Impact of tomographic reconstruction with Bayesian penalty in the quantification of PET/CT studies

Universidad de Navarra

Gabriel Reynés-Llompart

Supervisor: Dr. Josep M. Martí-Climent

Co-Supervisor: Dra. Cristina Gámez Cenzano

Faculty of Medicine

Universidad de Navarra

September 2019

Agradecimientos

Hay un gran número de personas a quien debería agradecer y seguro en este breve espacio me dejo a muchos por el camino, pero al igual que saben todos los que se han embarcado en similares caminos, este trabajo no hubiera sido posible sin la ayuda y soporte de muchos.

Quiero empezar por agradecer a mi supervisor y a mi co-supervisora, Josep M. Martí Climent y Cristina Gámez Cenzano, por darme esta oportunidad y guiarme en el camino. Al primero especialmente por todo el tiempo que me ha dedicado, pese a las distancias, y por todo el soporte y sabiduría aportados. A la segunda por haberme dado la oportunidad de poder trabajar en el Servicio de Medicina Nuclear de Bellvitge y todos los recursos que con ello me ha brindado.

Debo agradecer a todo el Servicio de Medicina Nuclear y la Unidad PET por toda la ayuda prestada siempre que han podido, y pese a toda la carga de trabajo que sostienen, por siempre encontrar un hueco para escucharme.

También cabe agradecer a Cristina Picón y al Servicio de Física Médica del Institut Català d'Oncologia con quien he trabajo estos últimos años, por ayudarme en lo material y mental de compaginar estos últimos meses un trabajo asistencial con la realización de la tesis, y por tener allí gente dispuesta de hablar de física en cualquier momento. A General Electric y en especial a David Álvarez y Antonio Alcántara, por creer en la importancia de tener a alguien investango con su nuevo equipamiento y ayudarme siempre con todo lo que necesitado de ellos. A Olivier Caselles del Institut Universitaire du Cancer de Toulouse, por haberme ensenñado todo lo que pudo y guiarme en el incio de este proyecto. A Elena Ponti

del Ospedale di San Marco en Monza por acogerme allí y discutir conmigo todas las ideas sobre el funcionamiento del Q.Clear. A Elena Prieto por haber corregido la tesis de forma tan detallista.

De entre todas las personas concretas, en especial debo agradecer a Aida por haberme ayudado en en cuanto ha tenido la oportunidad, haber aguantado mis bajos ánimos y haberme empujado hacia delante.

Para finalizar, un tipo distinto de agradecimiento debo darselo a mis padres, ya que por descontado, sin ellos nunca hubiera llegado aquí.

Table of contents

List of figures

List of tables

List of Abbreviations

Acronyms / Abbreviations

APD	**A**valanche **P**hoto**d**iode
BGO	**B**ismuth **G**ermanium **O**xide
BSREM	**B**lock **S**equential **R**egularized **E**xpectation **M**aximization
BV	**B**ackground **V**ariability
CNR	**C**ontrast-to-**N**oise **R**ratio
COV	**C**oefficient **O**f **V**ariance
CR	**C**ontrast **R**ecovery
CT	**C**omputer **T**omography
DI-Q-5R	**D**iscovery **IQ 5 R**ings
DI-Q	**D**iscovery **IQ**
DOI	**D**epth **O**f **I**nteraction
FBP	**F**iltered **B**ack**p**rojection
FDG	2-deoxy-2-**F**fluoro-**D**-**G**lucose
FOV	**F**ield **O**f **V**iew
FWHM	**F**ull **W**idth at **H**alf **M**aximum
GLCM	**G**ray-**L**evel **C**o-occurrence **M**atrix
GLRM	**G**ray-**L**evel **R**un Length **M**atrix
GLZM	**G**ray-**L**evel **Z**one **M**atrix
IQ	**I**mage **Q**uality

LBR	**L**esion-to-**B**ackground **R**atio
LOR	**L**ine **O**f **R**esponse
LSO	**C**erium-doped **L**utetium **O**xyorthosilicate
LYSO	**L**utetium-**Y**ttrium **O**xyorthosilicate
MAP	**M**aximum **A** **P**osteriori
MLEM	**M**aximum-**L**ikelihood **E**xpectation-**M**aximization
MTV	**M**etabolic **T**umor **V**olume
NEMA	**N**ational **E**lectrical **M**anufacturers **A**ssociation
OSEM	**O**rdered **S**ubset **E**xpectation **M**aximization
PET	**P**ositron **E**mitting **T**omography
PMT	**P**hoto**m**ultiplier **T**ubes
PSF	**P**oint **S**pread **F**unction
PVE	**P**artial **V**olume **E**ffect
ROI	**R**egion **O**f **I**nterest
SiPM	**Si**licon **P**hoto**m**ultipliers
SNR	**S**ignal-to-**N**oise **R**atio
SUL	**SU**V normalized for **L**BM
SUV	**S**tandardized **U**ptake **V**alue
TI	**T**extural **I**indices
TOF	**T**ime **O**f **F**light
VOI	**V**olume **O**f **I**nterest
VPHD	**V**UE **P**oint **H**D

Chapter 1

Introduction

Positron emission tomography (PET) is a medical diagnostic technique that provides functional and biochemical information. Throughout the last few decades, PET imaging has experienced tremendous advancements [1–3]. It started as a research tool mainly focused on neurological applications and now it is an integral and indispensable part in the daily practice medicine. The primary application of PET imaging is in clinical oncology, specially on the staging and restaging of tumours; however, as a quantitative tool it has a wide range of applications, such as the evaluation of treatment response or the definition of radiotherapy volumes [4].

The number of PET exams performed has not stopped growing and it is estimated to keep expanding with the introduction of new radiopharmaceuticals and the establishment of new diagnostic indications. This situation requires a thorough optimization of resources, such as reducing the occupation of the scanner or the total used radiopharmaceutical activity, which are factors that directly increase with the number of patients and affect the economical profitability of a PET unit. However, to be competitive with other emerging diagnostic techniques this optimization must be linked to an increase in image quality and quantitative value of PET studies.

Additionally, it is also necessary for all tests using ionizing radiation to be performed with the lowest possible radiation dose to the patient, compatible with the necessary diagnostic quality [5, 6]. The radiation dose received during a PET study can be assumed as directly proportional to the administered activity [7, 8]. Likewise, image quality of a PET study is directly related to the number of events recorded in the acquisition. Hence both parameters (image quality and radiation dose) depend on the activity administered to the patient. Consequently, the parameters of image quality and radiation dose are linked by the acquisition time and the administered activity of a PET study.

Since its introduction, PET scanners have undergone a continuous development to improve various aspects of imaging, ranging from fundamental changes in its design to new reconstruction and data processing methods [9]. From the design point of view, in the present generation of PET scanners, most manufacturers have chosen to include high temporal resolution detectors (LSO or LYSO) to increase the time-of-flight (TOF) correction capability [10–12]. However, GE Healthcare has chosen to maintain a series of PET scanners with BGO crystals, the Discovery IQ (D-IQ) series. A BGO scintillator, presents a lower temporal resolution but with greater sensitivity and a reduced price. Besides, the pricing allows an increase of tomograph sensitivity by expanding the axial field of view with an additional ring of detector blocks. The D-IQ tomograph in its configuration of 5 rings (D-IQ-5), and 26 cm of axial field, has a sensitivity of approximately 2.5 times superior to the 3-ring equipment, with 15.5 cm of axial field, measured according to the National Electrical Manufacturers Association (NEMA) standards. This gain in sensitivity can be translated into a reduction of the acquisition time while maintaining the number of detected photons.

On the other hand, the introduction of iterative reconstruction algorithms in PET imaging compared to filtered backprojection reconstructions (FBP), resulted in a substantial improvement in image quality and quantitative accuracy by successively incorporate modeling for the scattering corrections, random and, more recently, the point-spread function (PSF) [13].

The Ordered Subset Expectation Maximization (OSEM) algorithm is the most commonly used iterative reconstruction method, although it has some limitations, such as the limit to reach full quantitative convergence due to an increase in noise as the effective iterations are increased.

In order to improve the OSEM performance (i.e. preserving image quality and improving the PET quantitative value), GE Healthcare introduced commercially the Q.Clear algorithm, which is a Bayesian penalized reconstruction algorithm [14], where the penalty is controlled through a parameter β that is the only possible user input. The penalty function is designed such that the image edges are preserved while the background noise of the image remains low [15]. Thus, the regularization allows a commitment between image quality and quantification.

For the case of a PET tomograph with LYSO crystals and TOF correction, the behavior of the Q.Clear algorithm has been studied [16], observing that it improves the image quality of the study, with respect to an OSEM algorithm, and proposing an optimum β value. However, this optimization is not transferable to a reconstruction from different equipment with different characteristics such as the D-IQ.

Besides technical parameters, the methods to extract clinical relevant data from PET studies also evolved during the last years. For example, the emerging analysis of textures features on PET imaging consists of a set of post-processing mathematical tools for the evaluation of the heterogeneity of a certain volume. Texture analysis can generate hundreds of different parameters, which can be used in the application of a new medical imaging paradigm called radiomics [17]. Despite the recent interest in textural analysis, there is still a lack of consensuses to define the parameters that influence these [18]. As explained above, the Q.Clear algorithm acts as a selective filtering that differentiates it from common algorithms. There is no study yet that explains how this reconstruction can affect the morphology of the uptake in the lesions and therefore the analysis of the textures.

Despite all recent advances, there are still old problems in PET imaging that need to be addressed. Image quality in PET studies can be considered noisy when compared to other modalities, such as magnetic resonance or computed tomography. Under this circumstance, PET image quality should be evaluated thoughtfully in order to maximize the information extracted. Evaluation of image quality can be performed by different approaches. A common way to measure it is in terms of the signal-to-noise ratio (SNR), which for a simple linear reconstruction algorithm should be proportional to the squared number of recorded counts (assuming non-degrading effects). However, this relation does not hold for a modern iterative reconstruction algorithm, specially when incorporating PSF modelling [19, 20]. Moreover, reconstructed algorithms such as Q.Clear act as a selective filtering, changing the noise properties non uniformly, so the relation between counts and SNR could increase in complexity. Although it is common to use single parameters such as SNR to evaluate image quality, nowadays it is generally accepted that a good metric must be linked to a specific diagnostic task, as could be lung lesion detection. There are many different methods and metrics to evaluate image quality, and as stated previously, the Q.Clear reconstruction can change its relation to image properties. Thus, it is important to select the most relevant parameters for image quality assessment related to each different diagnostic task.

Chapter 2

Central Hypothesis and Purpose of this Thesis

Central hypothesis:

The characterization of the operation of a new PET/CT model and the optimization of the acquisition protocol together with the corresponding and adequate selection of a Bayesian penalty parameter in the reconstruction should lead to an improvement in the quantitation and image quality of clinical images.

Main goal:

To study the impact of a new reconstruction algorithm on the image quality and quantification of clinical PET studies performed with a high sensitivity BGO PET.

Specific aims:

1. Evaluate the PET imaging characteristics of the Discovery IQ 5-rings PET/CT.
2. Optimize the penalization factor for a Bayesian reconstruction algorithm (Q.Clear) under different levels of PET image noise.

3. Study the impact of Q.Clear reconstructions in different quantitative metrics, including the heterogeneity and morphology of lesions.

4. Evaluate the impact of Q.Clear reconstruction on the image quality metrics, and its relation to the subjective perception.

Chapter 3

Background

3.1 PET Imaging Physics

At a fundamental level, PET imaging is based in the use of proton rich radioactive nuclides, which decay through positron emission (also known as β^+ decay). This decay will produce a daughter nuclide containing one less proton so that its atomic number Z is decreased by one. Typical nuclides are ^{11}C, ^{13}N, ^{15}O, ^{18}F, ^{22}Na, ^{64}Cu,^{68}Ga, ^{82}Rb and ^{124}I; Table 3.1 presents some of its physical properties. The suitability of a nuclide for PET imaging depends on many factors, as could be its half-life, branching ratio, ease of production, the β^+ energy or its possible binding molecules. Thereby, from a practical clinical point of view, the most important nuclide is ^{18}F, which decays to ^{18}O emitting a positron of 633.5 keV.

Once the positron is emitted by spontaneous decay, it rapidly loses its kinetic energy in inelastic interactions with atomic electrons in the surrounding tissue, travelling a certain distance (*positron range*) depending on its energy and the tissue density; typical ranges are from 1 to several millimeters, increasing with the energy of the positron. When the positron has lost almost all its kinetic energy, it will interact with one electron of a surrounding atom, resulting in an annihilation of both the positron and the electron [22]. The annihilation results in the emission of 2 or more γ photons, with 2 photons accounting for almost all cases,

Table 3.1 Selected list of nuclides that decay by positron emission and are relevant to PET imaging. Data from Nuclides(www2.bnl.gov/ton). Adapted from Simon R. Cherry et al. [21].

Nuclide	Half-life	E_{max}(MeV)	Branching Ratio
^{11}C	20.4 min	0.96	1.00
^{13}N	9.97 min	1.20	1.00
^{15}O	122 s	1.73	1.00
^{18}F	109.8 min	0.63	0.97
^{22}Na	2.60 y	0.55	0.90
^{64}Cu	12.7 h	0.65	0.29
^{68}Ga	67.6 min	1.89	0.89
^{82}Rb	1.27 min	2.60/3.38	0.96
^{124}I	4.17 d	1.53/2.14	0.23

making the emission of more photons negligible. Due to the conservation of energy, the sum of the photon energies must be equal to the sum of the electron and positron masses, implying that each photon must be emitted with an energy of 511 keV.

If the annihilation of the positron and the electron occurs when the kinetic energy of both particles is exactly zero, the momentum in the centre of the masses is also zero. Due to the momentum conservation, the emitted photons should also preserve a zero momentum, which implies that both photons must be emitted in opposite directions, forming a 180° angle. However, the annihilation process does not occur when the total momentum is strictly zero, and the emitted photons will not depart in exactly opposite directions. This effect is called *acollinearity* of the photon pair. Although it is a small effect, it introduces an uncertainty on the departure angles, typically $180 \pm 0.25°$ [23].

Each photon will travel through the surrounding material and may interact with it. The interaction can be via photoelectric effect, Compton scattering, or Rayleigh scattering. In Rayleigh scattering the photon does not lose any considerable amount of energy, but considering the energy of the gamma photons, this kind of interaction could be neglected. For the other interactions, the mass attenuation coefficient of the 511 keV gammas in water

is $0.096\,\mathrm{cm}^2/\mathrm{g}$, low enough to allow the two photons escaping the body of the subject. The basis of PET imaging is the detection of these photons.

Surrounding the patient there is a ring of detectors; if both emitted photons are detected in a defined time window (usually in the order of nanoseconds), the process is called a coincidence event, and a line of response (LOR) is created joining the points of the detection (Fig. 3.1 A). The information recorded in every LOR is assembled and employed to produce an image of the activity uptake in the patient's body.

If two events are measured within the time window by two detectors, it is interpreted as a coincidence event. The time window τ applied to score coincidences depends on the time resolution $\delta\tau$ of the device. An electronic coincidence sorting unit filters the events through the application of a coincidence time window to the detected photons.

Each accepted LOR is binned in what is known as raw data, a collection of the number of detected events for each LOR, normally organized or stored as sinogram or projection data. If the data are stored in the list-mode each individual LOR is registered, and the time stamp is also recorded.

The desired coincidence events occur when both photons from the same annihilation are detected in coincidence without any kind of interaction prior to the detection, this is called a true coincidence or *true event*.

There are other kinds of coincidences. When at least one of the photons interacts with the surrounding matter by means of the Compton scattering, it changes the original direction of the photon, modifying the detection angle from the departure angle (180°) and assigning a wrong LOR to the coincidence event. These kind of coincidences are known as *scatter events*. The ratio of scatter events to the total coincidence events (*scatter fraction*) depends on the attenuation properties of the subject and of the detectors [24]. Also, part of the kinetic energy of the photon is lost in each scatter process. As common PET detectors measure the photon energy, scatter events could be minimized if all photons from different energies than

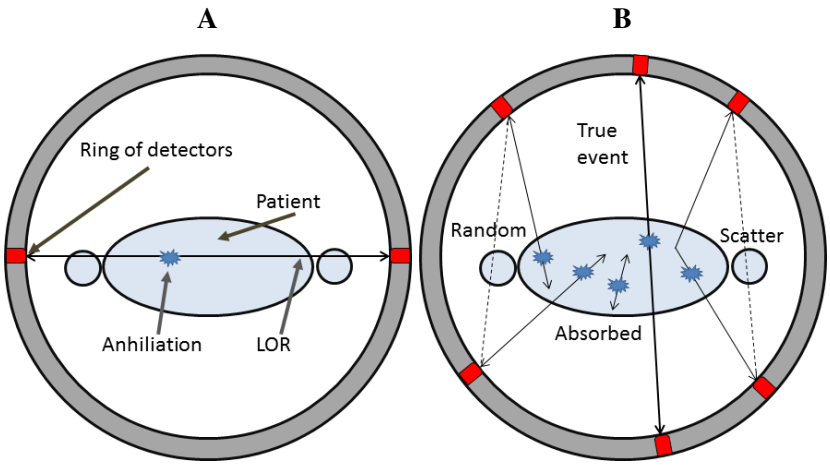

Fig. 3.1 (A) Definition of a line of response, and (B) physical effects in PET, representing attenuation, scatter and random events.

a defined energy window are discarded. However, there is an intrinsic limit to this caused by the energy resolution and by the need to maintain a minimum detection sensitivity.

Another type of events are the *random events*, which occur when two photons from different annihilation events are detected in the same time window. This process is more likely observed at high count rates, as it increases the probability between different but nearly simultaneous events. Then, random events would be more important in the presence of large activity concentrations. It is also dependent on the time window of the detection. If N_{d1} and N_{d2} are the total photon rates detected at detectors 1 and 2, respectively, and τ is the duration of the timing window, for a single event at detector d_1 at time t_1, all the events detected in the interval $[t_1-\tau, t_1+\tau]$ will be accepted. During this interval, in detector d_2 there will be $2\tau N_{d_2}$ photons detected. Thus, the mean rate of randoms will be given by:

$$\langle R \rangle = 2\tau N_{d_1} N_{d_2}. \tag{3.1}$$

Slower detection systems induce to increase the time window, which in turn increases the probability of random events [24]. Moreover, Eq. 3.1 is valid for all detector pairs accounted in the acquisition, thus increasing the possible detector pairs increases the random

coincidences generated. Nevertheless, this type of coincidences can arise from activity anywhere in the region between the detectors, including activity outside the useful FOV, making the relation between true events and the random counting rate dependent in a complex relation on both the source and the detector geometry [22, 25].

