Università degli Studi di Genova
Scuola di Dottorato in Neuroscienze
Curriculum di Neuroscienze cliniche e sperimentali

From clinics to methods and back: a tale of amyloid-PET quantification

Andrea Chincarini

Submitted in fulfillment of the requirements for the degree of
Doctor of Philosophy in Neuroscience
of the
Università degli Studi di Genova, May 22nd, 2018

Abstract

Background

The in-vivo assessment of cerebral amyloid load is taking a leading role in the early differential diagnosis of neurodegenerative diseases. With the hopefully near introduction of disease-modifying drugs, we expect a paradigm shift in the current diagnostic pathway with an unprecedented surge in the request of exams and detailed analysis.

Up-to-date clinical practices can rely either on a rather invasive cerebro-spinal fluid assay (CSF), or on a positron emission tomography (PET), where the availability of a small number of radiotracers were validated to be effective proxy of brain amyloidosis.

With the increasing experience in amyloid-PET, the scientific community realized that previous approaches to the analysis were not enough to capture the physiological variability and serve as a reliable tool for both medical report and clinical research.

Materials

In this work I developed new approaches to amyloid-PET analysis. They were validated and integrated in a common framework to be used in clinical practice. The integration of these new approaches also allows a new measure paradigm (the rank), paving the way for more robust evaluation of the PET scan.

Analyses were carried out on a total of $\approx$ 700 scans, coming from three dataset: the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Multicentric Pilot Dataset (MPD) and a smaller single-center dataset which was used for side analyses.

These dataset are very diverse in terms of size ($\approx$ 500, $\approx$ 150 and 40 respectively), inclusion/exclusion criteria (research, trial based or naturalistic population), data quality and number of participant centers. The tracer of choice was ^{18}F-Florbetapir.

Methods

Two novel semi-quantification methods (ELBA and TDr) were developed, tested and validated over the various dataset and versus the visual assessment. Test and validation cases included: comparison with CSF values, coherence among quantifiers, clinical follow-up, regional and longitudinal analysis.

A single quantifier (rank) was designed to summarize all results into a more robust index, to be tested in clinical setting.

Results

We show that independent quantification methods can indeed be useful and complementary to the visual dichotomic assessment. The high internal coherence among methods validated their applicability on a wide range of use cases, scanners and patient clinical status.

These analyses are now running without human supervision and are currently under testing in clinics thanks to a user-friendly web interface.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Prof. Flavio Nobili for his friendship, continuous inspiration and support.

I would also like to thank Prof. Nobili's team: their dedication and passion is a tribute to their mentor and a real drive for anyone who is working with them.

A special thank goes to Dr. Ugo Paolo Guerra, for most of this work would not have been possible without him.

Dedication

To my wife.

List of abbreviations

Aβ	*Amyloid beta*
AD	*Alzheimer's Disease*
ADNI	*Alzheimer's Disease Neuroimaging Initiative*
AIF	*Arterial Input Function*
APP	*Amyloid Precursor Protein*
AUC	*Area Under Curve*
CBF	*Cerebral Blood Flow*
CC	*Cross-Correlation*
CNS	*Central Nervous System*
CSF	*Cerebro-Spinal Fluid*
CT	*Computer Tomography*
DO	*Degree od Order*
ELBA	*EvaLuation of Brain Amyloidosis*
FDG	*Fludeoxyglucose*
FDOPA	*Fluorinated Levodopa*
fMRI	*Functional Magnetic Resonance Imaging*
FN	*False Negative*
FOV	*Field Of View*
FP	*False Positive*
FWHM	*Full Width at Half Maximum*
GM	*Gray Matter*
LOR	*Line Of Response*

ML-EM *Maximum Likelihood Expectation Maximization*

MNI *Montreal Neurological Institute*

MPD *Multicentric Pilot Dataset*

MRI *Magnetic Resonance Imaging*

MS *Mean Squares*

NC *Normalized Correlation*

OSEM *Ordered Subset Expectation Maximization*

PCA *Principal Component Analysis*

PET *Positron Emission Tomography*

PIB *Pittsburgh compound-B*

PSF *Point Spread Function*

PVE *Partial Volume Effect*

ROC *Receiver Operating Characteristic*

ROI *Region Of Interest*

RP *Reference PET*

SPECT *Single Photon Emission Computed Tomography*

SUV *Standardized Uptake Value*

SUVr *Standardized Uptake Value Ratio*

TAC *Time-Activity Curve*

TDr *Time-Delayed Ratio*

TN *True Negative*

TOF *Time Of Flight*

TP *True Positive*

Uptake *Quantity of ligand that binds to a specific site*

Voxel *Volumetric picture element*

WM *White Matter*

Contents

List of Tables

List of Figures

Chapter 1

Research context

1.1 Scope and intent of this work

The latest research criteria for diagnosing Alzheimer's Disease (AD) suggest that in-vivo quantification of brain amyloid-β (Aβ) deposition using positron emission tomography (PET) is emerging as a crucial tool in diagnosing or in excluding Alzheimer's disease (see Dubois et al. (2014), but also McKhann et al. (2011); Johnson et al. (2013a)) and is likely to play a pivotal role in upcoming clinical trials of disease modifying agents (Fagan et al., 2006; Klunk et al., 2004; Rowe et al., 2007). Biomarkers of cerebral β-amyloid are increasingly used in clinical trials (Karran and Hardy, 2014; Blennow et al., 2013) This stresses the need for reliable and available biomarkers of brain A β pathology. The foreseeable increase in the use of this testing methodology, the sensitive nature of the outcome for the patient prognosis and the availability of different radioligands pose new challenges to clinicians.

Currently, human readers who want to acquire the necessary expertise are advised to attend ligand-specific courses - often brought in by pharmaceutical companies - where visual evaluation procedures tend to favour a dichotomic reading (positive/negative). This approach is justified by the apparent scarcity of cases where the ligand uptake distribution cannot be easily identified or where its spatial extent is limited. Indeed, while most amyloid-PET images are rather easily evaluated by a trained eye, as amyloid-PET becomes a widespread tool uncertain instances are going to be met more and more frequently.

Because of the non trivial visual assessment in a non negligible number of cases, a more sophisticated approach is required, which provides quantification (and rank) besides classification.

This work deals with proxy measures of brain amyloidosis derived from clinical-grade positron emission tomography (PET) images. It involves significant methodological developments in order to deliver clinically usable indexes, with a significant impact on Nuclear Medicine practices, training and assessment.

The shortcomings of current clinical practice when dealing with amyloid-PET images are mainly due to the novelty of amyloidosis tracers and the uncertainties in the causative models linking amyloid presence, neuroinflammation, neurodegeneration and symptoms. This work examines these shortcomings and develops new approaches to overcome them.

This leads to the development of new tools for quantification, their validation and their implementation to enhance the clinical experience significantly.

The drastic reduction of the quantification uncertainty deriving from the conventional calculation (i.e. SUVr) is fundamental not only to select and study atypical subjects with borderline amyloid burden, but becomes necessary when considering longitudinal measures to validate and monitor the effectiveness of future anti-amyloid drugs.

1.2 Brain amyloidosis

Amyloidosis is a general term that describes a wide spectrum of diseases characterized by a deposit of amyloidogenic peptides in different organs. These peptides derive from proteolysis of a transmembrane glycoprotein which is present in all the cells – although predominantly in neurons – called amyloid precursor protein (APP).

The APP is processed by two secretase enzymes that act in sequence, leading to distinguished pathways (amyloidogenic and non-amyloidogenic) depending on which determines the production of the Aβ peptide. APP may be initially cut by the α-secretase (non-amyloidogenic pathway) or by the β-secretase (amyloidogenic pathway). Both secretases generate a soluble fragment – the sAPPα and sAPPβ respectively – which is released into the extracellular environment, and a C-terminal fragment which remains anchored to the membrane (figure 1.1). The latter is used as a substrate by γ-secretase to generate two other soluble fragments: the p3 peptide, that corresponds to the region 17-40/42 of Aβ sequence in the non-amyloidogenic pathway; and a Aβ fragment consisting of 40/42 amino acids, that comes from the amyloidogenic pathway (γ site) and has fibrillogenic characteristics.

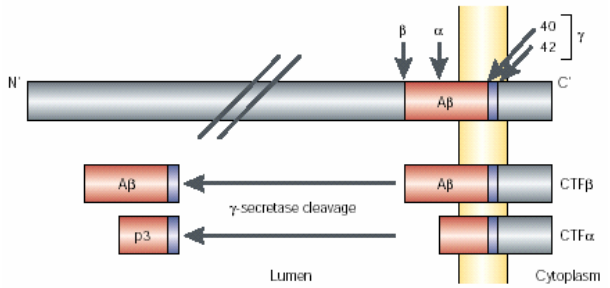

Figure 1.1: Formation of the peptide β amyloid. The γ-secretase in the amyloidogenic pathway can cut the βAPP CTF in two different points of its sequence: γ-40 and γ-42, and generates two different amyloid peptides: Aβ1-40 and Aβ1-42, which are able to be shaped into β-sheets and aggregate into fibrils.

As there are physiological mechanisms for the production of Aβ, there are clearance pathways. Two main mechanisms have been identified:

- the transport of Aβ (predominantly in the form Aβ1-40) across the blood brain barrier into the circulation (Zlokovic, 2004)
- the degradation of Aβ (predominantly in the form Aβ1-42) by means of peptidase (Guénette, 2003; Tanzi et al., 2004) such as NEP (neutral endopeptidase) and IDE (insulin degrading enzyme).

The first mechanism involves the transport mechanism in the blood which is mainly mediated by LRP-1 receptor (LDL receptor related protein-1) in the cerebral microvascular endothelium.

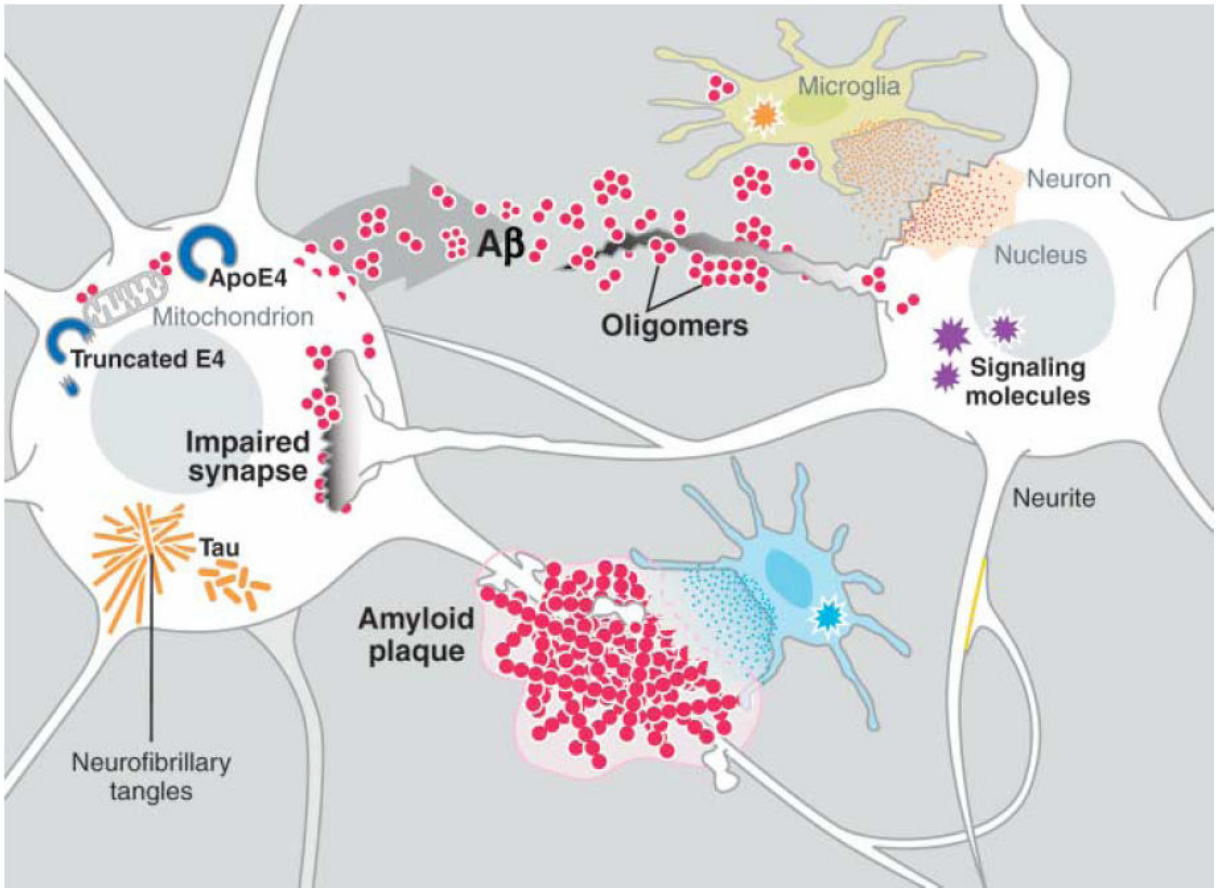

Figure 1.2: Aβ toxicity mechanisms.

The main forms of Aβ peptides consist of 40 or 42 amino acids (Aβ1-40 Aβ1 -42), distinguishable for the different COOH terminal.

The longest form (Aβ1-42) is equipped with two additional hydrophobic amino acid residues; it is able to aggregate more rapidly and to form fibrils, therefore it is the peptide that is initially deposited for the formation of plaques (Glabe, 2001).

The Aβ in the brain tissue exists in various states of aggregation as monomers, dimers and oligomers of high molecular weight (Fig. 1.3); the aggregation of oligomers eventually generates proto-fibrils and then the fibrils (Klein et al., 2004). That said, only the fibrils are going to form plaques.

The Aβ peptides are neurotoxic. Early in vitro studies compared the effect of the fibrils and monomers, finding that the latter gave no toxicity effect (Lorenzo and Yankner, 1994). Therefore the main object of the studies is Aβ in the fibrillar form, which is considered responsible for the cytotoxic effect observed also in-vivo conditions.

According to the amyloid cascade hypothesis, the accumulation of fibrillar Aβ in the form of plaques damages neurons through two different mechanisms (Fig 1.2):

- a direct mechanism, where Aβ interacts with components of the cell membrane and directly damages neurons and / or increases the susceptibility of neurons to a variety of damage factors such as excitotoxicity, hypoglycemia or peroxidative damage (Koh et al., 1990)

- an indirect mechanism, Aβ damages neurons through the activation of the microglia and astrocytes to produce toxic and inflammatory mediators, such as nitric oxide (NO), cytokines and reactive oxygen intermediates (Meda et al., 1995; Bianca et al., 1999), that cause neuronal death by apoptosis or necrosis.

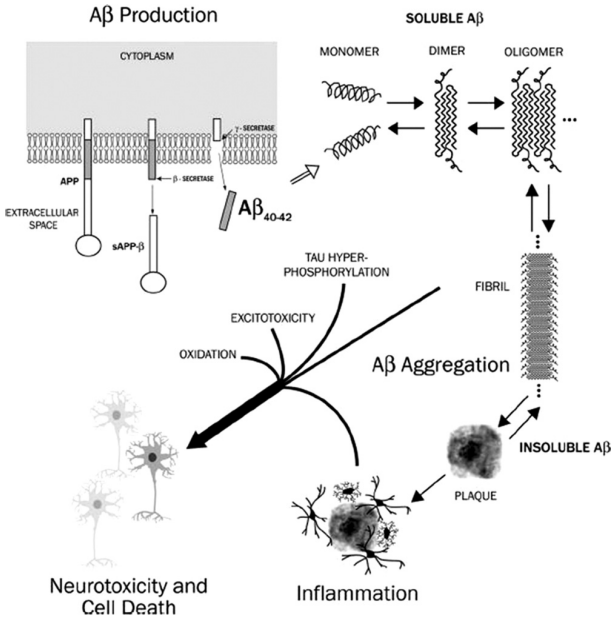

Figure 1.3: Overview of the Aβ production and aggregation.

Amyloid deposit in the central nervous system represents the most frequent form of amyloidosis in humans. When these deposits are initially localized in the walls of cerebral blood vessels the most common clinical manifestation is the stroke; when deposits extend through selected parenchymal areas, particularly the limbic system, these are associated with neurodegeneration leading to dementia.

1.3 Amyloid-related pathologies

Amyloid-β deposit is characteristic of some neurodegenerative diseases associated with dementia: the Alzheimer's disease (AD) and the Lewy bodies family (including dementia with Lewy bodies and Parkinson's disease).

Aβ is also associated with other disease processes that affect brain function such as cerebral amyloid angiopathy (CAA), brain trauma including traumatic brain injury (TBI) and chronic traumatic encephalopathy (CTE), Down's syndrome (DS) and meningiomas.

While amyloid accumulation can occur in the body too – for example systemic amyloidosis involving the heart (Ikonomovic et al., 2008) – its presence in the CNS is the most studied case as it is linked to AD.

Alzheimer's disease is a complex, irreversible neurodegenerative disease, characterized by an initial general cognitive decline and marked memory loss progressing to dementia and severe impairment of daily life functions. Typical presentations of the disease lead to the death of the patient after 7–10 years from diagnosis.

AD is a high-prevalence disease in the elderly population ($\simeq$ 30% for ages $>$ 75 years) with an estimated 14 million patients in Europe and the United States. The progressive nature of this particular neurodegenerative disorder is linked to the synaptic and neuronal damage in cortical areas devoted to memory (starting from the medial temporal lobe) and then affecting broader areas in the frontal and parietal lobes. Macroscopically, early stages are characterized by synaptic loss and neuronal damage, accompanied by a reactive gliosis and the deposition of neurofibrillary tangles (NFTs, consisting of hyperphosphorylated tau proteins P-tau) and Aβ plaques, while advanced stages are characterized by severe

neuronal loss leading to cortical and sub-cortical atrophy. To date, the available genetic, pathological and biochemical data support the thesis that the alteration of the balance between amyloid production and clearance is the main input in the cascade of events that leads to neurodegeneration and dementia.

1.4 Clinical criteria and diagnosis

The clinical diagnosis of AD has been linked to the presence of dementia for many years. The first criteria were introduced in 1984 by the National Institute of Neurological and Communicative Disorders and Stroke's Alzheimer's Disease and Related Disorders Association (ADRDA-NINCDS). They were still focused on the clinical presentation, saying that (1) diagnosis of AD could only be considered as *likely* until confirmed by autopsy; and (2) the clinical diagnosis could only be confirmed when the disease was at an advanced enough stage to cause significant functional disability.

In 1995 Braak and Braak worked out the staging of AD pathology severity based on neurofibrillary changes, that is by identifying six stages of disease progression which can be distinguished on the location and concentration of NFT (fig 1.4, Braak and Braak (1995)).

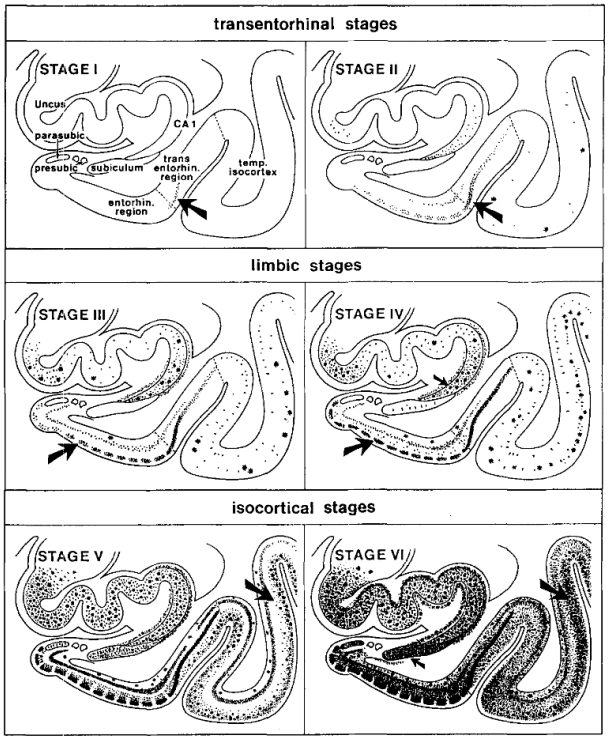

Figure 1.4: Braak stages.

The extensive research carried on over the next several years in the field of neuroimaging and biochemistry have led to the formulation of a new clinical entity: the mild cognitive impairment or MCI. MCI is a syndromic condition which identifies a population of patients with objective (i.e. confirmed by suitable neuropsychological examination) cognitive impairment in one or more cognitive spheres, while still being unimpaired in everyday life. Follow-up studies of this group have shown that MCI have a high risk of developing AD, with approximately 50% of them converting within 2-5 years after diagnosis. Other risk factors include genetic (ApoE ϵ genotype), lifestyle, history of psychiatric conditions and education, possibly modulated by systemic co-morbidities such as diabetes and vascular diseases.

In 2007, the International Working Group (IWG) for New Research Criteria for the diagnosis of AD and subsequently the National Institute on Aging-Alzheimer's Association (NIA-AA), have revisited the NINCDS-ADRDA guidelines and delivered new criteria introducing the biomarker concept, by which the diagnosis of AD is related to the presence of biological markers; thus the diagnosis started the transition from a clinical-pathological entity to a clinical-biological one.

These criteria required that AD can be recognized in vivo and regardless of the dementia symptoms if the presence of at least two essential criteria can be objectively measured. The first requirement is a *core clinical phenotype* that requires evidence of significant memory impairment, associated with a progressive worsening of the performance. In addition, the diagnosis of *probable* AD would require support features represented by objective evidence of alteration in one or more of the following biomarkers:

- presence of atrophy of the medial temporal lobe measured with Magnetic Resonance Imaging (MRI)
- anomalous protein concentration ($A\beta_{42}$) in the cerebro-spinal fluid (CSF)
- impaired glucose metabolism in the brain, measured by ^{18}F-fluorodesoxyglucose Positron Emission Tomography (FDG-PET).

The introduction of biomarkers as a support feature (although still not mandatory for the diagnosis) was the most important innovation introduced by the criteria of 2007. This radical change allowed the identification of the so-called *prodromal* AD, that is a condition *at risk of AD* where at least one biomarker shows typical alteration.

In 2010, the same group of authors published a second paper, redefining the lexicon of Alzheimer's pathology and disease. AD now becomes a life-long process, a continuum ranging from an (asymptomatic) pre-clinical stage, to mild cognitive impairment and – only in the end – to dementia. The Braak stages (transentorhinal I, II; limbic III, IV and neocortical V, VI) are then linked with pre-clinical – or asymptomatic – AD; prodromal AD, with symptoms or signs of cognitive impairment (one or more altered biomarkers); and symptomatic AD.

