Reliability of molecular imaging diagnostics

Synthese

Reliability of molecular imaging diagnostics

--Manuscript Draft--

Manuscript Number:	SYNT-D-18-01253R3
Full Title:	Reliability of molecular imaging diagnostics
Article Type:	S.I. : Reliability
Keywords:	philosophy of medicine; diagnostics; medical imaging; precision medicine; nuclear medicine; diagnostic tests; accuracy; validity; reliability; repeatability
Corresponding Author:	Elisabetta Lalumera, PhD Universita degli Studi di Milano-Bicocca Milano, ITALY
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universita degli Studi di Milano-Bicocca
Corresponding Author's Secondary Institution:	
First Author:	Elisabetta Lalumera, PhD
First Author Secondary Information:	
Order of Authors:	Elisabetta Lalumera, PhD
	Stefano Fanti, MD
	Giovanni Boniolo
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Advanced medical imaging, such as CT, fMRI and PET, has undergone enormous progress in recent years, both in accuracy and utilization. Such techniques often bring with them an illusion of immediacy, the idea that the body and its diseases can be directly be inspected. In this paper we target this illusion and address the issue of the reliability of advanced imaging tests as knowledge procedures, taking Positron Emission Tomography (PET) in oncology as paradigmatic case study. After individuating a suitable notion of reliability, we argue that i) PET is a highly theory-laden and non-immediate knowledge procedure, in spite of the photographic-like quality of the images it delivers; ii) the diagnostic conclusions based on the interpretation of PET images are population-dependent; iii) PET images require interpretation, which is inherently observer-dependent and therefore variable. We conclude with a three-step methodological proposal for enhancing reliability of advanced medical imaging.

Title Page with all author information

Elisabetta Lalumera (corresponding author), Dipartimento di Psicologia, Milano-Bicocca University, Milano, Italy, elisabetta.lalumera@unimib.it

Stefano Fanti, Policlinico S. Orsola, University of Bologna

Giovanni Boniolo, Dipartimento di Scienze Biomediche e Chirurgico Specialistiche, University of Ferrara

[1]

[1] * ** ***

Reliability of molecular imaging diagnostics

Abstract

Advanced medical imaging, such as CT, fMRI and PET, has undergone enormous progress in recent years, both in accuracy and utilization. Such techniques often bring with them an illusion of immediacy, the idea that the body and its diseases can be directly inspected. In this paper we target this illusion and address the issue of the reliability of advanced imaging tests as knowledge procedures, taking Positron Emission Tomography (PET) in oncology as paradigmatic case study. After individuating a suitable notion of reliability, we argue that i) PET is a highly theory-laden and non-immediate knowledge procedure, in spite of the photographic-like quality of the images it delivers; ii) the diagnostic conclusions based on the interpretation of PET images are population-dependent; iii) PET images require interpretation, which is inherently observer-dependent and therefore variable. We conclude with a three-step methodological proposal for enhancing the reliability of advanced medical imaging.

1. Introduction

In the last few decades, rapid progress in different converging scientific fields (molecular biology, medical genetics, computational and informational technology and biomedical technology) has gradually transformed the standard clinical practice into what is called *precision medicine*[1], namely, "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person" (National Institute of Health 2018). Especially thanks to the new sequencing technologies which allow to obtain individual information at *-omic* levels (genomics, epigenomics, transcriptomics, metabolomics, etc.), the goal of precision medicine is to substitute diagnoses and treatments for the "average" patient into well-defined individual patients belonging to a very narrowly defined groups of patient having the same -omic characteristics.

As an example, consider the case of a patient consulting her oncologist after breast cancer surgery. Twenty years ago, the only therapeutic options were hormonal suppression or broad-spectrum chemotherapy, while now, information about the molecular characteristics of her cancer, provided by a genetic test, can indicate a range of therapies focused on her particular tumour markers, with less side effects and better outcomes. However, genetics and -omics in general are not the sole important players. Consider a second patient consulting his oncologist after having been radically operated for a prostatic cancer and now experiencing a raise in PSA Prostate-Specific Antigen, a very sensitive blood test. Again, twenty years ago the options would have been

1 "Precision medicine" and "personalised medicine" are often used as synonyms; though with "precision medicine", the stress is on targeting a specific disease or malfunction with treatments and tests, rather than a larger category of similar diseases (i.e. triple negative breast cancer versus breast cancer, see Wu et al. 2018), whereas "personalised medicine" refers to the consideration of patient-specific factors in diagnosis and treatment (Desmond-Hellmann et al. 2011; National Institute of Health 2018). We acknowledge this distinction though it is not essential to the argument of this paper, (See https://ghr.nlm.nih.gov/ Precision Medicine; see also https://www.nih.gov/research-training/allofus-research-program).

hormonal blockade, or radiation therapy directed to the prostatic bed. Now, information about the site of recurrence of the disease, provided by an imaging test, can indicate several therapies depending on the extent of the disease that could be local, extended at the lymph nodes or spread to the bone. In both scenarios, therefore, diagnostic tests have a key role. Treatments are more directly associated with patients' health outcomes, but results of tests are necessary for the choice of such treatments, and therefore they are equally essential.

In this paper we address the epistemic credentials of advanced imaging tests. Sociologists and historians of science have pointed out that medical imaging in general has always carried with it a powerful illusion of immediate knowledge, namely, the idea that the body can be simply seen through and the diseases recognized by the doctor's impartial gaze (Joyce 2008, van Dijck 2011). As technology progresses, this conviction reinforces, permeates popular culture, and may result in inappropriate requests and overutilization of imaging tests by patients and practitioners, as we will illustrate in section 3 below (for overuse see Hendee et al 2010, Hofmann 2010). Obviously, however, advanced imaging tests are neither infallible nor immediate – they have limitations and boundaries of application. Some of these limitations and boundaries of application - such as sensitivity, specificity, and predictive positive and negative values - are clinical and scientific issues, while some others come from conceptual aspects of tests as knowledge procedures, and can be discussed by philosophers. Examples of conceptual matters related to (virtually any) test are: does the test measure what it is intended to measure? (is the operationalization proper, as Neopositivists would say?) What kind of knowledge is involved in the test (practical vs propositional, perceptual vs conceptual, direct vs indirect etc.)? Are test results true of what they measure, and in which sense of "true"? Is the test reliable, and in which sense of "reliable"? There is some interest in these themes in the philosophy of medicine now (Stegenga et al. 2016, pp. 353-354), even if dedicated studies are still sparse (Delehanty 2010, Lalumera and Fanti 2017)[2]. The general conviction behind this work is that understanding the conceptual characteristics of imaging tests as knowledge procedures may shed light on their appropriate use in medical practice.