Both scatter and random events assign a wrong LOR, resulting in a noisy background to the true coincidence distribution and decreasing overall image quality. To mitigate the contribution of random events, a usual technique is the delayed coincidence channel method. This method is usually applied by acquiring an additional sinogram with a delayed coincidence time window. Non-true coincidences can be evaluated in this sinogram, though, since the random events are uniformly distributed in time, the delayed coincidences have the same mean as the random events in a correct time window. This delayed sinogram is then subtracted from the real sinogram. On the other hand, scatter events are more complicated to model, and they are usually corrected from a model-based approach of the relation between the involved gamma photon scatter and the attenuation map [22, 26].

There is a certain time of flight for each photon, from the annihilation event to the detection process; if the detectors time resolution is low enough, the difference in detection time of both photons can be used to shorten the possible annihilation range along the LOR. This is known as TOF. In the absence of TOF information, it is assumed that the decay event could have occurred anywhere along the LOR, with uniform probability.

3.2 PET scanners

Although in recent years, new detector materials have been introduced [9], most modern PET scanners still use a pixelated array of scintillator crystals in order to detect both annihilation photons. Scintillator material emits visible light when a γ-photon energy is deposited in it, by the transition of electrons in an excited state and the posterior fast decay into the ground state by the emission of photons in the visible range [27]. An ideal detector would have high

Table 3.2 Properties of selected inorganic scintillators. Some of these specifications can be modified by the vendor specific impurity proportions and growth methods. Adapted from T. Lewellen [28].

	NaI(Tl)	BGO	LYSO	LSO
Chemical Formula	NaI(Tl)	$Bi_4(GeO_4)_3$	$Lu_{2-x}Y_xSiO_5$:Ce	$Lu_2(SiO4)O$:Ce
Effective atomic number (Z)	53	74	60	66
Density (g/cm^2)	3.7	7.1	7.1	7.4
Peak emission wavelength (nm)	410	480	420	420
Light yield (photons/keV)	38	7.2	30	34.8
Decay time (ps)	230	300	40	31

density and high atomic number to increase the absorption coefficient of the γ-photons, a fast decay time to increase the resolution time, a high luminosity to increase the number of output photons of the incident photon, a high spatial and energy resolution (to discriminate the scatter events), and it should be as inexpensive as possible to be affordable to extend as maximum the axial arrangement of the crystals [28].

Scintillators can be divided in two categories: organic and inorganic. Organic scintillators are based on hydrocarbon compounds and exhibit excellent time properties, but they also possess low density. On the other hand, inorganic scintillators are based on high atomic number (Z) elements, and exhibit better absorption coefficients and also a desirable more approximate linear response. Inorganic scintillators are only scintillators in crystal form, and mostly they are activated by impurity atoms of other elements, which causes a disturbance to the crystal matrix. Some of this common scintillators are sodium iodide activated with thallium (NaI(Tl)), bismuth germanate oxide (BGO), lutetium oxyorthosilicate (LSO), and lutetium yttrium oxyorthosilicate (LYSO). Table 3.2 summarizes some of its characteristics. Although the most common crystal used for a long time has been BGO, since the end of the last decade most scanners use LSO or LYSO crystals, due to their high luminosity and time resolution, suitable for TOF measurements.

The crystals are attached to a device to convert the emitted scintillation photons into an electrical signal. Most common devices are the Photomultiplier Tubes (PMT), made of a

vacuum glass tube, which houses a photocathode, an anode, and several dynodes. When the light from the scintillator enters the PMT through a glass entrance window, an electron may be generated by phoelectric effect. This electron is accelerated by a potential difference across the PMT towards a series of dynodes, at each of which more electrons are ejected in an avalanche effect, resulting in a detectable electric signal at the back of the PMT.

Although PMT was the dominant technology in previous PET generations, recently, new PMT alternatives are entering the market, as could be Avalanche Photodiodes (APD) or Silicon Photomultipliers (SiPM). SiPM comprise a robust alternative, more compact, with a high gain (similar to a PMT), good intrinsic timing resolution (under 200 ps), a higher value of photon detection efficiency than PMTs, and they are MRI compatible [9, 29, 30].

Table 3.3 summarizes the specifications of some modern PET/CT scanners. The only BGO PET presented is the Discovery IQ from GE Healthcare; all other relevant PET are LSO or LYSO systems.

Table 3.3 Specifications of different commercial PET/CT systems.

Manufacturer	**GE Healthcare**			**Siemens**		**Philips**		**Canon Medical**
System	**D-IQ-5**	**D-710**	**D-Mi**	**mCT Flow**	**Vision**	**Ingenuity**	**Vereos**	**Celestion**
References	[31]	[32]	[32, 33]	[34]	[35]	[9, 30, 36]	[9, 30, 36]	[9, 30, 36]
Detector material	BGO	LYSO	LYSO	LSO	LSO	LYSO	LYSO	LYSO
Crystal size (mm^2)	6.3x6.3x30	4.2x6.3x30	3.95x5.3x25	4x4x20	3.2x3.2x20	4x4x22	4x4x19	4x4x12
Total detector elements	11520		-	324448	-	28336	23040	30720
Photodetector	PMT	PMT	SiPM	PMT	SiPM	PMT	SiPM	PMT
Patient bore (cm)	74.0	70	70	78	78	71.7	70	88.0
TOF	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Transaxial FOV (cm)	70		-	81.5	-	67.6	-	70.0
Axial FOV (cm)	26	15.7	20	21.8	25.6	18.0	16.4	19.6
Energy window	425-650	-	-	435-540	-	440-66	-	425-650
Energy resolution (%)	-	-	-	11.5	-	11.1	11.1	12.4
Time coincidence window (ns)	9.5	8.2	-	4.1	-	4.5	4.0	1.6
Time resolution (ns)	N.A.	5.44	3.85	0.5	-	0.5	0.3	0.4
Axial resolution 1 cm	4.8	4.8	4.8	4.4	3.8	4.7	4.0	5.0
Axial resolution 10 cm	4.8	5.6	4.7	5.5	4.3	5.2	4.3	5.4
Transaxial resolution 1 cm	4.5		-	4.4	3.7	4.8	4.1	5.1
Transaxial resolution 10 cm	5.4		-	4.9	3.9	5.1	4.5	5.1
Sensititivy (kcps/kBq)	22.8	7.1	13.5	9.7	16.4	7.3	5.7	3.8
Scatter fraction (%)	36.2	-	-	33.2	38.7	36.7	30.0	37.3
Peak NEC (kcps/kBq/ml)	124/9.1	144/29	180/20	180/28	306/32	124/20.3	171/50	70/29.6

3.3 Physical factors affecting image quality

From its creation, the main limitations in PET/CT imaging have been the same: the relative poor spatial resolution and the generally low SNR ratio.

3.3.1 Spatial resolution

Spatial resolution is mainly limited by the physics of positron detection. Two fundamental limits exist, the positron range and the acollinearity, described earlier in Section 3.1. For ^{18}F, the positron range blurring is approximately 0.54 mm Full Width at Half Maximum (FWHM) [23]. The acollinearity effect introduces a small deviation of 0.25° around the 180° mean value. The impact of this effect on spatial resolution depends on the detector ring diameter, the FWHM is approximately $2.2 \cdot 10^{-3}$D [23], where D is the detector diameter in millimeters. Following this equation, for a 70 cm PET, the blurring effect will be about 1.5 mm. This effect is one limiting factor on the design of whole-body PET scanners, which require a large ring to position different patient sizes and accessories.

Combining only both effects, it can be estimated that a whole-body oncological PET scanner can achieve a spatial resolution of approximately 2 mm [23]. However, as it can be seen in Section 3.2, current PET scanners are still far from it. Mostly because the dominant factor is the loss of spatial resolution produced by the detectors. Which in turns is limited, in a pixelated detector design, by the cross section (width) of each individual crystal used in the PET detector. While a theoretical spatial resolution of half the crystal width can be achieved, the finite detector sampling over the FOV often degrades it [3, 37]. Traditional PET scanners use much larger photomultiplier tubes than the individual detectors in the matrix to which the tubes are coupled, so Anger logic must be employed to estimate the location of the photons original impact, limiting the further improvement of spatial resolution in PET imaging due to an added uncertainty in the localization of the events. Some progress has been made in recent years in the so called digital PET systems, by using the above commented SiPM

light detectors which are able to minimize the coupling between the detector crystal and light detector to a 1:1 ratio [9]. Unfortunately, the detection width can not be decreased as desired, since it is limited by the detected photons, i.e. the low SNR, which is the other main disadvantage of PET scanners [25]. One solution would be to increase the depth of each detector element to improve the counting efficiency. However, apart for the detectors width, its thick (typically 2–3 cm) leads to another geometric effect that degrades spatial resolution, known as depth of interaction (DOI) or parallax error, caused by the fact that the annihilation photons can interact at any depth in the scintillator material, resulting in a degradation of the spatial resolution as you move away form the centre of the field of view [21]. Several solutions have been proposed to minimize this error, by developing DOI encoding detectors [38].

Lastly, the resolution obtained in clinical images will be also affected by the reconstruction method and filter, which will add a blurring to suppress noise and, inevitably, will contribute to a spatial resolution decrease.

The lack of small spatial resolution will impact directly on the image quality, by means of an undesired blurring of small structures, but will also limit the ability to obtain accurate quantitative values. This is due to an undesired cross-contamination between adjacent structures with distinct activities, referred as the partial volume effect (PVE) [39, 40]. Structures smaller than two to three times of the FWHM, will be specially affected by the PVE. Spatial resolution, and in general spatial variations in PVE will be a confounding factor on many studies [41]. PVE effect is not only created for the intrinsic resolution of the system, and can also be created by the voxel size or by a decrease on resolution due to the movement of the structure.

3.3.2 Factors affecting Signal-to-Noise Ratio

Initial PET reconstruction algorithms could only acquire two dimension images, reconstructing each slice of the subject for each plane of the sinogram, and then obtaining the 3D tomographic image after stacking each plane. Thus, only those LOR corresponding to detector pairs in the same ring were necessary; these are the *direct planes*, although some limited cross ring LOR were allowed to enable planes between adjacent rings. These *cross planes* can be assumed as been acquired by a virtual ring of detectors [22]. To limit the number of undesired coincidences (random and scatter events), these systems usually have tungsten collimating septa around the detector, in order to reduce the view of one to the same ring detectors, and are usually known as 2D PET systems.

The removal of the septa allows a 3D data acquisition, employing any LORs between rings, even for larger ring differences. This leads to a substantial increase in sensitivity, but at the expanse of increasing scatter and random events. First PET systems that allowed 3D reconstruction also included removable septa to additionally operate as a 2D PET. As an example to illustrate the gain in sensitivity, the 2D/3D Discovery ST PET/CT system [11] has a sensitivity at the centre of the FOV of 1.9 cps/kBq and of 9.5 cps/kBq operating in 2D and 3D mode, respectively. On the other hand, the scatter fraction under the NECR peak is 16.3% in 2D and of 32.6% in 3D, measured according to the NEMA NU 2-2001 standards. Despite the increase of undesired events, the sensitivity gain of 3D PET clearly compensates its drawbacks, and modern systems do not include any 2D option.

Pure sensitivity and scatter fraction is difficult to correlate with image quality. In this sense, one parameter that somehow better describe the equipment performance is the SNR. In PET imaging and in a first approximation, the detection process follows a Poisson distribution and the SNR dependence on the number of detected events (**N**) can be modelled as

$$SNR \sim \sqrt{N}, \tag{3.2}$$

given a PET scan time per bed of T, and assuming it is sufficiently short compared with the half-life of the radionuclide being imaged, so that radioactive decay can be ignored, and that loss of events from detector or system dead time is negligible, as happens with ^{18}F radiopharmaceuticals. The number of events acquired in a PET scan is approximately given by

$$N \sim G\varepsilon^2 AT, \tag{3.3}$$

where G is the average geometric coverage of the scanner, ε is the efficiency of the detector, and A is the concentration of activity in the field of view [25]. The factor ε is squared as the photons must be detected by two detectors at the same time. Then, the SNR will depend on the activity in the field of view as

$$SNR \sim \varepsilon\sqrt{GAT}. \tag{3.4}$$

If the system incorporates TOF modeling, the relationship between SNR and the timing resolution $\delta\tau$, is given by $SNR \sim \sqrt{1/\Delta t}$ [42], and in this case the SNR can be approximated as

$$SNR \simeq K\varepsilon\sqrt{\frac{GAT}{\delta\tau}}, \tag{3.5}$$

where K is a factor that accounts for patient-specific factors such as attenuation and scatter of the photons inside the body and on the source and scanner geometry [22, 25].

As can be easily seen from the previous equations, under the same ratio of time and activity, an increase of the SNR can be reached by increasing the time resolution of the detector or by extending the axial FOV. Unlike other PET systems (see Table 3.3) the D-IQ-5 uses the latter method. Investing in increase of the axial FOV is not a dead-end, and some ambitious projects are currently under development to create a PET scanner with a FOV that covers the whole body in a single static acquisition [43, 44].

3.4 Iterative Reconstruction Algorithms

First PET tomographic reconstructions were routinely performed by analytical approaches, normally FBP. The main advantage of an analytical approach is the computation speed; however, the accuracy of the reconstruction is limited by the approximation of the method as well as the intrinsic PET limitations (undesired events, positron range, etc.). Nevertheless, in 3D PET the most common reconstruction is the iterative reconstruction [45, 26]. Iterative methods are more time-computing expensive, but have the power to adopt a model approach to incorporate the physics involved from the positron annihilation to the coincidence detection, and without requiring the solution of difficult analytical problems. Even more, a model-based iterative approach starts from the assumption of a statistical nature of the detection of radiation, while analytical methods can only control the noise by means of filtering the sinogram.

The modern era of the iterative reconstruction started with the works of Shepp and Vardi [46] and Lange and Carson [47] on the maximum-likelihood expectation-maximization (MLEM) framework, based in a Poisson model of the emission data. This is probably the most studied iterative reconstruction in medical imaging. The basic idea behind iterative reconstruction approaches is summarized in Figure 3.2. In short the algorithm starts with an initial estimation of the reconstruction (normally a uniform image), from where it computes the simulated projections (*forward projection*). The resulted projections are compared to the real ones, and a matrix of weight is obtained to update the reconstruction estimate *back projection*. This process is repeated until the iterations converge.

For a 3D PET, the image is binned into voxels, which can be represented by a vector $\mathbf{f}$ distributed in J voxels $(j = 1,..J)$. The goal of the reconstruction algorithm is to find the distribution f which has the highest probability to have generated the measured projection data $\mathbf{y}$, where if the system has I LORs $(i = 1...I)$, then y_i is the number of counts measured in each LOR.

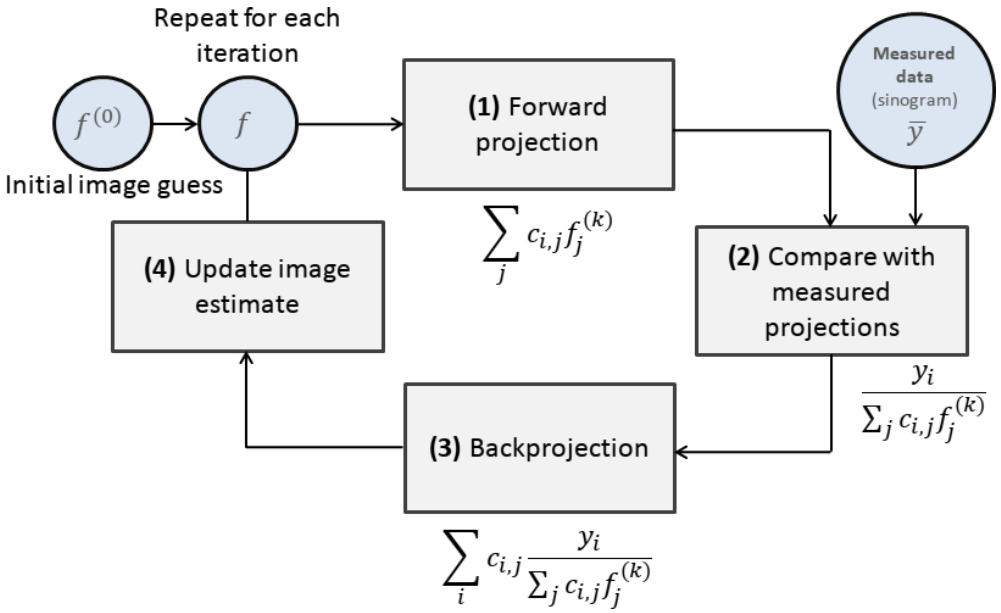

Fig. 3.2 Schematic diagram for the maximum-likelihood expectation-maximization method. The algorithm starts with an initial image guess in the upper left ($f^{(0)}$).