In 2011, experts from the National Institute of Aging (NIA) along with those of the Alzheimer's Association (AA) issued new clinical criteria (NIA-AA) in which various states of the pathology are defined, with and without the biomarker support. The NIA-AA criteria define three frameworks: asymptomatic (preclinical AD), pre-dementia (prodromal AD - MCI due to dementia), and dementia (due to AD). These criteria also introduce three different states for each of the three frameworks. The asymptomatic or pre-clinical stage was divided in

- Stage I: asymptomatic cerebral amyloidosis characterized by the positivity of amyloidosis biomarkers
- Stage II: confirmed amyloidosis accompanied by neurodegeneration markers, although still without significant cognitive symptoms
- Stage III: similar to stage II but with signs or symptoms, although still insufficient to diagnose the state of MCI

Both IWG criteria and NIA-AA recognize that the AD starts before symptoms occur, but the NIA-AA identify patients with evidence of $A\beta$ accumulation as asymptomatic AD, whereas in the IWG criteria the amyloid deposit is only identified as a risk factor.

For the stages of pre-dementia and dementia, NIA-AA criteria recognize three levels of probability (high, intermediate or unlikely) that the disease is due to AD based on the

information of biomarkers, which now increase the confidence level of diagnosis and are no longer a criterion of diagnosis by themselves.

In a recent position paper, the International Working Group (IWG) (Dubois et al., 2014) presented a review of the diagnostic criteria for Alzheimer's disease. Based on an extensive review of the recent literature (from 2007 to January 2013), the authors proposed a diagnostic algorithm that allows the application of the same criteria in different stages of the disease, further strengthening the understanding of Alzheimer's disease as dual clinical-biological entity.

These reviewed criteria also deals with typical, atypical and mixed AD. They require that in-vivo evidence of Alzheimer's-like pathology be demonstrated by positivity of at least one biomarker such as: reduction of the concentration of $A\beta_{1-42}$ in the CS together with increased concentration of phosphorylated-τ (alternatively the increased uptake of radiopharmaceuticals for amyloid at PET); the presence of autosomal dominant mutation typical of AD (in PSEN1, PSEN2 or APP genes).

An important departure from previous criteria is the appearance of the distinction between patho-physiological biomarkers – which are necessary for the diagnosis – and topographical or progression biomarkers that are no longer included in the diagnostic process but, albeit having a lower specificity for Alzheimer's Disease, are closely related to the clinical manifestations and its severity.

The presence of pathological amyloid load is not in and of itself diagnostic of AD, but the absence of amyloid in a patient who is having symptoms of either mild cognitive impairment or early dementia excludes AD as the basis of that dementia. While it doesn't give the precise alternative diagnosis, it does tell that AD in that specific clinical context – that otherwise fits with AD – is quite unlikely.

Further studies indicated that the relationship between AD, amyloid burden, neurodegeneration and NFT can be rather complicated. In most cases, AD-related symptoms follow a significant brain amyloid burden, but there are also a number of subjects in which neurodegeneration (and symptoms) is present without brain amyloidosis. These subjects are tagged Suspected non-Alzheimer patho-physiology (SNAP). In addition, there are also subjects showing NFT indistinguishable from those observed in AD, but still without brain amyloidosis; these are tagged Primary age-related tauopathy (PART) (Montine et al., 2011; Jack, 2014) Autoptic studies have indeed found correlation between neurofibrillary-mediated neurodegeneration and cognitive decline but not between neurodegeneration and amyloid plaques (Crary et al., 2014).

Chapter 2

PET fundamentals

Positron emission tomography is an imaging technique used in nuclear medicine that allows non-invasive visualization and quantification of biological and physiological processes. This technique provides three-dimensional images of the anatomical district under examination by exploiting the decay processes of radioactive isotopes, which were previously introduced into the patient's bloodstream by injection of a radiopharmaceutical (Bailey et al., 2005).

This methodology is used extensively both in clinical oncology (to grade tumors and for the research of metastases), and in neurology (for the study and clinical evaluation of various neurodegenerative pathologies) (Catafau and Bullich, 2015).

2.1 Physics principles

PET uses the decay processes of radioactive isotopes used to mark molecules called tracers (or radiopharmaceuticals) that are administered intravenously. PET tracers are marked with radioisotopes that decay β^+ (that is emitting a positron e^+ and a neutrino ν_e).

These molecules are engineered to take part in a specific biological process related to the pathology under consideration. After the tracer is administered, the patient is placed in the scanner for the acquisition.

In this framework, the measure consists in detecting the number of counts in a given volume, where the measure timings are short when compared to the half life of the radioisotope, and long with respect to the biochemistry kinetics that guide the tracer spread into the anatomical region under investigation.

The radioisotope marking the tracer decays through the emission of a positron (e^+) which, in organic tissues, has a mean free path of $1 \sim 2\,mm$. The positron then annihilates with an electron and two collinear photons of $0,511\,MeV$ are then emitted (figure 2.1).

As the photon direction is random, one needs to detect the coincidence of coplanar events. A series of scintillators fashioned in a circular array around a central axis detects a signal only when two collinear events are detected within a time-difference of $10 - 20\,ns$, thus revealing the emission line of sight. Significant post processing of this raw signal is needed though, to retrieve the emitter position.

Typical radioisotopes are ^{18}F and ^{11}C, due to the fact that they can be easily produced

in cyclotrons and that their half-life is rather short: ~ 110 min and ~ 20 min respectively (thus avoiding enduring radiation hazard for patients).

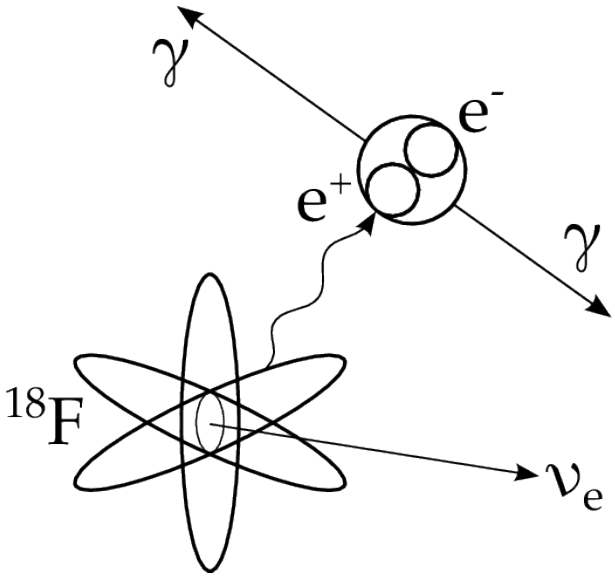

Figure 2.1: Decay of the radioisotope ^{18}F and annihilation of the positron with an electron, resulting in the emission of two γ photons.

The most common radioisotope is ^{18}F, which decays as follows:

$$^{18}_{8}F \rightarrow {}^{18}_{9}O + e^{+} + \nu_e \qquad (2.1)$$

2.2 Detection

The acquisition system in PET is a scanner consisting of a ring of scintillators coupled to photomultipliers. The patient is lying on a support which can slide within the ring.

The detection is obtained by means of scintillators, that convert high energy photons – generated by the annihilation of the positron – into photons of relatively low energy (in the visible spectrum). These photons reach the photocathode of the photomultipliers, which converts the light signal into an electrical signal.

To avoid counting photons not due to an annihilation event, it is necessary to energy-select the events. The energy range (or window) is usually chosen as $350 - 650\,keV$ (Saha, 2016).

Signal is then detected through a coincidence filter, which selects simultaneous events ($\Delta_t < 10 - 20\,ns$). When two photons are detected both within the time window and within the energy window, they are registered as a coincident count, which occurred along an imaginary line connecting the two scintillators that recorded the event and called the *line of response* (LOR).

The three-dimensional image is reconstructed by joining an ordered sequence of two-dimensional images obtained by sliding the support on which the patient is positioned.

These two-dimensional images are obtained through complex reconstruction algorithms that use the profiles of the counts detected on each acquisition plane. The final image consists of a three-dimensional matrix with scalar values, whose element intensity is proportional to the concentration of the radiopharmaceutical in the tissue.

The spatial resolution of PET images is generally $2 - 3\,mm$ and depends on many factors such as the mean free path of the positron, the structure of the detector and the size and shape of the scintillators (Bailey et al., 2005).

2.2.1 Detection issues

The mass of the parent nuclide in the decay exceeds the sum of the masses of the decay products, so the difference in mass is converted into energy and is available to the gener-

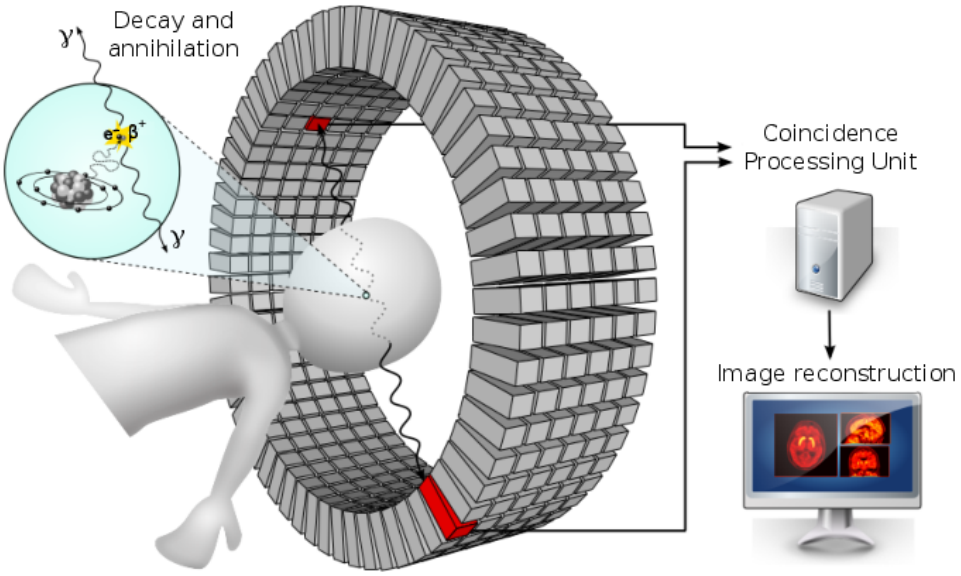

Figure 2.2: Pictorial representation of a typical PET detector, consisting of a ring of scintillators coupled to photomultipliers and a processing unit to retrieve the emitter position.

ated particles. Most of this energy is transferred as kinetic energy to the positron and to the neutrino, or converted into photons, and only a small part is transferred to the child nucleus (still as kinetic energy).

The positron resulting from the decay will interact with the atoms of the tissue, losing its kinetic energy and finally annihilating with an electron. The maximum traveled distance depends on the initial kinetic energy of the positron, and it is of the order of $2.4\,mm$ for ^{18}F (Cherry et al., 2004).

If the positron has not exhausted its kinetic energy before annihilation there is an energy excess, which – in addition to the fact that the electrons of the tissue are also moving – is manifested as a non-collinearity of the two photons produced by the annihilation. This non-collinearity is of the order of some fraction of a degree, which, in a typical PET scanner, amounts to an uncertainty in the annihilation cohordinates of $2 - 3\,mm$.

After the annihilation event, the photons will travel in the patient's tissues in opposite directions; in this phase, different types of interaction occur between the photon and the medium, which attenuate or block the intensity. The modulation property of the medium depends on its characteristics (such as the density), and is described by a linear attenuation coefficient μ. Mathematically, the fraction of photons that crosses a thickness x of matter with attenuation coefficient μ is expressed by the following relation:

$$\Gamma = e^{-\mu x}$$

which extends naturally to a sequence of materials with thickness x_i and coefficient μ_i as:

$$\Gamma = e^{-\sum_i \mu_i x_i}$$

The general formula for a detector measuring N_0 events from a radioactive source without attenuation (for example in air where $\mu \approx 0$), in the presence of matter becomes:

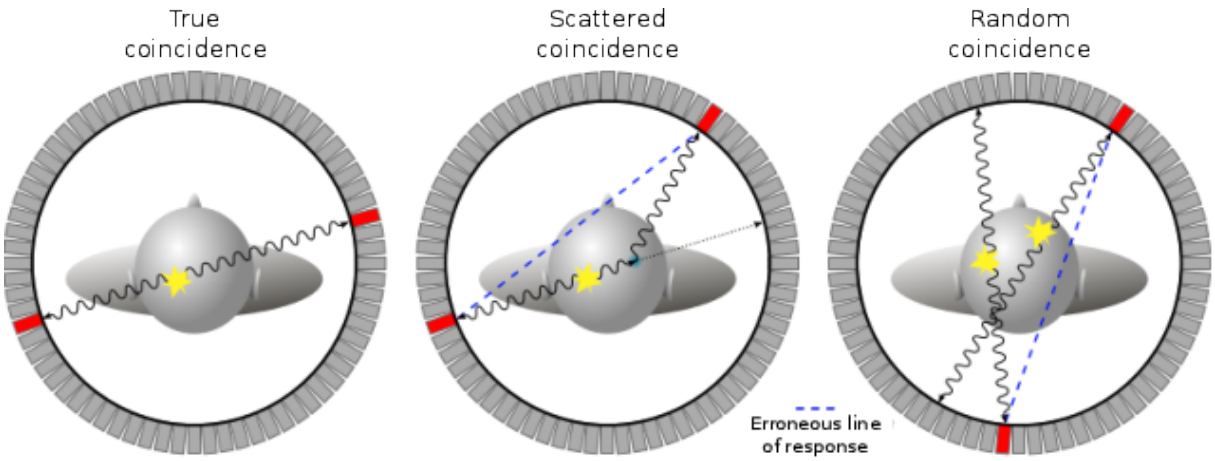

Figure 2.3: Types of coincidences in a PET scanner (courtesy of (Häggström, 2014))

$$N = N_0\, e^{-\int_0^d \mu(x)\, x\, dx}$$

where $mu(x)$ is the linear attenuation coefficient expressed as a function of the coordinate for the different types of tissues and d represents the total distance between the source and the detector.

The attenuation phenomenon in human tissues can be rather significant, so that it is necessary to know the magnitude of this effect to obtain the real distribution of the tracer within the patient. In modern scanners, PET is combined with computerized tomography (CT), which delivers morphological and tissue properties information. The CT is converted into an attenuation map and used to correct the intensities in the image (Lonsdale and Beyer, 2010; Bailey et al., 2005).

While the attenuation coefficient is a function of the tissue alone, the reduction of the photon count rate corresponding to a specific local distribution is also a function of the LOR on which it is detected. Ideally, the only events that should be recorded are those associated with the actual annihilation of positrons, called *true coincidence*; however, a series of fringe events which satisfy the coincidence detection criteria are still considered as annihilation events, resulting in noise and degradation of spatial resolution.

Most of the spurious coincidence events are classifiable in the three types (figure 2.3)

- **True coincidence**: associated with the detection of a pair of photons emitted by the annihilation event. The two photons are revealed within the temporal and spatial coincidence window, that is, between an opposed pair of scintillators.

- **Scattered coincidence**: occurs when one or both photons undergo Compton scattering. The Compton interaction causes both energy loss and change of direction, resulting in an incorrect location of annihilation, and ultimately in image degradation and a reduction in contrast.

- **Random coincidence**: when photons associated with two distinct annihilations are seen by the detection system as coming from a common annihilation event. Random coincidences are sources of noise and are generally proportional to the

square of the activity in the field of view (FOV), as well proportional to the time and energy window.

Acquisitions with PET tomographs in the presence of high counts rate are strongly degraded, so it is necessary to apply corrections for this type of error. A random event evaluation method is known as the coincidence cancellation method: the temporal coincidence window is not opened when the first photon hits the detector, but with a special delay to avoid any real coincidence. The ratio between the number of coincidences with and without an applied delay provides the relationship between real coincidences and random coincidences (Bailey et al., 2005; Saha, 2016).

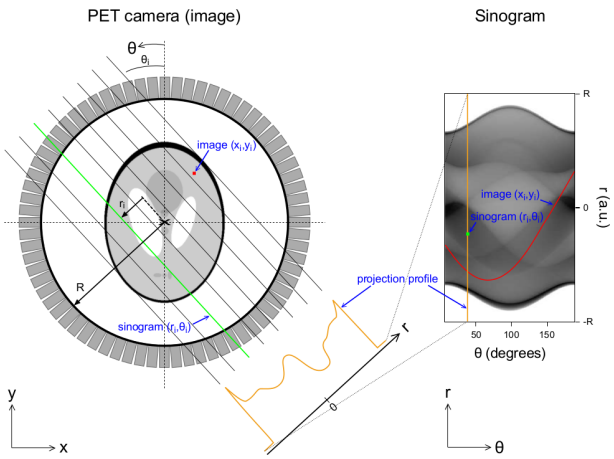

Figure 2.4: The image space (xy) is represented in sinogram space ($r\theta$). A group of LORs constitute a projection profile, corresponding to a column in the synogram. A point in the image space corresponds to a curve in the space of the sinogram, while a point in the space of the sinogram is a LOR in the image space (courtesy of Häggström (2014)).

A considerable improvement over the simple LOR reconstruction is the ability of locating the annihilation event along it; this is achieved thanks to the *time of flight* (TOF) technique.

Thanks to the recent introduction of new scintillator crystals and electronics with better temporal resolution, it is now possible to obtain information on approximate position of the annihilation event along the LOR by measuring the small time difference in the detection of the two photons.

For an annihilation event placed at Δx from the center of the FOV, the Δt between the two detections will be:

$$\Delta t = 2\frac{\Delta x}{c}$$

The time difference is really tiny: for a spatial offset of $9\,cm$ the required temporal resolution is $600\,ps$. TOF information considerably improves image quality, contrast and signal-to-noise ratio (Vandenberghe et al., 2016).

2.3 Image reconstruction

PET raw data are stored into two-dimensional matrices called the sinograms (fig 2.4). Each pair of detectors records the line or projection integral, i.e. the sum of the coincident events along the respective LOR, which is identifiable by two parameters: the perpendicular distance from the center of the FOV (r) and an angle (θ).

The goal of the reconstruction phase is to recover the radiopharmaceutical distribution starting from the sinogram; that is taking the synogram data and delivering a three-dimensional matrix whose element (called *voxel* [1]) is located in the appropriate anatomical

[1] A *voxel* (volumetric image element) is a volume element representing an intensity in a three-

position, and whose value gives information about the radiopharmaceutical uptake.

There are two main reconstruction methods: analytical algorithms, in which the mathematical model is analytically inverted (such as the Filtered backprojection, FBP), and iterative algorithms (such as the Ordered Subset Expectation Maximization, OSEM), in which the tomographic image is reconstructed using iterative statistical methods (Bailey et al., 2005; Hudson and Larkin, 1994). Given the proven superior performance of iterative methods, we shall only consider these for a brief introduction.

2.3.1 Iterative approach

Iterative methods estimate a series of tentative radioactivity distributions and compares the respective projections with those actually acquired, refining the former at each iteration until correspondence is satisfactory (Madsen, 2005). While computationally more expensive, it allows to modulate and account for the statistical fluctuations associated with noise, both on the reconstructed images and on the raw data side (Herman, 2009).

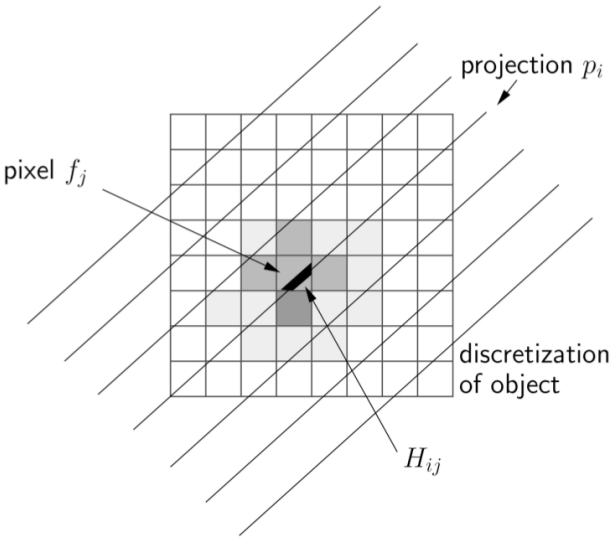

Figure 2.5: Pictorial representation of eq. 2.3.1 (courtesy of (Alessio and Kinahan, 2006))

In addition, estimation models can be complemented with a-priori information, such as the CT acquisition, which enhance the spatial resolution and greatly improves the reconstruction process.

Iterative methods employ two algorithms: one evaluates the projections of the estimated object versus the acquired one, and another that makes the new estimate based on these differences. The goal is to find $\bar{f}$ as solution of the equation

$$\vec{p} = \mathbf{H} \cdot \vec{f} + \vec{n} \tag{2.2}$$

where $\vec{p}$ is the vector of counts represented as a sinogram, $\mathbf{H}$ is a coefficient matrix, $\vec{f}$ is the unknown image vector and $\vec{n}$ is the noise.

An optimized solution method for eq. 2.2 is known as *maximum likelihood expectation maximization* (MLEM) (Dempster et al., 1977; Bruyant, 2002). This method exploits the fact that the processes of radioactive disintegration are described by the Poisson distribution (Hudson and Larkin, 1994). The MLEM algorithm finds the best estimate of the counts for each element of the image (henceforth labeled as f_k). This operation is performed by looking for, and mapping the average number of disintegrations that can most likely determine the counts acquired onto the projections (p_i).

Iterative methods consist of five basic steps (Alessio and Kinahan, 2006):

- the image model, which usually consists in the discretization of the space domain in N pixel voxel depending on the approach (2D or 3D);

dimensional space, similar to the *pixel* which represents a image element in a two-dimensional space.

- the system model **H**, that links the tentative image with the projections; the H_{ik} element of the system model is characteristic of the imaging system and represents the probability that an output from the volume element k is revealed in the projection i;
- the statistical model, which describes the relationship linking the measurements to their expected values. Most approaches consider photon detections to follow the Poisson statistics;
- the metric, that is the measure used to define the "best" image
- the maximum likelihood estimator, which optimizes the likelihood function L over the parameters to deliver the best estimate of the image at each iteration.

Equation 2.2 is expanded – neglecting noise – to write $\bar{p}_i$ explicitly as the average counts on the $i-th$ projection and f_k as the activity of element k (see figure 2.5)

$$\bar{p}_i = \sum_{k=1}^{K} H_{ik}\, f_k$$

Assuming every p_i is an independent Poisson variable, the likelihood (L) to observe $\vec{p}$ when the emission map is $\vec{f}$ is written as the product of the probabilities associated with each bin of the synogram as:

$$L(P = \vec{p}|\vec{f}) = \prod_{i=1}^{M} e^{-\bar{p}_i} \frac{\bar{p}_i{}^{p_i}}{p_i!} \tag{2.3}$$

where M is the total number of projections.

Maximum likelihood estimators are advantageous because they offer unbiased, minimum variance estimates as the number of measurements increases towards infinity. This means that as the number of measurements or projections becomes large the expected value of the image estimate approaches the true image. Due to the inherent noise in emission photon counting, methods often choose to allow some bias in the reconstructed image in return for reduced noise levels. This bias is introduced in the form of spatial smoothing effectively adding error to the image mean values while reducing overall noise levels. This smoothing is performed either implicitly by stopping the algorithm before reaching the ML solution or explicitly through some smoothing operation.

