- Shope T B, Gagne R M and Johnson G C 1981 A method for describing the doses delivered by transmission x-ray computed tomography *Med. Phys.* **8** 488–95
- Siewerdsen J H, Cunningham I A and Jaffray D A 2002 A framework for noise-power spectrum analysis of multidimensional images *Med. Phys.* **29** 2655–71
- Singh S, Kalra M K, Do S, Thibault J B, Pien H, Connor O J and Blake M A 2012 Comparison of hybrid and pure iterative reconstruction techniques with conventional filtered back projection: dose reduction potential in the abdomen *J. Comput. Assisted Tomography* **36** 347–53
- Smith W E and Barrett H H 1986 Hotelling trace criterion as a figure of merit for the optimization of imaging systems *J. Opt. Soc. Am. A* **3** 717–23
- Stabin M 2006 Nuclear medicine dosimetry *Phys. Med. Biol.* **51** R187
- Stevens L G 1896 Injurious effects on the skin *Br. Med. J.* **1** 998
- Strohmeier C F and Julesz B 1972 Spatial-frequency masking in vision: critical bands and spread of masking *J. Opt. Soc. Am.* **62** 1221–32
- Surti S, Scheuermann J, El Fakhri G, Daube M E, Abi-Hatem N, Moussallem E, Lim R, Benard F, Mankoff D and Karp J S 2011 Impact of TOF PET on whole-body oncologic studies: a human observer lesion detection and localization study *J. Nucl. Med.* **52** 712–9
- Swank R K 1973 Absorption and noise in x-ray phosphors *J. Appl. Phys.* **44** 4199–203
- Swets J A 1964 *Signal Detection and Recognition by Human Observers* (New York: Wiley)
- Tanner W P, Wilson P and Swets J A 1954 A decision-making theory of visual detection *Psychol. Rev.* **61** 401–9
- Tapiovaara M J and Wagner R F 1985 SNR and DQE analysis of broad spectrum x-ray imaging *Phys. Med. Biol.* **30** 519–29
- Thilander-Klang A C, Ackerholm P H, Berlin I C, Bjurstam N G, Mattsson S L, Mansson L G, von Schéle C and Thunberg S J 1997 Influence of anode-filter combinations on image quality and radiation dose in 965 women undergoing mammography *Radiology* **203** 348–54
- Thomas D C, Blettner M and Day N E 1992 Use of external rates in nested case-control studies with application to the International Radiation Study of Cervical Cancer Patients *Biometrics* **48** 781–94
- Thornbury J R 1994 Eugene W Caldwell lecture: Clinical efficacy of diagnostic imaging: love it or leave it *Am. J. Roentgenol.* **162** 1–8
- Tixier F, Hatt M, Le C C, Le Pogam A, Corcos L and Visvikis D 2012 Reproducibility of tumor uptake heterogeneity characterization through textural feature analysis in 18F-FDG PET *J. Nucl. Med.* **53** 693–700

- Tsui B M W, Metz C E, Atkins F B, Starr S J and Beck R N 1978 A comparison of optimum spatial resolution in nuclear imaging based on statistical theory and observer performance *Phys. Med. Biol.* **23** 654–76
- Tward D J, Siewerdsen J H, Daly M J, Jaffray D A and Paul N S 2007 Soft-tissue detectability in cone-beam CT: evaluation by 2AFC tests in relation to physical performance metrics *Med. Phys.* **34** 4459–71
- Tward D J and Siewerdsen J H 2008 Cascaded systems analysis of the 3D noise-power spectrum of flat-panel cone-beam CT *Med. Phys.* **25** 5510–29
- Tward D J and Siewerdsen J H 2009 Noise aliasing and the 3D NEQ of flat-panel cone-beam CT: effect of 2D/3D apertures and sampling *Med. Phys.* **36** 3830–43
- Tylavsky D J and Sohie G R L 1986 Generalization of the matrix inversion lemma *Proc. IEEE* **74** 1050–2
- UNSCEAR 1993 *Sources and Effects of Ionizing Radiation: UNSCEAR 1993 Report to the General Assembly, with Scientific Annexes* (New York: United Nations) pp 1–922
- UNSCEAR 2001 *Hereditary Effects of Radiation: UNSCEAR 2001 Report to the General Assembly, with Scientific Annex* (New York: United Nations) pp 1–160
- UNSCEAR 2008a. UNSCEAR 2006: Report annex B: Epidemiological evaluation of cardiovascular disease and other non-cancer diseases following radiation exposure (New York: United Nations) pp 325–83
- UNSCEAR 2008b. UNSCEAR 2006: Report annex A: Epidemiological studies of radiation and cancer (New York: United Nations) pp 13–322
- Van de Sompel D, Brady S M and Boone J 2011 Task-based performance analysis of FBP, SART and ML for digital breast tomosynthesis using CNR and channelised Hotelling observers *Med. Image Anal.* **15** 53–70