The probability of finding y given f, will be given by a Poisson mass function

$$p(y|f) = \prod_{i=1}^{I} e^{-\bar{y}_i} \frac{\bar{y}_i^{y_i}}{y_i!}, \tag{3.6}$$

where $\bar{y}_i$ is the expected value of y_i ($E[y_i]$), and the expected count of an i LOR (forward projection) will be given by

$$\bar{y}_i = \sum_{j=1}^{J} c_{i,j} f_j, \tag{3.7}$$

where c_{ij} is the so-called system matrix, which represents the probability that an emission from voxel j will be detected in LOR i. The first part of the MLEM algorithm is given by the maximum likelihood (ML), which is just the logarithm of the Poisson distribution

$$L(y|f) = \sum_{i=1}^{I} y_i \log(\bar{y}_i) - \bar{y}_i - log(y_i!). \tag{3.8}$$

The ideal solution will be given by the maximization of this likelihood function given f_j. Substituting the equation 3.7, and noting that the last term of the previous summation does

not contain the f_j term, the equation to maximize is given by

$$L(y|f) = \sum_{i,j} y_i \log(c_{i,j} f_j) - c_{i,j} f_j. \tag{3.9}$$

The second part of the MLEM algorithm is the maximization of the previous equation, which is out of the scope of the present introduction; see references [26, 45] for a detailed description. In short, Sheppard and Verdi obtained the MLEM equation using the maximization expectation of the likelihood, which (in its sinogram-based implementation) has the closed form

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_j c_{i,j}} \sum_i c_{i,j} \frac{y_i}{\sum_j c_{i,j} f_j^{(k)}}, \tag{3.10}$$

where y_i is the number of counts collected in LOR i. If the measured counts are corrected by random and scatter events, then it will not follow a Poisson distribution and Eq. 3.10 is not a valid assumption. That is because the expected value, instead of Eq. 3.7 is given by

$$E[y_i] = \sum_{j=1} c_{i,j} f_j + r_i + s_i, \tag{3.11}$$

where r and s are the random and scatter events. The subtracted randoms modify the expected value, but not its variance, and for a Poisson distribution the variance and the expected value must be the same. The likelihood usually is modified to account for this effect, and the MLEM has the form

$$f_j^{(k+1)} = \frac{f_j^{(k)}}{\sum_j c_{i,j}} \sum_i c_{i,j} \frac{y_i}{\sum_j c_{i,j} f_j^{(k)} + s_i + r_i}. \tag{3.12}$$

The MLEM algorithm has some particular useful properties. First, the total counts in each iteration are preserved; second, it guarantees positive values for each voxel (negative concentrations are not expected); and last, it assures convergence to the maximum likelihood estimate.

On the other hand, the main disadvantage which limits the application of the MLEM method in clinical practice is the slow convergence. Each iteration performs one forward projection of the estimate and one backprojection of the whole data set, thus a method as the FBP (on backprojection) reconstructs all the dataset with approximately half the time needed for one MLEM iteration [48]. Many attempts have been developed to speed the reconstruction time of the MLEM. The most frequently used method in nuclear medicine reconstruction is the Ordered Subsets Expectation Maximization (OSEM) method by Hudson and Larkin [49]. In the OSEM algorithm projection data are divided into ordered subsets, with each subset containing an equal number of projections. Splitting the measured dataset into different subsets and using only one subset for each iteration speeds up the algorithm with a factor equal to the number of subsets. Therefore, the product of iterations and subsets is commonly refereed as *effective iterations*. However, in comparison to an MLEM this method will not guarantee to converge to the maximum likelihood (and will not happen in general [26]), which is a desired condition for an optimal tracer quantification. In any case, Hudson and Larkin proved that image quality for the same number of iterations in MLEM and effective iterations in OSEM is comparable if the number of subsets is not too high [49].

A faster reconstruction time with a comparable image quality, and the method simplicity in the form of a closed solution, makes it enough to widespread the use of OSEM in PET reconstruction. For example, the algorithm VPHD by GE Healtchare which is usually referred in the present dissertation is a commercial variation of and OSEM algorithm. For most situations, OSEM provides accurate quantitative results within 3%, but bias up to 50% can be achieved for lesions with hotter backgrounds. This effect can partially be explained by differences in the convergence rate of different regions [50].

The other main disadvantage of the MLEM algorithm (and of all ML methods in general, as the OSEM algorithm) is that the convergence solution is highly dependent on the initial data noise (the so called ill-conditioning problem), which can be appreciated in Fig 3.3 as

a noise magnification with the number of iterations. With each iteration the reconstructed image will be closer to the convergence, but the reconstruction also increases in noise. This is due to the initial data measurements, which are Poisson random variables. One approach to solve this problem is to stop the iteration process after a desired level of convergence and smoothness [51, 52]. Another approach, which is the one commonly applied in clinical practice, is to post-filter the reconstruction after the desired convergence. This has the advantage to simply choose the desired image quality by modifying a post-filtering cut-off frequency, and easily adjusting the desired level of noise to the nuclear medicine physicians, in the other hand, though, the post-filtering will modify the image quantification values. Moreover, this approach has been exploited by clinical trials and international initiatives [53, 54], to equal the quantitative parameters of different PET machines and centres.

Another approach, is to add a smooth penalty factor in the likelihood function. If we have some prior information regarding how the image is expected to be (in terms of smoothness or other metrics), then this information can be used in a Bayesian formulation, performing a regularized iterative reconstruction [55].

In a pure Bayesian regularization, the prior distribution of the image $p(\mathbf{f})$ is introduced in the likelihood trough the Bayes rule

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{y})}. \tag{3.13}$$

If we are interested in the maximum a posteriori (MAP) estimate of the image, the objective function has the form

$$\Phi(\mathbf{f}) = L(\mathbf{y}|\mathbf{f}) + p(\mathbf{f}). \tag{3.14}$$

The block sequential regularized expectation maximization (BSREM) is a convergent OSEM-like algorithm to solve the MAP problem. It is an example of a penalized likelihood

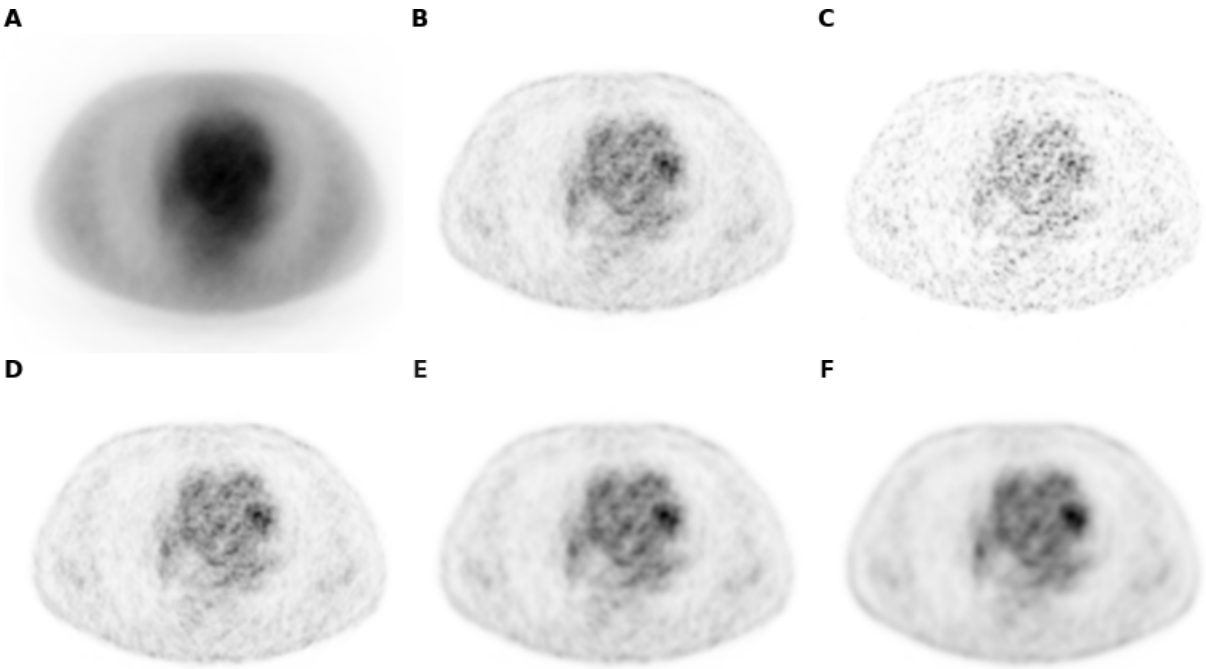

Fig. 3.3 Effect of reconstruction parameters in ML methods. A) MLEM with 2 iterations, B) 48 iterations and C) 192 iterations. The details of the image increase with the iterations, however the image ends dominated by noise. D) An OSEM reconstruction using 12 subsets and 4 iterations (48 effective iterations) has a similar aspect as the corresponding MLEM reconstruction. E) The same OSEM reconstruction as D with a 4.8 Gaussian post-filtering, which is more suitable for clinical practice. F) the latter reconstruction adding PSF modeling.

image reconstruction method, which adds a penalty term to the likelihood function that controls image quality [26, 56, 57].

Q.Clear is a commercial implementation of a BSRM reconstruction. It includes a point-spread function (PSF) modeling and controls the noise through the use of a penalty term [57, 16], which imposes more smoothing in lower activity regions and less smoothing in higher activity regions, resulting in smoother cold backgrounds and improved hot lesions signal-to-noise ratios. The behaviour of Q.Clear is controlled by a parameter β. Image reconstruction is performed by maximizing the following penalized-likelihood objective function over all non-negative images, assuming a Poisson model for the data [58]:

$$\Phi(x) = \sum_i y_i log([\mathbf{cf}]_i + r_i) - ([\mathbf{cf}]_i + r_i) - \beta R(f). \tag{3.15}$$

The $R(f)$ term is the regularization [59], which has the form

$$R(f) = \sum_j \sum_{k \subset N_j} w_j w_k \frac{(f_j - f_k)^2}{f_j + f_k + \gamma |f_j - f_k|}, \tag{3.16}$$

where N_j is the subset of neighbors of voxel j and both w_j and w_k are penalty terms that controls for the position dependent of the local smoothing, because in a PET scanner the resolution is not uniform around the FOV [15]. The parameter γ controls over the importance of the relative difference of pixels, to avoid cases where $|f_j - f_k| \gg f_j + f_k$, principally on the image edges; normally γ is set equal to 2.

3.5 Quantitative parameters

An absolute PET quantification should be extracted by means of kinetic analysis, in the form of relevant metabolic parameters, for example glycolysis absorption in the case of FDG studies. However, these studies are difficult to perform due to high demands in time

a non-trivial computation of the data. One alternative are the Standardized Uptake Value (SUV) related metrics, usually referred as semi-quantification to differentiate them from the parametric imaging obtained from kinetic analysis. SUV is defined as

$$SUV(t) = \frac{C(t)}{A'(t)/W} \tag{3.17}$$

where $C(t)$ is the concentration in Bq/ml of each voxel, $A'(t)$ is the injected activity in Bq decay corrected to the beginning of the PET acquisition, and W is the patient mass in grams. From this equation one can see that SUV is just an escalation of the pixel values, to allow its comparison for different injection activities and patient weights. If we assume that the radiopharmaceutial is distributed throughout all the body and that the body density is equal to the water density, a SUV equal to 1 means an homogeneous distribution of the radioparmaceutial and an SUV bigger than 1 means as a increased uptake. Computing SUV gives units of g/ml or equivalent, although it could be considered a unitless parameter, as it is used in soft tissue where it can be assumed that the body densitiy is equal to the water density. Still, it is customarily to be presented in units of g/ml.

SUV has several limits that affects its reliability, from physical factors to biological ones [60, 61]. As can be deduced from Section 3.4, the SUV value will depend on the reconstruction algorithm and its characteristics, specially if the iterations are not enough to achieve the convergence or if there is any post-filtering. For example, Q.Clear, by using the penalty function, allows an effective SUV convergence, providing more accurate values. On the other hand, from a biological point of view, one of the main limitations of all SUV related parameters is that $C(t)$ is a function of time, and it also depends of each person and tissue, and then the SUV value is much dependent on the PET acquisition time after the radiopharmaceutical injection.

Other similar definitions quantitative metrics derived from SUV exist. The most used in clinical practice is the SUV normalized to the lean body mass (LBM) instead of the body mass, referred as SUL.

Radiomics and texture features

Radiomics is the use of quantitative imaging features extracted from medical images to characterize tumor pathology or heterogeneity. The potential of such an approach is to quantify properties of tissues and/or organs beyond the capability of visual interpretation or simple metrics [17]. The majority of features currently used in radiomics studies were designed to evaluate visual perception characteristics quantitatively, such as contrast, complexity, etc. On the present dissertation, the words radiomics, texture or heterogeneity features will be mostly used indistinctly, although the former can include more parameters that are not strictly measures of heterogeneity.

First-order features estimate properties of the histogram of individual voxel intensities; parameters such as SUV_{mean} or SUV_{max} can be considered as first-order features, along with other less common ones such as distribution deviation, skewness or kurtosis, and histogram entropy. First order features ignore the spatial arrangement between pixels, and as such cannot be considered as true texture or heterogeneity measures [18].

Second-order or higher-order features include the ones derived from the gray-level co-occurrence matrix (GLCM), Gray-Level Run Length Matrix (GLRM) or the the Gray-Level Zone Matrix (GLZM). These features are considered second-order features because they depend on the spatial position of each gray level, and hence can capture more information from the image than first-order features.

Most of the features were originally created to solve non-PET imaging problems, as the ones derived from the Harlick matrix [62] which was originally created to solve a spatial map detection problem. Only in the last couple of years texture analysis was applied to

PET and other radiological images (such as CT or magnetic resonance imaging), so the field of research is in its early days, still expanding and adapting its methodologies. In the present section we only focus in the description of these features used in our in-house code (written in python 3.6); however these features are general and are also included in other codes as the pyradiomics package [63], event if their name convention or some aspects of their mathematical definition could be slightly different. Due to the vast number of textural indices (TI) we focused in a subset of indices described in previous publications [64]. Our in-house code only uses features extracted from the Co-Occurence matrix, Run Lenght Matrix and Zone Lenght Matrix, although there are other common features extracted from the Neighborhood Gray-tone Matrix (NGTM), wavelet analysis or fractal analysis. Each used feature category included is explained in more depth below.

Co-ocurrence matrix: the gray-level co-occurrence matrix (GLCM), defined by Robert Haralick et al. in 1973 [62], represents the frequency at which each gray-level i intensity appears adjacent to each other gray-level intensity j for a given direction, which in a 3D matrix corresponds to 13 different angles. From the resulting matrix, several heterogeneity parameters can be extracted. The parameters were obtained for each possible angle and then the mean value was computed.

- The second order parameter of homogeneity, was defined as

$$GLCM_{Homogeneity} = mean_{angle}\left(\sum_i\sum_j \frac{GLCM(i,j)}{1+|i-j|}\right). \tag{3.18}$$

- The entropy of the GLCM was defined as

$$GLCM_{Entropy} = mean_{angle}\left(\sum_i\sum_j GLCM(i,j)\log_{10}(GLCM(i,j)+\varepsilon)\right), \tag{3.19}$$

where ε is the smallest possible variable to account for cases where $GLCM(i,j)$ is zero.

- Correlation was defined as

$$GLCM_{Correlation} = mean_{angle}\left(\sum_i\sum_j \frac{(i-\mu_i)(j-\mu_j)GLCM(i,j)}{\sigma_i\sigma_j}\right), \tag{3.20}$$

where μ_i/σ_i and μ_j/σ_j are the mean/variance of rows i and column j, respectively.

Run Length Matrix: the Gray-Level Run Length Matrix (GLRM), defined by Mary Gallowey in 1974 [65] gives the size of homogeneous runs for each grey level, where a run is collinear image pixels with the same gray level intensity value. The same procedure for averaging was used for all angles as in the GLCM matrix. For a GLRM matrix, with j voxels of homogeneous runs with intensity i.

- The Short-Run Emphasis (SRE) is the distribution of the short homogeneous runs defined as

$$SRE = mean_{angle}\left(\frac{1}{N}\sum_i\sum_j \frac{GLRM(i,j)}{j^2}\right), \tag{3.21}$$

where N corresponds to the total number of homogeneous runs in the GLRM.

- The Long-Run Emphasis (LRE) is the distribution of the long homogeneous runs defined as

$$LRE = mean_{angle}\left(\frac{1}{N}\sum_i\sum_j j^2 GLRM(i,j)\right). \tag{3.22}$$

- The High Gray-level Run Emphasis (HGRE) is the distribution of high grey-level runs defined as

$$HGRE = mean_{angle}\left(\frac{1}{N}\sum_i\sum_j i^2 GLRM(i,j)\right). \tag{3.23}$$

- The Low Gray-level Run Emphasis (LGRE) is the distribution of low grey-level runs defined as

$$LGRE = mean_{angle}\left(\frac{1}{N}\sum_i\sum_j \frac{GLRM(i,j)}{i^2}\right). \tag{3.24}$$

Zone Matrix: the Gray-Level Zone Matrix (GLZM), defined by Guillaume Thibault in 2003 [66], measures the distribution of clusters of grays levels. A gray level zone is defined as the number of connected voxels that share the same gray level intensity. A voxel is considered connected if there is another voxel of the same gray levels in the 26 neighborhood voxels in a 3D VOI. Contrary to the GLCM or GLRLM, the GLZM does not depend of any direction and only one matrix is computed. Given the number of homogeneous zones of j voxels with the intensity i:

- The Short-Zone Emphasis (SZE), which represents the distribution of the short homogeneous zones in an image is defined as

$$SZE = \frac{1}{N}\sum_i\sum_j \frac{GLZM(i,j)}{j^2}. \tag{3.25}$$

- The Long-Zone Emphasis, representing the distribution of long homogeneus zones is defined as

$$LZE = \frac{1}{N}\sum_i\sum_j j^2 GLZM(i,j). \tag{3.26}$$

- The Low-Level Zone Emphasis (LGZE), the distribution of the low grey-level zones, is defined as

$$LGZE = \frac{1}{N}\sum_i\sum_j \frac{GLZM(i,j)}{i^2}, \tag{3.27}$$

where N corresponds to the total number of homogeneous runs in the GLRM.

- The High-Level Zone Emphasis (HGZE), the distribution of the high grey-level zones, was computed as

$$HGZE = \frac{1}{N}\sum_i\sum_j i^2 GLZM(i,j). \tag{3.28}$$

Chapter 4

Performance Characteristics of the Whole-Body Discovery IQ PET/CT System

Reynés-Llompart, G., Gámez-Cenzano, C., Romero-Zayas, I., Rodríguez-Bel, L., Vercher-Conejero, J. L., & Martí-Climent, J. M. (2017). Journal of Nuclear Medicine, 58(**7**), 1155-1161.

In this chapter, we present the results for Specific Aim 1 of this thesis: evaluate the PET imaging characteristics of the Discovery IQ 5-rings PET/CT.

Reynés-Llompart G, et al. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. Journal of Nuclear Medicine, 58(7): 1155-1161.
http://doi.org/10.2967/jnumed.116.185561

Chapter 5

Phantom, clinical and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q.Clear) on a BGO PET/CT scanner

Reynés-Llompart, G., Gámez-Cenzano, C., Vercher-Conejero, J. L., Sabaté-Llobera, A., Calvo-Malvar, N., & Martí-Climent, J. M. (2018). Medical physics, 45(**7**).