The maximum likelihood solution of 2.3 gives the iterative equation:

$$\hat{f}_j{}^{(m+1)} = \frac{\hat{f}_j{}^{(m)}}{\sum_{i=1}^{n} H_{ij}} \left(\sum_{i=1}^{n} H_{ij} \frac{p_i}{\sum_{k=1}^{K} H_{ik}\, \hat{f}_k{}^{(m)}} \right)$$

where $\hat{f}_j{}^{(m+1)}$ is the m+1 iteration estimate of voxel j in iteration m.

The MLEM algorithm usually requires from 20 to 50 iterations to reach an acceptable solution, starting from an initial estimate $\hat{f}_j{}^{(0)}$ (which is often a constant image).

Since MLEM requires a projection and a rear projection at each iteration, the total processing time is considerably greater than that required by any analytical algorithm. To improve on this, current standards employ the ordered-subset version (OSEM) of the MLEM. This slight modification of ML-EM uses subsets of the entire data set (for instance by dividing the projections into sets with different views, or azimuthal angles) to ensure a rapid convergence of the algorithm (Hudson and Larkin, 1994).

The OSEM methods are a family of algorithms where bias, a-priori knowledge, system-specific issues and statistical models can vary quite a bit, so that not all algorithms are alike in terms of performance and image quality.

2.3.2 The point spread function

Many common figures of merit are used to describe the spatial resolution of reconstruction methods. Resolution is loosely defined as the level of reproduction of spatial detail in the imaging system.

This is often expressed though the point spread function (PSF). Imagine a single point source is placed in the system and reconstructed into an image: the reconstructed image is a spread out version of the point. In PET, the PSF is frequently characterized by the transaxial and axial Full Width at Half Maximum (FWHM) term, which describes the width at half of the maximum value of the PSF of a Gaussian shaped function fitted to the PSF (in the transaxial or axial direction). This term captures some of the behavior of the resolution of the method, but does not do a good job describing the lower values of the PSF, considering the Gaussian function has very long tails. Other descriptive resolution values include the Full Width at Tenth Maximum (FWTM, computed similarly as FWHM) and properties of the modulation transfer function (Fourier transform of PSF).

Considering that the smoothing required to reduce the variance in images leads to degraded resolution, the bias-variance trade-off also translates to a inherent resolution-variance trade-off. Therefore, it is not adequate to simply state method A yields better resolution than method B, but must be shown that at an equal variance level, method A has better resolution than method B. While it is advantageous to have an imaging system with as low a FWHM as possible, this often affects the system sensitivity (count rate).

The PSF can also be derived analytically (Schmitt et al., 1988), with a Monte-Carlo simulation (Alessio et al., 2006) or by direct measurement (Panin et al., 2006). The inclusion of the PSF in the reconstruction methods improves the spatial resolution as well as the contrast (Panin et al., 2006).

2.3.3 Partial volume effect

The PSF concept is closely tied to another PET image issue: the *partial volume*(PV) effect (figure 2.6). PV refers to the fact that a pixel includes a mixture of signals coming from different sources. Two phenomena cause this mixture: the limited spatial resolution of the imaging system and the image sampling. Owing to the image blur introduced by limited spatial resolution, the signal coming from a point source will be detected not only in one pixel, but also in neighboring pixels. When considering a functional structure of interest, part of the activity from the structure will thus also be detected outside of a VOI drawn around the structure, an effect that is called spill-out. Conversely, activity

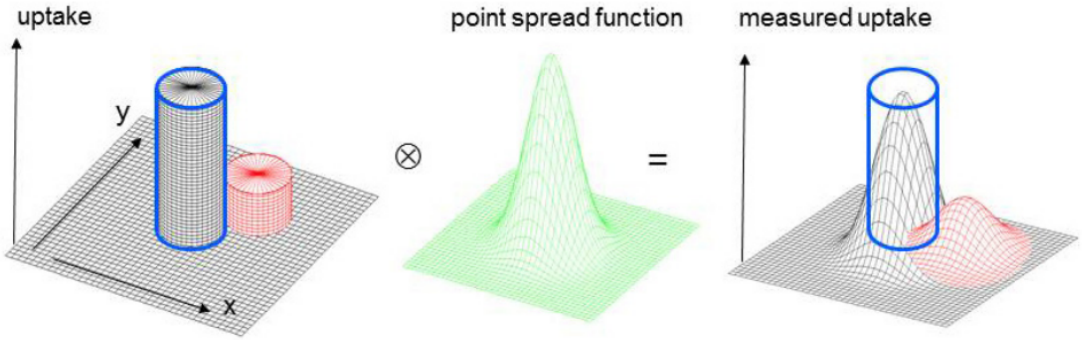

Figure 2.6: Partial volume effect due to the spatial resolution of the imaging system.

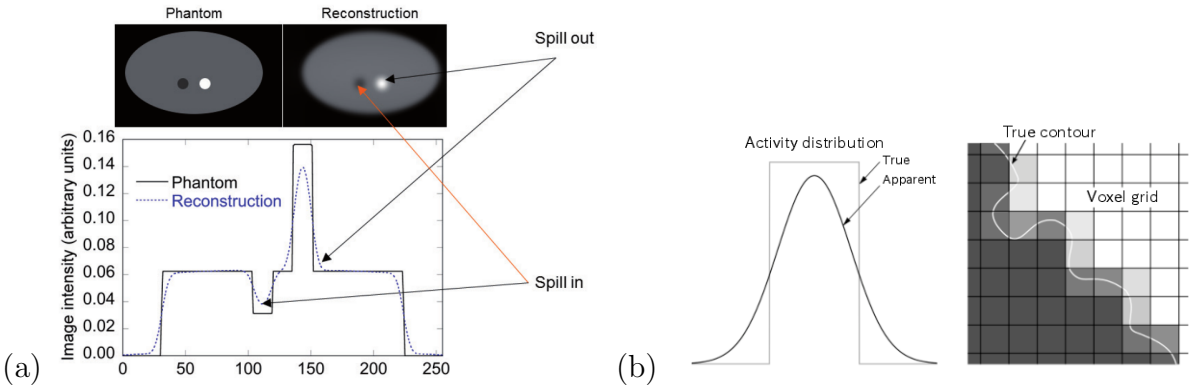

Figure 2.7: (a) Illustration of spill-out and spill-in PV effects. Spill-out results in estimating reduced activity in a region that has higher activity than its surroundings; spill-in results in overestimating the activity in a region that has lower activity than its surroundings. (b) This effect has also a significant impact on the contrast.

from structures close to the structure of interest will spill-in to the structure of interest. Therefore, if one considers only the most intense pixel to estimate the source activity, this activity will be underestimated but this underestimation might be partially compensated for by spill-in. Activity in neighboring pixels will be overestimated. Overall, PV thus introduces quantitative biases. (Matsubara et al., 2016). Both effects are shown in figure 2.7.

2.3.4 Image quality

The definition of image quality in nuclear medicine (and hence that of 'best' image) is still rather open. An objective comparison of image quality is often difficult and can only be performed in the context of a specific application, or task. The working definition is that image quality must be assessed on the basis of average performance of some task of interest by some observer or decision maker (Barrett, 1990).

Two of the major imaging task categories are classification tasks and estimation tasks. Classification tasks categorize the image or features in the image into one or more classes. A common classification task is detection (a binary classification task). Some other examples of classification tasks include signal detection, image segmentation and diagnosis. On the other hand, estimation tasks seek numerical parameters from an imaging system such

as the quantitation of physiological parameters or the estimation of features for pattern recognition. Specific examples of estimation tasks include a cardiac ejection fraction or the value of tracer flux in a tissue compartment model.

Perhaps the most common task in nuclear medical applications is the detection of tumors. Since a physician assesses images for this classification task, the best method for evaluating reconstruction methods for detection is a human-observer study. Unfortunately, thorough human- observer studies are time consuming and often impractical because they require the evaluation of numerous reconstructed images and ideally numerous trained observers.

The theoretical approach to image quality shall not be investigated in this work. Its consequences shall be seldom looked into, although quality assessment will always be a tangible issue and sometimes a real nuisance for the quantification tasks. The practical definition of a “good” PET scan – and the minimal hardware and software requirements to achieve it – is outside the scope of this work and is currently debated in Imaging Study Groups of Nuclear Medicine Societies.

Chapter 3

Amyloid PET

The first amyloid-specific PET ligand was developed by Mathis and Klunk at Pittsburg University. It was derived by a modification of tioflavin-T, a fluorescent colorant used in neuropathology to mark amyloid plaques in brain tissue. The modified molecule (Pittsburgh Compound B, PIB) was marked with ^{11}C, and the first human studies have been carried out since 2002 in collaboration with the Uppsala University (Sweden). PIB has since been comprehensively studied at many institutions around the world that have compared the amount of amyloid in the brain.

Numerous studies are now available which look at different aspects of ^{11}C-PIB PET: from tracer build-up as a function of age to the comparison across different disease states. They have also investigated patients who were amyloid positive and followed them over time to see whether they develop typical clinical patterns of Alzheimers disease. So amyloid imaging has been well established as a robust biomarker of the presence of abnormal amyloid deposition in the brain, which is one of the key pathologic features of AD.

Based on the results with PIB, there was a great deal of interest in development of PET agents labeled with ^{18}F. It first began with Avid radiopharmaceuticals (Eli Lilly), which received U.S. Food and Drug Administration (FDA) approval in 2012 for the first PET imaging agent, Amyvid. There have been two subsequent approvals: GEs Vizamyl, and Piramals Neuraceq.

Because of considerable clinical interest in amyloid imaging, the Society of Nuclear Medicine and Molecular Imaging and the Alzheimers Association developed appropriate use criteria with well-defined indications for clearing up confusing clinical situations and aiding in patient management. The criteria also provided guidance on when use of amyloid imaging was inappropriate.

3.1 Radiotracers

The amyloid-specific radiopharmaceutical must possess some key properties: they must be a small lipophilic, non-toxic molecule, with a high affinity binding, specificity and selectivity for Aβ and reactive to relatively long-lived radioisotopes like ^{18}F or ^{11}C. In addition, no radio-labelled metabolites that bind to the brain should be formed.

The first and most studied molecule is the Pittsburgh Compound, a derived compound of the tioflavine-T. It is a small, neutral and quite lipophilic molecule with formula 2 –(4

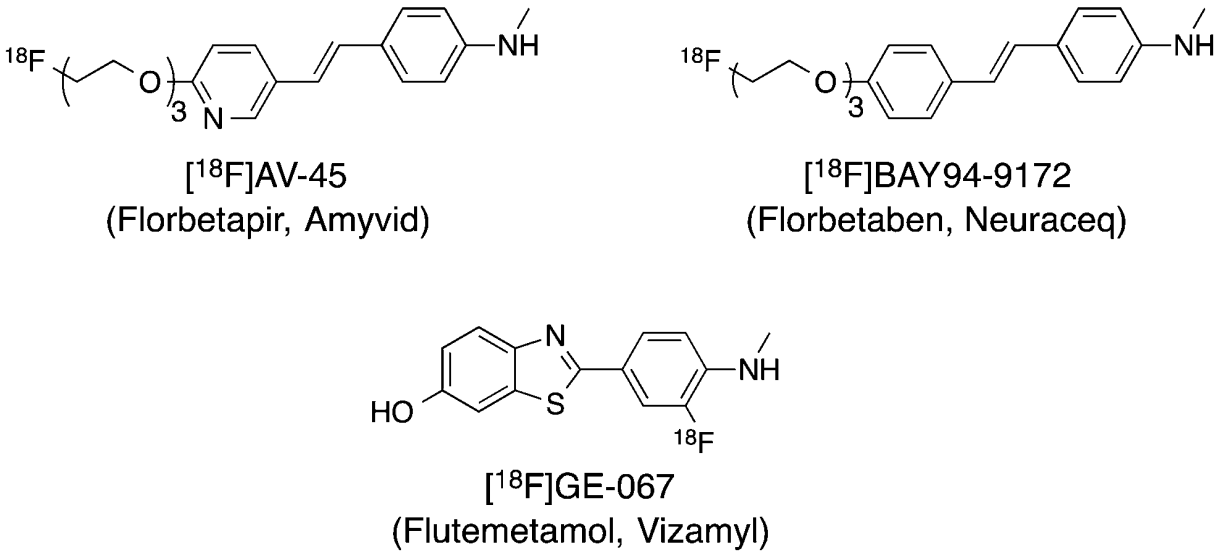

Figure 3.1: Fluorinated radiotracers

–[11C] –(methylamino)phenyl) benzo[d]thiazol –6 –ol. Its fluorinated analog is the 2 -(3 -[18F] -fluoro -4 -(methylamino)phenyl)benzo[d]thiazol -6 -ol, known with the name of ^{18}F–Flutemetamol. Finally there are two more compounds derived from the Polietilen Glicole Stilbene: (E) -4 -(2 -(6 -(2 -(2 -(2 -18F -fluoroethoxy) ethoxy) ethoxy) pyridine -3 -yl) vinyl) -N -methylbenzenamine, ^{18}F-Florbetapir and ^{18}F-Florbetaben (figure 3.1).

Overall, binding affinity and lipophilicity are the most important properties for in vivo tracing of neuroimaging. Although lipophilicity is needed for the tracer to cross the blood-brain barrier (BBB), if the radiopharmaceutical is too lipophilic it may deliver signal (i.e., binding) in sites other than those containing Aβ.

Preclinical and autoptic studies have shown that the cerebral cortical uptake of in-vivo tracers is correlated with the load of neuritic plaques observed post-mortem in the same regions. Tracers rapidly enter the brain through passive diffusion of the BBB and the fraction reaching the site of interest is very high, mostly depending on the regional blood flow. An important source of variability in the extent of the amyloid load stained by the different tracers derives from the time the different tracers spend in the white matter and the absorption in the gray matter. While the high affinity binding to amyloid fibrils results in a much slower clearance from the amyloid plaques in the gray substance, clearance from the white substance depends on the degree to which non-specific low-affinity links occur; for example, more lipophilic tracers may have a high distribution volume in the lipid-rich white substance.

3.2 Appropriate clinical use

Several Working Groups worldwide considered a range of clinical scenarios in which amyloid PET should be recommended. The guidelines are based on peer-reviewed, published literature, to ascertain available evidence relevant to the recommendations.

Although empirical evidence of impact on clinical outcomes is not yet available, a set of specific recommended use criteria were agreed to define the types of patients and clinical

circumstances in which amyloid PET could (and should not) be used.

As both dementia care and amyloid-PET technology are in active development, recommendations are subject of periodic reassessment. The reader is referred to (Johnson et al., 2013a) and (Guerra et al., 2015) for more insight into the criteria.

To date, a succint review of the guidelines implies that amyloid-PET is appropriate when these conditions are satisfied:

- a cognitive complaint with objectively confirmed impairment is present;
- Alzheimers disease as a possible diagnosis, but when the diagnosis is uncertain after a comprehensive evaluation by a dementia expert;
- when knowledge of the presence or absence of amyloid-beta pathology is expected to increase diagnostic certainty and alter management.

Typical patients are those:

- with persistent or progressive unexplained mild cognitive impairment;
- satisfying core clinical criteria for possible Alzheimers disease because of unclear clinical presentation, either atypical clinical course or etiologically mixed presentation;
- with progressive dementia and atypically early age of onset (usually defined as 65 years or less in age).

Conversely, amyloid imaging is inappropriate in the following situations:

- patients with core clinical criteria for probable Alzheimers disease with typical age of onset;
- to determine dementia severity;
- as surrogate of genotyping for suspected autosomal mutation carriers;
- in asymptomatic individuals;
- non-medical usage (e.g. legal, insurance coverage, or employment screening);

The natural extension of the appropriateness guidelines consists in contextualizing the amyloid PET into the general diagnostic tree and the decision-making process. The guidelines therefore should be seen outside the nuclear medicine setting and should span a far broader scope involving neurologists, dementia experts and neurophychologists. For instance, dementia experts – and in general the clinical team that deals with these patients – should provide proper documentation to demonstrate the medical necessity of an amyloid PET scan.

In addition, all pertinent clinical expertises should help in identifying a specific subset of individuals with mild cognitive impairment for whom an amyloid PET scan is appropriate. Finally, it is the greater health policy infrastructure (in charge of developing educational programs) who must increase the awareness on the amyloid PET appropriate use, and who is in charge of providing instructions on how this test should be used in the clinical decision-making process.

Tracer	Dose	Time after injection (min)	Acquisition (min)
^{18}F-Florbetapir	370 MBq (10 mCi)	30–50	10
^{18}F-Flutemetamol	185 MBq (5 mCi)	90	20
^{18}F-Florbetaben	300 MBq (8 mCi)	90	20

Table 3.1: Injection parameters.

3.3 Evaluation in clinics

The administration of the tracer and the reading of the PET scans must be performed by a trained nuclear medicine doctor, as requested and recommended by the respective companies (table 3.1). Criteria for interpretation may differ among the available tracers and readers should apply the specific FDA and EMA recommendations for a given ligand.

Although the assessment should be given independently from previous neuroimaging and clinical data, in general one should consider:

- the clinical question;
- possible comorbidities and alterations of the cerebral structure
- the acquisition protocol and the scanner characteristics
- the physiological distribution of the tracer
- the anatomical location of the altered concentration
- the (optional) use of semiquantitative analysis
- correlation with other clinical and instrumental data
- possible causes of false negatives / positives

The examination also depends on a the image quality. While doctors should consider the efficiency and characteristics of their own scanner, recommended minimum parameters are: 140 kV, 60 - 100 mA, scan time $>$ 10 s for CT; Field of View of 30 cm, 256×256 (pixel 1-2 mm) matrix for PET; reconstruction algorithm: iterative methods (OSEM or similar) with a Gaussian filter FWHM 1-2 mm; resolution recovery protocols should be used when available. If the patient is not collaborative, it is suggested that the acquisition be performed in "list-mode", with a-posteriori reconstruction and with appropriate time-frames selection.

3.3.1 Visual assessment

The maximum uptake of radiotracer at the encephalic district occurs within a few minutes after injection, followed by rapid clearance in the following 30 minutes.

Healthy control subjects show relatively low levels of accumulation in the cortical and cerebellar gray matter (with significant contrast between the white and the gray substance). Regional analyses show slightly higher levels of accumulation in the caudate, in the putamen and in the hippocampus.

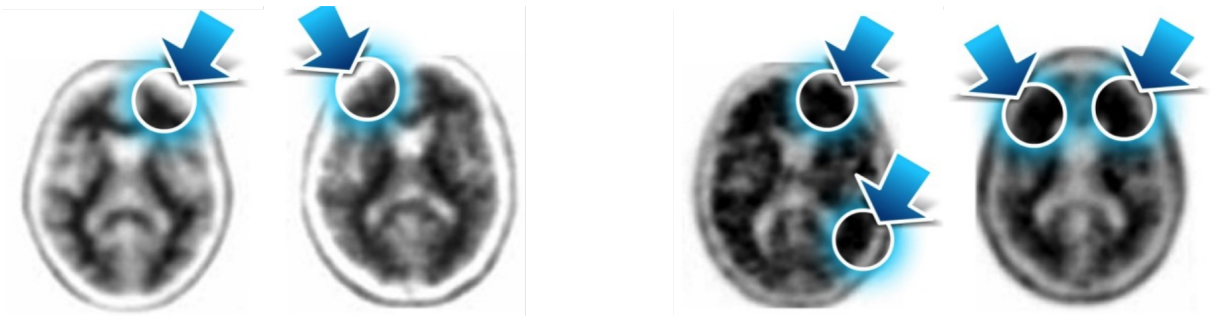

Figure 3.2: Transaxial sections of amyloid PET scans showing WM/GM contrast for a "textbook" negative (left) and a positive case (right). Arrows mark some of the most usual areas to look for contrast loss. Color scale is inverse (dark/light colors = high/low uptake)

The typical negative tests present high uptake only in correspondence of the white matter, and maintain white / gray contrast in the whole brain. The typical positive test exhibit two or more areas (lobes or significant portion of them) of reduced contrast between the uptake of the white matter and that of cortical gray matter. The typical reference region for contrast comparison is the cerebellum. In this site the contrast between white and gray matter is generally maintained even in patients with AD diagnosis.

The assessment of the contrast must be carried out on all lobes (temporal, occipital, frontal and parietal). An examination is declared as positive when at least two lobes show significant contrast reduction.

The cerebellum generally maintains the same contrast between WM and GM, both in patients with low and with high Aβ (Nelissen et al., 2009). For this reason, the contrast measured in the cerebellum is usually kept as a reference.

An appropriate gray/color scale and intensity reference should be used in order to facilitate the constrast assessment. Manufacturers recommend either color or gray palette depending on the tracer (for instance, for Flutemetamol it is recommended to set the scale intensity at $80 - 90\%$ in the pons region).

Negative scans show relatively low levels of tracer accumulation located in the cortical and cerebellar gray matter, with a significant contrast between the high-intensity uptake of the WM and the low uptake of the GM. So the typical negative test shows high uptake only at WM, where the tracer binding is non-specific. In contrast, the typical positive test shows two or more areas of reduced contrast between the uptake in WM in the cortical GM (figure 3.2). Thus, a key feature to distinguish Aβ positive from negative scans is the loss of contrast gray / white matter, with the uptake of the tracer extending to the edge of the cerebral cortex. The gaps between the two hemispheres can no longer be defined. The scan is evaluated as positive in the presence of at least two lobes in which there is an evident loss of contrast between WM and GM (Eli Lilly Inc., 2012; GEhealthcare, 2012).

The contrast assessment must be performed in all the lobes (temporal, occipital, frontal and parietal) starting from the cerebellum, to get a clear picture of the base contrast in the reference region.

There are numerous elements to be checked in the visual evaluation of amyloid PET images, in addition to the contrast between the WM and GM signals.