In what follows, we will focus on imaging diagnostics, and take molecular imaging tests in oncology, specifically FDG-PET[3], as a paradigmatic case study. Our goal is to assess their epistemic appropriateness, or in other words, their *reliability as knowledge procedures* – a notion that we will specify in section 2. Imaging is particularly interesting in this respect, as we will illustrate in section 3, because of the high expectation of immediacy it carries – patients tend to believe that with a PET (Positron Emission Tomography) or a CT scan (Computed Tomography scan), doctors "just see" the disease, as from a transparent magnifying lens. In fact, we will remind in section 3 that though the data output is a naturalistic image, molecular diagnostic

2 A notable exception is Megan Delehanty's PhD's dissertation on the epistemic credentials, and especially on the reliability, of PET images in clinical oncology. Delehanty focuses on the question of whether and how PET as a data-generating process produces reliable knowledge, while we enlarge the picture and consider the reliability of the technology together with the way it is usually employed within the medical community. In broadening our scope with such socio-epistemological question, we think our work completes Delehanty's excellent analysis.

3 As we will see in a while, FDG stands for fludeoxyglucose, that is, the usual radiotracer used for PET neuroimaging and cancer patient management.

images are far from being immediate perceptions of inner bodily features[4]. In addition to this intrinsic characteristic, we will show that there are two further epistemic aspects of molecular imaging tests that may impact their status as knowledge procedures. The first concerns the *reference-class dependence*, and the second the *observer-dependence*. In section 4, we will argue that reference-class dependence affects the way in which imaging tests are evaluated as accurate. In section 5, we will discuss observer-dependence and argue that a certain amount of observer-dependent variability is ineliminable, even when standardised technological equipment is guaranteed, because images are inherently *asemantic*, in the sense that they require interpretation in order to be meaningful. Differently said and using a semiotic jargon, the image produced by a PET is the sign which stands for something else (in this case the anatomo-physiology of the patient) but which needs a subject (in this case the clinician) to be interpreted. We will also consider approaches, such as standard-fixing consensus conferences, which try to govern (but which do not eliminate) observer-dependence. Section 6 contains a summary and our concluding remarks about reliability rethought.

Before starting our discussion, we specify that we chose PET as a case study for two reasons. First, as it is mostly recommended at the moment for staging cancers and monitoring treatments, it qualifies as especially suitable for precision medicine. Second, PET is in many cases the newest diagnostic tool, and this means the most requested by patients and family doctors, sometimes independently of a proper evaluation of its appropriateness in the specific case, as we discuss below, from section 3. In the standard procedure for many cases, the diagnosis is obtained after a cross observation of both a CT morphological scan and a PET scan. It might seem that there is an intrinsic difference between morphological imaging, such as CT, ultrasound and MRI, and functional imaging, such as SPECT, PET, and FMRI. CT uses a derived X-ray technology and CT uses a derived X-ray technology and provides images of tissue density, while PET employs radiotracers and delivers 3-dimensional metabolic images. However, in the context of our paper, the difference is not relevant, as both methods derive medical conclusions from a processed image via an interpretative theory. For example, in CT the likely malignant nature of a lesion is based on several features, including lesion density, aspects, and enhancement after contrast media. In FDG PET it is based on the degree if tracer uptake. Whether the data are about function or about morphology, they have to be interpreted in order to derive significant conclusions. Thus, the technological difference between these techniques does not bring with it a different notion or evaluation of reliability[5].

2. Reliability of diagnostic tests as knowledge procedures

Usually, "reliability" has at least three connected, but different meanings: one concerns epistemology in general, one the philosophy of the empirical sciences, and one the

4 The non-immediacy of medical imaging in the philosophy of neuroscience has been studied extensively, see Bogen (2001).

5 For an introduction to the differences between PET and CT, see for example RSNA 2019. For a philosophical illustration of CT and "seeing styles", see Friedrich (2010). We thank one of the reviewers for pressing us on this point.

methodology of the empirical sciences. The aim of this paragraph is to show that these three meanings complement each other in characterising *reliability of clinical tests,* in particular *reliability of molecular imaging tests as knowledge procedures*.

The epistemological meaning of reliability strongly intersects the debate about the nature of knowledge. As known, its core question is: What characteristics should have a belief in order to be qualified as knowledge? Here, the standard reliabilist idea is that a belief is knowledge whenever it is justified by means of a reliable process, namely, one that is truth-conducive or truth-tracking, in the sense that most of the beliefs it produces are true, either in terms of frequency, or in terms of propensity (see, for example, Goldman 1979 and Leplin 2007). Reliable processes are, for example, human perception in normal conditions, introspection and face recognition, whereas wishful thinking, clairvoyance and guessing are not. This idea of knowledge, whose birth can be dated back to Plato, has been criticised (mainly) for not being able to provide a satisfying answer to the sceptical challenge: a reliable process such as perception, for example, would fail to provide knowledge in a Matrix scenario (Goldman and Beddor 2016). Though this criticism is central to the epistemological debate, we will bracket it here. On the other hand, our aim is not to challenge the tenability of that concept of knowledge, as the sceptic does, but to clarify the notion of reliability. From this point of view, we could retain reliability in terms of truth-conduciveness or, if we prefer to weaken the claim, in terms of validity-conduciveness. Thus, according to this first epistemological meaning, a molecular imaging test is reliable if it is validity-conducive.