In this chapter we present the results for Specific Aims 2 & 3 of this thesis: optimize the penalization factor for a Bayesian reconstruction algorithm (Q.Clear) under different levels of PET image noise and to study the impact of Q.Clear reconstructions in different quantitative metrics, including the heterogeneity and morphology of lesions.

Reynés-Llompart G, Gámez-Cenzano C, Vercher-Conejero JL, Sabaté-Llobera A, Calvo-Malvar N, Martí-Climent JM. Phantom, clinical and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q.Clear) on a BGO PET/CT scanner. Medical physics, 2018, 45(7):3214-3222.
http://doi.org/10.1002/mp.12986

Chapter 6

Image quality evaluation in a modern PET system: impact of new reconstructions methods and a radiomics approach

Gabriel Reynés-Llompart, Aida Sabaté-Llobera, Elena Llinares-Tello, Josep M Martí-Climent, Cristina Gámez-Cenzano (2019). Scientific Reports, 1-**9**, 10640.

In this chapter, we present the results for Specific Aim 4 of this thesis: evaluate the impact of Q.Clear reconstruction on the image quality metrics and its relation to the subjective perception.

SCIENTIFIC REPORTS

OPEN

Image quality evaluation in a modern PET system: impact of new reconstructions methods and a radiomics approach

Received: 26 March 2019
Accepted: 4 July 2019
Published online: 23 July 2019

Gabriel Reynés-Llompart[1,2], Aida Sabaté-Llobera[2], Elena Llinares-Tello[2], Josep M. Martí-Climent[3] & Cristina Gámez-Cenzano[2]

The present work investigates the influence of different biological and physical parameters on image quality (IQ) perception of the abdominal area in a modern PET scanner, using new reconstruction algorithms and testing the utility of a radiomics approach. Scans of 112 patients were retrospectively included. Images were reconstructed using both OSEM + PSF and BSRM methods, and IQ of the abdominal region was subjectively evaluated. First, 22 IQ related parameters were obtained (including count rate and biological or mixed parameters) and compared to the subjective IQ scores by means of correlations and logistic regression. Second, an additional set of radiomics features was extracted, and a model was constructed by means of an elastic-net regression. For the OSEM + PSF and especially for the BSRM reconstructions, IQ parameters presented only at best moderated correlations with the subjective IQ. None of the studied parameters presented a good predictive power for IQ, while a simple radiomics model increased the performance of the IQ prediction. These results suggest the necessity of changing the standard parameters to evaluate IQ, particularly when a BSRM algorithm is involved. Furthermore, it seems that a simple radiomics model can outperform the use of any single parameter to assess IQ.

Positron emission tomography (PET) with ^{18}F-2-fluoro-2-deoxy-D-glucose (FDG) has become a routine image procedure for the management of oncological patients. Compared to other imaging modalities, PET exams are limited by their low spatial resolution and signal-to-noise ratio (SNR)[1]. However, there is still an interest in decreasing as much as possible the administered activity dosage, both for patient safety and economic concerns, though images must maintain a certain level of diagnostic accuracy, not only in clinical research and trials, but also for medical diagnostic purposes.

The assessment of image quality (IQ) in PET is a challenging task affected by biological and physical factors[2]. It can be studied using phantoms or human examinations by means of different quantitative metrics. Some of the measurements could be considered as standard IQ parameters, such as the variance, SNR or contrast-to-noise ratio (CNR) of a target region (commonly a lesion or part of the healthy liver when human beings are involved)[3–5]. These parameters are an objective measurement that allows automation, though there can lack a connection between them and how IQ is perceived by the physician in some defined tasks (e.g. organ definition).

Beyond the reconstructed image, there is another set of IQ parameters derived from the count statics of a PET study. The measurement of the patient noise equivalent count rate (NECR) is a promising idea in order to predict IQ, providing a measure of the image count statistics corrected by the degrading scatter and random events. Several studies have reported a relationship between NECR and other IQ parameters[6–10], including IQ perception[11]. However, in everyday practice, clinical NECR is not used to optimize clinical protocols or to establish a minimum level of IQ for clinical trials, probably due to the high uncertainties involved in its IQ prediction.

[1]Medical Physics Department, Institut Català d'Oncologia, L'Hospitalet de Llobregat, Barcelona, Spain. [2]PET Unit. Nuclear Medicine Dept, IDI. Hospital U. de Bellvitge-IDIBELL, L'Hospitalet de Llobregat, Barcelona, Spain. [3]Medical Physics Department, Clínica Universidad de Navarra, Pamplona, Spain. Correspondence and requests for materials should be addressed to J.M.M.-C. (email: jmmartic@unav.es)

Characteristics	
Age (years), median (range)	66 (19–86)
Sex, no. (%)	
Male	58 (52%)
Female	54 (48%)
Referral reason, no. (%)	
Lung	25 (20%)
Gynecologic	25 (20%)
Colorectal	21 (19%)
Lymphoma	12 (11%)
Skin Cancer	9 (8%)
Head and Neck	5 (4%)
Unknown Primary	4 (3%)
Hepatobiliary	4 (3%)
Urologic	4 (3%)
Breast	3 (3%)

Table 1. Clinical characteristics of the studied population and referral reason for the PET/CT scan.

Another form of IQ assessment is via model-based tasks using automated models. Nevertheless, qualitative tasks are difficult to automate, as they involve a subjective human assessment, and models are usually limited to lesion detectability and conspicuity[12,13].

In recent years, one of the main gains in IQ comes from advances in reconstruction methods. The inclusion of point-spread function (PSF) modeling in the iterative methods supposed an improvement in terms of diagnostic performance, though the relation between count statistics and IQ increased in complexity[7]. Additionally, penalized reconstruction methods were recently introduced into clinical practice. In contrast to ordered subset expectation maximization (OSEM), block sequential regularized expectation maximization (BSREM) methods can run until full convergence while controlling noise levels[14–16]. The penalization acts as a selective filtering and the level of noise or IQ could be rather different than OSEM with PSF algorithms. The reliability of predicting IQ using the aforementioned assessment methods, such as the SNR or NECR, has never been tested on these new reconstruction algorithms.

Despite all factors that could affect IQ in PET studies, dosage optimization of the administered activity is usually calculated only in terms of patient weight[17]. Once the acquisition starts, the only relevant parameters that have a direct impact on image quantification are the acquisition time and image reconstruction settings, the latter being delicate to modify[18].

A fast and automated model to predict IQ could optimize the acquisition and reconstruction parameters in real time, or serve as a basic metric to compare acquisitions in multicentric studies. Thus, this task could benefit the emerging field of radiomics, which intends to extract and process a large number of quantitative features from radiological images[19]. Automated IQ evaluation using these methodologies has been developed for brain and liver magnetic resonance imaging (MRI)[20,21]; however, there is still a lack of research on this topic in nuclear medicine imaging.

The present study has two main aims, and hence the manuscript is divide in two parts: the first one is to investigate the influence of different biological and physical parameters on IQ perception of the abdominal area using new algorithms (OSEM + PSF and BSREM) and a modern PET scanner (Discovery IQ); the second one aims to test the utility of a radiomics approach in the first task. The study is focused on the abdominal region as the presence of different anatomic structures, sometimes with low SNR and definition, makes it a complex area to evaluate in PET studies.

Material and Methods

We obtained approval from the Bellvitge University Hospital Institutional Review Board. All work was done in accordance with institutional guidelines and regulations. This manuscript has been revised for its publication by the Clinical Research Ethics Committee of Bellvitge University Hospital. Written informed consent was waived by this Committee, as it was a retrospective analysis of our usual everyday work. The data of the patients were anonymized for the purposes of this analysis. The confidential information of the patients was protected according national normative.

Patient selection, image acquisition and reconstruction. A total of 112 patients were retrospectively included. Patients were selected sequentially from torso oncological FDG PET/CT studies; a detailed description of its referral reason can be seen on Table 1. Exclusion criteria were: a blood glucose level higher than 200 mg/dl, an uptake time outside the range of 60–100 min after FDG injection, and any abnormal condition such as artifacts or lesions which averted a correct evaluation of the abdominal region.

PET/CT acquisitions were performed according to the EANM 2.0 guidelines[17]. Patients were injected with 2.7 MBq/kg and scanned at 2 min/bed position. All data were acquired on a Discovery IQ 5-ring PET/CT[22] (GE Healthcare, Waukesha). Mean injected activity was 194 MBq (range 97–374 MBq) and mean uptake time was 68 min (range 60–98 min). Overlap between beds was 19%.

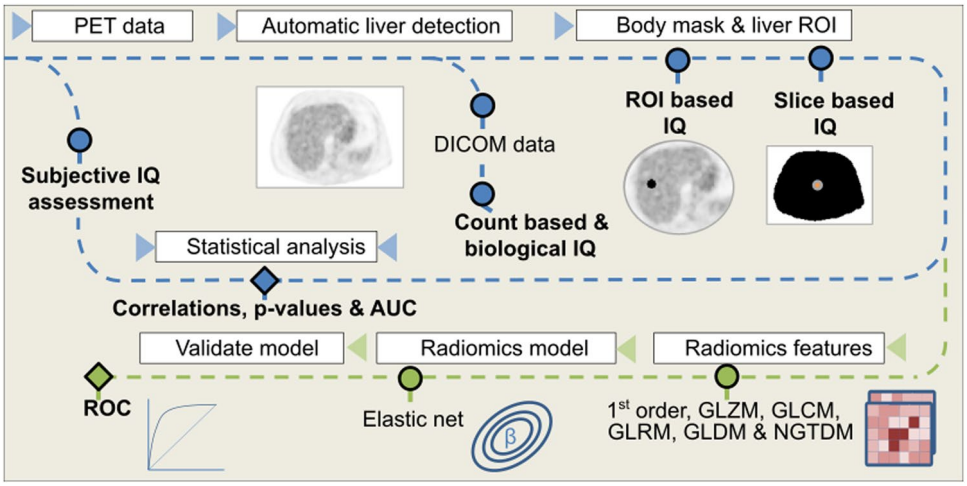

Figure 1. All image quality features were extracted and processed using an automatic pipeline. Blue line describes the first phase of the methodology: image is converted to SUV units and an automatic algorithm detects the slice including more liver parenchyma. Then, all DICOM data are extracted from the bed corresponding to this slice and a region of interest is placed on the liver to extract ROI-based image quality metrics. From a body mask, all slice-based image quality parameters are extracted. The green line describes the second phase: all common radiomics features are also extracted from the selected slices, as well as from its surrounding volume. Next, an elastic-net model is fitted selecting the relevant features. Results are compared in both lines with the subjective assessment.

Biological	Count related	Mixed
Pre-image reconstruction		
Age, glucose level, body weight, body height, BMI, LBM	Activity at scan time, true count rate, random count rate, scatter rate, NECR, PNECR	Uptake time, R_{DW}, R_{DBMI}, R_{DLBM}
Post-image reconstruction		
Noise related	**Radiomics**	**Others**
$Variance_{ROI}$, SNR_{ROI}, $Variance_{Slice}$, SNR_{Slice} and CNR	1st order, GLCM, GLZM, GLRM, GLDM, NGTDM	Patient position misplacement (center shift)

Table 2. Pre- and post-image reconstruction considered IQ parameters.

Two different reconstructions were used: an OSEM iterative reconstruction with modeling PSF (OSEM + PSF), commercial name VUE Point HD-Sharp (VPHD-S, GE Healthcare, Waukesha), using 12 subsets, 4 iterations and a 4.8 Gaussian post-filtering; and a BSRM penalized algorithm with PSF correction, Q.Clear (GE Healthcare, Waukesha), using a β value of 350, which is a validated penalization value for torso oncological examinations[16]. Both algorithms used an image matrix of 256 × 256 and CT based attenuation correction, as well as dead time, random, and scatter events corrections.

Subjective image quality evaluation. Images were transferred to a dedicated review platform (AW Server 2.0) (GE Healthcare, Waukesha). IQ perception was evaluated by two different expert nuclear medicine physicians; both rankers had more than two years of clinical experience using the BSRM and OSEM + PSF reconstructions in the PET/CT system. Figure 1 summarizes the workflow for the extraction and processing of all data. Physicians were asked to evaluate the IQ of the axial slices of the abdominal area (IQ_{ABD}) considering the conspicuity of the structures and the apparent noise. The score was ranked from 1 to 5 (1 non-diagnostic IQ, 2 poor IQ for diagnosis, 3 acceptable IQ but could lead to some undetermined judgment, 4 good IQ, and 5 excellent IQ). Moreover, all images were visualized in a randomized order mixing both reconstructions. The IQ_{ABD} was also grouped between low diagnostic quality (LQ) (1–3.5 score) and high diagnostic quality (HQ) (>3.5 score) to obtain a binary problem.

Image quality features extraction. All data were processed with an in-house software programmed using Python 3.7 that automatically detected the bed containing more liver parenchyma, which was defined as the abdominal bed. As the Discovery IQ has an axial field-of-view (FOV) of 26 cm, it is a reasonable assumption that a single bed will include a major part of the liver. Table 2 presents all studied variables. All parameters were obtained from the data available in the DICOM header and from the image.

First, the image was loaded and converted to SUV units (Fig. 1). The slice containing the most liver parenchyma was automatically detected by using some heuristics on the suspected position range and the expected SUV values from healthy liver. More details of the used method can be found on supplemental data. Once the slice was defined, all count data and patient (biological) related data were extracted from the corresponding slice DICOM data. Additionally, an automatic region-of-interest (ROI) was placed in the healthy liver to account for SUV variance and SNR.

Next, a segmentation of the patient body in the liver slice was performed using a thresholding method followed by a morphological processing, which provides a mask used to perform all non-ROI based measurements. The same mask was used to fit the minimum circle around the abdominal surface and find the patient position misplacement (center shift).

Biological parameters that could potentially affect IQ included the age of the patient, glucose level at the injection time, patient height and weight, and uptake time. Body mass index (BMI) and lean body mass (LBM) were also computed, the second one defined as recommended by EANM 2.0 guidelines[17], according to Janmahasatian equation[23], which depends on patients' sex.

NECR was computed directly over the total prompts, the random events, and the scatter factor extracted from manufacturer's data inside the relevant DICOM tags, using the formulation provided by the NEMA standards[24], defined as

$$NECR = \frac{T^2}{T + S + R}. \tag{1}$$

No additional corrections were used over these data, as could be the extraction of all count outside the body[6,9]. A metric closely related to the NECR was also used, called pseudo-NECR (PNECR) which was directly obtained from the sinogram and proportional to the NECR[6], and defined as

$$PNECR = \frac{(T + S)^2}{T + S + R/2}. \tag{2}$$

An additional set of mixed parameters that combine count and biological parameters was considered. The ratios between activity at the acquisition start time and patient weight, BMI and LBM were also computed, defined as R_{DW}[11], R_{DBMI}[11] and R_{DLBM}, respectively.

The mean value and variance were measured in the healthy liver ROI and in the body mask. SNR was measured dividing the mean value and standard deviation. CNR was measured using the ROI mean value and the mean and standard deviation of the mask. Lastly, from the body mask, the minimum surrounding circumference was extracted to compute the patient positioning shift (center shift, to abbreviate).

Radiomics features were extracted using the pyradiomics package[25] from the same body mask described in the previous paragraphs. Moreover, the mask was extended to the two consecutive slices in both cranial and caudal directions to obtain a 3-dimensional mask, which will be referred as zone features. The extracted features are defined in compliance with feature definitions as described by the Imaging Biomarker Standardization Initiative (IBSI)[26]. A fixed bin number of 64 was used for feature extraction, employed in previous studies showing good reproducibility[27].

Investigations using standard parameters. All parameters listed in Table 2, excluding the radiomics features, were correlated to the IQ_{ABD} for both algorithms. Furthermore, this metrics were also compared using a two tailed Wilcoxon signed-rank test. Next, we studied the predictive power of all relevant parameters. All data was randomly split in a training ($n = 73$) and a test ($n = 39$) set. For each statistically significant value ($p\text{-value} < 0.05$) a logistic regression was fitted on the train data. Predicted IQ_{ABD} was computed for the test and train datasets, and the area under the curve (AUC) was obtained from the Receiver Operating Characteristic (ROC) curve.

Building a Predictive Radiomics Model. To build a radiomics model to assess the IQ perception, the same train and test datasets were used. All Table 2 parameters were initially included in the radiomics model. This model consists of all IQ related parameters, as well as the texture parameters from the 1st order statistics (19 features), Gray Level Co-occurrence Matrix (GLCM, 24 features), Gray Level Size Zone Matrix (GLZM, 16 features), Gray Level Run Length Matrix (GLRM, 16 features), Gray Level Dependence Matrix (GLDM, 14 features), and Neighbouring Gray Tone Difference Matrix (NGTDM, 5 features). Non-normal features were log transformed. All features were standardized, by subtracting to each value the variable mean and dividing by the standard deviation. As the number of patients is limited, we used an algorithm to perform a feature reduction. First, Spearman's rank correlation coefficients were calculated to examine the internal correlation between individual features. Redundant features with linear correlation coefficients >0.95 were removed. Then, an elastic-net feature selection approach and model building was adopted, which is a combination of the least absolute shrinkage selection operator (LASSO) and the Ridge Regression, and is suitable for the regression of high-dimensional data[28]. The LASSO shrinks all regression coefficients towards zero to set the coefficients of non-contributing features to exactly zero. To find an optimal penalization terms, a ten-fold cross validation with minimum criteria was used in the elastic-net parameter tuning. The retained features with non-zero coefficients were used for regression model fitting and combined into a radiomics signature. Different models were computed for the OSEM and BSRM algorithms.