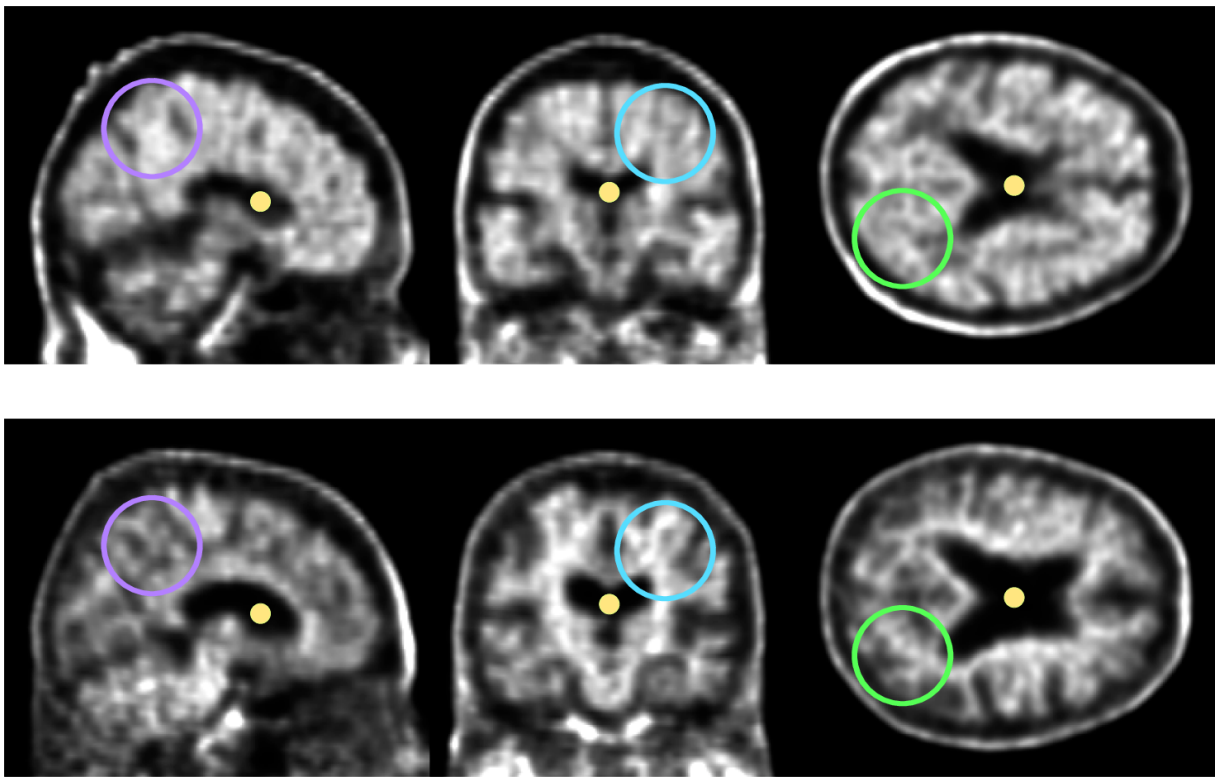

Figure 3.3: Real example of a positive (top) and negative (bottom) case. Contrast is less pronounced than the "textbook" example. Orthogonal planes examination show the precuneus (purple), fronto-orbital (azure) and occipital (green) regions.

In principle each tracer has its own positivity criteria, although most of them overlap and, in general, the outcome of the examination does not depend on the specific choice of the criteria. Particular attention should be paid to the analysis of regions known to be typical sites of accumulation in AD: fronto-orbital, precuneus and cingulate posterior, lateral temporal, parietal and caudate nuclei. It is appropriate that these areas be carefully examined on the three orthogonal planes.

3.3.2 Limitations

A confounding factor to be taken into account is the presence of atrophy, where the uptake is affected by PVE (Schmidt et al., 2015). The evaluation of scans with significant global or local atrophy is difficult and cannot in general be supported by the amyloid PET alone. For this reason it is recommended to have an recent MRI scan to complement the reading with morphological information.

Another difficulty is due to the peculiar anatomy of some patients, often due to excessive ventricles dilatation or due to traumas, vascular complications or past neurosurgery, which make the association of the signal to the various brain structures particularly complicated.

Other difficulties may arise from noisy images, artifacts or inappropriate reconstruction (blurring) (Schmidt et al., 2015). Suboptimal acquisitions should always be avoided as they can compromise the examination in all but the most trivial cases.

An additional variable to consider in the visual evaluation is the different presentation of the available tracers: in facts, tracers behave slightly different from each other, especially with respect to the non specific uptake in the WM and the clearance in both GM and WM (Schmidt et al., 2015). Because of these differences, clinicians are recommended to attend specific courses on each of the tracers they intend to use.

Visual evaluation is a fast and widely applicable procedure and most scans are easily evaluated and categorized into negative / positive after a proper visual assessment. Still, there is a fraction of cases ($10 - 20\%$, but it can be as high as 30% depending on the inclusion criteria) where this qualitative approach is not trivial.

3.4 Relationship with liquor-based amyloidosis markers

Two Aβ measure modalities have been established: cerebro-spinal fluid (CSF) Aβ42 dosage and amyloid PET. Both highly correlate with brain biopsy findings, and this has led to consider both amyloid PET and liquor biomarkers as pathophysiological or diagnostic markers of disease.

A potential advantage of amyloid PET over CSF Aβ42 as an early diagnostic marker is the relative non-invasivity of the procedure and the possibility to detect regional Aβ depositions that might occur before the global neocortical signal becomes pathologic. On the other hand, CSF analysis has the advantages that it may easily incorporate assessments such as *tau* (a measure of neuronal degeneration (Blennow et al., 2010)) and hyperphosphorylated tau (p-tau; a potential marker of tau pathology (Buerger et al., 2006)).

There is a significant agreement between amyloid PET and β-amyloid liquor measurements, and the liquor biomarkers show a diagnostic accuracy similar to that of amyloid PET in the diagnosis of AD (Fagan et al., 2006). However, the liquor values of Aβ42 and the results of amyloid PET show a nonlinear association.

Several studies have examined the link between amyloid PET and CSF Aβ42 (see for instance Fagan et al. (2006); Grimmer et al. (2009); Palmqvist et al. (2015)). For instance, while mean CSF Aβ42 is decreased in AD, this decrease may reflect plaques acting as an Aβ42 "sink", hindering transport of soluble Aβ42 between brain and CSF.

In terms of discrimination potential for MCI due to AD, the diagnostic accuracy of CSF and PET biomarkers is similar, although CSF markers show an earlier response (Palmqvist et al., 2016). The most frequent cases of mismatch are those in which amyloid PET is negative, compared to pathological values of Aβ42 in the liquor. This effect underlines the greater precocity of CSF positivization - which captures all the aspects of amyloid mis-metabolism - with respect to amyloid PET, which instead shows the cerebral deposits of amyloid when they already are in the fibrillar phase.

These results therefore suggest that the two biomarkers measure different aspects of the amyloid pathology. In a longitudinal perspective, the liquor Aβ42 and the amyloid PET show different trajectories along the course of the disease, where the amyloid liquor reaches the plateau early, while the signal of the amyloid PET continues to increase dynamically even in the later stages of disease (Palmqvist et al., 2016; Mattsson et al., 2014).

This is of great importance in a clinical setting, because the Aβ42 is a more sensitive

marker in the earliest stages of the disease, while amyloid PET becomes a more sensitive marker with the progression of the disease, even in the asymptomatic or paucisymptomatic phase. Accordingly, the Aβ42 is closely linked to the status of ApoE ϵ4 carrier, whereas amyloid PET is more related to t-tau and p-tau in the liquor.

The Aβ42 /40 ratio is a more accurate measure of amyloidosis than just Aβ42 since it reduces the variability due to the overall production and elimination of amyloid, thus significantly reducing the "false positives" and "false negatives" obtained by measuring only the Aβ42 . This is confirmed by consistent evidence showing a greater concordance between the Aβ42 /40 liquor ratio and the amyloid PET compared to the Aβ42 alone. It is obvious, however, that the outcome of these correlations strictly depends on the system used to quantify both the amyloid of the liquor and the cerebral amyloidosis seen on PET.

In a routine clinical setting, and in the context of early diagnosis of AD, it may be important to answer the question: which biomarker is better / which marker should be performed first. Liquor analysis has the advantage of including both amyloidosis and neurodegeneration parameters, it requires less complex and less expensive tools than amyloid PET, but it requires a relatively invasive procedure such as rachicentesis.

Amyloid PET is less invasive and appears to have greater reliability in longitudinal evaluations and among different centers. To date, the main criterion of choice is the effective availability of a reference center with the required expertise and that regularly uses one of these markers.

Whatever the relationship between cerebral amyloidosis measured with CSF and PET (linear or nonlinear), it is evident that both modalities are expected to provide markedly altered results in already symptomatic patients (with MCI or mild dementia), if the pathology in question is AD.

Chapter 4

Quantification and semi-quantification

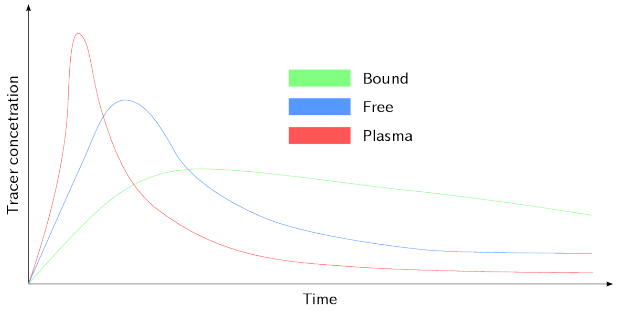

Figure 4.1: Qualitative representation of the tracer activity as function of time in three compartments or binding states: bound (the state of interest, specific binding), the plasma concentration (used in the Arterial Input Function) and the free (or non specific) state.

PET quantification basically consists in applying an appropriate mathematical model to the collected data, in order to extract numerical estimates of the parameters linked to the physiological processes. This approach brings great advantages compared to the qualitative evaluation, as it allows to estimate the quantity of ligand bound in a specific state and its spatial distribution. At any given time after the tracer injection, the tracer can be detected in several different bound states, whose proportions are a function of the radiopharmaceutical kinetics and of the biochemistry involved in the transport. So the typical intensity collected by PET is a mix of overlapping signals, of which only one is of interest. To isolate the desired signal component – in our case the one given by the concentration of the tracer bound to the aggregates of Aβ – it is necessary to use a mathematical model that links the tracer dynamics to the intensity shown in the PET image. In the kinetic model, each of these layers is known as a compartment (figure 4.1).

The gold standard for the quantification is the complete kinetic model with arterial blood sample: the scan must be acquired as a continuous series of snapshots with short temporal intervals, starting from the injection and until the concentrations reaches an equilibrium (dynamic scanning). This protocol allows to measure the time-activity curve, that is the trend of the radioactivity signal in every point (Schmidt et al., 2015).

The quantification (or absolute quantification) then represents the ideal method for the in vivo measurement of the specific binding of a PET tracer to the molecular target to be studied (Ichise et al., 2001). It is based on the identification of the compartment model that best describes the states of the radiotracer, from the blood compartment (site of administration) to the molecular target of interest at the cerebral level.

In general, an absolute quantitative model allows to measure not only a simple uptake or concentration but a series of parameters that correlate with the density of the receptors and their affinity (Binding Potential) without the influence of the variations of regional cerebral blood flow (rCBF) or other factors that can modify the retention and brain wash-out.

Dynamic scanning must include the few minutes that occur after the injection, during which the signal is dominated by the tracer flow from the plasma to the tissues, and it is highly correlated with the brain's regional blood flow (Gjedde et al., 2013). Then the intensity undergoes a non-equilibrium situation, during which the specific tracer binding approaches a plateau, followed by a period of relative equilibrium. The time required to achieve equilibrium depends on many factors, such as the type of tracer, the regional blood flow and the local amyloid load (Nelissen et al., 2009; Wong et al., 2010; Price et al., 2005). By knowing the concentration of the tracer in the plasma and its temporal trend it is possible to solve the system of differential equations which characterizes the flow among the compartments.

In order to reduce the invasiveness due to the arterial blood sampling, simplified models have been developed. These models exploit the intensity ratio between target regions and a reference one, which should exhibit negligible specific uptake. Moreover, an approach of this type allows the development of automated methods for analysis and allows the comparison between data collected in different centers.

4.1 Compartmental models

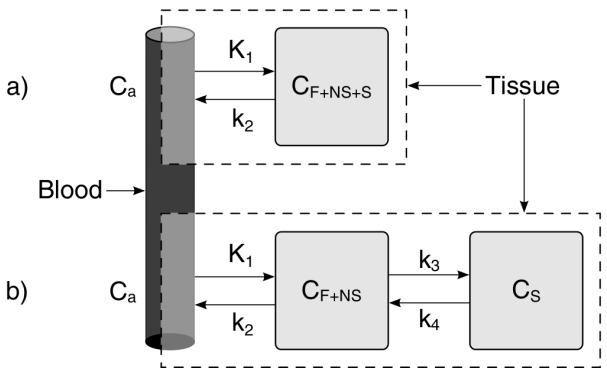

Figure 4.2: Example of 1-compartmental (a) and 2-compartmental models (b). Each compartment is characterized by the tracer concentration C in its free (F), non-specific (NS) and specific (S) binding state.

Compartmental approaches are mathematical models used for the description and quantification of the uptake and the clearance of tracers in tissues (Innis et al., 2007; Gunn et al., 2001; Watabe et al., 2006).

These models are used to describe systems that vary over time, but not in space, so that it is assumed that there are no concentration gradients within the sampled volume. In addition, the concentration of the tracer in the blood stream is treated as a input to the system, instead of being represented by a compartment. The tracer in the blood flux has a distinct time-activity curve, called arterial input function (AIF). Its typical representation is shown in figure 4.1.

Under these hypotheses, the typical compartmental model takes into consideration three states: a free compartment, in which the tracer has passed the blood-brain barrier, but is not yet bound; the bound compartment, in which the tracer is specifically linked to the amyloid aggregates; a compartment in which the tracer is bound, but not to the amyloid (a-specific bound state) (Morris et al.

(2004), see figure 4.2).

Each compartment is characterized by the concentration of the tracer as a function of time. These concentrations are connected to each other by a group of ordinary differential equations, which express the balance between the tracer mass entering and leaving each compartment. By solving these equations it is possible to determine the quantities of interest.

Tracer molecules are free to move among different compartments with rates determined by a number of constants k, describing the fraction of molecules traveling from one compartment to another per unit of time. The temporal trajectory $C_i(t)$ of the tracer concentrations C in the compartments i is the variable to be solved. More sophisticated models include three compartments, with the separate determination of the specific and non-specific binding. This latter model needs 6 rate constants.

Radioactive decay is generally neglected and tracer concentration at any given time is determined by the exchange rates with the adjacent compartments.

The critical data in compartmental models is the accurate knowledge if the AIF. This information can be acquired as:

- Arterial blood sampling: 20 to 30 samples of arterial blood are taken from the patient, initially at short time intervals, and then at a lower rate. The samples radioactivity is measured by an ad-hoc γ detector (Matsubara et al., 2016). This method is rather invasive and entails a lengthy procedure, two issues that typically discourage the arterial sampling in the clinical routine. Another limitation is the fact that the concentration of the radiopharmaceutical is not guaranteed to be the same both in the blood collection site (typically a peripheral artery) and in the measurement site (in our case, the brain); in addition there can also be a time-delay between the peripheral measure and the real AIF (Contractor et al., 2012).

- Normative data: in practice, and when the injection protocol is well known and enforced, patients sharing common characteristics (age, weight, sex, ...) exhibit similar AIF. The real curve therefore can be estimated from a normative dataset. The patient needs only to undergo a few blood samples, which are used to calibrate the normative population AIF. The main limitations of this technique consists in neglecting the individual characteristics. In addition, the position and height of the AIF peak may differ between this method and the previous one (Contractor et al., 2012).

- Image-based methods: if the heart or a major artery (such as the aorta) is in the FOV, the AIF can be derived from a region of interest (ROI) of the image (Ikoma et al., 2008). This method is non-invasive and easy to implement; unfortunately, due to the reduced size of the ROI from which the AIF is derived, it is essential to address the problem of the PVE. In addition, AIFs computed on images reconstructed with iterative methods (such as OSEM) could be plagued with systematic deviations (Heurling et al., 2015).

AIF is considered to be the same for all the analyzed volume (in our case for the whole brain), so that the same AIF is used fit the model on all ROIs.

For each compartment, the net flow is expressed as a rate equation in units of concentration per unit of time (dC/dt). The kinetic parameter K_1 makes an exception, which has

the unit of $ml\ g^{-1}\ min^{-1}$; it indicates the rate of tracer exchange between the arterial plasma and the unbound compartment of the tissues. In the simple 1-compartment model K_1 enters as

$$\frac{dC_{\text{F+NS+S}}}{dt} = K_1\, C_a - k_2\, C_{\text{F+NS+S}} \tag{4.1}$$

The tracer concentration in the arterial blood C_a is known and given (the AIF). Solving:

$$C_{\text{pet}} = C_{\text{F+NS+S}} = K_1\, C_a \otimes e^{k_2\, t}$$

where $\otimes$ is the temporal convolution (Gunn et al., 2001).

In a 2-compartment model – where one compartment represents the free and non-specific concentration, while the other represents the specific-binding concentration – the rate equations are:

$$\begin{array}{rcl} \frac{\mathrm{d}}{\mathrm{d}t} C_{\text{F+NS}} & = & K_1\, C_a - (k_2 + k_3)\, C_{\text{F+NS}} + k_4\, C_S \\ \frac{\mathrm{d}}{\mathrm{d}t} C_S & = & k_3\, C_{\text{F+NS}} - k_4\, C_S \end{array}$$

The signal coming from the two compartments cannot be separated in the PET scan (Gunn et al., 2001; Watabe et al., 2006) so that C_{pet} is the sum of the concentrations:

$$C_{\text{pet}} = C_{\text{F+NS}} + C_S = \frac{K_1}{\alpha_1 - \alpha_2} \left[(\alpha_1 - k_3 - k_4) e^{-\alpha_1\, t} - (\alpha_2 - k_3 - k_4) e^{-\alpha_2\, t} \right] \otimes C_a \tag{4.2}$$

and the α coefficients are:

$$\alpha_{1,2} = \frac{1}{2} \left(k_2 + k_3 + k_4 \pm \sqrt{(k_2 + k_3 + k_4)^2 - 2 k_2\, k_4} \right)$$

The calculation of the kinetic parameters can be directly addressed by solving the convolution integral with iterative algorithms (Marquardt, 1963), although there exist simplified techniques that reduce the computational cost, see for instance (Patlak et al., 1983; Logan et al., 1990a). These latter approaches attempt to transform a non-linear problem into a linear one, simplifying the regression procedure.

Other approaches do away with the AIF. The trick is to find a reference region for which the specific binding is small enough to be negligible; calculating the time activity curve (TAC) of this region from the image and comparing it to the TAC of relevant ROIs, it is possible to deduce the ratios between the parameters of the kinetic model (simplified reference tissue model, SRTM). The advantage of methods based on the reference region is to be able to avoid blood sampling and analysis of metabolites, thus avoiding all errors related to these procedures.

Typically the target region is represented by a two-compartment model, while the reference region is a single-compartment model (i.e. the amount of specific binding is considered to be negligible). The AIF in equation 4.2 is then replaced with the activity calculated in the reference tissue, directly obtaining the kinetic parameters (Ikoma et al.,

2008; Lammertsma et al., 1996).

Although these techniques show different advantages compared to those using AIF – especially in terms of invasiveness – it is necessary to use them with caution due to the assumptions implied in their methodology. For example, the existence of a specific binding of the tracer in the reference region causes an underestimation of the specific binding in the target region.

4.2 Quantification drawbacks and amyloid-PET specific issues

The main problem in absolute quantification methods is the general invasiveness and patient discomfort, which makes them impractical in a clinical routine. Besides the invasive procedure of AIF, all quantification techniques require a longer acquisition time (0-60/90 minutes, depending on the tracer). This increases patient discomfort and the concrete possibility of movement artifacts. In addition, quantification models not relying on AIF need some key assumptions to be verified, which in most cases can lead to errors in evaluating the kinetic parameters.

Another class of problems comes from the compartmentalized model, which often involves a simplification of the underlying physiological process. For instance, it is necessary to exclude the metabolites from the activity count otherwise one overestimates the AIF.

Previous studies in patients with AD dementia or MCI and in control subjects suggest that absolute modeling of amyloid-β accumulation in the brain is feasible for ^{11}C-PIB (Price et al., 2005; Lopresti et al., 2005), ^{18}F- Florbetapir (Joshi et al., 2012; Becker et al., 2013b) and ^{18}F- Flutemetamol (Nelissen et al., 2009; Heurling et al., 2015). In these studies it is shown that a two-compartment model (free concentration (F) + non-specific (NS) and specific binding (S)) and 4 constants (k_1-k_4) can fit experimental data and yield the BP or the total distribution volume. However, the same studies have proposed simplified methods, and they have validated them with the complete compartmental kinetic model. Each simplification though can introduce a bias. Simplifications involve the reduction of the acquisition time and the use of the input function obtained from PET images (ROI localized in the internal carotid artery) instead of arterial samples. The reduction of acquisition time seems to be more critical, while the use of simplified quantification such as Logan's graphical analysis (Logan et al., 1990b) and reference region (cerebellum, SRTM), are rather robust.

A third class of problems lies in the fact that parameter estimation is influenced by many factors related to image reconstruction (PVE for instance), acquisition methods and rCBF. This latter is a crucial point in particular in the study of neurodegenerative diseases because the reduction of rCBF can affect the radiotracer brain kinetics and result in an erroneous measurement of simplified parameters, such as the Standardized Uptake Value ratio (SUVr), as has been elegantly demonstrated for PIB in one longitudinal study (Becker et al., 2013a). Thanks to simulation procedures, it has been demonstrated that the reduction of SUVr observed in patients with AD during the 2-4 year follow-up, which is difficult to explain in the absence of anti-amyloid therapy, was due to the reduction of rCBF. The results were corroborated by the absence of reduction of the Binding Potential (BP) obtained by the application of a complete compartmental kinetic

model (Lammertsma, 2017).

It is clear that further studies are needed, that target absolute quantification in a large number of patients. This is particularly true for the validation of simplified methods.

The correct use of quantification in patients assumes an ethical value particularly in view of the upcoming anti-amyloid drugs, and in order to reduce the number of patients in trials, reduce radiation exposure, etc.

4.3 Semi-quantification

Another possible approach for the extraction of meaningful quantities from PET scans involves the use of the so-called semi-quantification methods (sometimes called relative-quantification). Compared to absolute quantification, semi-quantification represent a leaner alternative even though it is less accurate. The main advantage of semi-quantification methods is that they do not require dynamic scans or the AIF, the drawback is that they only provide simplified measures with higher uncertainties.

Semi-quantification methods rely on static scans, which consist of a single acquisition (typical duration ≈ 20 minutes) that starts after a relatively long time from the injection (depending on the tracer but with typical delays of ≥ 40 minutes). Static scans minimize the time the patient has to stay in the scanner, reducing the complexity of acquisition and analysis.

Most studies on the quantification of cerebral amyloid load in literature use semi-quantitative methods; from oncology (Lucignani et al., 2004) to the quantification of the cerebral amyloid load (Wong et al., 2010; van Berckel et al., 2013; Bourgeat et al., 2015; Camus et al., 2012; Klunk et al., 2015; Thurfjell et al., 2014). Among semi-quantification methods, the Standardized Uptake Value ratio (SUVr) is the most used and validated, compared to the binary reading.