In the second sense of "reliability", what is reliable or not are the *data* obtained by a certain procedure. This debate belongs to the philosophy of experimental sciences, where Woodward (2000) and Bogen (2002) introduced the distinction between data and phenomena, and within this framework, they talked of "reliability of data" (Woodward 2000). In Bogen's words, data are "records of things which investigators perceive, or which register on their equipment" and "the crucial epistemic questions of empirical epistemology have to do with how conclusions about effects are supported by reasoning from data". Reliability of data is then characterised as how precise, accurate and epistemically appropriate these are, "with respect to the features upon which the investigator's reasoning to conclusions about the effect of interest depends" (Bogen 2002, p. 135). So, for example, if our effect of interest is the presence or absence of prostate cancer in the lymph nodes of a patient, imaging data from a PET scan will count as reliable in this sense if it is clear enough to provide sufficient information to settle this issue. More specifically, in her work on PET imaging, Delehanty proposes that reliability of data in general depends both on "preservation of the structure or features of the object and a match between the granularity of the world at which a particular question is directed and the granularity of the instrument" (Delehanty 2005, p. 02). Thus, according to this second meaning, a molecular imaging test is reliable if it is precise, accurate and epistemically appropriate. Note that this aspect is strictly correlated with the capacity of diagnostic tests correctly individuating all and only the diseased subjects. That is, it is correlated with what is called *diagnostic accuracy*, and it is measured in terms of *sensitivity* and *specificity* of the test. Sensitivity is the ability of the test to identify correctly those who have the disease (it is also known as the true positive proportion). Specificity is the ability of the test to identify correctly those who do not have the disease (it is also known as the true negative proportion). In the assessment of a new diagnostic procedure, these values are individuated experimentally

– ideally via RCTs, more usually via non-randomised cohort studies – by a confrontation with the best test currently available, called the gold standard or reference standard (Peters et al. 2015).

Let us come now to the third meaning of "reliability", which is not philosophical, but belongs to the methodology of the empirical sciences. In this third sense, "reliability" means both *repeatability* and *reproducibility*. Usually a test or an experiment is repeatable if its outcomes can be found again (within a certain interval of error) by the same experimenter in the same lab by means of the same technique or instrument but in different times. On the other hand, a test or an experiment is reproducible if its outcomes can be found again (within a certain interval of error) by different experimenters in different labs by means of different techniques or instruments in different times[6]. Thus, a test or an experiment is reliable if it is repeatable and reproducible. In particular, a clinical test is reliable if it is repeatable and reproducible within certain boundaries due to, for example, the variation in conditions of patients (a glucose blood test after a meal may give a different result when applied to the same person in fasting condition). Notice that reliability as repeatability and reproducibility does not influence the accuracy of a test (neither vice versa). Many palm-readers may well agree that my life-line is short, and yet I will not worry. Nevertheless, reliability is essential for assessing the status of a test as a knowledge procedure. If a diagnostic test has a low reliability score, it is impossible to set up a multicentre clinical trial to assess its validity, and patients that are tested in different imaging sites cannot be managed adequately (Boellaard et al. 2015)[7].

Summing up, *the reliability of molecular imaging tests as knowledge procedures should be characterised by the fact that their outcomes (data and images) are validity-conducive, precise, accurate and epistemically appropriate, but also repeatable and reproducible.* Once arrived at this point, the real problems with the reliability, in particular the clinical reliability, of an outcome of a molecular imaging test should be faced.

3. PET and the illusion of immediate vision

After the above clarification of the concept of reliability, this section summarizes the basic functioning of molecular imaging diagnostics and starts assessing its epistemic status as a knowledge procedure.

PET is based on revelation of photons emitted by radioactive isotopes bound to molecules, called radiotracers. The general underlying theory, beyond physics, is that different cells or cellular activities have a different uptake of radiotracers, so data about the uptake of radiotracers can be evidenced for the presence of a certain kind of cells or activities. PET has important clinical applications in neurology and cardiology, but at

6 See the position and the terminology suggested by the National Institute of Standards and Technology (https://www.nist.gov/pml/nist-technical-note-1297)

7 We believe that the distinction between reliability of process (first sense, epistemology), reliability of data (second sense, philosophy of science) and reliability as repeatability (third sense, scientific methodology) can be useful in the paper because different readers can be more familiar with one or the other of the three senses. We thank one of the reviewers for pressing us on this point.

least 90% of the clinical workload is in oncology. The most widely used radiotracer in oncology is 18F-FDG, which is essentially radioactive glucose. For oncological diagnosis, the radiotracer is injected into the patient's body, and its metabolisation enables to identify regions of hyperglycolysis, which is associated with various malignant tumours, according to Warburg's theory of cancer (Warburg 1956). More specifically, as the radiotracer decays, it emits photons. When the patient enters the scanner, special cameras detect photon collision events, and a computer converts the occurrence of such events into numbers, and then into pixels of an image. A higher concentration of pixels indicates "spots" of "foci" where a larger quantity of radiotracer has accumulated, that is, where there is a higher level of metabolic activity of glucose, as in tumour cells, or in tissues that physiologically tend to accumulate it (brain, heart, kidneys and urinary tract). The image is then interpreted by a specialist (nuclear physician or radiologist), who classifies the spots as normal, anomalous or pathologic, and when pathological, specifies its localisation and stage (Alavi and Reivich 2002; Waterstram-Rich and Gilmore 2016).

The special role of PET in oncological diagnosis is accountable to the fact that while conventional radiological methods (such as CT) are based on identification of morphological features of organs, PET allows to identify functional characteristics, as based on the properties of the radiotracers. 18F-FDG PET is thus capable of revealing the status of disease process, providing unique information on tumour staging and response to therapy. Knowing exactly at what stage the disease has arrived at is an essential precondition for treatment of patients, as it helps with critical decisions about interventions. For example, staging of lung cancer to establish the surgical resectability is based on CT for the local extent of the tumour (T staging), but on 18F-FDG PET, it is the lymph nodal involvement (N staging). The molecular imaging test brings better outcomes for the patients, in terms of avoiding futile surgery, specifically thoracotomies (Fischer et al. 2009; Hicks et al. 2001). In personalised medicine[8], FDG-PET allows earlier determination of the effectiveness of standard treatments in individual patients and, if necessary, enables the patient to start an alternative treatment earlier. It may also facilitate evaluation of the effectiveness of experimental treatments, thereby speeding their entry into clinical practice (National Research Council 2007).

Molecular imaging diagnostic tests are increasingly requested by family doctors, oncologists and by patients themselves, and there is a rising trend of utilisation in many countries (for US data see Mitchell 2008; for EU see Eurostat 2017). There is also a rising trend of overutilisation, as sometimes a PET exam is requested even if it is not appropriate, namely, not recommended by the evidence-based guidelines for a certain pathology[9]. For example, nuclear imaging tests are often requested by patients for post-treatment surveillance of asymptomatic prostate cancer and breast cancer, though there is no evidence that they provide benefits, according to guidelines (Boellaard et al. 2015). Overutilisation of diagnostic imaging is influenced by diverse socio-economic, structural and psychological factors, such as attraction for the newest

8 Here we use personalised medicine since we are in the situation indicated as such by National Research Council in footnote 2.