- Van Trees H L and Bell K L 2013 *Detection, Estimation and Modulation Theory, Part I* (Chichester: Wiley)
- Wang W and Gindi G 1997 Noise analysis of MAP-EM algorithms for emission tomography *Phys. Med. Biol.* **42** 2215–32
- Warren L M, Mackenzie A, Cooke J, Given-Wilson R M, Wallis M G, Chakraborty D P, Dance D R, Bosmans H and Young K C 2012 Effect of image quality on calcification detection in digital mammography *Med. Phys.* **39** 3202–13
- Whitaker M K, Clarkson E and Barrett H H 2008 Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods *Opt. Express* **16** 8150–73
- Wickens T D 2002 *Elementary Signal Detection Theory* (Oxford: Oxford University Press)
- Willemink M J, Leiner T, de Jong P A, de Heer L M, Nievelstein R A J, Schilham A M R and Budde R P J 2013 Iterative reconstruction techniques for computed tomography part 2: initial results in dose reduction and image quality *Eur. Radiol.* **23** 1632–42
- Wilson D W, Tsui B M W and Barrett H H 1994 Noise properties of the EM algorithm: II Monte Carlo simulations *Phys. Med. Biol.* **39** 847–72
- Wrixon A D 2008 New recommendations from the International Commission on Radiological Protection—a review *Phys. Med. Biol.* **53** R41–60
- Wunderlich A and Abbey C K 2013 Utility as a rationale for choosing observer performance assessment paradigms for detection tasks in medical imaging *Med. Phys.* **40** 111903
- Xu X G 2014 An exponential growth of computational phantom research in radiation protection, imaging, and radiotherapy: a review of the fifty-year history *Phys. Med. Biol.* **59** R233–302
- Xu X G, Chao T C and Bozkurt A 2000 VIP-man: an image-based whole-body adult male model constructed from color photographs of the visible human project for multi-particle Monte Carlo calculations *Health Phys.* **78** 476–86
- Xu X G and Eckerman K F 2010 *Handbook of Anatomical Models for Radiation Dosimetry* (London: Taylor and Francis)
- Xu J and Tsui B M W 2014 Quantifying the importance of the statistical assumption in statistical x-ray CT image reconstruction *IEEE Trans. Med. Imaging* **33** 61–73
- Yaffe M J and Rowlands J A 1997 X-ray detectors for digital radiography *Phys. Med. Biol.* **42** 1–39
- Yendiki A and Fessler J A 2007 Analysis of observer performance in unknown-location tasks for tomographic image reconstruction *J. Opt. Soc. Am. A* **24** B99–109
- Young S, Badal-Soler A, Myers K J and Park S 2012 A task-specific argument for variable-exposure breast tomosynthesis *11th Int. Workshop on Digital Mammography (Philadelphia, PA, USA, 8–11 July 2012)*

- Young S, Bakic P, Myers K J, Jennings R J and Park S 2013 A virtual trial framework for quantifying the detectability of masses in breast tomosynthesis projection data *Med. Phys.* **40** 051914–1
- Yu L, Leng S, Chen L, Kofler J M, Carter R E and McCollough C H 2013 Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms *Med. Phys.* **40** 041908
- Zankl M, Schlattl H, Petoussi-Henss N and Hoeschen C 2012 Electron specific absorbed fractions for the adult male and female ICRP/ICRU reference computational phantoms *Phys. Med. Biol.* **57** 4501–26
- Zankl M, Veit R, Williams G, Schneider K, Fendel H, Petoussi N and Drexler G 1988 The construction of computer tomographic phantoms and their application in radiology and radiation protection *Radiat. Environ. Biophys.* **27** 153–64
- Zhang B, Ma J, Liu L and Cheng J 2007 CNMAN: a chinese adult male voxel phantom constructed from color photographs of a visible anatomical data set *Radiat. Prot. Dosim.* **124** 130–6
- Zhou L and Gindi G 2009 Collimator optimization in SPECT based on a joint detection and localization task *Phys. Med. Biol.* **54** 4423–37
- Zhou L, Khurd P, Kulkarni S, Rangarajan A and Gindi G 2008 Aperture optimization in emission imaging using ideal observers for joint detection and localization *Phys. Med. Biol.* **53** 2019–34
- Zubal I G, Harrell C R, Smith E O, Rattner Z, Gindi G and Hoffer P B 1994 Computerized three-dimensional segmented human anatomy *Med. Phys.* **21** 299–302