The SUV is defined as

$$SUV = \frac{C_{\text{pet}}}{A/W} \tag{4.3}$$

where C_{pet} is the activity concentration in a ROI (measured by a calibrated PET scanner), A is the injected dose (in MBq) and W is the subject's weight. When computing the SUV, one often considers either the average or the maximum activity in the ROI. The ratio between two SUV is then simply

$$\text{SUVr} = \frac{\text{SUV}_{\text{target}}}{\text{SUV}_{\text{ref}}}$$

4.4 SUVr implementation

Following the visual evaluation guidelines for amyloid PET, it is clear that the uptake ROIs should be placed in cortical regions. The reference region choice though is quite arbitrary and there are several possible choices (Schmidt et al., 2015; Klunk et al., 2015): the encephalic trunk (Klunk et al., 2015), the cerebellum GM (Price et al., 2005; Klunk et al., 2015), or the entire cerebellum (both GM and WM) (Clark et al., 2011; Klunk

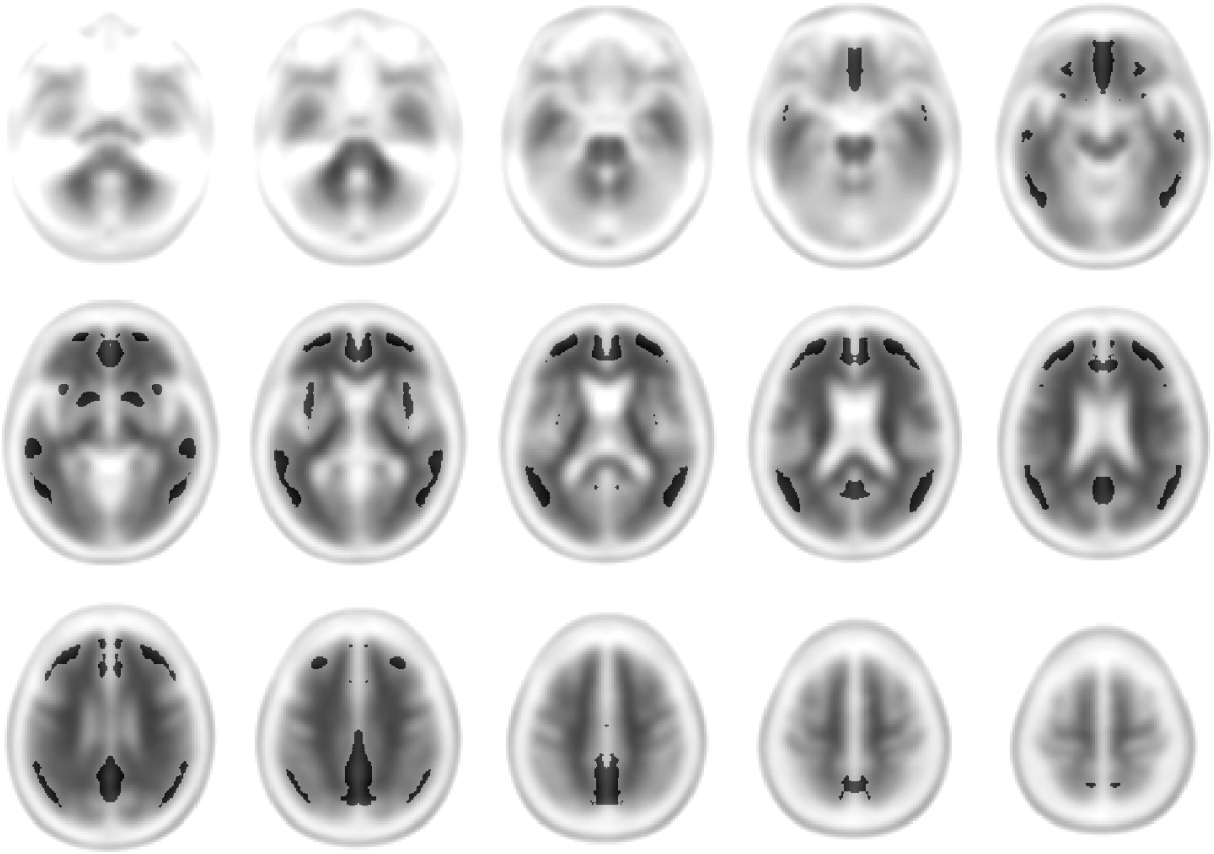

Figure 4.4: Uptake regions for the computation of SUVr values superimposed on the reference PET. The counts normalization region (not shown here) is the whole cerebellum.

et al., 2015; Mattsson et al., 2014).

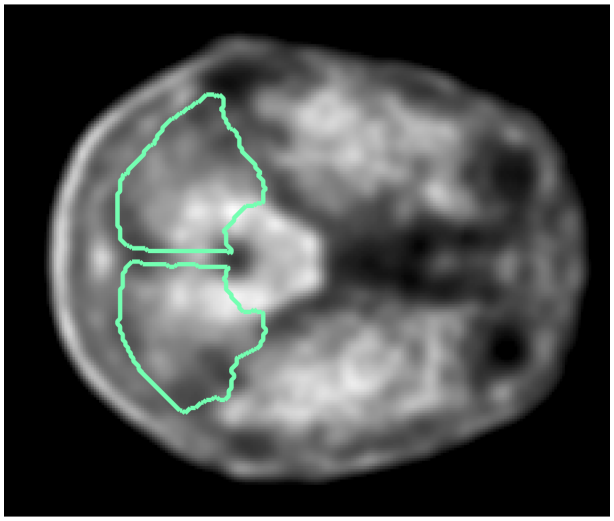

Figure 4.3: Sample axial section of the segmented cerebellar ROI

In this study we calculated the SUVr using the whole cerebellum as reference region (figure 4.3). The reason is that it is less prone to segmentation errors than the selection of the cerebellum gray matter alone or the brain stem, as reported in some works in literature (Schmidt et al., 2015).

We calculated the average cortico-cerebellar SUVr on all scans, to compare the ELBA score to this widely used semi-quantification method. We used a data-driven approach with the whole cerebellum (white and gray matter) as reference and a number of cortical ROIs as uptake regions, as displayed in figure 4.4.

The SUVr information was used neither in the blind nor in open validation phase.

The cortical regions were: medial frontal gyrus, lateral frontal cortex (middle frontal

Figure 4.5: Data-driven SUVr probability map.

gyrus), lateral temporal cortex (middle temporal gyrus), lateral parietal cortex (inferior parietal lobule), insula, caudate nucleus, and precuneus-posterior cingulate region. They were obtained by a data-driven approach similar to the one described in (Klunk et al., 2015).

Briefly, we took 50 negative and 50 positive subjects from the N and P set (i.e. those subjects tagged *negative* and *positive* independently by all readers). Each image was spatially normalized into MNI space and intensity normalized by the mean counts in the whole cerebellum. Then a positive and negative mean image was generated. The negative mean was subtracted from the positive one, the result was left-right symmetrized and smoothed with a 3D-Gaussian filter ($\sigma = 3$ mm, figure 4.5). We found an optimal threshold by maximization of the area under the ROC curve of the SUVr between the 50 positive and 50 negative scans.

4.5 SUVr: pros and cons

The SUVr normalizes the activity in a ROI (up to the single voxel) with the (average) activity in a reference region (reference ROI). The reference ROI is usually chosen among those regions for which there is minimal specific binding of the tracer. The approach used in the SUVr is a simplified version of the SRTM and it in general good agreement with it (Villemagne et al., 2011; Wong et al., 2010).

In amyloid PET, SUVr results were in good agreement with histopathological measures of the density of neuritic plaques (Clark et al., 2012; Thurfjell et al., 2014; Clark et al., 2011). In addition, by making an accurate choice of the cut-off – for example by maximizing the accuracy of the classification with the SUVr compared to that made by visual evaluation – binary classification based on SUVr was comparable with that obtained by CSF analysis (Mattsson et al., 2014). Finally, the thresholded SUVr evaluation proved to be comparable to the visual assessment (Thurfjell et al., 2014).

That said, there are some significant drawbacks in the SUVr method, the most prominent one is that SUVr values are quite variable. The heterogeneity of SUVr values (and the related cut-off) is due to many factors, such as the tracer used, the image acquisition parameters and the analysis methods. All can have a significant impact on the actual calculated values.

The selection of the reference region, for instance, is very delicate, both because of the anatomical choice and because of its segmentation, which can be responsible for significant uncertainty in the counts.

The different tracers are also markedly different with respect to specific and non-specific uptake and clearance. The SUVR values calculated from acquisitions obtained with the different tracers are therefore not compatible (Landau et al., 2014; Klunk et al., 2015). Obviously, also cut-off values are influenced by the tracer, even when using the same uptake and reference ROIs. The unfortunate result is that direct comparison among SUVr computed on scans acquired with different tracers is not possible.

SUVr values are also sensitive to the PVE: this effect tends to increase the activity in the GM with the non-specific high uptake of adjacent WM, thus causing loss of contrast between the two. Consequently, this effect leads to overestimation of SUVr values, especially in those subjects with low amyloid load and significant atrophy (Schmidt et al., 2015).

Another important parameter to consider in the SUVr measurement is the acquisition timing, that is the choice of the time interval in which to perform the PET scan. Signal acquisition should be performed when the balance between the uptake and the clearance of the tracer in the tissues has been reached. The achievement of this balance depends on the tracer, but it typically occurs after about 40 minutes from the injection of the radiopharmaceutical (Wong et al., 2010), therefore it is suggested to perform a static acquisition of 20 minutes after the achievement of the equilibrium (van Berckel et al., 2013; Klunk et al., 2015). In practice though, the real steady-state occurs at far longer times and it is not truly reached at the typical acquisition starting times. For instance, it is known that a delayed measurements results in a SUVr overestimation (Nelissen et al., 2009; van Berckel et al., 2013). When protocols are not precisely followed the variation in SUVr values can be significant. This is particularly important in multicentric studies, where each center might not follow some agreed guidelines on protocols for internal or practical reasons.

In addition, the use of SUVr in longitudinal studies has proved to be impaired by the multiple factors that can undermine this measure (Schmidt et al., 2015; van Berckel et al., 2013). In these studies, where the typical protocol involves a follow-up at 1–3 years, changes in amyloid load are expected to be small. Measures however, indicate that the variability of the signal in the reference region and the variability in regional blood flow can cause considerable uncertainty in SUVr values, resulting in a significant percent of subjects exhibiting a decrease in amyloid load (in absence of any treatment) (van Berckel et al., 2013). Indeed, the uptake and clearance of the tracer are related to the blood flow; this depends not only on the type of tissue and on inter-individual physiological differences, but also from advanced neurodegeneration (Zhang et al., 2014). Therefore some authors even suggest that longitudinal analysis with SUVr is not recommended (Villemagne et al., 2013).

As last remark, SUVr definition and implementation require accurate image registration and segmentation of the cortical and reference ROIs. In general the ROI placement can be a non trivial task due to the possible template choice, the image quality and reconstruction paramenters, and the transformation used to implement the registration. Even though some of these issues can be relatively compensated, analyses are prone to errors due to the PVE when using pre-defined ROIs, particularly on the small cortical regions.

Some studies improved on this issue by complementing the analysis with the subject MRI (T1-weighted, volumetric and isotropic scan) (Bourgeat et al., 2015). This multi-modal approach leads to good results at the expense of complexity and the availability of a volumetric MRI. At the far end of multi-modal studies we should mention the PET-MRI semiquantitative method implemented by (Cecchin et al., 2017). In this case, the SUVr values are corrected not only by atrophy (with the direct co-acquisition of a 3T-MRI) but also for the rCFB, thanks to the dynamic PET acquisition.

4.5.1 More on ROIs

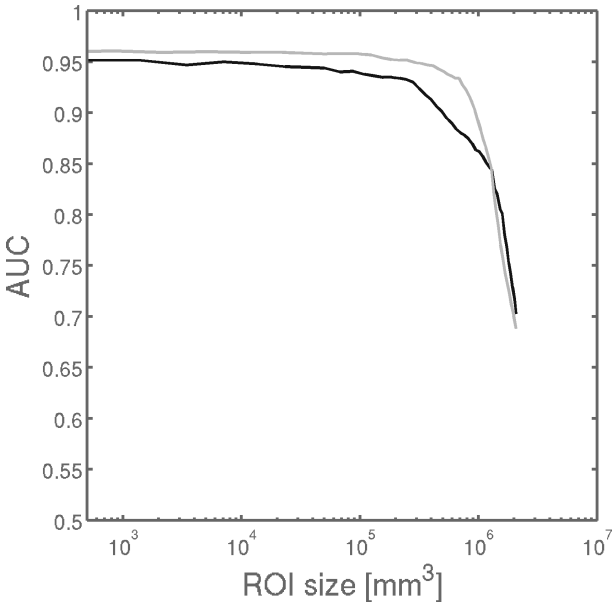

Figure 4.6: AUC for SUVr discrimination versus total ROI volume (mm^3) and registration technique (affine, black; non-linear, gray)

The selection of uptake ROIs even in absence of the subject's MRI is, in practice, less dramatic than this section leads to believe. When image registration to a template is sufficiently accurate, we can show that there is actually a rather broad range of selections, all providing comparable SUVr results. This is not to say that the actual SUVr values are comparable, as they indeed depend on the ROIs. Rather, it means that the discrimination ability – say – with respect to the visual assessment is maintained.

In this work I performed a ROC analysis study on SUVr discrimination ability by varying the image registration technique and the size of the cortical ROIs. The "gold standard" in this case was the visual evaluation (negative/positive) on 100 subjects (50 negative and 50 positive). The size of all the ROIs were adjusted by decreasing the threshold on the probability map (figure 4.5).

It turns out that whole-brain SUVr is quite accommodating in terms of ROI size and it gives consistent results over a very wide range of ROI total volumes (the x-axis in figure 4.6 is logaritmic). We also see that non-linear transformation (i.e. of the scan onto the PET template) gives significantly and consistently higher performances. This was actually to be expected as the non-linear spatial registration allows a more accurate placement of the anatomical districts onto the MNI space, from where we derive the SUVr ROIS.

4.6 Image analysis

Semi-quantitative analysis of amyloid PET images are prone to be implemented with automatic algorithms. Since the analysis generally consists in the evaluation of the intensity in specific anatomical regions, all images need to undergo some processing steps to make them comparable to each other (spatial and intensity normalization); without these steps

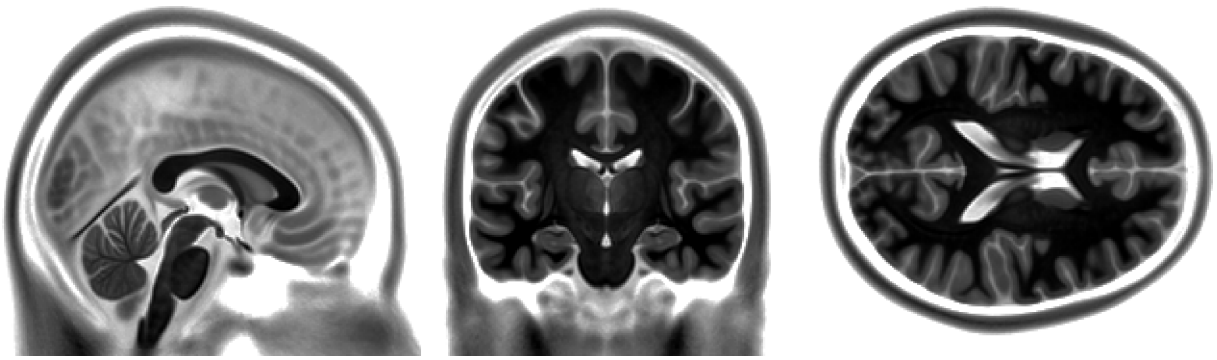

Figure 4.7: Sagittal, coronal and axial section of the MNI ICBM 152 template (MRI)

the signal would be dominated by noise due to different orientations, morphology and other physiological differences, and acquisition protocols. So all analysis presented in this work involve some preprocessing steps, in order to make the images comparable.

The first step is a quality control; reliable measure require good quality images, with reasonable count rates, positioning in the field of view (FOV) and optimized reconstruction algorithms. The quality control is – at present – a human-driven task. It was necessary to exclude extremely noisy images or images that showed important artifacts from the analysis.

The second step is the spatial normalization onto a normative space, to ensure the alignment of all brain structures and for all the subjects. Spatial normalization is the most important step to reduce the inter-individual variance. This is typically implemented with image registration techniques and the use of a reference template, mapped onto the normative space. The normative space used in this work is the Montreal Neurological Institute (MNI) reference.

After spatial normalization, an optional intensity one can be applied. This can improve certain processing steps by improving general conformity in the intensity of the analyzed group.

Finally, there are many choices for the extraction of relevant information. A very common strategy is to use ROIs, but in this work we shall see that other options are possible. Statistical and geometric tools can also be applied in order to reduce inter-individual variability or improve cohort-wise uniformity. These details will be given when appropriate in the following chapters.

In general, all scans in this work underwent the spatial and intensity normalization. I tried to use the bare minimum of more sophisticated pre-processing techniques with the intent to increase usability and understanding.

4.6.1 PET dataset and subject selection

Analyses in this work were performed using three dataset: the Alzheimer's Disease Neuroimaging Initiative (ADNI [1]); a selection of patients from the AVID-18 clinical trial,

[1] Alzheimer's Disease Neuroimaging Initiative (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the Na-

courtesy of Fondazione Poliambulanza Istituto Ospedaliero Brescia, Italy; and a multicentric pilot study of 147 subjects from 4 centers.

ADNI dataset

I downloaded ^{18}F-Florbetapir scans of 244 subjects from the the ADNI archive in the most fully processed format (series description in LONI Advanced Search: AV45 Coreg, AVG, Std Img. and Vox Siz, Uniform Resolution, subjects identification in table 4.4). Subjects were selected to have at least two scans (at baseline and after an approximately 2 years of follow-up), and 16 subjects came with three scans for a total of $n_i = 504$ PET images (i.e. 228 subject with two scans and 16 subjects with 3 scans). The ensemble properties of these images are shown in table 4.1.

Subjects clinical evaluation was taken to be the closest diagnosis to the baseline PET scan date. Cohorts were grouped by the ADNI core clinical criteria ((Aisen et al., 2010)) as: normal subjects (NS, N=70), early mild cognitively impaired (EMCI, N=86), mild cognitively impaired (MCI, N=26), late mild cognitively impaired (LMCI, N=51) and Alzheimer's diseases (AD, N=11).

FP dataset

This dataset consists of 40 subjects acquired during the AVID-18 clinical trial at one centre (Fondazione Poliambulanza Istituto Ospedaliero Brescia, Italy, FPA-BS) with ^{18}F-Florbetapir tracer and the following acquisition parameters: injected dose = 370 MBq, acquisition time = 10 min (50 min. after the injection), image reconstruction on a 256×256 matrix with 4 iterations, 21 subset, gaussian filter with FWHM = 2 mm.

These scans were assessed to be a mix of negative and positive scans (14 and 26 respectively), and the evaluation was visually confirmed by one of the expert readers (UG).

Multicentric pilot dataset (MPD)

A set of 147 subjects from a naturalistic population (age 54-87) were acquired in four clinical centers with ^{18}F-florbetapir PET, producing two scans per subject: an early (0-5 min) and a late scan (50-70 min). The participating institutions were: Nuclear Medicine Unit, Department of Health Sciences (DISSAL), University and IRCCS AOU

tional Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.;NeuroRx Research; Neurotrack Technologies;Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics.TheCanadianInstitutes of Health Research is providing funds to support ADNI clinical sites in Canada.Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, SanDiego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Table 4.1: Subjects and scanner ensemble statistics

Maker	Model	N. of scans	N. of subj	M/F	Age
SIEMENS	1093	16	8	3/5	77.3 (8.0)
SIEMENS	1094	34	21	6/15	70.4 (6.7)
Siemens/CTI	ACCEL	12	6	5/1	72.3 (7.7)
GEMS	Advance	16	8	3/5	73.9 (4.3)
Philips Medical Systems	Allegro Body(C)	4	4	3/1	74.0 (7.3)
SIEMENS	Biograph20 mCT	1	1	1/0	72.0 (0.0)
SIEMENS	Biograph64	5	5	2/3	73.5 (7.4)
GE MEDICAL SYSTEMS	Discovery 600	4	2	1/1	71.5 (11.2)
GE MEDICAL SYSTEMS	Discovery 710	2	2	0/2	68.2 (1.3)
GE MEDICAL SYSTEMS	Discovery LS	19	11	7/4	71.2 (7.2)
GE MEDICAL SYSTEMS	Discovery RX	8	4	2/2	71.5 (4.2)
GE MEDICAL SYSTEMS	Discovery ST	32	23	14/9	75.1 (6.6)
GE MEDICAL SYSTEMS	Discovery STE	74	43	31/12	72.6 (7.3)
Philips Medical Systems	GEMINI TF Big Bore	6	5	4/1	75.2 (6.6)
Philips Medical Systems	GEMINI TF TOF 16	25	12	5/7	71.5 (6.8)
Philips Medical Systems	GEMINI TF TOF 64	14	9	5/4	72.7 (13.6)
Philips Medical Systems	Guardian Body(C)	10	5	3/2	75.6 (5.1)
Siemens/CTI	HR+	103	54	26/28	72.9 (7.7)
Siemens ECAT	HRRT	59	26	14/12	73.5 (8.6)
Philips	Ingenuity TF PET/CT	3	3	2/1	70.6 (18.1)
CPS	LSO PET/CT HI-REZ	45	24	11/13	70.2 (7.6)
Philips Medical Systems	NULL	2	1	0/1	70.7 (0.0)
SIEMENS	SOMATOM Definition AS mCT	10	10	5/5	75.1 (7.3)
-	-	504	244†	128/116	72.6 (7.6)

†The number of subjects in the total is the number of unique subject identifiers. This is not the sum of the respective column because 38 subjects were scanned on a different system at baseline and follow-up.

Center	# subjects	Gender M / F	Age [95% CI]	Visual assessment Neg / Pos
FPA-BS	55	24 / 31	71 [69 72]	29 / 26
HUG-GE	44	20 / 24	72 [69 74]	27 / 17
FSM-PV	19	9 / 10	76 [73 79]	7 / 12
HSM-GE	29	15 / 14	75 [72 78]	10 / 19

Table 4.2: Multicentric Pilot Dataset demographics.

S.Martino-IST, Genova, Genova, Italy (HSM-GE, 29 subjects); Department of Nuclear Medicine, Fondazione Poliambulanza, Brescia, Italy (FPA-BS, 55 subjects); University Hospitals and University of Geneva, Geneva, Switzerland (HUG-GE, 44 subjects) and Nuclear Medicine Unit, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy (FSM-PV, 19 subjects). Demographics and pre-test diagnosis are provided in table 4.2 and 4.3.

Reference PET

The reference PET (*RP*) is a mean image of the FP dataset (figure 4.8). These scans were used to generate a spatial reference only and were not included in any other development (i.e. the ELBA and TDr semi-quantification methods).

The mix of positive and negative subjects was only used to provide a balanced template,

Figure 4.8: Sagittal, coronal and axial section of the amyloid-PET template in MNI space.

with the aim to provide an average image, sampled from a typical population, which is used as registration target. In this respect, the evaluation of the single reader needs not to be confirmed, as it only serves to generate an average, smoothed reference.

Generally speaking, the registration process between same-modality images is more robust than the cross-modality counterpart. The main benefit of a *RP* therefore, was to relax the need of matching PET scans directly onto the MNI-MRI template, either directly or having the subject's MRI as guide.

To generate the *RP*, the 40 scans underwent a recursive registration process, each step delivering a mean template. The first step consisted of an affine registration onto the ICBM-152 MRI template using a mutual-information metric. A intermediate PET template was generated by averaging over the registered images. The intermediate template was used as reference for another registration batch to generate a second reference. The subsequent steps used an affine plus a weak non-linear registration – that is a non-linear warping using a large (12 *mm*) smooth operator on the deformation field (Thirion, 1998) – to improve on the reference image generation. The iterative process ended after no more than 5 steps, when the generated reference did not show significant changes with respect to the previous step.