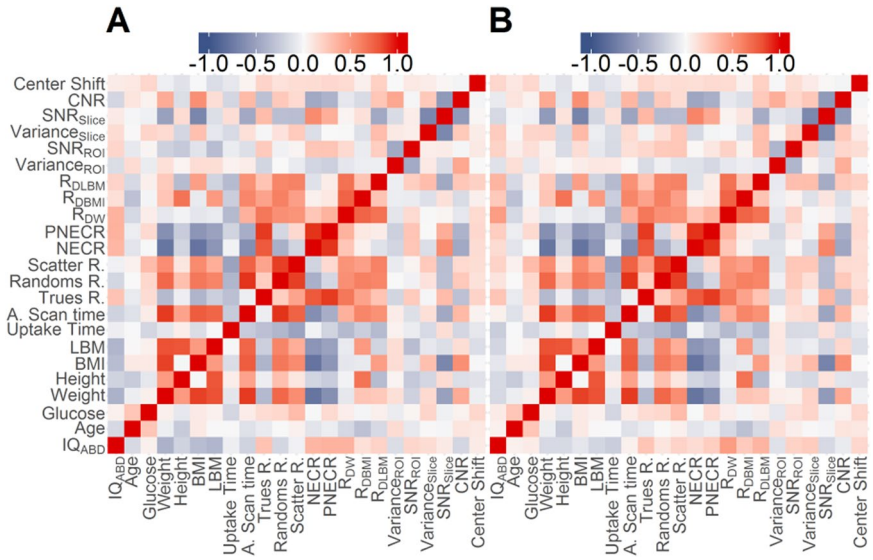

Figure 2. Pearson's r correlations between all studied variables for OSEM + PSF (A) and BSRM (B) reconstructions. Statistically significant correlations ($p < 0.05$) are the ones with $|r| > 0.2$.

The performance of the model was reported using the ROC methodology and AUC values in the training and test sets. The regression and its validation were performed using the R software version 3.4.4 (the R foundation) through the caret and glmnet packages.

Results

Mean IQ_{ABD} was 3.0 ± 0.8 and 3.2 ± 0.8 for the OSEM + PSF and BSRM reconstructions, respectively ($p = 0.006$ using a paired t-test). Weighted Cohen's kappa coefficient between rankers was 0.46. Figure 2 presents the correlation matrix between all studied variables. For the OSEM + PSF the three IQ parameters presented only at best moderated correlations. The highest correlation coefficients were found with patient weight ($r = -0.574$), LBM ($r = -0.48$), BMI ($r = -0.41$), activity at scan time ($r = -0.37$), and NECR ($r = 0.37$). For the BSRM algorithm, correlations between variables and IQ_{ABD} score were lower: R_{DW} ($r = 0.43$), weight ($r = -0.24$), and LBM ($r = -0.22$). For NECR the correlation was also reduced ($r = 0.12$). On the other hand, despite some observable degree of heteroscedasticity in the data, there was a clear positive correlation between NECR and SNR^2 ($r = 0.54$ and 0.56, for the OSEM + PSF and BSRM algorithms, respectively).

Supplemental Fig. 1 shows the relation of NECR with BMI, and Supplemental Fig. 1 the relation of SNR^2_{Slice} with NECR. It should be noted that R_{DW} presented a highly non-normal distribution of values, and hence the validity of the regression coefficient is limited. Figures 3, 4 present the relation between IQ_{ABD} and some selected variables for both reconstructions.

For the discretized analysis, in the case of the OSEM + PSF reconstruction, the parameters that present statistically significant differences ($p < 0.05$) are patient LBM ($p = 0.0005$), R_{DW} ($p = 0.007$), weight ($p = 0.001$), height ($p = 0.003$), CNR ($p = 0.01$), BMI ($p = 0.02$), NECR ($p = 0.02$), and PNECR ($p = 0.04$). For BSRM the parameters which present lower p-values are patient R_{DW} ($p = 0.006$), height ($p = 0.03$), CNR ($p = 0.04$), and LBM ($p = 0.05$). For further details, see Supplemental Table 1.

Table 3 shows the AUC obtained from fitting a logistic regression to each statistically significant variable for the train and test dataset. For all parameters, the OSEM + PSF reconstruction presented higher AUC values than the BSRM reconstruction, and from both reconstruction methods, the R_{DW} followed by the LBM parameters presented the highest AUC values for the test data.

Regarding the radiomics model, Supplemental Fig. 1 presents the parameter tuning of the elastic-net model. The resulting ROC can be seen in Fig. 5 for the test and train datasets. The resulting AUC is greater for the OSEM + PSF compared to the BSRM reconstruction. Also, for both algorithms, the radiomics AUC values were higher than the single parameter logistic regressions. The model selected variables and their importance are shown in Supplemental Fig. 4.

Discussion

The assessment of PET IQ is a complex task, as is highly subjective and depends on many different parameters. This study demonstrates its difficulty when single parameters are used, and aims to point the necessity of adopting an alternative model, as could be a radiomics model, especially when considering the increasing tendency of using penalized algorithms in modern systems. Our work shows how when using modern reconstruction algorithms and clinical acquisition settings most common single parameters are not correlated with the evaluation of IQ by physicians.

Research methodology in the present study is similar to that of Queiroz *et al.*[11], and in a similar fashion, we found an expected relation between NECR and SNR^2. Even if our settings were different, especially, to the best of our knowledge, this is the first study evaluating the relationship between NECR and IQ using a BSRM

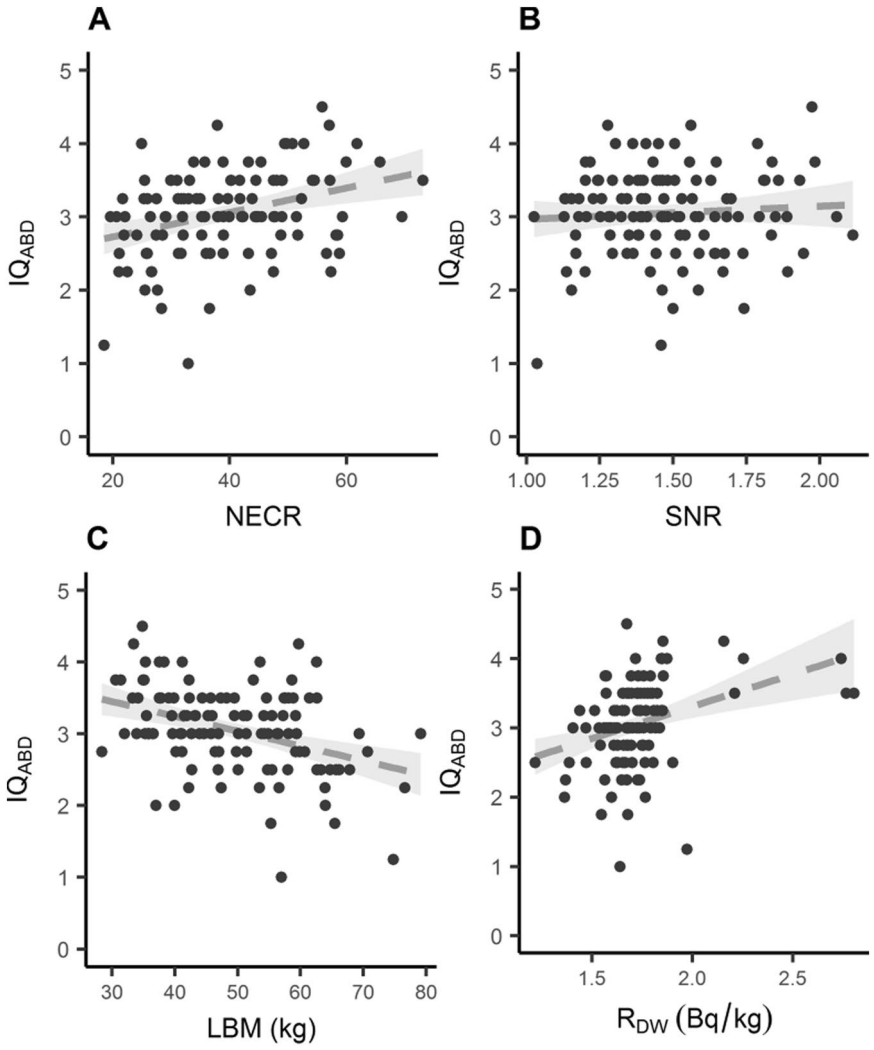

Figure 3. Selected relevant studied variables for the OSEM + PSF (VPDH-S) reconstruction method. The dotted line represents an adjusted linear regression and its 95% confidence interval.

reconstruction algorithm. Despite all this, we did not find any relevant relation between the NECR and IQ scores, particularly for the BSRM reconstruction method. A possible explanation is that our work is restricted to patients with an uptake time below 100 min, instead of the 128.3 min in average in the mentioned paper. Increasing the uptake time could increase the range of NECR values. Our purpose was to use a clinical relevant setting and we restricted our data accordingly. Furthermore, our injected activity is 2.7 MBq/kg instead of 4.3 MBq/kg of Queiroz *et al.*

In accordance to previous publications[15,22], in our study IQ was also ranked higher for BSRM than for OSEM + PSF. However, when comparing IQ scores with most IQ parameters, lower correlations and higher p-values were found for the BSRM algorithm. This is partially explained by the higher and less variable IQ_{ABD} scores, which limit a possible correlation. Yet, the non-linear reconstruction possibly dismisses the effect of external causes in IQ. The single parameter presenting a higher AUC value for the BSRM algorithm is R_{DW}. This result must be taken with caution though, as it presents a non-uniform distribution, as can be seen in the linear regression figures. Thus, the good results in predicting HQ and LQ images could be due a discretization effect, so further work should be performed to confirm its utility. Aside from R_{DW}, among all studied parameters, LBM is the only one that shows a lower p-value and a higher AUC in both reconstruction methods.

Most publications about PET IQ using clinical data only focus in lesion conspicuity[3–5], although there are other independent diagnostic tasks. When dealing with the abdominal zone, parameters such as SNR (extracted from a ROI in the healthy liver) are often used as a measure of IQ[4], but according to our results, they may have limited value differentiating between LQ and HQ images.

In contrast, we present a simple radiomics model as a proof of concept that a different paradigm can be applied on IQ evaluation, increasing the AUC presented by any single parameter. The present model has several limitations, though. First, this is a retrospective study with a relatively small sample size, even if an independent validation cohort from our institution was used. In the future, a large-scale multicenter study would be convenient to fully assess the generalization ability of the model. Second, it uses extracted data from a single slice

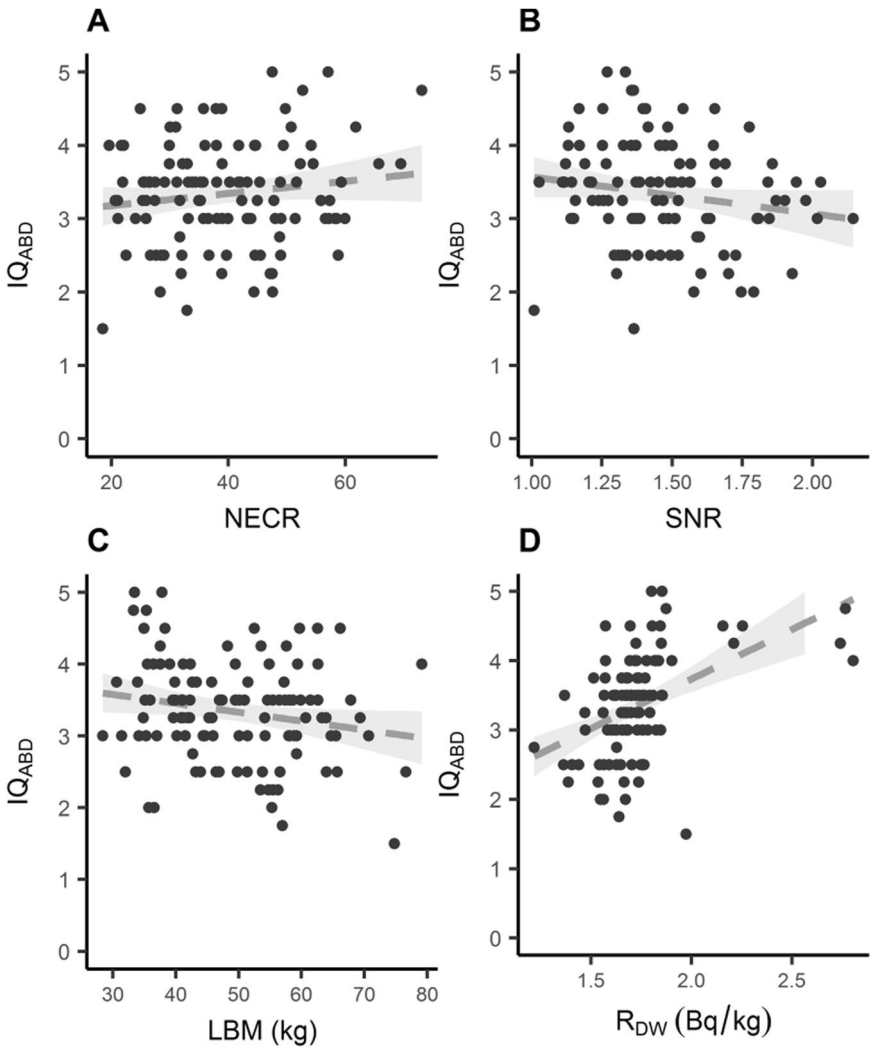

Figure 4. Selected relevant studied variables for the BSRM (Q.Clear) reconstruction method. The dotted line represents an adjusted linear regression and its 95% confidence interval.

	OSEM + PSF		BSRM	
	Train	Test	Train	Test
Weight (kg)	0.67 (0.50–0.84)	0.64 (0.51–0.78)	0.57 (0.38–0.76)	0.58 (0.44–0.71)
Height (cm)*	0.56 (0.36–0.75)	0.59 (0.44–0.73)	0.51 (0.32–0.70)	0.62 (0.48–0.75)
BMI (kg/m^2)	0.67 (0.50–0.84)	0.64 (0.49–0.79)	0.56 (0.37–0.75)	0.55 (0.42–0.69)
LBM (kg)*	0.63 (0.44–0.82)	0.65 (0.50–0.80)	0.51 (0.33–0.72)	0.60 (0.47–0.73)
NECR	0.74 (0.57–0.91)	0.65 (0.51–0.78)	0.62 (0.44–0.79)	0.56 (0.42–0.69)
PNECR	0.76 (0.58–0.94)	0.66 (0.53–0.80)	0.60 (0.42–0.78)	0.59 (0.47–0.73)
R_{DW}*	0.75 (0.60–0.91)	0.73 (0.61–0.86)	0.75 (0.59–0.91)	0.64 (0.59–0.78)
CNR*	0.54 (0.36–0.73)	0.63 (0.48–0.77)	0.58 (0.39–0.76)	0.53 (0.39–0.66)

Table 3. Calculation of the AUC and the 95% confidence interval for all significant variables when using the OSEM + PSF algorithm. Variables that were also significant for the BSRM reconstructions are marked with an asterisk.

and the surrounding slices as different inputs, although the evaluation was performed in the entire abdominal area. Despite the abdominal slice selected was manually verified, in order to increase the number of slices better algorithms for detecting the abdominal area should be applied, as a miss-selection of the abdominal zone could include undesired structures (such as the heart), which could potentially affect any feature values. Third, the model uses an elastic-net algorithm, but other more sophisticated models, such as neural networks could be applied[21]. Moreover, it would be interesting to mix different reconstruction methods and settings in the same model, although that would require a completely different study design, out of the present scope. Lastly, some

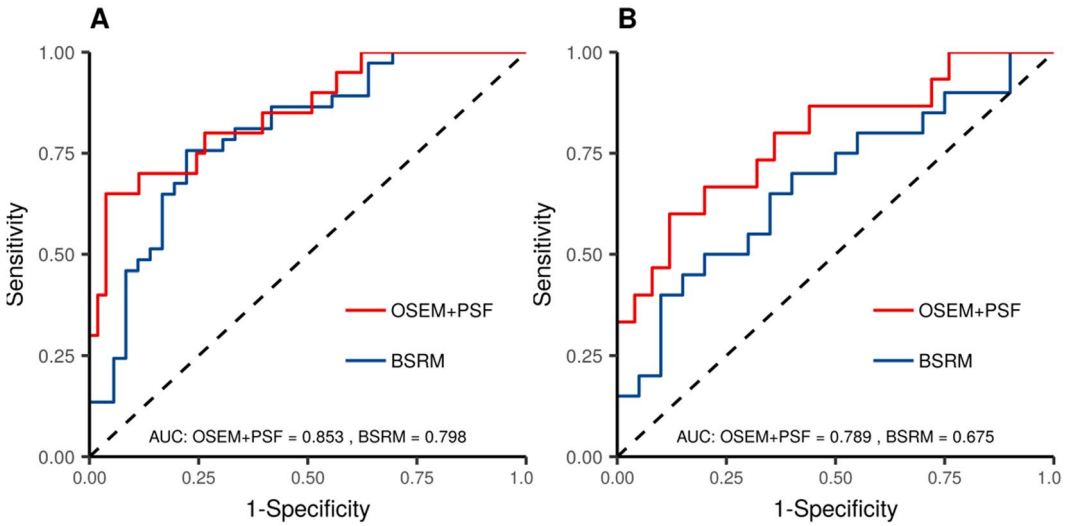

Figure 5. Model performance by means of a ROC curve for (**A**) train and (**B**) test datasets.

radiomics features could have a direct interpretation in terms of some IQ traits, such as lesion conspicuity or structure definition. Even if we have treated the model as a black box, it will be still useful to interpret the relation of each radiomics feature with a specific aspect of IQ. It should be noted that to achieve this goal, a current limitation of the present approach is the difficulty to obtain higher correlations between IQ rankers, the present study shows a rather moderate correlation, more work should be done extending redesign the study to include more rankers, ideally from different institutions.

Furthermore, PET IQ is potentially dependent on many pre-imaging parameter conditions[17], some of which were considered in the present manuscript (i.e. glucose level or uptake-time), but others are difficult to quantify, as could be other metabolic conditions. Additionally, PET imaging has the possibility to modify IQ by changing the acquisition time or reconstruction settings. Beyond the clear advantages of obtaining an objective IQ score, an IQ radiomics model could be performed during the PET scan, by applying a fast OSEM reconstruction during the acquisition, and modifying the duration of the scan or reconstruction settings according to the results. Moreover, our methods are easily extensible to other anatomical areas, such as the brain, where a correct definition of the structures could be of special importance for multicentric clinical trials[20].

Conclusion

The present work is a first step to a comprehensive analysis of the abdominal area IQ, pointing the necessity of changing the standard parameters to evaluate IQ, particularly when a BSRM algorithm is involved. Moreover, the promising role of a radiomics approach to assess IQ has been investigated, and according to our results a simple model can outperform the use of any single parameter.