9 Alongside campaigns promoted by scientific societies and institutional and private healthcare providers, there is a philosophical debate on overutilisation and medical futility, addressing both the definition of the phenomenon and ethical consequences. See Hofmann (2010).

technology, availability of facilities, patients' independent search of information in the web, advertising campaigns from hospitals and private medical centres and physicians' self-referral (Lysdahl and Hofmann 2009; Kilani et al. 2011).

Among the factors behind the tendency to overutilize, one is especially relevant to our assessment of nuclear imaging tests as knowledge procedures, namely, the conviction that with a PET scan[10], the body is just *seen through*, not simply tested (as in a blood test, for example). The illusion seems to resist evidence. As an example of experimental studies on the impact of the illusion of immediate vision, Yasunaga and colleagues (2006, 2008) found that providing patients with information about test accuracy was shown to have no impact on willingness to pay for prostate cancer screening or whole-body PET scans for cancer. Studies show that incomplete knowledge of the techniques, appropriate utilization and limitations is partially shared by primary care pratictioners (Han et al. 2014).

The phenomenal character of the diagnostic image, which is photography-like, sustains the illusion. It is part of the everyday conception of seeing it as an immediate process, namely, one not mediated by theories or background knowledge, that is, not "theory-laden". As we know from the history of the contemporary philosophy of science (Losee 2001; Oldroyd 1986), this conception was present in many positivist and neo-positivist's epistemological accounts, where seeing – when compared to interpreting, decoding and inferring – was taken to be the paradigm of reliability in the sense of validity-conduciveness. Note that this approach went almost in parallel with the idea that what we can observe with our eyes – the observable entities – have a special role with respect to hypotheses about phenomena. To briefly complete the framework, it is worth recalling that medical images have been considered as the new canon of seeing and objectivity (see Daston and Galison 1992). Unfortunately, on the one hand, a PET image is not a photography and on the other hand, neither scanning nor reading images are unmediated processes, as we know from what was called the "new philosophy of science" of the seventies (see Brown 1979).

A quick look at the description of the process of image production provided above is sufficient to dispel the illusion of immediacy. First, there is a *theory*, albeit a well-confirmed one, mediating the inference from the quantity of radiotracer accumulated in a certain kind of cell – due to its glycolytic activity – to the conclusion that it is a malignant cell. Knowledge of the theory is not a precondition for extracting information from a PET image (just like knowledge of the laws of optics is no precondition for observing planets with a telescope), but it grounds the reliability of the whole procedure. Second, there are algorithms converting positron emission data to an image format, which involve various normalisations and corrections. Leaving normalisations and corrections aside, as Delehanty (2010) persuasively explained, the image format *per se* is essentially a pragmatic choice in the presentation of such data: "the final conversion of this data into the form of a vaguely naturalistic image is simply a matter of assigning a color (or gray level) to particular ranges of numerical values and then displaying the data in a 2-D or 3-D array. It could just as easily be represented in

10 The psychological allure of images is of course not unique to molecular imaging tests, of course, but is common to all medical imaging diagnostic tests (such as CT for example, see footnote 4 above). What is specific to molecular imaging, we believe, is that immediacy of images is additionally difficult to defend.

other ways" (Delehanty 2010, p. 161). For instance, the variation in FDG uptake within some defined tissue could, in principle, be displayed in graphical format, or as a list of numbers, though it would be much less cognitively accessible for a human reader in such formats[11]. Third, there is the interpretation stage, in which (in positive cases) an image with black or more intensely coloured areas is *seen as* the image of a tumour. Seeing-as is an interpretive and theory-laden process as argued, for example, by many contemporary philosophers of science such as Hanson (1958, 2001) and Kuhn (1990)[12], and it involves background knowledge and/or implicit instructions[13]. We will return on the interpretation or reading of PET images in section 5, when discussing observer dependence.

In light of the above considerations, we have to conclude that molecular imaging is not just like "seeing" in a trivial way, that is, not just like the conception of seeing that involves immediacy and absence of interpretation, and which is not theory-laden. This of course neither makes PET scans unreliable as knowledge procedures for detecting or staging tumors, nor diminishes their reliability – but it suffices for concluding that the illusion of immediate vision that they carry with, is misplaced. This is something worth noting both for a philosophical analysis of their epistemic status, and for its practical applications. Given that the illusion of immediate vision is among the factors that cause overutilization and overprescription, then dispelling the illusion with some amount of correct information to the non-experts (patients, families, and to a certain extent family doctors) may help promoting a more fair and appropriate utilization of such technologies.

4. Reference-class dependence of test accuracy

Philosophers of medicine have pointed out recently that the choice of reference classes has a key role in many areas of medicine and medical research and that values, preferences, and chance ineliminably affect such choice. To put it simply: Whether drug A is better than drug B, or whether condition C is to be considered a disease of not, crucially depends on the groups of subjects or patients that were selected for the study (Kingma 2007, Ashcroft 2004, Stegenga 2018). Here we illustrate how this general point can be particularized to the case of the evaluation of PET. In this respect, our discussion here confirms a recent trend of attention in the philosophy of medicine, whereas our qualitative research in the relevant medical literature suggests that in the imaging field the problem is often overlooked.

Consider an oncological patient, Ms P, who is told by her oncologist Dr O that

---

11 Nonetheless, medical images as numbers can be more easily read by software, and this is the basis of Radiomics. See Gillies et al. (2015).

12 As known, under this tradition in the philosophy of science, there was the hidden figure of Kant and of a form of neo- or post-Kantism, see Boniolo (2007).

13 The competence and expertise of an expert reader of medical images are a research field in itself, and it is especially debated now that artificial readers become increasingly available. See Krupinski (2010), Samei and Krupinski (2010) and Shiraishi et al. (2011).

she will undergo a PET scan, as it is the most accurate diagnostic technique that can be used to determine whether her lung cancer has spread to the lymph nodes or not. In an ideal evidence-based medicine scenario, Dr O is following the guidelines for the diagnosis of lung cancer issued by an international scientific association, which are written by panels of experts on the basis of various kinds of evidence, among which meta-analyses of diagnostic test assessment studies have a top role. Let us now zoom in and consider one of these meta-analyses. Gould et al. (2003) compared CT and FDG-PET for the mediastinal staging of non-small cell lung cancer, and concluded that the latter was more accurate, as the FDG-PET median sensitivity and specificity were 81 and 90%, respectively, while for CT, they were 59 and 79%, respectively. This average result is sufficient for evaluating PET as preferable to CT in that clinical context, and to go back to our imagined case, to answer Ms P's clinical question.