The MNI-provided lobes, ventricles and sub-cortical regions segmentations were mapped onto the *RP* and visually inspected for consistency.

Center	Diagnosis	Subtype		
	NA / MCI / Dem.	NA / Dep. / SMC	MCI / MCI-AD / AD	FTD / VD / CBS / MSA / SNAP
FPA-BS	9 / 12 / 34	6 / 2 / 1	12 / 0 / 19	6 / 6 / 2 / 1 / 0
HUG-GE	13 / 23 / 8	13 / 0 / 0	15 / 8 / 5	2 / 0 / 0 / 0 / 1
FSM-PV	5 / 4 / 10	5 / 0 / 0	4 / 0 / 9	1 / 0 / 0 / 0 / 0
HSM-GE	9 / 6 / 14	11 / 4 / 0	6 / 3 / 11	0 / 0 / 0 / 0 / 0

Table 4.3: Legend: NA, Normal Aging; MCI, Mild Cognitive Impairment; MCI-AD, MCI due to Alzheimer's Disease; SMC, Subjective Memory Complaint; AD, Alzheimer's Disease; FDT, Frontotemporal Dementia; VD, Vascular Dementia; CBS, Cortico-basal Syndrome; MSA, Multisystemic atrophy; SNAP, Suspected non-Alzheimers Pathology; Dem, Dementia; Dep, Depression.

Table 4.4: ADNI subjects id.

002_S_0729	002_S_1155	002_S_2010	002_S_2073	002_S_4473	002_S_4521	002_S_4654	002_S_4746	002_S_4799	003_S_1122	003_S_4555	005_S_0572	005_S_0602	005_S_0610
006_S_4485	006_S_4515	006_S_4679	006_S_4713	007_S_4568	007_S_4611	007_S_4620	007_S_4637	009_S_0751	009_S_0842	009_S_1030	009_S_2208	009_S_4388	009_S_4543
009_S_4612	009_S_4741	011_S_0021	011_S_0023	011_S_4893	012_S_4545	012_S_4643	013_S_1186	013_S_4579	013_S_4580	014_S_4576	016_S_1117	016_S_2031	016_S_4584
016_S_4591	016_S_4902	018_S_2155	018_S_2180	018_S_4349	018_S_4400	018_S_4597	018_S_4696	018_S_4809	018_S_4868	019_S_4680	019_S_4835	021_S_2077	021_S_2100
021_S_2142	021_S_4857	022_S_0096	022_S_2087	022_S_4805	023_S_0031	023_S_0042	023_S_0061	023_S_0887	023_S_0926	023_S_2068	023_S_4501	023_S_4502	024_S_0985
024_S_2239	024_S_4674	027_S_0074	027_S_1045	027_S_2183	027_S_2219	027_S_2245	027_S_4757	027_S_4873	027_S_4919	027_S_4936	029_S_4585	031_S_0618	031_S_0867
031_S_1066	031_S_2018	031_S_2022	031_S_2233	031_S_4590	032_S_2119	032_S_4386	033_S_0906	033_S_0920	033_S_0923	033_S_1016	033_S_1098	035_S_2061	035_S_2074
035_S_4414	035_S_4582	036_S_4538	036_S_4562	036_S_4714	036_S_4715	036_S_4736	036_S_4878	037_S_4381	037_S_4750	041_S_4874	041_S_4876	052_S_4626	052_S_4807
053_S_4557	053_S_4661	053_S_4813	057_S_1269	057_S_4888	067_S_0056	067_S_0059	067_S_2195	067_S_2196	067_S_4767	067_S_4782	072_S_2026	072_S_2027	072_S_2037
072_S_2072	072_S_2083	072_S_2093	072_S_2116	072_S_2164	072_S_4522	072_S_4613	072_S_4769	072_S_4871	073_S_0089	073_S_4552	073_S_4559	073_S_4614	073_S_4739
073_S_4762	073_S_4777	082_S_2121	082_S_4090	082_S_4428	094_S_2201	094_S_4503	094_S_4560	094_S_4630	094_S_4649	098_S_2047	098_S_2052	098_S_2079	098_S_4275
098_S_4506	100_S_4469	100_S_4512	100_S_4556	109_S_2200	109_S_4380	109_S_4455	109_S_4499	109_S_4531	109_S_4594	116_S_4635	116_S_4732	116_S_4855	116_S_4898
123_S_2055	123_S_4780	123_S_4806	126_S_4675	126_S_4686	126_S_4712	126_S_4891	127_S_0112	127_S_0684	127_S_0925	127_S_1427	127_S_2213	127_S_2234	127_S_4624
127_S_4645	127_S_4765	127_S_4843	127_S_4844	127_S_4928	128_S_0135	128_S_0200	128_S_0225	128_S_0227	128_S_0272	128_S_0545	128_S_2002	128_S_2045	128_S_2123
128_S_2130	128_S_2151	128_S_2220	128_S_4571	128_S_4586	128_S_4599	128_S_4607	128_S_4636	128_S_4653	128_S_4742	128_S_4842	129_S_0778	130_S_4542	130_S_4641
130_S_4660	130_S_4817	130_S_4883	135_S_4566	135_S_4598	135_S_4657	135_S_4676	135_S_4689	135_S_4722	135_S_4723	137_S_0668	137_S_0722	137_S_0800	137_S_0972
137_S_0994	137_S_1414	137_S_4520	137_S_4587	137_S_4596	137_S_4623	137_S_4631	137_S_4632	137_S_4672	137_S_4678	137_S_4815	141_S_0717	141_S_0767	141_S_2333
141_S_4426	141_S_4711	153_S_4621	941_S_4365	941_S_4376	941_S_4764								

Chapter 5

Independent methods

There is room to improve the PET image reading by relaxing some of the constraints imposed by the SUVr approach (such as accurate image registration and the use of predetermined uptake and reference ROI) while providing robust ranking among subjects and proportionality to the visual assessment. With that, I do not wish to replace or belittle the established visual and SUVr-based semi-quantifications. Rather, the intent is to complement those with a novel and independent approach, with the goal of providing more robust and diversified knowledge on difficult-to-read cases.

This chapter describes two novel methods: ELBA (**E**va**L**uation of **B**rain **A**myloidosis) and TDr (**T**ime-**D**elayed **r**atio). Both were developed on images acquired with one of the new ^{18}F ligands (i.e. ^{18}F-Florbetapir).

ELBA and TDr are both intended to be proxy measures of the brain amyloid load. Together they can provide a ranking system to aid in the visual assessment.

5.1 ELBA: SUVr-independent evaluation of brain amiloidosys

ELBA is designed to deliver a whole-brain amyloid-burden estimation. Comparison to SUVr semi-quantification, clinical evaluation at follow-up visits and cerebro-spinal fluid (CSF) analysis is provided to complement the method validation.

The ELBA method was developed on scans currently available in the ADNI database and acquired with ^{18}F-Florbetapir, which was chosen by ADNI to be the reference radioligand in the evaluation of brain amyloidosis (Jagust et al., 2010). The analysis procedure is automatic and does not need any human supervision save for an optional check after the spatial registration process, to ensure that the processed image is consistent and has acceptable characteristics. In this study I first introduce the processing steps to characterize a PET scan using two features which are combined to give the ELBA score, next I proceed with visual assessment in blind and open sessions, finally I use the consensus visual assessment to set a cut-off value for ELBA and SUVr measures, and compare results.

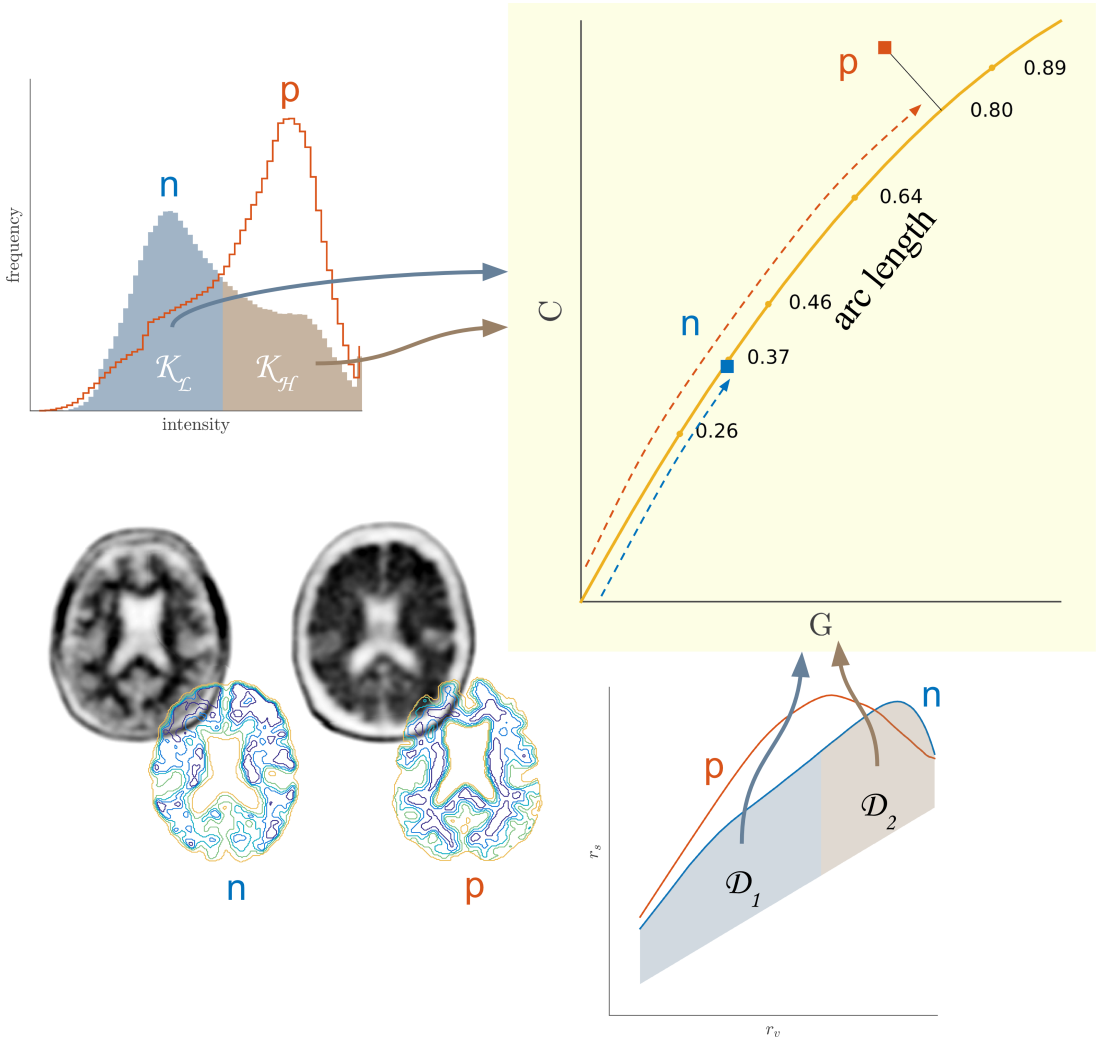

Figure 5.1: Graphical representation of the ELBA method. From iso-intensity curves to the geometric and intensity (contrast) features. These are combined to get a single score, which is the arc length (curvilinear abscissa) of a curve fitted on a large dataset.

5.1.1 Working principles

The intent of ELBA is to capture intensity distribution patterns rather than actual counts in specific ROIs. Considering the brain as a whole, I observed that geometrical appearances of iso-intensity surfaces are rather characteristic in typical negative and positive subjects, the latter showing a sparser and more convoluted appearance than the former. In addition, whole-brain intensity histograms appear to be skewed towards higher intensities in positive subjects.

As qualitative interpretation, the PET signal clusters onto the gray matter patches with significant amyloid load, often surpassing the adjacent non-specific white matter intensity. The presence of higher intensity patches biases the counts statistics and - when a threshold is applied - gives a more complex surface (with notches and several non-connected components).

To capture and quantify these characteristics I developed two features: one that gauges the iso-intensity surface complexity and another that assesses the histogram propensity towards higher values. These features are global properties of the whole brain and do not

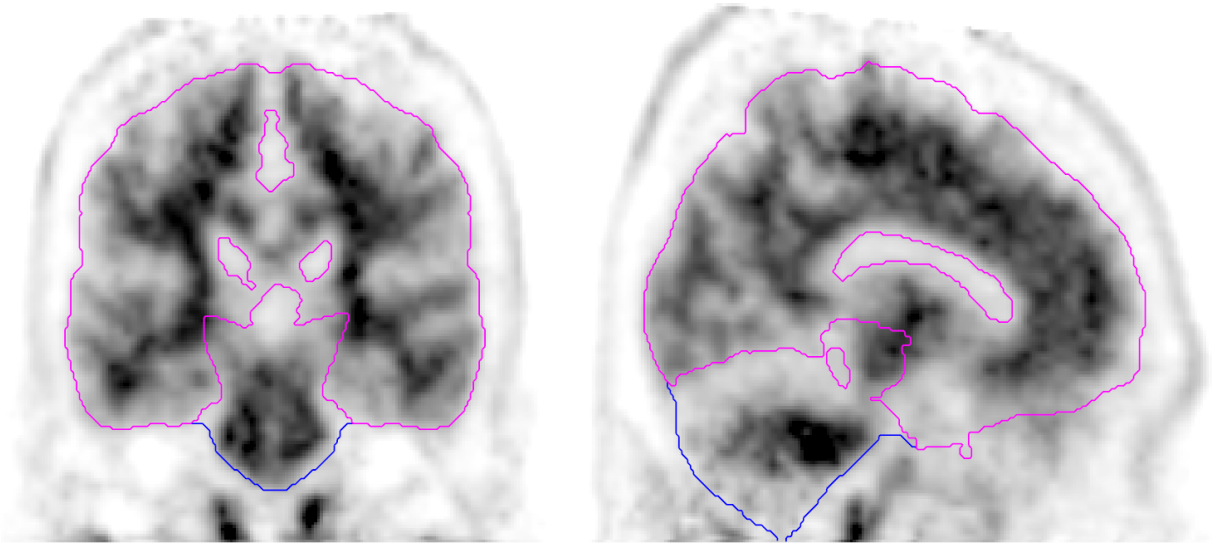

Figure 5.2: Intensity and spatially normalized scan with automatic segmentations: the whole brain (blue line 'a') and the region of interest segmentation (magenta line 'b') where the cerebellum and brain stem have been removed. Coronal view (left), sagittal view (right)

require a reference ROI. An infographic showing these steps is provided in figure 5.1.

5.1.2 ADNI image processing

Basic image processing consisting in a spatial registration to MNI coordinates is performed to allow the segmentation of the brain from the head. This is necessary because non-cortical regions like ventricles, cerebellum and scalp do not carry information specific to the amyloid burden. All downloaded scans (labelled with i, $i = 1..n$) were spatially registered with an affine transformation onto the RP, delivering p_i MNI-aligned images (see paragraph 4.6.1). I then segmented the cortical surfaces (c_i) and ventricles regions (v_i) from each p_i by means of non-linear mapping of the available pre-segmented masks on the reference image RP: the RP was registered with a non-linear transformation onto the target PET and the deformation field was applied to the segmented masks.

I extracted the brain ROI b_i considering all brain lobes delimited by the cortical surface c_i, neglecting the cerebellum, the brain stem and the ventricles. A sample of the image processing result is illustrated in figure 5.2. To reduce the processing errors, longitudinal scans from the same subject were treated as a batch from the beginning (i.e. sharing registration parameters and masks).

Images were then measured with two methods: intensity-based and geometry-based, each delivering a characteristic feature.

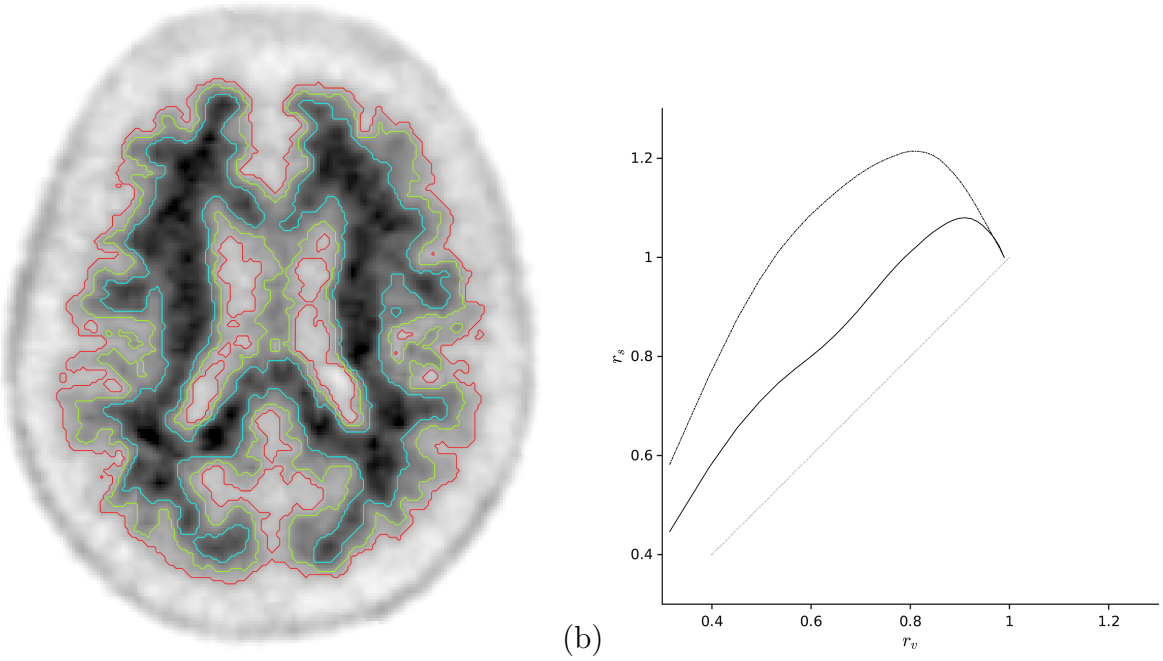

Figure 5.3: Plot a: iso-intensity partition illustration on a axial projection, related to three percentile values of the intensity counts within the brain ROI (0.25, 0.50 and 0.75; red, yellow and blue). Plot b: characteristic curve (r_v,r_s) for two typical cases: a low and a high amyloid burden scan (thick curve / dotted curve respectively). Values are normalized to the respective brain volume and boundary. The thin line is the bisector.

5.1.3 Implementation

Geometric feature

The selected brain volume b_i is partitioned into $n_L = 48$ iso-intensity levels $0 < l_j < 1$, $j = 1..n_L$ taken at equal quantile distances of the whole intensity distribution (i.e. n_L quantiles in the interval $[1 - 1/n_L, n_L]$). The number of levels is of little consequence provided it is chosen $n_L \gtrsim 10$ to ensure adequate sampling of the distribution (I tested the feature outcome with $n_L = 16, 24, 32, 48$ levels). Partitions consist of s_j surfaces and V_j enveloped volumes defined as

$$V_j = \{num.\ of\ voxels \in b_i, voxel\ intensity \geq l_j\}$$

$$s_j = \sum_{\partial V_j} 1$$

where the ∂ symbol denotes the boundary, that is s_j counts the number of voxels on the perimeter. The s_j and the enveloped volumes V_j are not required to be a connected set.

Each partition is characterized by two numbers: one representing the radius r_j^s of an equivalent sphere having the same surface extent as s_j, and another is the radius r_j^v of the equivalent sphere of volume V_j, that is

$$r_j^s = \left(\frac{s_j}{4\pi}\right)^{\frac{1}{2}}, \quad r_j^v = \left(\frac{3V_j}{4\pi}\right)^{\frac{1}{3}}$$

If I plot r_j^v and r_j^s for all $j = 1..n_L$ on a cartesian plane I get a characteristic curve inferiorly bounded by the bisector line $r^v = r^s$, which is the limit for all s_j being actual spheres. The characteristic curve distances itself from the bisector line the most when the s_j is rough and notched. Because of the peculiar spatial distribution of counts in the brain, typical appearance of the characteristic curves is rather different for amyloid-positive and negative scans (figure 5.3).

When we subtracted the trivial bisector line, typically positive scans show a higher surface-to-volume ratio on the higher intensity levels (low r^v) with respect to the lower intensity levels (high r^v), and viceversa for negative scans.

The characteristic curve is integrated without the bisector area on the lower and higher half of its domain D (i.e. the range of r^v) to deliver the geometric feature G_i:

$$G_i = \frac{\int_{D_1} (r^s(r) - r)\ dr}{\int_{D_2} (r^s(r) - r)\ dr}, \quad D_1 = [min(r^v), r^v/2], D_2 = [r^v/2, max(r^v)]$$

Intensity feature

This feature gauges the intensity and contrast values in b_i and divides them into clusters. The chosen clustering method was *kmeans* (MacQueen, 1967; Seber, 1984) with two classes ($High$, Low). Since *kmeans* uses an iterative algorithm starting form a random sample, to ensure reproducibility we ran it for 10 repetitions, then choosing the one with minimum within-cluster sums of point-to-centroid distances.

In each class we computed the number of elements K_{High} and K_{Low} and the class median intensity value I_{High}, I_{Low}. Following convention, we linearly scale the intensity histogram so that the values corresponding to the 1% and the 99% percentiles are mapped onto the $[0, 1]$ interval. I then defined the intensity feature C_i as

$$C_i = ln\left(\frac{K_{High}}{K_{Low}}\frac{I_{Low}}{I_{High}}\right)$$

The intensity feature modulates the relative number of elements in the classes with their contrast. As in the geometric feature, this latter formulation is expressed as a ratio too, so that both features are internally (intra-subject) normalized.

ELBA score

The two image features G_i and C_i were plotted on a Cartesian plane and used to fit a quadratic polynomial. Each point can be projected onto the curve to get two new coordinates: a curvilinear abscissa x_c (arc length) and a curvilinear ordinate y_c (see figure 5.4).

The ELBA score is simply x_c after a linear scaling and a shift to conveniently place the origin at the cut-off between negative and positive scans.

Up to this point the construction of the ELBA score did not require any indication on the subject amyloid burden, nor any hint about its age or clinical status. It was merely a way to combine information on the geometrical distribution of PET counts in the brain and information on the contrast between the brightest and darkest intensity components.