References

1. Berg, E. & Cherry, S. R. Innovations in instrumentation for positron emission tomography. *Semin. Nucl. Med.* (2018).
2. Fukukita, H. *et al.* Japanese guideline for the oncology FDG-PET/CT data acquisition protocol: Synopsis of Version 2.0. *Ann. Nucl. Med.* **28**, 693–705 (2014).
3. Lois, C. *et al.* An Assessment of the Impact of Incorporating Time-of-Flight Information into Clinical PET/CT Imaging. *J. Nucl. Med.* **51**, 237–245 (2010).
4. Yan, J., Schaefferkoette, J., Conti, M. & Townsend, D. A method to assess image quality for Low-dose PET: Analysis of SNR, CNR, bias and image noise. *Cancer Imaging* **16**, 1–12 (2016).
5. Amakusa, S. *et al.* Influence of region-of-interest determination on measurement of signal-to-noise ratio in liver on PET images. *Ann. Nucl. Med.* **32**, 1–6 (2018).
6. Carlier, T. *et al.* Clinical NECR in 18F-FDG PET scans: Optimization of injected activity and variable acquisition time. Relationship with SNR. *Phys. Med. Biol.* **59**, 6417–6430 (2014).
7. Chang, T. *et al.* Reliability of predicting image signal-to-noise ratio using noise equivalent count rate in PET imaging. *Med. Phys.* **39**, 5891–5900 (2012).
8. Karakatsanis, N. A., Fokou, E. & Tsoumpas, C. Dosage optimization in positron emission tomography: state-of-the-art methods and future prospects. *Am. J. Nucl. Med. Mol. Imaging* **5**, 527–47 (2015).
9. Watson, C. C. *et al.* Optimizing injected dose in clinical PET by accurately modeling the counting-rate response functions specific to individual patient scans. *J. Nucl. Med.* **46**, 1825–34 (2005).
10. Mizuta, T. *et al.* NEC density and liver ROI S/N ratio for image quality control of whole-body FDG-PET scans: Comparison with visual assessment. *Mol. Imaging Biol.* **11**, 480–486 (2009).
11. Queiroz, M. A., Wollenweber, S. D., von Schulthess, G., Delso, G. & Veit-Haibach, P. Clinical image quality perception and its relation to NECR measurements in PET. *EJNMMI Phys.* **1**, 103 (2014).
12. Schaefferkoetter, J. D., Yan, J., Townsend, D. W. & Conti, M. Initial assessment of image quality for low-dose PET: Evaluation of lesion detectability. *Phys. Med. Biol.* **60**, 5543–5556 (2015).
13. Wangerin, K. A. *et al.* Effect of 18F-FDG Uptake Time on Lesion Detectability in PET Imaging of Early-Stage. *Breast Cancer. Tomography* **1**, 53–60 (2015).
14. Ahn, S. & Fessler, J. A. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imaging* **22**, 613–626 (2003).

15. Teoh, E. J., McGowan, D. R., Macpherson, R. E., Bradley, K. M. & Gleeson, F. V. Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System. *J. Nucl. Med.* **56**, 1447–1452 (2015).
16. Reynés-Llompart, G. *et al*. Phantom, clinical, and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q. Clear) on a BGO PET/CT scanner.
17. Boellaard, R. *et al*. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur. J. Nucl. Med. Mol. Imaging* **42**, 328–354 (2015).
18. Boellaard, R., Krak, N. C., Hoekstra, O. S. & Lammertsma, A. A. Effects of noise, image resolution, and ROI definition on the accuracy of standard uptake values: a simulation study. *J. Nucl. Med.* **45**, 1519–1527 (2004).
19. Lambin, P. *et al*. Radiomics: The bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
20. Esteban, O. *et al*. MRIQC: Predicting Quality in Manual MRI Assessment Protocols Using No-Reference Image Quality Measures. *bioRxiv* 1–18, https://doi.org/10.1101/111294 (2017).
21. Esses, S. J. *et al*. Automated image quality evaluation of T2-weighted liver MRI utilizing deep learning architecture. *J. Magn. Reson. Imaging* **47**, 723–728 (2018).
22. Reynés-Llompart, G. *et al*. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. *J. Nucl. Med.* **58**, 1155–1161 (2017).
23. Janmahasatian, S. *et al*. Quantification of lean bodyweight. *Clin. Pharmacokinet.* **44**, 1051–1065 (2005).
24. *National Electrical Manufacturers Association. Performance measurements of positron emission tomographs. NEMA Standards Publication NU 2-2012* (2012).
25. Van Griethuysen, J. J. M. *et al*. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).
26. Zwanenburg, Alex and Leger, Stefan and Vallières, Martin and Löck, S. and others. Image biomarker standardisation initiative. *arXiv Prepr.* arXiv1612.07003, https://doi.org/10.17195/candat.2016.08.1 (2016).
27. Presotto, L. *et al*. PET textural features stability and pattern discrimination power for radiomics analysis: An "ad-hoc" phantoms study. *Phys. Medica* **50**, 66–74 (2018).
28. Society, R. S. Regression Shrinkage and Selection via the Lasso Author (s): Robert Tibshirani Source: Journal of the Royal Statistical Society. *Series B (Methodological), Vol. 58, No. 1 Published by: Blackwell Publishing for the Royal Statistical Society Stabl.* **58**, 267–288 (2010).

Author Contributions

G.R.L. participated in the design of the study, carried out the technical data acquisition, in-house programs, data analysis, and drafted the manuscript. A.S.L. participated in the design of the study and the subjective image quality data acquisition, data manipulation, and critical review of the manuscript. E.L.T. participated in the subjective image quality data acquisition. J.M.M.C. participated in the design of the study, and was responsible for the technical support, data interpretation, and critical review of the manuscript. C.G.C. was responsible of a critical review of the manuscript. G.R.L., A.S.L., J.M.C. and C.G.C. were responsible for the study concept and design. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-46937-8.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Chapter 7

General discussion

The main goal of this thesis was to study the impact of a new reconstruction algorithm on the image quality and quantification of clinical PET studies performed with a high sensitivity BGO PET scanner. To do so we started with Aim 1, by studying the PET imaging characteristics of the D-IQ-5R PET/CT. As it was a new model, the study of the system performance it is necessary in order to establish a ground base to compare any result with other previous research based on a different PET/CT system. We followed with Aims 2 and 3, which are the core of the present dissertation, as they solve the image optimization problem under the penalized reconstruction algorithm. Aim 2 studies how as different reconstruction parameters affect the quantification accuracy of different lesion sizes, as well as their impact on the image quality evaluation on real patients. Aim 3 moves beyond classical PET quantification to study the impact of different reconstruction settings on lesion heterogeneity and morphology measurements in form of texture indices. Finally, Aim 4 tries to account for the impact of Q.Clear reconstruction on different image quality metrics and its relation to the subjective image quality perception.

In Aim 1, which is mainly covered in Chapter 4, we investigated the different performance characteristics of the D-IQ-5R, primarily based on the NEMA standards. Yet we extended the tests in some aspects, especially referring to image quality, as the BSRM reconstruction algorithm is the most innovative aspect in comparison to the older BGO-based PET systems.

Beyond this, from a physical point of view, the sensitivity gain could be considered one of the main advantages of the D-IQ-5R, compared to the other PET/CT systems currently available. As an example, the Siemens mCT Flow [34], which is a TOF PET/CT that could be considered competing at a similar market segment, has a sensitivity of 7.6 cps/kBq, a 33% of the D-IQ-5R one. This also holds for the high-end digital PET/CT systems; the GE Healthcare D-MI [33] has a sensitivity of 12.65 cps/KBq and the Philips Vereos [67] has a sensitivity of 5.7 cps/kBq, both values at the centre of the ring. On the other hand, the major drawback of the D-IQ could be considered the lack of TOF reconstruction, due to the BGO detector, which still presents some advantages in terms of a gain of reconstruction information [42]. Furthermore, the worse time resolution of BGO detectors could impact in image quality of more uncommon PET nuclides, such as ^{90}Y, which possess high contamination of non-desired photons that could increase the noise and would especially benefit from a TOF reconstruction [68]. In any case, the D-IQ is intended to be a multipurpose PET/CT, though its performance is optimized for ^{18}F radiopharmaceuticals, as other nuclides normally involve higher concentrations, i.e. ^{15}O exams. This can be seen in the NECR peak, which is located at 9.1 kBq/ml, closer to typical FDG concentrations. On the other hand, the mCT Flow has its NECR peak at 22.5 kBq/ml or the D-MI has its peak at 29.0 kBq/ml, which could be an advantage when higher radiopharmaceutical concentrations are used.

The image quality test is usually the most difficult to interpret of the NEMA tests, as it can not always be compared in a straightforward way with other scanners, due to differences in image processing and or acquisition time. Using the standard approach or a more realistic condition (with a worse combination of acquisition time and phantom activity concentration), the introduction of the Q.Clear reconstruction noticeably improves image quality, increasing the CR coefficients and decreasing the BV. These are consistent with previous results from the Discovery 960 scanner based on LYSO crystals [16], and posterior results using other settings [69–74]. Also the images of the Jaszczak and NEMA image quality phantoms reconstructed with the Q.Clear algorithm showed a higher contrast, with less apparent noise.

In the last section of Chapter 4, in order to translate the phantom image quality value to patient images, we studied different quantification parameters on lesions. One can observe that the differences

found between the Q.Clear reconstruction and the VPHD-S are not linear; even more, although in most lesions the BSRM algorithm tended to increase lesion quantification, in some of them it decreased. These last results directly link with the next aim of the present dissertation.

In Aim 2, which is mainly addressed in Chapter 5, we investigated the optimization of a penalization factor for a Bayesian reconstruction algorithm (Q.Clear) under different levels of PET image noise. In order to do so, we started by studying the behaviour of the Q.Clear algorithm on phantom data. As the gold standard of PET reconstructions are the different variations of the OSEM method, it is convenient to compare the Q.Clear algorithm behaviour to the expected for an OSEM reconstruction. However a direct comparison of both algorithms is not straightforward. On the one hand, by decreasing β values on the Q.Clear algorithm comparable quantitative metrics (measured as CR) as in an OSEM reconstruction can be achieved if there is an increase on the iterations or subsets, or a reduction of the total post-filtering. On the other hand, our study shows that the image quality improvement is not homogeneous, depending on the LBR and total administered activity. In general terms, as the β value increases, the quantification capacity (measured as CR and contrast) decreases, as well as the noise, leaving a total increase on detectability (in terms of CNR or SNR both for phantoms or patients). When comparing patient data, no relevant differences of noise measured in the liver were observed, although Q.Clear reconstructions presented a slightly lower value. For lesion quantification, the Q.Clear reconstruction increased lesions SUV mean and maximum. Moreover, when using the Q.Clear reconstruction algorithm in a BGO scanner, and combining phantom with patient data, a β value of 350 and 200 appears to be the optimal value for ^{18}F-FDG oncology and brain PET/CT, respectively. Our results show that for torso scans, this value assures a compromise in image quality and quantitation accuracy. However, the changes in PET quantification introduced by the algorithm selection and settings should be considered with care, especially when standardization and harmonization are required as is the case of multicenter studies [53].

In Aim 3, covered in Chapter 4, we investigated the direct impact of a BSRM reconstruction in the stability of heterogeneity features. Texture indices show the effect of lesion homogeneity as β value increases. From a practical point of view we found that most features are stable (COV < 10%) when using the recommended range of β values. However, that is not the case when the range of β

values is extended. Thus it will be important for any future studies seeking to use radiomics features from a BSRM reconstruction to keep their imaging protocols as uniform as possible.

In Aim 4, covered in Chapter 6, we studied how common quantitative parameters behave when used to measure clinical image quality on Q.Clear reconstructions. The study was limited to the abdominal area, and we were able to show that not a single commonly used parameter presented a good correlation with image quality perception. All correlations where lower in the BSRM reconstruction than in the OSEM + PSF one. This is an important landmark as it is crucial to modify how to evaluate image quality with the incorporation of new penalized reconstruction methods in clinical practice. When dealing with the abdominal area, parameters such as SNR (extracted from a ROI in the healthy liver) are often used as a measure of image quality [75], but according to our results, they may have limited value in distinguishing among different image quality studies. The last part of the chapter deals with a proof of concept of radiomics model, which, if extended, could be useful to solve this problem.

The last two aims involve the use of heterogeneity features. As explained during the background section 3.5, the majority of TI currently used in radiomics studies were designed to quantitatively evaluate visual perception characteristics in images, such as background contrast or some pattern of complexity. Despite the enthusiasm for radiomics and the exponentially growing number of studies using FDG PET heterogeneity prameters, there is still work to do in order to harmonize its methodological application. As an example, after the initial publications showing some kind of relation between TI and different tumour progressions, the first meta-analysis showed that none of the considered publications (18 studies from 2009 to 2013) presented enough statistical rigour to pass a minimum cut-off required to be considered in the meta-analysis [76], hence part of the methodology used by these pioneer studies should be discouraged in a modern project. Consequently, at the initial steps of answering Aim 2 texture analysis was an incipient discipline and many aspects of its computations were unknown, but by the time of working in the specific Aim 3, several other initiatives had emerged into the goal of standardizing all the methodology involving radiomics studies.

The evolution in the computation of radiomics features explains the two different approaches used during this dissertation. First, in Chapter 5, an in-house software was written using Python 3.6.

Developing our own software allowed a deep understanding of the different aspects that could affect the heterogeneity computation, as well as the caveats of different methodological approaches followed by other authors.

However, during the process of answering the subsequent specific aim (Aim 4), we decided to use an external package to compute all radiomics features, known as pyradiomics [63]. One advantage of using this external tool is that it includes more features than the initially planned in our in-house computations, but more important is its participation in an initiative to standardize the naming and computation of different features [77], allowing us to focus in the answer of the aims of the thesis without losing time in other factors. The software also works as an external library to Python 3.6, thus the changes in the general workflow were minimal. Additionally, the code is open-source, supporting the revision by the scientific community in search of possible computational errors.

7.1 Future Directions

The present study deals with the optimization of the Q.Clear penalization factor under different levels of PET image noise, but we restricted our study to a general oncological and neurological case. However, there is still work to do in order to find the optimal parameters for other cases, but more importantly to clinically evaluate the new reconstruction method in particular settings. Specifically, in the last years, Q.Clear has been proven of utility in different ^{18}F-FDG oncological settings, as in the detection of lung nodules [78], lung nodules by dual-time point [79], mediastinal lymph nodes in lung cancer [78] and colorectal liver cancer metastases [80]. However, there is some criticism on the use of Q.Clear for some settings. For example, Devriese et al. [81] studied the impact of Q.Clear reconstructions in the assessment of solid tumour response criteria (PERCIST), and they recommend the use of an harmonization reconstruction protocol (without PSF modelling and non-penalized) [53] as it overestimates true SUL values. Hence, there is a future direction of work to find a subset of Q.Clear penalization factors compatible with the EARL harmonization criteria.

Another controversy is in lymphoma assessment. In a letter to the editors, Barrington et al. [82] note that Q.Clear increases the SNR of lesions but do not modify the uptake in reference of the

mediastinum and liver (as we discussed previously), and hence its use can lead to different scoring of the Deauville criteria and response interpretation using the Lugano Classification. Therefore, there is a need to validate BSRM algorithms in lymphoma patients before its adoption in routine practice.

In general terms, when the goal is to detect a lesion with high LBR, Q.Clear demonstrated to be superior, and the present dissertation proposed a optimization criteria. However, when considering to use normal regions, small lesions in low LBR settings, or quantitative parameters involving ratios to normal structures, there is still work to perform for a clinical validation and optimization. These include some oncological cases as liver metastases, or non-oncological cases as could be the study of epilepsy focuses.

A point of further research stated in Chapter 4, is the possible benefit of using different optimization parameters on different anatomical areas, where count statistics and resolution demands are different. This necessity can be clearly seen on brain images, where the required β values are much lower than those that could be considered clinically acceptable on a torso acquisition. As a general rule, as image statistics, the β value can be decreased. Using this criteria, a future optimization can lead to whole-body acquisitions with different β values depending on the anatomical zone, clinical goal or count statistics in the same reconstruction. Furthermore, it could also be studied on dynamic tracer quantification, where yet any study evaluated the impact of Q.Clear reconstructions.

Another interesting area of future research is the definition of a set of new metrics to study the IQ performance of PET reconstructions. Furthermore, before applying different IQ criteria on clinical images, it will be beneficial to apply this metrics on phantom data to know how your algorithm in combination of your PET equipment behaves. This is performed in our study (Aim 2 & 3) but a set of definite parameters should be standardised in order to evaluate any new equipment when using a penalized algorithm. For example, in CT imaging a new protocol has been recently published proposing a series of performance metrics to evaluate image quality when iterative reconstruction algorithms are used in this modality [83]. It points the need to use some complex metrics as modulation transfer function (MTF) and task-based related parameters, in a similar fashion as shown in the present dissertation, as also most common parameters has a limited utility to quantify IQ of CT studies when using modern reconstruction algorithms. A similar work should be done for PET imaging, to compare

and harmonize image quality among different centres when penalized reconstruction algorithms are involved.

In addition, in our study we proposed a radiomics model that could serve as a final IQ evaluation. However, in its current state it just works as a proof of concept, and further work must be done in order to be clinically adopted. The most important step is a large-scale multicenter study, comparing its IQ assesment with more physician and a larger cohort of patients, allowing the use of more complex algorithms as neural nets. Moreover, the algorithm should be extended to different anatomical areas or clinical goals.

Lastly, beyond the clear advantages of obtaining an objective IQ score, an improved IQ radiomics model could be used during the PET scan, by applying a fast OSEM reconstruction during the acquisition, and modifying the duration of the scan or reconstruction settings on the fly according to a desired IQ.

Chapter 8

Conclusions

In the present work we studied how a new implemented BSRM reconstruction algorithm (Q.Clear) impacts on image quality and quantification of clinical PET studies performed with a high sensitivity BGO PET scanner.