However, if we look carefully into the meta-analysis, we notice that the data on sensitivity and specificity show a relevant variability: PET sensitivity has an interquartile range of 67 to 91%[14]. Why is that? The reason is that patients enrolled for the different studies considered in the meta-analysis (39 non-randomised cohort studies) belonged to different reference classes. In particular, the prevalence of malignant lymph nodes ranged from 5 to 64% (with a median prevalence of 32%). As we know from Bayes theorem, prevalence strongly influences the resulting sensitivity and accuracy values, for it alters the pre-test probability of finding malignant cells. By the way, this is exactly the reason why, from a clinical point of view, the positive and negative predictive values are more important than the simple values concerning sensitivity and specificity.

Is such variability eliminable? Not really, as a meta-analysis necessarily involves many different studies run in different hospitals or research centres. Whereas sensitivity and specificity of a tests are objective values, the choice of the reference class, that is, the choice of sub-population at stake and thus the prevalence, is never a completely objective matter. Once minimal methodological criteria for a study design are met, considerations of availability of patients in one's facility, time and experimenter's own guiding hypotheses play a role.

It is also worth noting that, from the same meta-analysis, we learn that the specificity of PET in staging lung cancer seems to decline as time passes (Gould et al. 2003, p. 85, fig. 3). From 1994 to 1998, specificity value was 100% in 4 out of 11 studies, whereas from 1999 to 2003, only 2 studies out of 22 reported a specificity value of 100%. This may strike as surprising, for we tend to suppose that a diagnostic test performs better as its execution procedures becomes more familiar to technicians and physicians. In fact, this is again due to the selection of reference classes of patients in the studies. Initially, it is likely that the new test is assessed in patients that are already treated for the pathology one is testing for (in this case, lung cancer with mediastinal metastases), and provided in addition to, rather than in alternative to, the standard test. Such patients are likely to test positive, and for the right reason. Later, the imaging test is employed in less clinically homogeneous populations, and accuracy values are likely

14To put it very simply, the interquartile range of a data set is where is a measure of where the bulk of the values lie.

to decrease.

Reference-class dependence and resulting variability of accuracy values are not a weird result of the meta-analysis we focused on. Rather, the fact is widely acknowledged in the literature, and considered a problem for the evaluation of new imaging tests (see, for example, Hunik and Krestin 2002). It could be circumvented by substituting randomised trials to cohort studies – where, for example, a patient may be assigned either to the CT group or to the PET group, and voluntarily renounces a choice. Such studies for diagnostic tests, however, raise both practical and ethical problems. They are time-consuming in a context where technological progress is very fast, and potentially clash with the ethical principle of non-malevolence, as usually there is already some evidence that the new competitor may be better, although of lower quality according to the EBM ranking of evidence, (Jarvik 2002; Lalumera and Fanti 2017).

Where does this leave us with respect to Ms P's case? It is still true that Dr O is right in scheduling a PET scan for her, as the molecular imaging test is on average more accurate. However, its assessment has flaws, which are difficult to eliminate, for they are inherent to the very procedure of assessment currently performed within the medical community. Personalised medicine rests on tests that are evaluated for the average patient, which is evidently different from our Ms P.

Again, this does not make molecular imaging tests unreliable as knowledge procedures, but it signals that we have to take into consideration their being strongly dependent on the population we are considering, and on the prevalence in that population of the pathology under consideration. This is a general problem for all evidence-based medicine and research, but as we noticed in the opening of this session, it is rarely if ever addressed in the field of advanced imaging.

5. Observer-dependence

We noted above that image reading is not naïve seeing, but rather it is a form of seeing-as, in the sense introduced by Hanson and Kuhn, for it is theory-laden. Moreover, we are adding here that the seeing-as of diagnostic images is observer-dependent, namely, different readers may (and frequently do) issue different diagnostic judgements (say, positive or negative) from the perception of the same image. In other words, the inter-reader disagreement is quite frequent. This is a well-known fact in radiology and nuclear medicine, and various strategies are being devised to reach higher concordance rates.

What are the determinants of such disagreement? One is the variability of expertise of readers. Studies show that image readers' performance of accuracy show a learning curve, to the effect that experts are better than novices, as their error rate is lower (van Westreenen et al 2003). In the case of FDG-PET, for a correct, interpretation, it is important to be aware of benign variants that demonstrate high glycolytic activity, and pathologic lesions which may not be FDG-avid, as well as to understand the physiologic and biochemical basis of these findings (Hofman and Hicks 2016). As some of this knowledge usually comes with experience, it is not surprising that novices tend to commit more errors.When disagreement comes from errors, it can be avoided, to some extent, by improving medical training and communication between centres and hospitals, favouring knowledge of standard reading conditions (e.g. lighting, monitors) specified in the guidelines (Taylor 2007).

There is however another determinant of inter-reader disagreement about the interpretation of PET images, which is more philosophically interesting: it is the fact that criteria of interpretation, in some cases, are not settled yet, especially when a certain technique or application is completely new. In this case, even expert readers disagree, because they disagree on the criteria, or apply different ones. When criteria are to be settled, the epistemic goal is consensus on the semantics of the image. A specific example can illustrate the point. FDG-PET can be usefully employed to evaluate early response to therapy in Hodgkin's lymphoma. In 2009, however, there were no shared interpretation criteria, and several international meetings of experts (consensus conferences) were organised in order to settle the issue. However, in a Delphi study[15] where the explicit criteria proposed in the meetings were provided, agreement between pairs of expert reviewers never reached perfection, but only good results (from 0.69 to 0.84 as measured with the Cohen kappa, 0.76 as measured with the Krippendorf alfa). One difficult issue, among others, is the definition of "PET good responders" versus "non-good responders", which plays a role in image interpretation. The degree of reduction in tracer uptake is used to categorize patients as "PET good responders to therapy" versus "PET non-responders"; however, this is just the subjective evaluation of a tone of dark grey to a tone of light grey, and this is the core of image interpretation. It is clear that this is influenced not only by the operator's experience and skill, but also by their awareness of clinical data (the eye will tend to search for the known likelihood of response to therapy, thus a modest change of grey shade is frequently reported as "good response" in patients with elevate probability of response to therapy), by personal attitude (some readers are keen to be optimist, thus reporting as much cases as possible in the group of "good responders"), and by perception of referring physician expectations (cases referred for suspected poor response are more likely to be accordingly reported) A controversial case is reproduced in FIG. 1 taken from a Delphi study (Biggi et al. 2013, fig. 4).