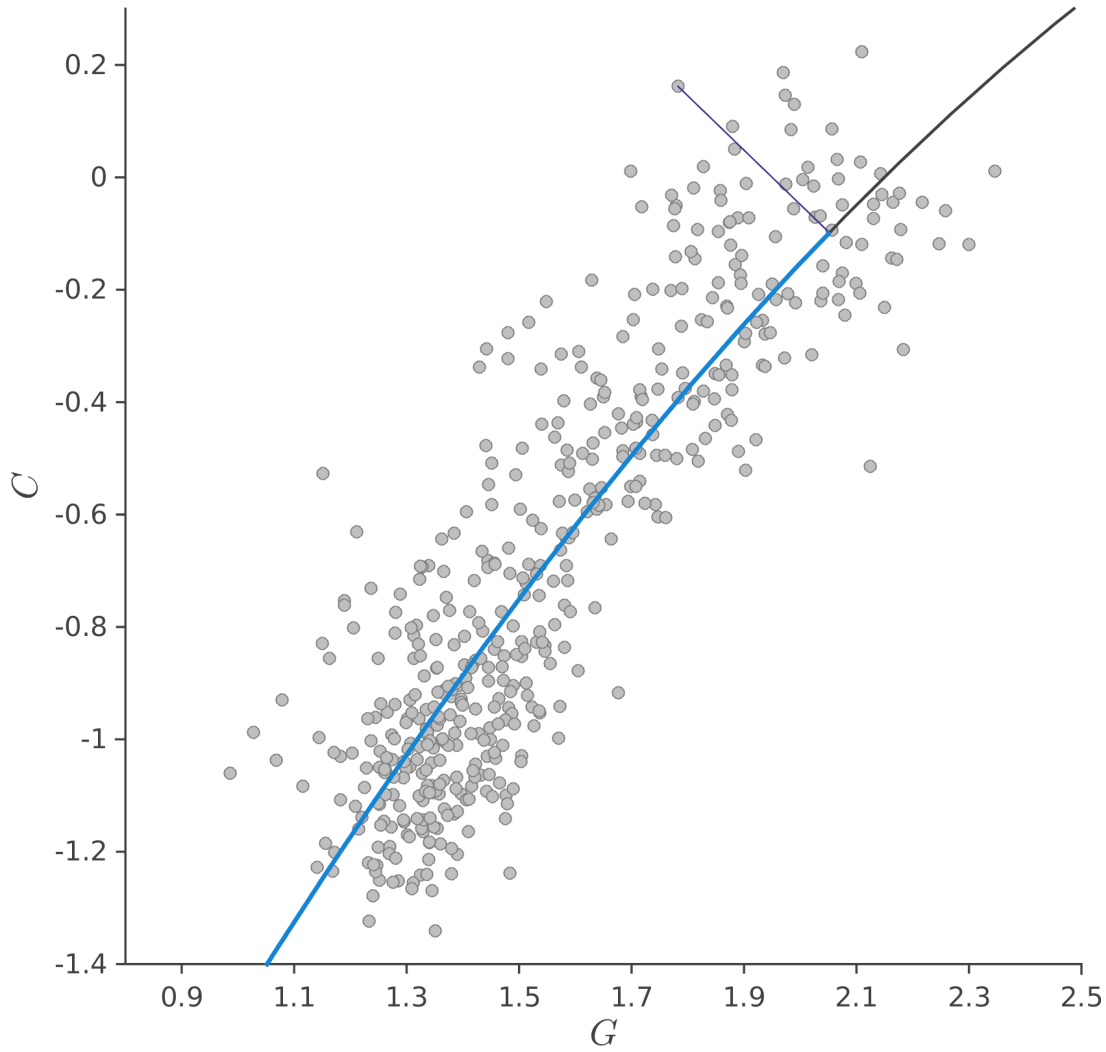

Figure 5.4: ELBA scatter plot (intensity feature C vs. geometric feature G). The black line is the quadratic model. For each scan (dots) the blue part of the line represents the curvilinear abscissa x_c (arc length), and the perpendicular line is the curvilinear ordinate y_c.

Alternative implementations

There can be equivalent implementations of the ELBA score which can replace the formulas in section 5.1.3. For instance, any formulation which differentiates highly notched vs. smoother surfaces can be used in place of the geometric feature. Similarly, intensity bias in a scan histogram can be equivalently assessed by the following ratio

$$I_b = \frac{< I > - q_1}{q_2 - < I >}$$

where $< I >$ is the mean intensity and q_1 and q_2 are the 1% and the 99% intensity percentiles.

Even the use of the curvilinear abscissa described in section 5.1.3 is not mandatory and

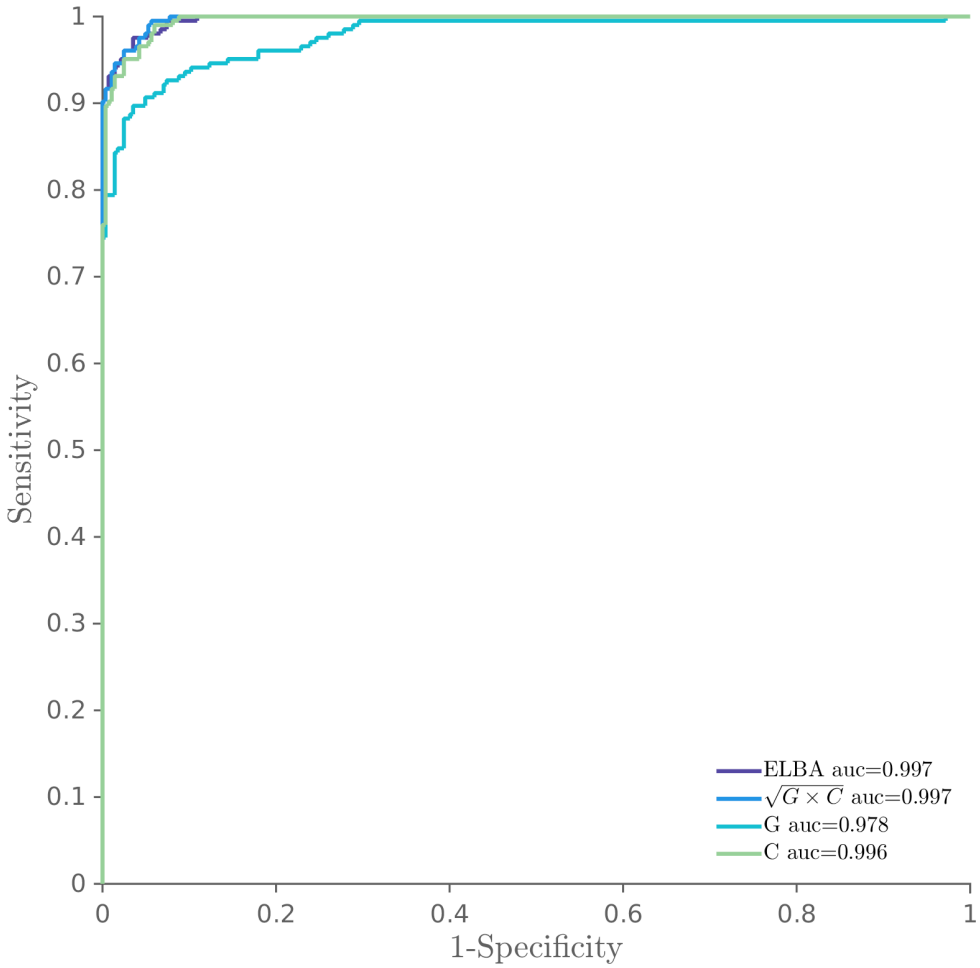

Figure 5.5: ROC curves for the ELBA score, the geometric (G) and intensity (C) features (as computed in section 5.1.3), and the geometric mean score ($\sqrt{G \times C}$, section 5.1.3). Positive/Negative labels used to construct the ROC curves are those given following the open session (section 6.1.2).

an alternative score E' can be formulated with a simple geometric mean[1] as

$$E' = \sqrt{G \cdot C}$$

In addition, given the apparent high correlation between the two features G and C (figure 5.4) one might also be tempted to use one feature only rather than a combination of the two.

As example of possible different implementations I show in figure 5.5 the performance comparison among the ELBA score, the geometric and intensity features alone and the alternative score E'.

The analysis shows the equivalence of the curvilinear abscissa vs. the geometric mean approach, whereas the single feature comparison seems to favour the intensity feature C. For the sake of robustness, I kept both features in the ELBA score.

[1]The geometric mean is often used when comparing different items to find a single "figure of merit" when each item can span different numeric ranges.

5.2 Time-Delayed ratio

While SUVr is based on the intensity ratio and ELBA relies on the spatial distribution, I now construct a new measure called Time-Delayed ratio (TDr), which is a semi-quantification technique based on the tracer kinetics. Minimal kinetic information is derived by two static scans, named "early" (E) and "late" (L).

The early scan is an acquisition starting right before the injection and lasting for ≈5 minutes (0-5 min). The late scan is the standard static scan used to evaluate the amyloid PET (i.e. 50-70 min or 90-110 min depending on the tracer).

In the first 5 minutes, the signal is dominated by the tracer in the blood flux, so that this scan can also be taken as a proxy of the perfusion (Wong et al., 2010; Price et al., 2005; Contractor et al., 2012). The TDr method is similar to the SUVr but it differs from it in terms of the definition of the uptake and reference ROIs. In its simplest form, TDr is defined as:

$$TDr = \frac{< I^L > |_{D_E}}{< I^L > |_{D_L}} \tag{5.1}$$

where $< I^L >$ is the **mean intensity** measured on the late scan and averaged over the domains D_E and D_L. The domains are a collection of ROIs defined on the early and late scans as follows:

$$D_E = v \in E \mid I_v \geq I_0^E \tag{5.2}$$

$$D_L = v \in L \mid I_v \geq I_0^L \tag{5.3}$$

where v represents the voxels in the scan and I_v is the intensity in v. In words, D_E and D_L are a collection of voxels whose intensities are greater than a specified value (see figure 5.7 and 5.8). So, TDr is an intensity ratio like the SUVr but the ROIs used in the computation are neither predefined nor data driven. They are instead tailored onto the patient, and different from patient to patient (see for instance figure 5.6) and from scan to scan even in the same patient (in case of a longitudinal study).

5.2.1 Working principles

TDr is inspired by the absolute quantification approach. The tracer initially flows into the blood stream ad then slowly migrates into the tissues. From the early scan it is then possible to select high-flux tissues (restricting the search to cortical areas). Should the β-amyloid be present in these tissue, the tracer will bind to it and we have the advantage of knowing that it is delivered appropriately and in relatively large quantity (high uptake). Conversely, should the β-amyloid be absent, then the tracer is most likely to be washed out as the tissue is well perfused (good clearance).

The numerator in equation 5.1 then describes the intensity we get in the *late* scan averaged on the voxels with high blood flux defined in the *early* scan (eq. 5.2). Clearly, a high value of $< I^L > |_{D_E}$ corresponds to a high uptake and vice-versa.

Now, as the absolute value of this intensity can vary among subjects for a variety of reasons (injected activity, blood flow, physiological characteristics, etc.), this intensity must be normalized. The denominator in equation 5.1 describes the intensity we get in

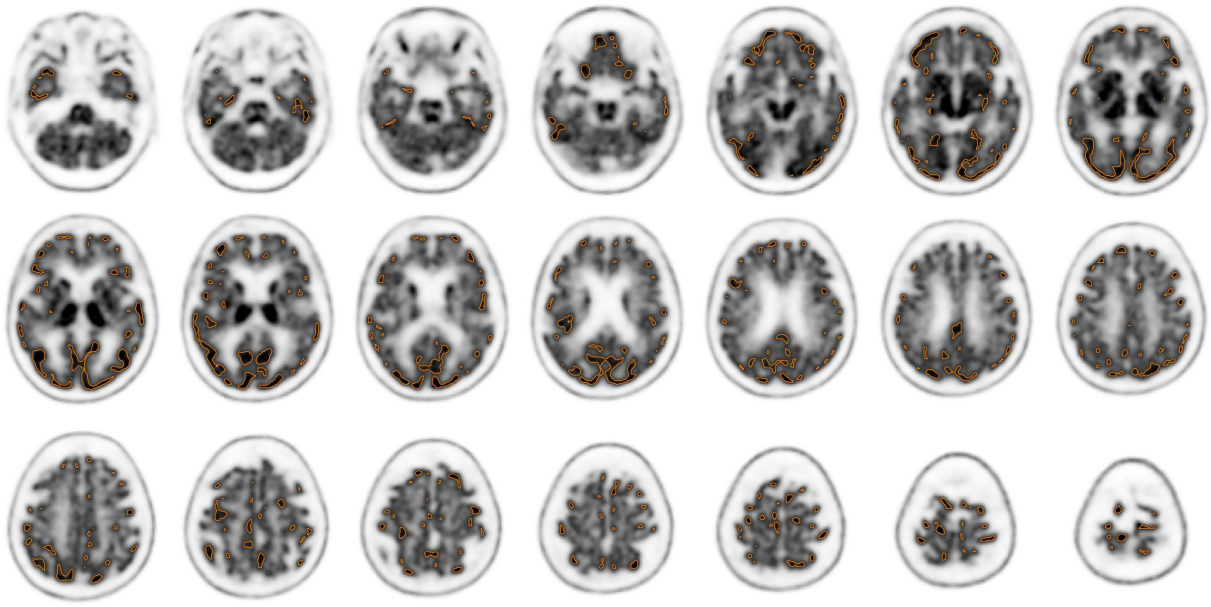

Figure 5.7: Early scan (E) and highest flow domain (D_E, red outline).

the *late* scan averaged on the voxels with high uptake (specific or non-specific) defined in the same *late* scan (eq. 5.3). Therefore, instead of using a fixed reference tissue (like the cerebellum), I used the highest uptake domain, otherwise known as the hot spot, which can be located in the gray or in the adjacent white matter.

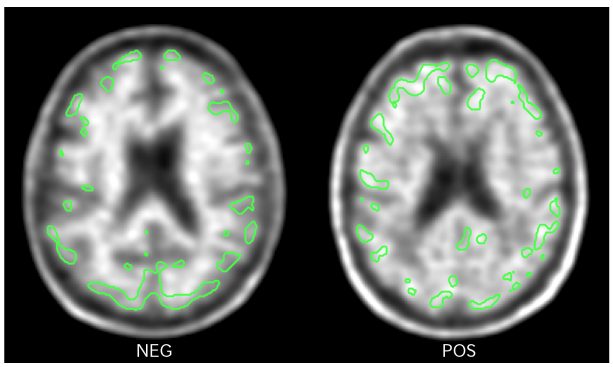

Figure 5.6: Uptake domain D_E on two different subjects (one amyloid-negative and one amyloid-positive).

This normalization frees the TDr from the constraints of accuracy in image registration (although we shall see that there is still a limit there) and in the definition of a reference tissue. Another immediate benefit is that D_E and D_L are defined on images acquired with the same scanner, and reconstructed with the same method. This means that the PVE – although still present – is partially "compensated" by selecting both domain in the same space. Because of this definition and the selection of high flow regions, the measurement of amyloid load in these areas is more accurate, since the contrast between the load and the hot spot is the highest.

The D_E and D_L domains are function of the patient characteristics and no reference regions or cortical atlases are needed. Domains typically consists of several distinct ROIs, providing both local and global information. Therefore TDr values are similar to SUVr in that they can be combined or averaged to give regional and whole-brain information.

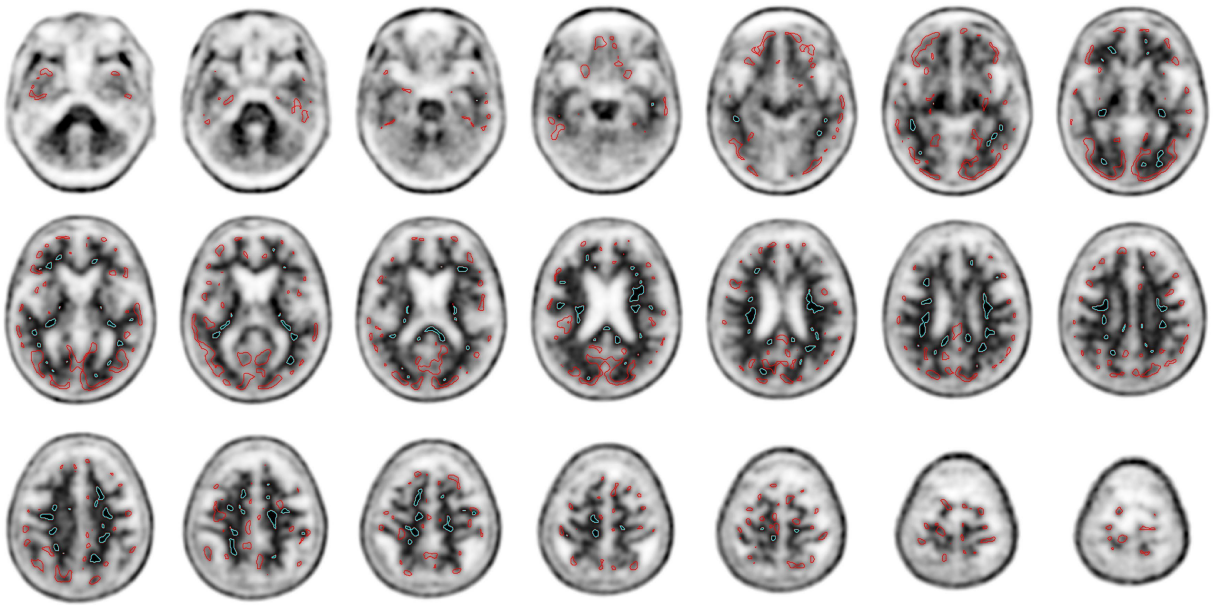

Figure 5.8: Late scan (L), the highest flow domain (D_E, red outline) and the highest uptake domain (D_L, cyan outline).

5.2.2 MPD image processing

TDr requires only a basic image processing, consisting in a spatial registration to MNI coordinates. Contrary to ELBA, it needs no the segmentation of the brain from the head. The only TDr peculiar requirement is that the early scan (E) be registered onto the late scan (L) in the native space prior to any other spatial or intensity normalization. This is in order to minimize spatial resampling errors (figure 5.9).

Once the E and L scans are aligned in the native space, the L scan guides the transformation to put scans into the MNI space, in the same fashion as with ELBA. Therefore all MPD scans were spatially registered with an affine transformation onto the RP.

Here too, we segmented the cortical surfaces (c_i) and ventricles regions (v_i) from each p_i by means of non-linear mapping of the available pre-segmented masks on the reference image RP. We then extracted the brain ROI b_i considering all brain lobes delimited by the cortical surface c_i, neglecting the cerebellum, the brain stem and the ventricles. The b_i is the confidential volume within which the domains D_E and D_L are defined.

5.2.3 Implementation

TDr implementation is rather straightforward. It only needs an accurate registration of the early scan onto the late scan, so that D_E ROIs can be properly applied on the latter. For this reason the early image must be spatially normalized; this was done by a 6-parameter linear transformation, with the L scan as fixed image. All subsequent transformations (i.e. onto the MNI space) are applied to both E and L images alike.

After proper image registration and once in the MNI space, we apply a loose brain masks to cut off the scalp, brain stem, cerebellum and basal ganglia, and estimate the highest counts (D_E and D_L) following equations 5.2 and 5.3 with $I_0^E = 0.85$ percentile and

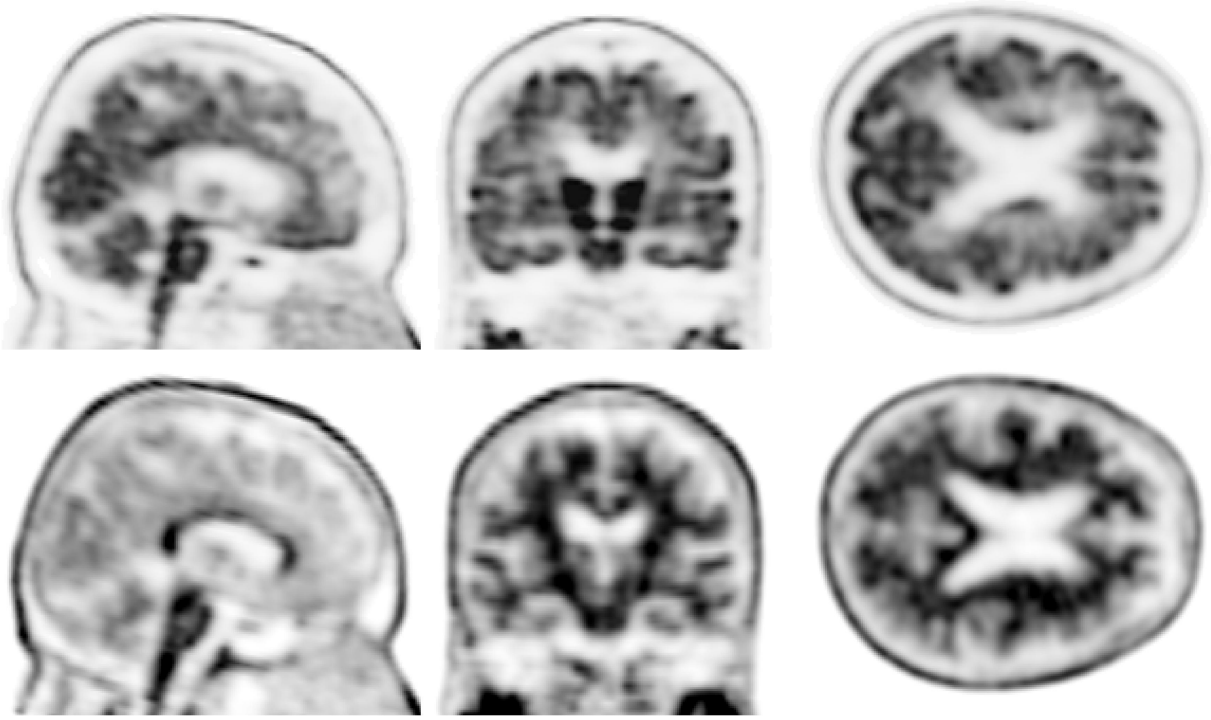

Figure 5.9: Sagittal, coronal and axial section of a sample early (E) and late (L) scan, both spatially normalized and aligned onto the MNI space

$I_0^L = 0.99$ percentile.

These values were chosen as first working hypotheses to select the high-flow cortical regions only (I_0^E) and the hot-spot (I_0^L). Future work shall investigate the effect of these two parameters and provide optimized values based on contrast measured on a normative population.

Chapter 6

Validation

6.1 ELBA Validation

All PET images were evaluated by five independent readers: two nuclear medicine (NM) physicians and one neurologist who have been trained to give teaching courses to NM physicians and read more than 200 scans with supervision (expert readers) and two moderately expert readers (NM physician) who read more than 200 images under supervision.

Upon coarse data examination, the readers noticed an apparent quality difference among scans. They agreed therefore on an operational definition of "*sub-optimal image quality*" for the purpose of keeping track of quality-induced mis-readings during the validation.