The main results obtained are the following:

- The D-IQ PET/CT with a 5 ring block of detectors has the highest overall performance of the Discovery BGO-based scanners and of the rest of the current commercially available scanners that was compared with, showing improved sensitivity and count rate performance.
- Compared to standard OSEM reconstructions, Q.Clear increases contrast recovery and decreases background variability, producing an overall increase of contrast-to-noise ratio in phantom studies, and increasing the signal-to-noise ratio in patient lesions. Image quality assessment of torso and brain studies also improves: a β value of 350 and 200 appears to be the optimal value for ^{18}F-FDG oncology and brain PET/CT studies, respectively. For torso scans, this value assures a compromise in image quality and quantitation accuracy.
- Texture indices show the effect of lesion homogeneity as β value increases. In the oncological clinical proposed β range (between 300 and 450) almost all the texture indices presented coefficients of variance under a 10% threshold.

- None of the studied IQ parameters correlated with the clinical IQ assessment, pointing the necessity of changing the standard parameters to evaluate IQ, particularly when a penalized algorithm as Q.Clear is involved. Moreover, the promising role of a radiomics approach to assess IQ has been investigated, and according to our results a simple model can outperform the use of any single parameter.

References

[1] Ronald Nutt. The history of positron emission tomography early medical applications for the positron. *Molecular Imaging and Biology*, 4(1):11–26, 2002.

[2] Terry Jones and David Townsend. History and future technical innovation in positron emission tomography. *Journal of Medical Imaging*, 4(1):011013, 2017.

[3] Magdy M. Khalil. *Basic Science of PET Imaging*. Springer, 2016.

[4] Cristina Gámez-Cenzano and Francisco Pino-Sorroche. Standardization and quantification in FDG-PET/CT imaging for staging and restaging of malignant disease. *PET clinics*, 9(2):117–127, 2014.

[5] Ministerio de la Presidencia. Real decreto 1132/1997 por el que se establecen los criterios de calidad en medicina nuclear. *BOE*, 1997.

[6] Council Directive. 59/euratom of 5 december 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing directives 89/618/euratom, 90/641/euratom, 96/29/euratom, 97/43/euratom and 2003/122/euratom. *Official journal of the European Union*, 17(1):2014, 2013.

[7] J Valentin. Radiation dose to patients from radiopharmaceuticals:(addendum 2 to icrp publication 53) icrp publication 80 approved by the commission in september 1997. *Annals of the ICRP*, 28(3):1–1, 1998.

[8] Wesley E Bolch, Keith F Eckerman, George Sgouros, and Stephen R Thomas. Mird pamphlet no. 21: a generalized schema for radiopharmaceutical dosimetry—standardization of nomenclature. *Journal of Nuclear Medicine*, 50(3):477–484, 2009.

[9] Piotr J Slomka, Tinsu Pan, and Guido Germano. Recent advances and future progress in PET instrumentation. In *Seminars in Nuclear Medicine*, volume 46, pages 5–19. Elsevier, 2016.

[10] JM Martí-Climent, E Prieto, I Domínguez-Prado, MJ García-Velloso, M Rodríguez-Fraile, J Arbizu, C Vigil, C Caicedo, I Penuelas, and JA Richter. Contribution of time of flight and point spread function modeling to the performance characteristics of the PET/CT Biograph mCT scanner. *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)*, 32(1):13–21, 2013.

[11] V Bettinardi, L Presotto, E Rapisarda, M Picchio, L Gianolli, and MC Gilardi. Physical performance of the new hybrid PET/CT Discovery-690. *Medical physics*, 38(10):5394–5411, 2011.

[12] Jeffrey A Kolthammer, Kuan-Hao Su, Anu Grover, Manoj Narayanan, David W Jordan, and Raymond F Muzic. Performance evaluation of the ingenuity tf pet/ct scanner with a focus on high count-rate conditions. *Physics in medicine and biology*, 59(14):3843, 2014.

[13] Adam M Alessio, Charles W Stearns, Shan Tong, Steven G Ross, Steve Kohlmyer, Alex Ganin, and Paul E Kinahan. Application and evaluation of a measured spatially variant system model for pet image reconstruction. *IEEE Transactions on Medical Imaging*, 29(3):938–949, 2010.

[14] Johan Nuyts, D Beque, Patrick Dupont, and Luc Mortelmans. A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Transactions on Nuclear Science*, 49(1):56–60, 2002.

[15] Evren Asma, Sangtae Ahn, Steven G Ross, Anthony Chen, and Ravindra M Manjeshwar. Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images. In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2012 IEEE*, pages 4062–4066. IEEE, 2012.

[16] Eugene J Teoh, Daniel R McGowan, Ruth E Macpherson, Kevin M Bradley, and Fergus V Gleeson. Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system. *Journal of Nuclear Medicine*, 56(9):1447–1452, 2015.

[17] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2015.

[18] Mathieu Hatt, Florent Tixier, Larry Pierce, Paul E Kinahan, Catherine Cheze Le Rest, and Dimitris Visvikis. Characterization of PET/CT images using texture analysis: the past, the present: any future? *European Journal of Nuclear Medicine and Molecular Imaging*, 44(1):151–165, 2017.

[19] Tingting Chang, Guoping Chang, John W. Clark, Rami H. Diab, Eric Rohren, and Osama R. Mawlawi. Reliability of predicting image signal-to-noise ratio using noise equivalent count rate in PET imaging. *Medical Physics*, 39(10):5891–5900, 2012.

[20] T. Carlier, L. Ferrer, H. Necib, C. Bodet-Milin, C. Rousseau, and F. Kraeber-Bodéré. Clinical NECR in 18F-FDG PET scans: Optimization of injected activity and variable acquisition time. Relationship with SNR. *Physics in Medicine and Biology*, 59(21):6417–6430, 2014.

[21] Simon R Cherry and Magnus Dahlbom. PET: physics, instrumentation, and scanners. In *PET*, pages 1–117. Springer, 2006.

[22] Simon R Cherry, James A Sorenson, and Michael E Phelps. *Physics in Nuclear Medicine, 3rd edition*. Elsevier Health Sciences, 2012.

[23] William W Moses. Fundamental limits of spatial resolution in PET. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648:S236–S240, 2011.

[24] Pat Zanzonico. Positron emission tomography: a review of basic principles, scanner design and performance, and current systems. 34(2):87–111, 2004.

[25] Eric Berg and Simon R Cherry. Innovations in instrumentation for positron emission tomography. *Seminars in Nuclear Medicine*, 48(4):311–331, 2018.

[26] Jinyi Qi and Richard M Leahy. Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine & Biology*, 51(15):R541, 2006.

[27] Glenn F Knoll. *Radiation detectors and measurements*. New York: Wiley, 2000.

[28] Tom K. Lewellen. Recent developments in PET detector technology. *Physics in Medicine and Biology*, 53(17), 2008.

[29] David S, Georgiou M, Fysikopoulos E, and Loudos G. Evaluation of a SiPM array coupled to a $Gd_3Al_2Ga_3O_{12}$: Ce (GAGG:Ce) discrete scintillator. *Physics in Medicine and Biology*, 31(7):763–6, 2015.

[30] Jacobo Cal-Gonzalez, Ivo Rausch, Shiyam Sundar, Lalith Kumar, Martin Lyngby Lassen, Otto Muzik, Ewald Moser, Laszlo Papp, and Thomas Beyer. Hybrid imaging: Instrumentation and data processing. *Frontiers in Physics*, 6:47, 2018.

[31] Gabriel Reynés-Llompart, Cristina Gámez-Cenzano, Inmaculada Romero-Zayas, Laura Rodríguez-Bel, José L. Vercher-Conejero, and Josep M. Martí-Climent. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. *Journal of Nuclear Medicine*, 58(7):1155–1161, 2017.

[32] Alexander M. Grant, Timothy W. Deller, Mohammad Mehdi Khalighi, Sri Harsha Maramraju, Gaspar Delso, and Craig S. Levin. NEMA NU 2-2012 performance studies for the SiPM-based ToF-PET component of the GE SIGNA PET/MR system. *Medical Physics*, 43(5):2334–2343, 2016.

[33] David F.C. Hsu, Ezgi Ilan, William T. Peterson, Jorge Uribe, Mark Lubberink, and Craig S. Levin. Studies of a Next-Generation Silicon-Photomultiplier. *Journal of Nuclear Medicine*, (9):1511–1518.

[34] Ivo Rausch, Jacobo Cal-González, David Dapra, Hans Jürgen Gallowitsch, Peter Lind, Thomas Beyer, and Gregory Minear. Performance evaluation of the Biograph mCT Flow PET/CT system according to the NEMA NU2-2012 standard. *EJNMMI Physics*, 2(1):26, 2015.

[35] Joyce J van Sluis, Johan de Jong, Jenny Schaar, Walter Noordzij, Paul van Snick, Rudi Dierckx, Ronald Borra, Antoon Willemsen, and Ronald Boellaard. Performance characteristics of the digital Biograph Vision PET/CT system. *Journal of Nuclear Medicine*, page jnumed.118.215418, 2019.

[36] Charlotte S van der Vos, Daniëlle Koopman, Sjoerd Rijnsdorp, Albert J Arends, Ronald Boellaard, Jorn A van Dalen, Mark Lubberink, Antoon TM Willemsen, and Eric P Visser. Quantification, improvement, and harmonization of small lesion detection with state-of-the-art pet. *European Journal of Nuclear Medicine and Molecular Imaging*, 44(1):4–16, 2017.

[37] Rodney A Brooks, Victor J Sank, Alan J Talbert, and Giovanni Di Chiro. Sampling requirements and detector motion for positron emission tomography. *IEEE Transactions on Nuclear Science*, 26(2):2760–2763, 1979.

[38] Mikiko Ito, Seong Jong Hong, and Jae Sung Lee. Positron emission tomography (PET) detectors with depth-of-interaction (DOI) capability. *Biomedical Engineering Letters*, 1(2):70, 2011.

[39] Kjell Erlandsson, Irène Buvat, P Hendrik Pretorius, Benjamin a Thomas, and Brian F Hutton. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Physics in Medicine and Biology*, 57(21):R119–R159, 2012.

[40] Arman Rahmim, Jinyi Qi, and Vesna Sossi. Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls. *Medical Physics*, 40(6):064301, 2013.

[41] Mark van Heijl, Jikke M Omloo, Mark I van Berge Henegouwen, Jan J van Lanschot, Gerrit W Sloof, and Ronald Boellaard. Influence of ROI definition, partial volume correction and SUV normalization on SUV–survival correlation in oesophageal cancer. *Nuclear Medicine Communications*, 31(7):652–658, 2010.

[42] Takehiro Tomitani. Image reconstruction and noise evaluation in photon time-of-flight assisted positron emission tomography. *IEEE Transactions on Nuclear Science*, 28(6):4581–4589, 1981.

[43] Xuezhu Zhang, Jian Zhou, Simon R Cherry, Ramsey D Badawi, and Jinyi Qi. Quantitative image reconstruction for total-body pet imaging using the 2-meter long explorer scanner. *Physics in Medicine & Biology*, 62(6):2465, 2017.

[44] Paweł Kowalski, Wojciech Wiślicki, Roman Shopa, Lech Raczyński, Konrad Klimaszewski, Catalina Curceanu, Eryk Czerwiński, Kamil Dulski, Aleksander Gajos, Marek Gorgol, et al. Estimation of the nema characteristics of the j-pet tomograph using the gate package. *Physics in medicine and biology*, 2018.

[45] Gengsheng Lawrence Zeng. *Medical image reconstruction: a conceptual tutorial.* Springer, 2010.

[46] Lawrence A Shepp and Yehuda Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.

[47] Kenneth Lange, Richard Carson, et al. Em reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–16, 1984.

[48] Stefaan Vandenberghe, YVES D'Asseler, Rik Van de Walle, T Kauppinen, Michel Koole, Luc Bouwens, Koen Van Laere, Ignace Lemahieu, and RA Dierckx. Iterative reconstruction algorithms in nuclear medicine. *Computerized medical imaging and graphics*, 25(2):105–111, 2001.

[49] H Malcolm Hudson and Richard S Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609, 1994.

[50] Ronald Boellaard, Arthur Van Lingen, and Adriaan A Lammertsma. Experimental and clinical evaluation of iterative reconstruction (OSEM) in dynamic PET: quantitative characteristics and effects on kinetic modeling. *Journal of Nuclear Medicine*, 42(5):808–817, 2001.

[51] Eugene Veklerov and Jorge Llacer. Stopping rule for the mle algorithm based on statistical hypothesis testing. *IEEE Transactions on Medical Imaging*, 6(4):313–319, 1987.

[52] Valen Earl Johnson. A note on stopping rules in em-ml reconstructions of ect images. *IEEE Transactions on Medical Imaging*, 13(3):569–571, 1994.

[53] Andres Kaalep, Terez Sera, Wim Oyen, Bernd J Krause, Arturo Chiti, Yan Liu, and Ronald Boellaard. Eanm/earl fdg-pet/ct accreditation-summary results from the first 200 accredited imaging systems. *European Journal of Nuclear Medicine and Molecular Imaging*, 45(3):412–422, 2018.

[54] Charline Lasnon, Mohamed Majdoub, Brice Lavigne, Pascal Do, Jeannick Madelaine, Dimitris Visvikis, Mathieu Hatt, and Nicolas Aide. 18 F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer. *European Journal of Nuclear Medicine and Molecular Imaging*, 43(13):2324–2335, 2016.

[55] Erkan Ü Mumcuoglu, Richard M Leahy, and Simon R Cherry. Bayesian reconstruction of pet images: methodology and performance analysis. *Physics in Medicine & Biology*, 41(9):1777, 1996.

[56] Sangtae Ahn and Jeffrey A Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. 2003.

[57] Sangtae Ahn. *Convergent algorithms for statistical image reconstruction in emission tomography*. University of Michigan, 2004.

[58] Evren Asma, Sangtae Ahn, Hua Qian, Girishankar Gopalakrishnan, Kris Thielemans, Steven G Ross, Ravindra M Manjeshwar, and Alexander Ganin. Quantitatively Accurate Image Reconstruction for Clinical Whole-Body PET Imaging. 2012.

[59] Steve Ross. Q.Clear white paper. *GE Healthcare*, 2015.

[60] Michael C Adams, Timothy G Turkington, Joshua M Wilson, and Terence Z Wong. A systematic review of the factors affecting accuracy of suv measurements. *American Journal of Roentgenology*, 195(2):310–320, 2010.

[61] Ronald Boellaard, Nanda C Krak, Otto S Hoekstra, and Adriaan A Lammertsma. Effects of noise, image resolution, and roi definition on the accuracy of standard uptake values: a simulation study. *Journal of Nuclear Medicine*, 45(9):1519–1527, 2004.

[62] Robert M Haralick, K Shanmugam, Its'Hak Dinstein, et al. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[63] Joost JM van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.

[64] Fanny Orlhac, Michaël Soussan, Jacques-Antoine Maisonobe, Camilo A Garcia, Bruno Vanderlinden, and Irène Buvat. Tumor texture analysis in 18f-fdg pet: relationships between texture parameters, histogram indices, standardized uptake values, metabolic volumes, and total lesion glycolysis. *J Nucl Med*, 55(3):414–22, 2014.

[65] Mary M Galloway. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*, 75, 1974.

[66] Guillaume Thibault, Bernard Fertil, Claire Navarro, Sandrine Pereira, Pierre Cau, Nicolas Levy, Jean Sequeira, and Jean-Luc Mari. Shape and texture indexes application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(01):1357002, 2013.

[67] Michael Miller, Jun Zhang, Katherine Binzel, Jerome Griesmer, Thomas Laurence, Manoj Narayanan, Deepa Natarajamani, Sharon Wang, and Michael Knopp. Characterization of the vereos digital photon counting pet system. *Journal of Nuclear Medicine*, 56(supplement 3):434–434, 2015.

[68] Maurizio Conti and Lars Eriksson. Physics of pure and non-pure positron emitters for PET: A review and a discussion. *EJNMMI Physics*, 3(1), 2016.

[69] Sofia Michopoulou, Emma O'shaughnessy, Katharine Thomson, and Matthew J Guy. Discovery molecular imaging digital ready pet/ct performance evaluation according to the nema nu2-2012 standard. *Nuclear medicine communications*, 40(3):270–277, 2019.

[70] Nathaniel P Scott and Daniel R McGowan. Optimising quantitative 90 y pet imaging: an investigation into the effects of scan length and bayesian penalised likelihood reconstruction. *EJNMMI research*, 9(1):40, 2019.

[71] Tomoaki Otani, Makoto Hosono, Mitsunori Kanagaki, Yasuyuki Onishi, Naoko Matsubara, Kazuna Kawabata, and Hiroyuki Kimura. Evaluation and optimization of a new pet reconstruction algorithm, bayesian penalized likelihood reconstruction, for lung cancer assessment according to lesion size. *American Journal of Roentgenology*, pages 1–7, 2019.

[72] Mimmi Björsdorff, Jenny Oddstig, Nina Karindotter-Borgendahl, Helén Almquist, Sophia Zackrisson, David Minarik, and Elin Trägårdh. Impact of penalizing factor in a block-sequential regularized expectation maximization reconstruction algorithm for 18 f-fluorocholine pet-ct regarding image quality and interpretation. *EJNMMI physics*, 6(1):5, 2019.

[73] Youngho Seo, Mohammad Mehdi Khalighi, Kristen A Wangerin, Timothy W Deller, Yung-Hua Wang, Salma Jivan, Maureen P Kohi, Rahul Aggarwal, Robert R Flavell, Spencer C Behr, et al. Quantitative and qualitative improvement of low-count [68 ga] citrate and [90 y] microspheres pet image reconstructions using block sequential regularized expectation maximization algorithm. *Molecular Imaging and Biology*, pages 1–9, 2019.

[74] Edwin EGW ter Voert, Urs J Muehlematter, Gaspar Delso, Daniele A Pizzuto, Julian Müller, Hannes W Nagel, and Irene A Burger. Quantitative performance and optimal regularization parameter in block sequential regularized expectation maximization reconstructions in clinical 68 ga-psma pet/mr. *EJNMMI research*, 8(1):70, 2018.

[75] Jianhua Yan, Josh Schaefferkoette, Maurizio Conti, and David Townsend. A method to assess image quality for Low-dose PET: Analysis of SNR, CNR, bias and image noise. *Cancer Imaging*, 16(1):1–12, 2016.

[76] Anastasia Chalkidou, Michael J O'Doherty, and Paul K Marsden. False Discovery Rates in PET and CT Studies with Texture Features: A Systematic Review. *PloS one*, 10(5):e0124165, 2015.

[77] Alex Zwanenburg, Stefan Leger, Martin Vallières, and Steffen Löck. Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*, 2016.