15 The use of consensus conferences and Delphi procedures is widespread in the social and life sciences, whenever evidence underdetermines the answer to a given scientific or policy question, and experts disagree. For example, they are often employed in psychiatry, in order to decide whether a certain condition is to be considered a disease or not, and they were used in astronomy to assess the status of Pluto as a planet. In philosophy, their epistemic pedigree has been analysed by philosophers like Miriam Solomon and Jakob Stegenga. In advanced diagnostic imaging, they are often utilized in order to write and publish guidelines for the appropriate use of tests (see eg. In philosophy of science, Miriam Solomon has studied the role of consensus conferences and Delphi studies in the making of medical knowledge (Solomon 2007, 2015).

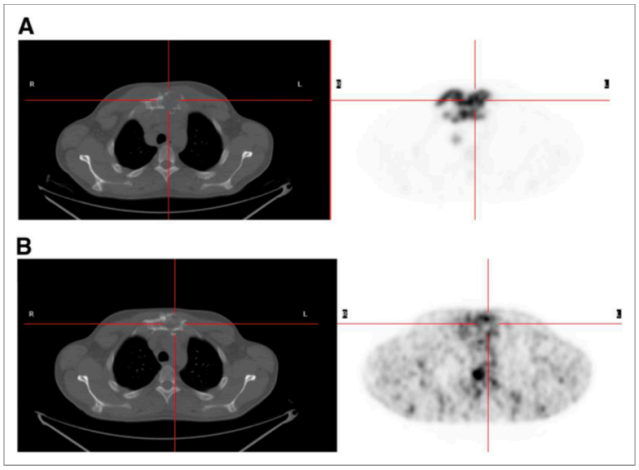

FIGURE 4. Patient 168, discordant case: baseline PET/CT (A); interim PET/CT (B). Increased uptake seen in sternum at site of pathologic fracture could be due to healing of fracture. Final consensus was score 3 residual lesion. Patient was alive in complete remission after 35 mo.

The situation gets clearly worse when local readers – not world-leading experts – are reporting, leading to major clinical implications. A very recent study shows that time apparently cannot modify such limitations, as after almost 10 years of use of the mentioned international interpretation criteria, the positive agreement among readers remains around 75%, posing clinical problems in daily practice (Burggraaff et al. 2018).

The example shows two aspects. First, that molecular diagnostic images are inherently asemantic: like signs of an uninterpreted language, they need to be conventionally associated to meanings. Neither dark spots on a screen nor a number that indicates the uptake value of the tracer is meaningful yet. The image-like character of such diagnostic tests tends to cloud this aspect. Analogously, the level of mercury in a thermometer is asemantic unless a correlation between its values and bodily temperature is fixed, and more importantly, the threshold for fever versus non-fever is set (at least approximately). The difference with the thermometer case is that the correlation between temperature values and bodily temperature (and its implication) is very widely agreed on and shared, whereas the semantics of advanced imaging is largely work-in-progress, and not always uncontroversial. This is the second aspect that the example shows.

To go back to our initial question: does observer-dependence make nuclear imaging tests unreliable? What should Ms P do, then, if she has just been treated for Hodgkin's lymphoma? Should she trust Dr O who tells her that the PET shows everything is fine now, or should she better not? As in the other two situations considered in the sections above, even this reader-dependence does not make PET unreliable. Nevertheless, we have to take into consideration that the clinical decision associated with the interpretation of the photography-like outcome of a PET is a difficult issue that needs great expertise on the part of the clinician. Nevertheless, from a patient's point of view, it would be better that this interpretation was not in the hand of a single clinician, whatever his/her expertise were, but in the hand of a team of expert clinicians which could arrive at a conclusion after a collective debate. We know that this is virtually impossible, taking into consideration the cost of such a collective decision, but probably this is the sole practicable way at least from an epistemological perspective.

6. Rethinking reliability of molecular imaging tests

In the paper we discussed the reliability of molecular imaging tests in oncology as

knowledge procedures, in light of their increasing importance, especially in precision and personalised approaches. After having briefly recalled what could be reliability as knowledge procedure in terms of validity-conduciveness, accuracy, repeatability and reproducibility, we, by discussing the PET case, have highlighted three different epistemic issues, which render the matter more complex. First, in spite of the illusion of immediacy they bring about, PET images are highly theory-laden and mediated artefacts, and their reliability is grounded on different levels of theory and computation. Second, the normal procedure employed to evaluate accuracy of PET diagnostic tests implies a reference-class dependence: sensitivity and specificity values are not enough since we have to take into consideration the population under analysis and the prevalence of the pathology we are interested in. Third, we illustrated the observer-dependence of diagnosis, which in the case of nuclear imaging, is not just due to the different distribution of error frequency among physicians, but also, at least in some cases, to the need of a semantics, that is, criteria of interpretation that maps what is seen in a clinical condition.

The three epistemic issue mentioned do not question the reliability of a molecular imaging test, but render the question more complex than it might appear at first sight, and contribute to dispel the illusion of immediacy of vision that these procedures might produce. The problem, therefore, is how to enhance their reliability, transforming them into something which is validity-conducive, precise and accurate, repeatable and reproducible.

In light of the considerations above, we suggest three complementary directions for action by the medical and healthcare community.

a) Artificial intelligence. Whenever interpretation is involved, objective criteria are needed. In advanced medical imaging there are basically two ways to improve the objectivity of interpretation criteria. The first is to enhance the use of AI as an aid to diagnosis. Machine-aided diagnosis, where the machine learns how to detect basic cases of lesions, may help alleviate the burden of routine exams to specialists. This would not depersonalize or dehumanize the diagnostic process - as the literature shows, the tendency now seems to favour mixed approaches, where the automation intervenes in just one step of a process involving human doctors (Gandhi et al 2018, Hosni et al. 2018).
b) One conceptual limit of AI, however, is that it cannot fix interpretive standards, but only apply them. In order to fix the standards of interpretation for nuclear imaging – what counts as a pathological finding, given a particular diagnostic question and radiotracer – the consensus of experts is needed. In section 5 above we illustrated the case of a Delphi study aimed at reaching consensus on the interpretation of FDG-PET scans for Hodgkin lymphoma (Biggi et al. 2013). Though in that specific case the consensus was hardly reached, due to further terminological disagreements, we suggest that the use of similar epistemic procedures be enhanced in imaging, whenever the semantic problem of interpretation opens up. This, in turn requires that the scientific community acknowledges that interpretation standards are indeed a problem, and we hope that our work in this article could help in this direction.
c) A parallel course of action that may help promote a better utilization of advanced medical imaging is patient communication. In a scenario of personalized medicine and active involvement of patients in the decisions about

their health management, it is important that s/he is aware of the possibilities and limitations of each step of the process of care, including diagnostic and screening tests (see, i.e. Ferretti, Linkeviciute, Boniolo 2017. This is especially crucial given the fact that advanced medical imaging diagnostic tests are very expensive, and therefore involve considerations of fairness and life values. Approaches that take into consideration the patient's "personal philosophy" may help in such situations (Boniolo and Sanchini 2016). Indeed, the imaging community is in need of directions for handling communication end ethical issues involved in the relationship with patients (Gonzalez et al. 2018).

As a general concluding remark, precision medicine and advanced diagnostic techniques clearly hold tremendous promise. Practically, they will save more lives, and ameliorate the quality of even more. Methodologically, they will bridge the gap between objective, but standardised evidence-based medicine, and individual patient's care. Nevertheless, a careful assessment of what current practices and instruments can and cannot do is functional to the attainment of the final goal. Knowing one's limits is knowing one's power.

References

Alavi, A., & Reivich, M. (2002). Guest editorial: the conception of FDG-PET imaging. *Seminars in Nuclear Medicine*, *32*(1), 2–5.

Ashcroft, R. (2004) Current epistemological problems in evidence-based medicine. *Journal of Medical Ethics* 30, 131-135.

Boellaard, R.,et al. (2015). FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *European Journal of Nuclear Medicine and Molecular Imaging*, *42*(2), 328–354.

Bogen, J. (2001). Functional imaging evidence: Some epistemic hot spots. In P. K. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 173–199). Pittsburgh: University of Pittsburgh Press.

Bogen, J. (2008). Experiment and observation. In P. Machamer & Silberstein, M. (Eds.), *The Blackwell guide to the philosophy of science* (Vol. 19, pp. 128–148). Oxford: Blackwell

Boniolo, G. (2007). *On scientific representation: From Kant to a new philosophy of science*. Houndmills: Palgrave Macmillan.

Boniolo, G., Sanchini, V. (eds) (2016), *Ethical counselling and medical decision-making in the era of personalized medicine*, Heidelberg: Springer.

Brown, J. (1979) Perception, Theory and Commitment: New Philosophy of Science, The University of Chicago Press

Burggraaff, C. N., Cornelisse, A. C., Hoekstra, O. S., Lugtenburg, P. J., De Keizer, B., Arens, A. I., et al. (2018). Interobserver agreement of interim and end-of-treatment

18F-FDG PET/CT in diffuse large B-cell lymphoma (DLBCL): Impact on clinical practice and trials. *Journal of Nuclear Medicine*, online first

Daston, L., & Galison, P. (1992). The image of objectivity. *Representations*, 40, 81–128.

Delehanty, M. (2010). Why images? *Medicine Studies*, *2*(3), 161–173.

Delehanty, M. C. (2005). *Empiricism and the epistemic status of imaging technologies*. Doctoral dissertation, University of Pittsburgh).

Desmond-Hellmann, S., Sawyers, C. L., Cox, D. R., Fraser-Liggett, C., Galli, S. J., Goldstein, D. B., et al. (2011). *Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease*. Washington, DC: National Academy of Sciences.

Eurostat (2017). https://ec.europa.eu/eurostat/statistics-explained/index.php?title=File:Use_of_imaging_equipment_—_number_of_PET_scans,_2010_and_2015_(per_100_000_inhabitants)_HLTH17.png. Accessed 18 October 2018.

Ferretti, G., Linkeviciute, A., Boniolo, G. (2017). Comprehending and Communicating Statistics in Breast Cancer Screening. Ethical Implications and Potential Solutions. In M. Gadebusch-Bondio, F. Spöring, J.-S. Gordon (eds), *Medical Ethics, Prediction and Prognosis: Interdipliplinary Perspectives*, New York: Routledge, pp. 30-41

Fischer, B., Lassen, U., Mortensen, J., Larsen, S., Loft, A., Bertelsen, A., et al. (2009). Preoperative staging of lung cancer with combined PET–CT. *New England Journal of Medicine*, *361*(1), 32–39.

Friedrich, K. (2010). 'Sehkollektiv': Sight Styles in Diagnostic Computed Tomography. *Medicine studies*, *2*(3), 185-195.

Gandhi, S., Mosleh, W., Shen, J., & Chow, C. M. (2018). Automation, machine learning, and artificial intelligence in echocardiography: A brave new world. *Echocardiography*, *35*(9), 1402-1418.

Gillies, R. J., Kinahan, P. E., & Hricak, H. (2015). Radiomics: Images are more than pictures, they are data. *Radiology*, *278*(2), 563–577.

Goldman, A. (1979). What is justified belief? In G. S. Pappas (Ed.), *Justification and knowledge* (pp. 1–25). Dordrecht: Reidel. Reprinted in A. I. Goldman (Ed.), *Reliabilism and contemporary epistemology* (pp. 29–49). New York: Oxford University Press, 2012.

Goldman, A. & Beddor, B. (2016). Reliabilist epistemology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/reliabilism/. Accessed October 20, 2018

Gould, M. K., Kuschner, W. G., Rydzak, C. E., Maclean, C. C., Demas, A. N., Shigemitsu, H., et al. (2003). Test performance of positron emission tomography and computed tomography for mediastinal staging in patients with non-small-cell lung cancer: A meta-analysis. *Annals of Internal Medicine*, *139*(11), 879–892.

Han, P. K., Klabunde, C. N., Noone, A. M., Earle, C. C., Ayanian, J. Z., Ganz, P. A., ... & Potosky, A. L. (2013). Physicians' beliefs about breast cancer surveillance testing are consistent with test overuse. *Medical care*, *51*(4), 315.

Hanson, N. R. (1958). Observation. In N. R. Hanson (Ed.), *Patterns of discovery: An inquiry into the conceptual foundations of science* (pp. 4–30). Cambridge: Cambridge University Press.

Hanson, N. R. (2001). Seeing and seeing as. In Y. Balashov & A. Rosenberg (Eds.), *Philosophy of science: Contemporary readings* (pp. 321–339). London: Routledge. Originally published in N. R. Hanson (Ed.), *Perception and discovery: An introduction to scientific inquiry* (pp. 91–110). San Francisco: Freeman, 1969.