[78] Eugene J Teoh, Daniel R McGowan, Kevin M Bradley, Elizabeth Belcher, Edward Black, and Fergus V Gleeson. Novel penalised likelihood reconstruction of pet in the assessment of histologically verified small pulmonary nodules. *European radiology*, 26(2):576–584, 2016.

[79] T Vieira Sampaio, D Faria Borges, F Silva Azevedo, Francisco Pimentel, and J de Oliveira Pereira. The impact of a bayesian penalized likelihood reconstruction algorithm on the evaluation of indeterminate pulmonary nodules by dual-time point 18f-fdg pet/ct. *Clinical nuclear medicine*, 42(7):e352–e354, 2017.

[80] Nassim Parvizi, James M Franklin, Daniel R McGowan, Eugene J Teoh, Kevin M Bradley, and Fergus V Gleeson. Does a novel penalized likelihood reconstruction of 18f-fdg pet-ct improve signal-to-background in colorectal liver metastases? *European Journal of Radiology*, 84(10):1873–1878, 2015.

[81] Joke Devriese, Laurence Beels, Alex Maes, Christophe Van de Wiele, and Hans Pottel. Impact of PET reconstruction protocols on quantification of lesions that fulfil the PERCIST lesion inclusion criteria. *EJNMMI Physics*, 5(1):1–13, 2018.

[82] Charline Lasnon, Blandine Enilorac, and Nicolas Aide. Reply to: "All that glitters is not gold – new reconstruction methods using Deauville criteria for patient reporting". *European Journal of Nuclear Medicine and Molecular Imaging*, 45(5):878–881, 2018.

[83] Ehsan Samei, Kirsten L Boedeker Donovan Bakalyar, Samuel Brady, Jiahua Fan, Shuai Leng, Kyle J. Myers, Lucretiu M. Popescu, Frank Ranallo Giraldo, Juan Carlos Ramirez, Justin Solomon, Jay Vaishnav, and Jia Wang. *AAPM TG 233: Performance Evaluation of Computed Tomography Systems*. Number 233. 2019.

Appendix A

Supplemental material: Performance Characteristics of the Whole-Body Discovery IQ PET/CT System

Reynés-Llompart, G., Gámez-Cenzano, C., Romero-Zayas, I., Rodríguez-Bel, L., Vercher-Conejero, J. L., & Martí-Climent, J. M. (2017). Journal of Nuclear Medicine, 58(**7**), 1155-1161.

*This appendix corresponds to the supplemental material added in the original publication.

Reynés-Llompart G, et al. Performance Characteristics of the Whole-Body Discovery IQ PET/CT System. Journal of Nuclear Medicine, 58(7): 1155-1161.
http://doi.org/10.2967/jnumed.116.185561

Jaszczak Phantom

In the context of this work, the Jaszczak phantom (Data Spectrum Corporation, Durham, NC, USA) was used in order to assess the image quality and resolution when different reconstructions are applied. This phantom is a cylinder fillable with water, with an internal diameter of 21.6 cm. The lower portion of the cylinder contains 6 sets of acrylic rods arranged in a pie-shaped pattern and with the following diameters: 4.8, 6.4, 7.9, 9.5, 11.1, and 12.7 mm. This allows the evaluation of spatial resolution in transaxial direction. The phantom was filled with a total activity of 25 MBq, providing a concentration of approximately 4.2 kBq/cm3, and acquired during 350 s. Then data was reconstructed using VHD, VPHD-S and Q.Clear algorithms, using a total of 70 cm FOV and the reconstruction parameters specified in the image quality section.

Supplemental results

Appendix B

Supplemental material: Phantom, clinical and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q.Clear) on a BGO PET/CT scanner

Reynés-Llompart, G., Gámez-Cenzano, C., Vercher-Conejero, J. L., Sabaté-Llobera, A., Calvo-Malvar, N., & Martí-Climent, J. M. (2018). Medical physics, 45(**7**).

*This appendix corresponds to the supplemental material added in the original publication.

Reynés-Llompart G, Gámez-Cenzano C, Vercher-Conejero JL, Sabaté-Llobera A, Calvo-Malvar N, Martí-Climent JM. Phantom, clinical and texture indices evaluation and optimization of a penalized-likelihood image reconstruction method (Q.Clear) on a BGO PET/CT scanner. Medical physics, 2018, 45(7):3214-3222.
http://doi.org/10.1002/mp.12986

Appendix C

Supplemental material: Image quality evaluation in a modern PET system: impact of new reconstructions methods and a radiomics approach

Gabriel Reynés-Llompart, Aida Sabaté-Llobera, Elena Llinares-Tello, Josep M Martí-Climent, Cristina Gámez-Cenzano (2019). Scientific Reports, 1-**9**, 10640.

*This appendix corresponds to the supplemental material added in the original publication.

Methodology to detect the liver slice

A heuristic method was used to batch process al PET scans to select the slice comprising a major part of the liver. An initial guess of the slices comprising the live was made (from the second bed to the third). An initial thresholding segmentation was performed to only consider the regions comprising a SUV between 0.1 and 2.6, to avoid lesions or physiologic glucose avid regions (such as heart or bladder). Next, the slice with the maximum mean SUV was found and the row and columns with the maximum mean SUV where selected as regions of the liver. The correct selection of the liver ROI placement was visually inspected for each patient.

Additional table and figures

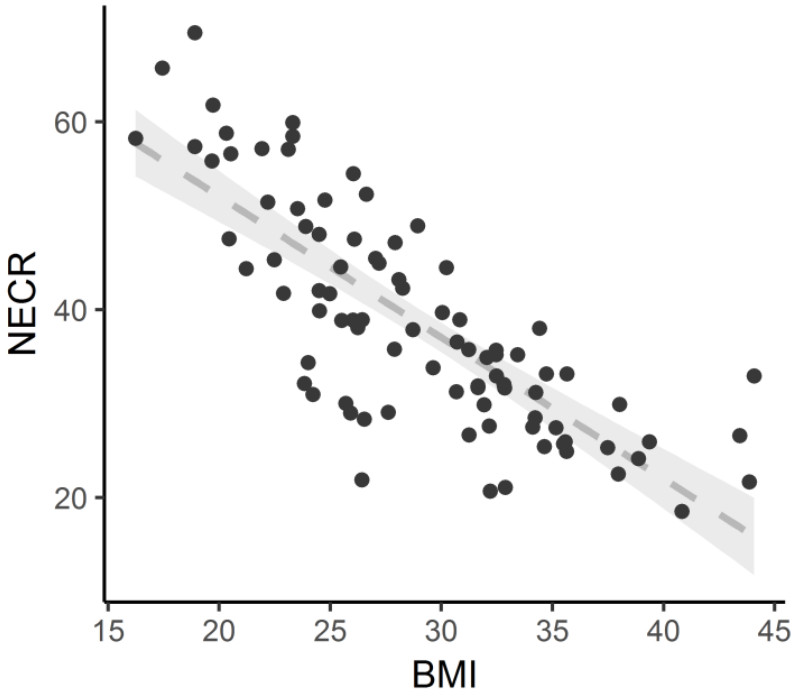

Fig. C.1 Relation of NECR and BMI. Dotted line represents an adjusted linear regression and its 95% confidence interval.

Table C.1 Studied variables mean and standard deviation for low (LQ) and high (HQ) image quality groups.

	VPHD-S			Q.Clear		
	LQ	HQ	p-value	LQ	HQ	p-value
Patients (n)	56	57		78	35	
Age (y)	67 ± 14	63 ± 13	0.06	67 ± 16	64 ± 12	0.1
Glucose level (mmol/l)	6 ± 2	6 ± 2	0.4	6.4 ± 1.6	6 ± 2	0.2
Weight (kg)	85 ± 21	70 ± 14	0.001	83 ± 20	74 ± 17	0.1
Height (cm)	165 ± 8	160 ± 10	0.003	165 ± 8	160 ± 10	0.03
BMI (kg/m^2)	31 ± 7	27 ± 6	0.02	30 ± 6	28 ± 6	0.17
LBM (kg)	55 ± 11	46 ± 10	> 0.001	53 ± 11	48 ± 11	0.05
Uptake Time (min)	71 ± 8	71 ± 9	0.6	72 ± 9	70 ± 8	0.19
Tracer Activity (MBq)	140 ± 38	123 ± 30	0.07	136 ± 38	127 ± 33	0.4
True events	$(11 \pm 2) \cdot 10^4$	$(12 \pm 1) \cdot 10^4$	0.12	$(11 \pm 20) \cdot 10^4$	$(12 \pm 1) \cdot 10^4$	0.17
Random events	$(40 \pm 15) \cdot 10^4$	$(34 \pm 12) \cdot 10^4$	0.11	$(38 \pm 16) \cdot 10^4$	$(36 \pm 1 \cdot 10^4$	0.8
Scatter events	$(9 \pm 1.8) \cdot 10^4$	$(9 \pm 1.5) \cdot 10^4$	0.48	$(9 \pm 2) \cdot 10^4$	$(9 \pm 12) \cdot 10^4$	0.9
NECR	35 ± 12	41 ± 12	0.02	36 ± 11	39 ± 12	0.3
PNECR	111 ± 20	121 ± 19	0.04	112 ± 20	118 ± 19	0.2
R_{DW}	1.6 ± 0.1	1.7 ± 0.2	0.007	1.6 ± 0.1	1.7 ± 0.2	0.006
R_{DBMI}	4.5 ± 0.6	4.5 ± 0.8	0.2	4.5 ± 0.5	4.5 ± 0.8	0.7
R_{RDLBM}	2.54 ± 0.4	2.7 ± 0.4	0.09	2.5 ± 0.4	2.7 ± 0.5	0.3
$Variance_{ROI}$	0.1 ± 0.2	0.05 ± 0.03	0.6	0.1 ± 0.3	0.05 ± 0.03	0.2
SNR_{ROI}	10 ± 3	11 ± 3	0.6	11 ± 4	11 ± 4	0.6
$Variance_{Slice}$	0.6 ± 0.3	0.7 ± 0.2	0.3	0.7 ± 0.3	0.7 ± 0.3	0.6
SNR_{Slice}	1.5 ± 0.2	1.5 ± 0.3	0.4	1.4 ± 0.3	1.5 ± 0.3	0.6
CNR	1.6 ± 0.43	1.3 ± 0.4	0.01	1.5 ± 0.4	1.3 ± 0.4	0.04
Center Shift	16 ± 10	20 ± 12	0.2	15 ± 9	19 ± 11	0.7

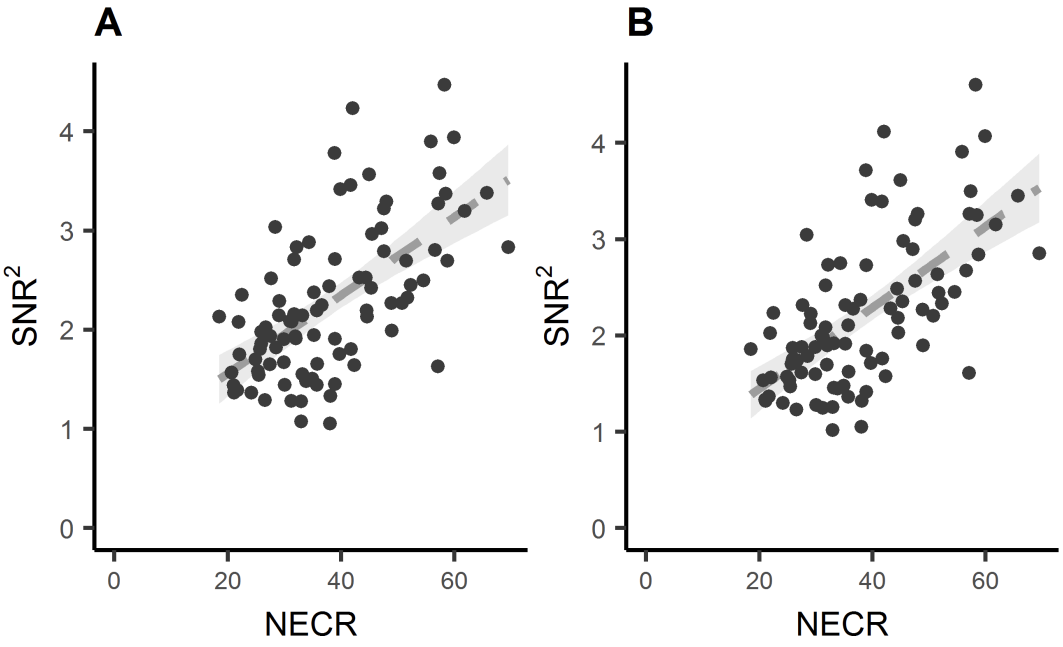

Fig. C.2 Relation of NECR and $SNR_{Slice}{}^2$, for a) OSEM + PSF and b) Q.Clear reconstruction methods. Dotted line represents and adjusted linear regression and its 95% confidence interval.

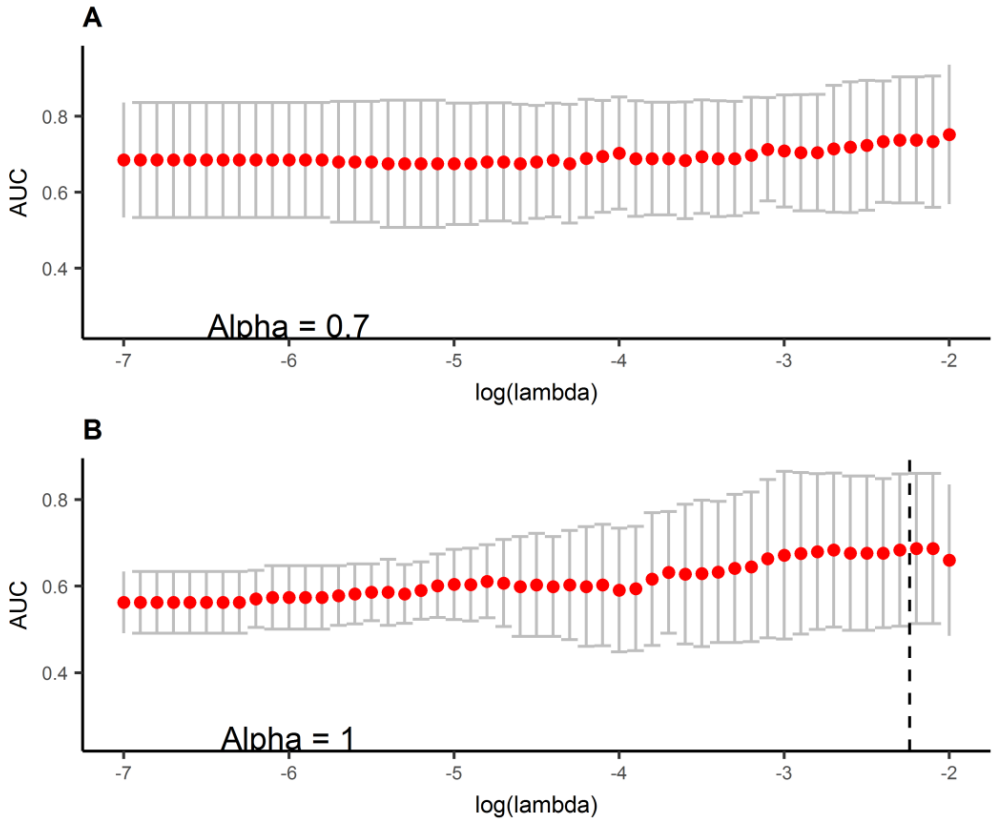

Fig. C.3 Elastic net lambda optimization by cross-validation.

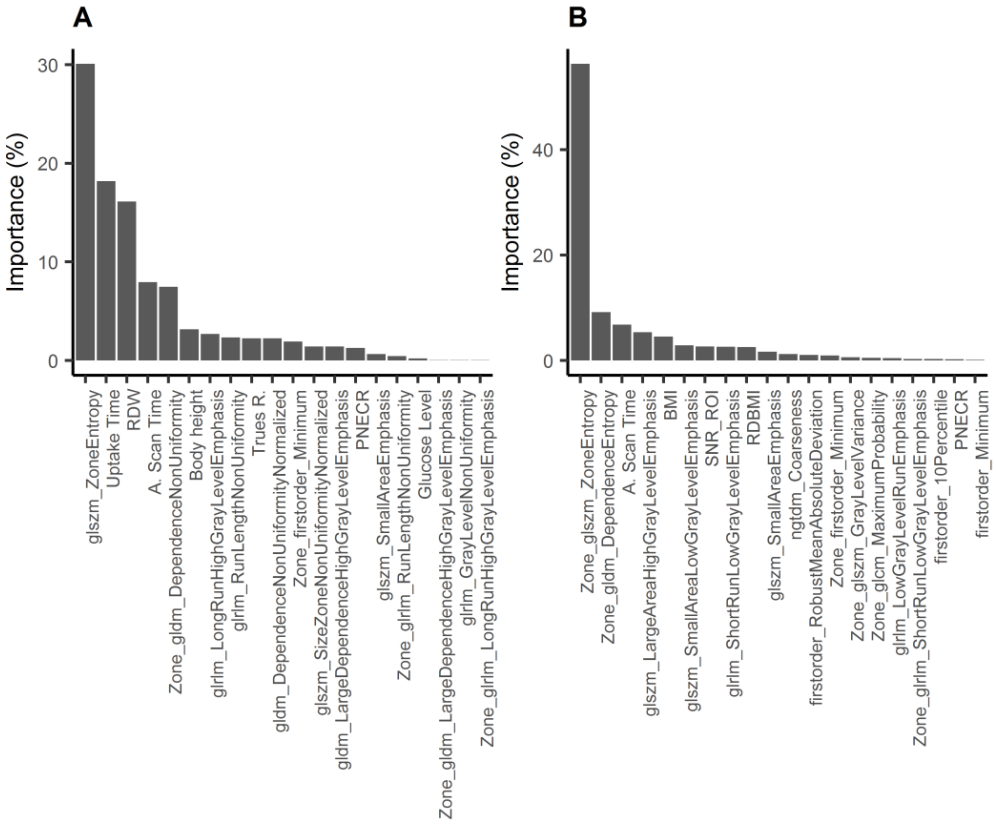

Fig. C.4 Importance of the selected features for A) OSEM+PSF and B) Q.Clear.