Hendee, W. R., Becker, G. J., Borgstede, J. P., Bosma, J., Casarella, W. J., Erickson, B. A., ... & Wallner, P. E. (2010). Addressing overutilization in medical imaging. *Radiology*, *257*(1), 240-245.

Hicks, R. J., Kalff, V., MacManus, M. P., Ware, R. E., Hogg, A., McKenzie, A. F., et al. (2001). 18F-FDG PET provides high-impact and powerful prognostic stratification in staging newly diagnosed non-small cell lung cancer. *Journal of Nuclear Medicine*, *42*(11), 1596–1604.

Hofmann, B. (2010). Too much of a good thing is wonderful? A conceptual analysis of excessive examinations and diagnostic futility in diagnostic radiology. *Medicine, Health Care and Philosophy*, *13*(2), 139–148.

Hofman, M. S., & Hicks, R. J. (2016). How we read oncologic FDG PET/CT. *Cancer Imaging*, *16*(1), 35.

Hosny, A., Parmar, C., Coroller, T. P., Grossmann, P., Zeleznik, R., Kumar, A., ... & Aerts, H. J. (2018). Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLoS medicine*, *15*(11), e1002711.

Hunink, M. M., & Krestin, G. P. (2002). Study design for concurrent development, assessment, and implementation of new diagnostic imaging technology. *Radiology*, *222*(3), 604-614.

Jarvik, J. G. (2002). Study design for the new millennium: Changing how we perform research and practice medicine. *Radiology*, *222*(3), 593–594.

Joyce, K. A. (2008). *Magnetic appeal: MRI and the myth of transparency*. Cornell University Press.

Kilani, R. K., Paxton, B. E., Stinnett, S. S., Barnhart, H. X., Bindal, V., & Lungren, M. P. (2011). Self-referral in medical imaging: A meta-analysis of the

literature. *Journal of the American College of Radiology*, *8*(7), 469–476.

Kingma, E. (2007) What is it to be healthy? *Analysis*, 67, pp. 128-133.

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, *72*(5), 1205–1217.

Kuhn, T. S. (1990). The road since structure. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Vol. 1990, pp. 3-13). Chicago: Philosophy of Science Association.

Lalumera, E., & Fanti, S. (2017). Randomized controlled trials for diagnostic imaging: Conceptual and practical problems. *Topoi*, online first

Leplin, J. (2007). In defense of reliabilism. *Philosophical Studies*, *134*(1), 31–42.

Losee, A. (2001). *Historical introduction to the philosophy of science*. Oxford: Oxford University Press.

Lysdahl, K. B., & Hofmann, B. M. (2009). What causes increasing and unnecessary use of radiological investigations? A survey of radiologists' perceptions. *BMC Health Services Research*, *9*(1), 155.

Mitchell, J. M. (2008). Utilization trends for advanced imaging procedures: Evidence from individuals with private insurance coverage in California. *Medical Care*, *46*(5), 460-466

National Research Council (US) and Institute of Medicine (US) Committee on State of the Science of Nuclear Medicine (2007). *Advancing nuclear medicine through innovation*. Washington, DC: National Academic Press.

National Institute of Health (2018). What is precision medicine? https://ghr.nlm.nih.gov/primer/precisionmedicine/definition. Accessed 30 September 2018.

Oldroyd, D. (1986). *The aArch of knowledge: An introductory study of the history of the philosophy and methodology of science*. London: Routledge Kegan & Paul.

Parkkinen, V.-P. and Williamson, J. (2017). Extrapolating from model organisms in pharmacology. In Osimani, B., editor, Uncertainty in pharmacology: epistemology, methods, and decisions. Springer, Dordrecht.

Peters, M. D. J., Godfrey. C. M., McInerney, P., et al. (2015). *Methodology for JBI scoping reviews. The Joanna Briggs Institute reviewers' manual 2015*. Adelaide, South Australia: The Joanna Briggs Institute.

Samei, E., & Krupinski, E. (Eds.) (2010). *The handbook of medical image perception and techniques*. Cambridge: Cambridge University Press.

Shiraishi, J., Li, Q., Appelbaum, D., & Doi, K. (2011). Computer-aided diagnosis and artificial intelligence in clinical imaging. *Seminars in Nuclear Medicine*, *41*(6), 449–462.

Solomon, M. (2007). The social epistemology of NIH consensus conferences. In Solomon, M. (Ed.), *Establishing medical reality: Essays in the Metaphysics and Epistemology of Biomedical Science*, pp. 167–177, Dordrecht: Springer.

Solomon, M. (2015). *Making medical knowledge*. Oxford: Oxford University Press, USA.

Stegenga, J. Et al. (2016). New directions in philosophy of medicine. In *The Bloomsbury Companion to Contemporary Philosophy of Medicine*, 343-366.

Stegenga, J. (2018). *Medical nihilism*. Oxford University Press.

Taylor, P. M. (2007). A review of research into the development of radiologic expertise: Implications for computer-based training. *Academic radiology*, *14*(10), 1252–1263.

Van Dijck, J. (2011). *The transparent body: A cultural analysis of medical imaging*. University of Washington Press.

van Westreenen, H. L., Heeren, P. A., Jager, P. L., van Dullemen, H. M., Groen, H., & Plukker, J. T. M. (2003). Pitfalls of positive findings in staging esophageal cancer with F-18-fluorodeoxyglucose positron emission tomography. *Annals of surgical oncology*, 10(9), 1100-1105.

Warburg, O. (1956). On the origin of cancer cells. *Science*, 123(3191), 309-314.

Waterstram-Rich, K. M., &, D. (2016). *Nuclear medicine and PET/CT E-Book: Technology and techniques*. St. Louis, MO: Elsevier Health Sciences.

Woodward, J. (2000). Data, phenomena, and reliability. *Philosophy of Science*, 67,3 S163-S179.

Wu, N., Zhang, J., Zhao, J., Mu, K., Zhang, J., Jin, Z., et al. (2018). Precision medicine based on tumorigenic signaling pathways for triple-negative breast cancer. *Oncology Letters*, *16*(4), 4984–4996.

Yasunaga, H. (2008). Willingness to pay for mass screening for prostate cancer: a contingent valuation survey. *International Journal of Urology*, *15*(1), 102-105.

Yasunaga, H., Ide, H., Imamura, T., & Ohe, K. (2006). The measurement of willingness to pay for mass cancer screening with whole-body PET (positron emission tomography). *Annals of nuclear medicine*, *20*(7), 457-462.