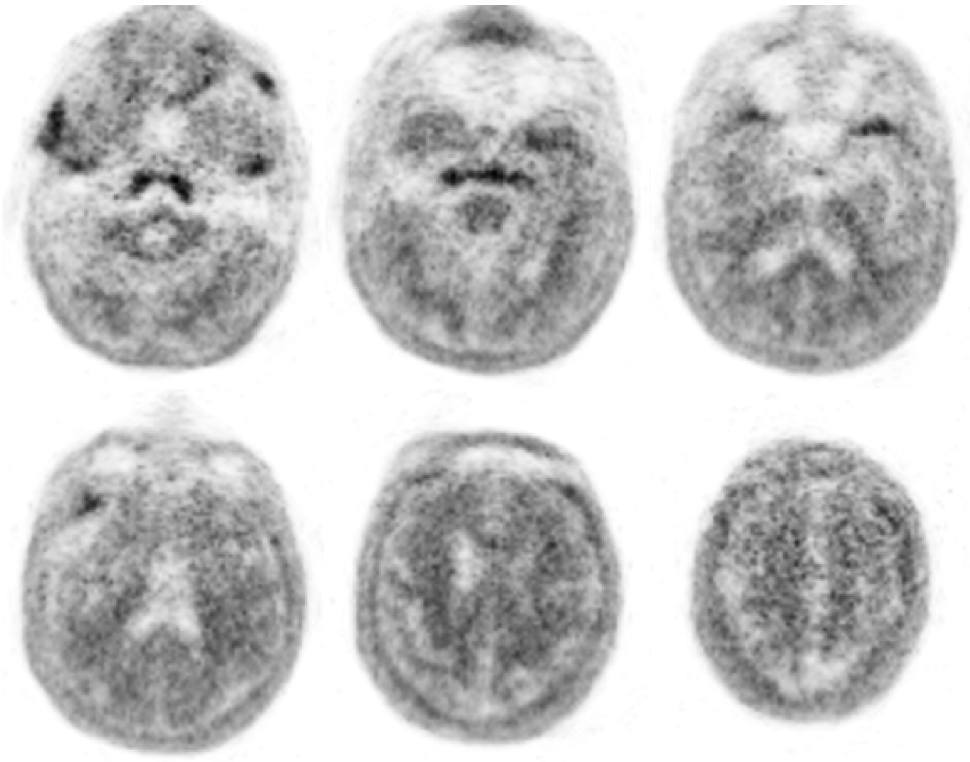

Figure 6.1: Transaxial picture of a sample scan considered "sub-optimal quality" by all readers.

They considered a subjective evaluation of low quality reconstruction, motion artifacts or low signal-to-noise ratio according to each own clinical experience. The aim was to tag scans whose characteristics could interfere in (or impair) the visual assessment. I provide an example of a low quality image which was flagged by all readers in figure 6.1.

Regardless of the their quality, all scans were processed and visually evaluated. The quality label was used for retrospective analyses only, to keep track of possible grounds for difficult cases.

The 504 images were divided into 488 among baseline and first follow-up scans, plus 16 second follow-up scans. For the validation purpose we used the 488 baseline and follow-up images, while the additional 16 scans were used in section 8.2.1 only.

Baseline and follow-up scans were read as independent images, all mixed together with random order so that evaluators were very unlikely to see the same subject twice during the reading sessions.

6.1.1 Blind phase

Images were presented after the preprocessing steps described in section 5.1.2. The blind evaluation was carried out by each reader without interaction, without support from any automatic analysis software, blind to the clinical data, blind to the ELBA output and according to each reader's own practice and experience.

Readers were initially allowed to judge on a three classes base: *negative*, *positive* and *uncertain*. Readers were asked to use *negative* and *positive* tags on images where they were absolutely confident of the visual assessment. Scans whose evaluation implied a more elaborated visual inspection and where the possibility of doubt existed should have been initially tagged as *uncertain*. In addition, readers were also asked to add a tag on the perceived image quality.

Besides the individual evaluations, the analysis of the blind phase delivered 4 image set: the P set and the N set, that is scans which were consistently marked as *positive* and *negative* by all readers; the U set, that is scans which received an *uncertain* comment from at least one reader; and the C set, that contains those scans which received contrasting judgement (*positive* and *negative* together from one or more readers).

More specifically, the U set consists of scans with at least one *uncertain* comment but otherwise no contrasting labels, whereas the C set consists of scans which received contrasting evaluations but which may also have had one or more *uncertain* comment.

6.1.2 Open phase

In this phase all scans in the U and C set were presented again. Each scan was evaluated by all five readers in an open session, with interaction, where they were invited to reach a consensus on either *negative* or *positive* label.

This time though the ELBA output was partially used to aid in the analysis. Readers were not made aware of the ELBA score but the image to evaluate was shown side-by-side with two other images, ordered on the ELBA score scale: the nearest one from the P set and the nearest one from the N set (see figure 6.3 for an example). This visualization

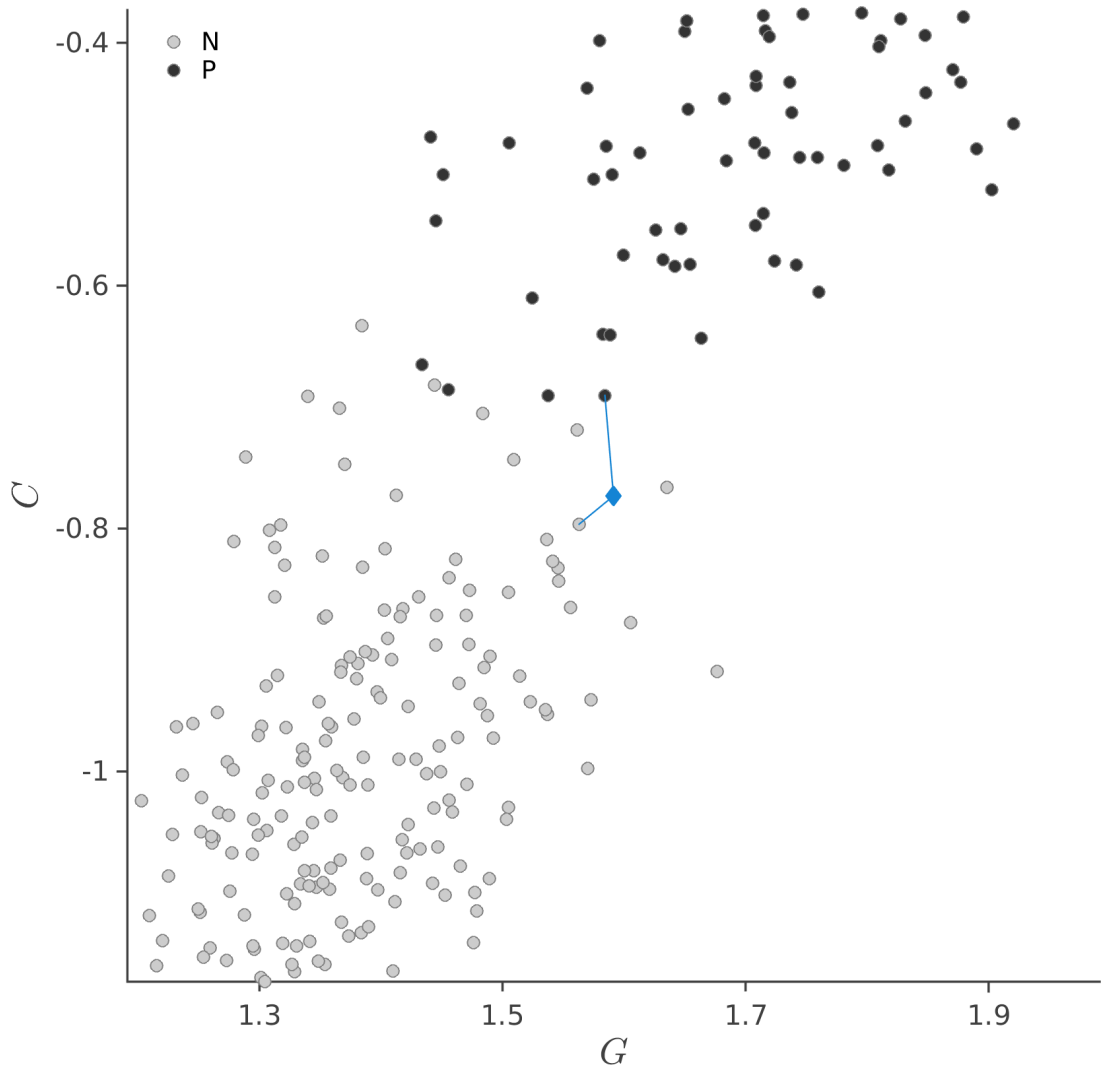

Figure 6.2: Zoom on the geometric (G) and intensity feature (C) plane around a new scan to be evaluated (blue diamond). The scan is compared to the nearest tagged cases taken from the N and P set only, i.e. scans with concordant independent evaluation by all readers given in the blins session (section 6.1.1). The distance is evaluated on the ELBA plane, where the nearest positive and negative scans are indicated by the blue line.

was meant to help in the assessment, by comparing the scan under scrutiny to the most similar, validated assets.

The original scans (i.e. not spatially normalized) were also available for consultation. They were used to cross-check the consensus evaluation during the open discussion.

This phase shall also be used to define the ELBA score cut-off, that is by maximizing the accuracy on the consensus-labelled dataset.

6.1.3 Comparison with SUVr-based methods

We calculated the average cortico-cerebellar SUVr on all scans, to compare the ELBA score to this widely used semi-quantification method. We used a data-driven approach with the whole cerebellum (white and gray matter) as reference and a number of cortical

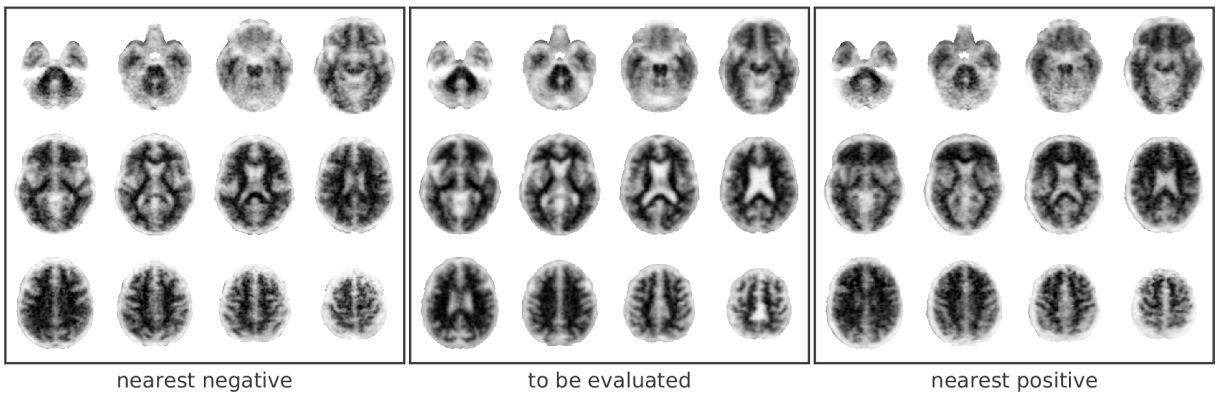

Figure 6.3: A trans-axial representation of the three scans from figure 6.2 as sample of the additional information used in the open phase session (section 6.1.2).

ROIs as uptake regions, as displayed in figure 4.4. The SUVr information was used neither in the blind nor in open validation phase.

SUVr measures were provided by ADNI too, but only on a fraction of the baseline PET scans (111). The ADNI-provided values are calculated according to protocols described in (Jagust et al., 2010) and they are the average cortical-cerebellar SUVr computed with two methods: one with the syngo PET Amyloid Plaque (sPAP) software and another with the Avid Semi-Automated Method (AVID, (Hutton et al., 2013)). We checked that SUVr values computed with the data-driven approach agreed with those already provided on the subset of baseline scans using correlation and linear regression analysis.

The final SUVr cut-off was chosen to maximize accuracy using the consensus negative/positive labels after the open phase session.

6.1.4 Comparison with CSF Aβ_{42} quantification

CSF was acquired by lumbar puncture following procedures and criteria identified by ADNI (analysis details and quality control procedures are available at `http://adni.loni.usc.edu/`). The biomarker data set used in this study was taken from the file series upennbiomk4.csv to upennbiomk8.csv. We considered CSF and PET data whose measurements were closest in time, restricting to lumbar punctures and PET measurements performed within 100 days of each other, which resulted in a selection of 203 subjects.

The cut-off on CSF Aβ_{42} values was 174 ng/L (from (Mattsson et al., 2014)), a value Mattsson found to be optimal to maximize accuracy between stable and progressive MCI.

We compared CSF with ELBA and SUVr scores and with visual assessment. We also evaluated the diagnostic performance for NS versus AD (57 vs. 51 subjects respectively), where clinical assessment was taken at the latest possible follow-up visit.

6.1.5 Clinical follow-up

The latest clinical evaluation was checked and compared to the initial assessment. We found 78 subjects who had their assessment reviewed from baseline (latest clinical data sheet downloaded on May 10, 2016). The average time to conversion was 35 $\pm$ 10

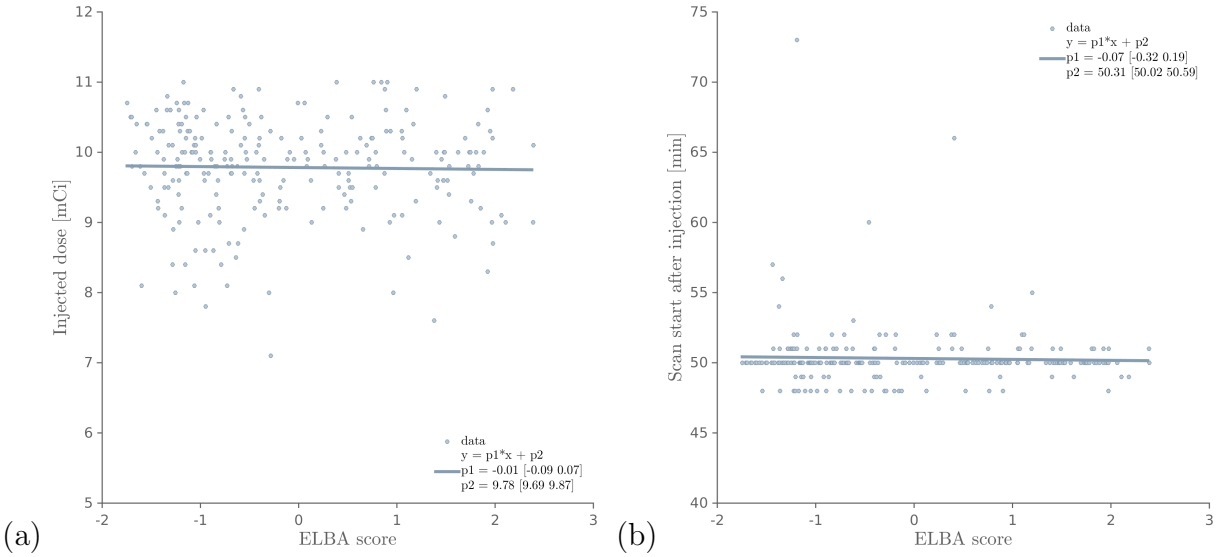

Figure 6.4: Injected dose (a) and scan start time after the injection (b) versus ELBA score. A linear model was used to fit the data. 95% conf. lev. on the line slope shows that there is no significant dependence of both variables on the score.

months (mean and standard deviation). They were divided into 4 classes according to the baseline/follow-up assessment: MCI→AD (46), MCI→NS (16), NS→MCI (10) and NS→AD (6).

Their ELBA and SUVr scores at baseline were used to measure consistency between semi-quantification methods and the agreement with the diagnosis at baseline and follow-up. When present, CSF Aβ_{42} level was used to help discussing borderline cases and evaluate diagnosis agreement with the biomarkers.

6.1.6 Further methodological considerations

We checked for dependency on tracer doses administered, scan start time after the injection and white matter integrity as possible factors which might influence the intensity mapping. Data set containing scan information were downloaded from ADNI (files av45meta.csv, ucd_adni1_wmh.csv and ucd_adni2_wmh.csv, additional analysis details are available at `http://adni.loni.usc.edu/`). These additional data were available for 242 scans.

All parameters were linearly regressed against the ELBA score, showing no significant trend (95% CL on the line slope include the zero). A visual representation of this analysis is found in figure 6.4 and 6.5.

In addition, we estimated the generalized accuracy, sensitivity and specificity of ELBA and SUVr scores vs. the visual assessment (open phase) by means of an iterative procedure. Briefly, we randomly selected 50% of the PET datasets as the training group – for the coefficient estimation for SUVr and ELBA – and 50% of the PET datasets as the testing group (for the cross-validation of the estimated coefficients). Each group contained approximately an equal number of positive and negative scans, based on the concordant judgement of the experts (i.e. after the open session). Cut-off values for SUVr and ELBA

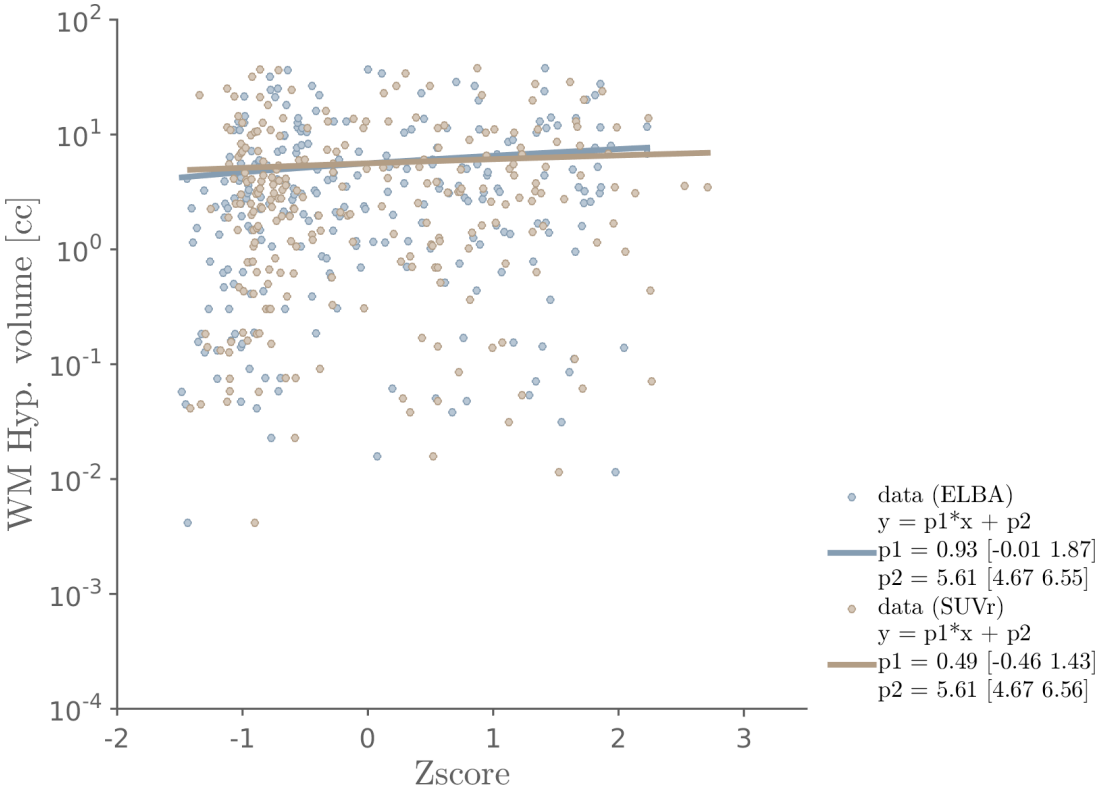

Figure 6.5: White matter (WM) hyperintensity volumes versus ELBA score and SUVr score (both plotted as zscores). 95% conf. lev. on the line slope shows that there is no significant dependence of the WM hyperintensity volumes on both scores.

scores were computed on the training group (maximizing accuracy) and applied to the testing group. The procedure was iterated 500 times.

6.2 TDr Validation

A thorough validation of TDr is still missing. In particular, we have no hystopathological study nor a direct confrontation with CSF values. This is due to the fact that TDr require a specific acquisition protocol which, to date, is only available at few high-level research centers. Therefore large dataset like ADNI, which embed much of the needed information, could not be used.

The requirement of two acquisitions per subject is not incompatible with clinical practice *per se* but it assumes well trained nuclear medicine physicians as well as technical support to properly manage the scanner timing, acquisition and patient management.

The early acquisition timing is particularly delicate because it influences the choice of D_E and therefore the result of semi-quantification with the TDr method; a 10-minute delayed early acquisition from the agreed protocol (0-5 min from the injection) may not provide the correct information, dominated by the blood flow signal.

Thus TDr has only been partially validated, that is by comparison with visual assessment and the two other validated semi-quantification methods: SUVr and ELBA.

We used the Multicentric Pilot Dataset (MPD, section 4.6.1), whose PET evaluation

labels (negative / positive) and clinical data were provided independently and blind to the quantification analysis. In addition, no consensus label was applied to the scans so that visual assessment reflects each center's expertise and criteria. From the point of view of TDr though, we can consider the visual evaluation as blind (in a similar fashion as in ELBA, section 6.1.1).

6.2.1 Comparison to SUVr and ELBA

We compared TDr with respect to the two other semi-quantification methods using Pearson correlation and the residual analysis. No scans were excluded and the correlation was computed on the whole dataset (no single-center evaluation).

We also computed all the semi-quantification cut-offs on the MPD dataset, following the criteria of maximum accuracy on the visual assessment (similarly to what was done in ELBA on the ADNI dataset).

Chapter 7

Results

7.1 ELBA

7.1.1 Blind and open phase performance

The score distribution was grouped by visual assessment and results are displayed in figure 7.1, where we show the single reader evaluation and the combined set after the blind and the open sessions (sections 6.1.1 and 6.1.2). An equivalent plot using the curvilinear ordinate y_c is provided in figure 7.2.

Of the 488 scans, 186 were consistently marked as *positive* by all readers (P set), 217 marked as *negative* (N set), 63 received an *uncertain* comment from at least one reader (U set) and 22 received contrasting judgement (*positive* and *negative* together, C set). The agreement among the readers after the blind session was measured by the intraclass correlation coefficient (ICC) and was found to be ICC=0.94, ($p < 10^{-4}$).

Scans labelled *uncertain* were rather consistent both in number – 59, 45, 49, 42 and 48 for AC, IB, UG, SM and FN respectively – and ELBA score. Moreover all readers consistently tagged 34 scans as *uncertain*. Not surprisingly, 45 out of 63 (71%) of all *uncertain* scans were also flagged as "sub-optimal quality" by at least one reader.

To define the cut-off value on the score we used the open phase results. The cut-off was chosen to maximize accuracy and the original curvilinear abscissa (x_c) was scaled and shifted to have cut-off = 0 and the mean score on the negative scans = -1 (lower x-axis in figure 7.1). The linear scaling is not a necessary step *per se*, it is applied only to facilitate the score interpretation.

The discriminating power was measured by the area under the receiver operating characteristic curve (AUC), giving AUC= 1.000 for N vs. P in blind condition (i.e. on the scans on which all readers independently concurred), and AUC= 0.997 [0.993 - 0.999] for *negative* vs. *positive* (accuracy= 0.97) after the open phase discussion (CL=0.95 within brackets).

We also show the distribution of the score grouped by clinical cohorts in figure 7.3; 244 subjects are plotted, grouped by their clinical classification at baseline. The values on the y axis are the average between the baseline and the follow-up ELBA scores. To enhance the reading, subjects in figure 7.3 are also grouped by age: for each cohort, the leftmost and rightmost dots represent subjects of age < 70 and > 70 years respectively.

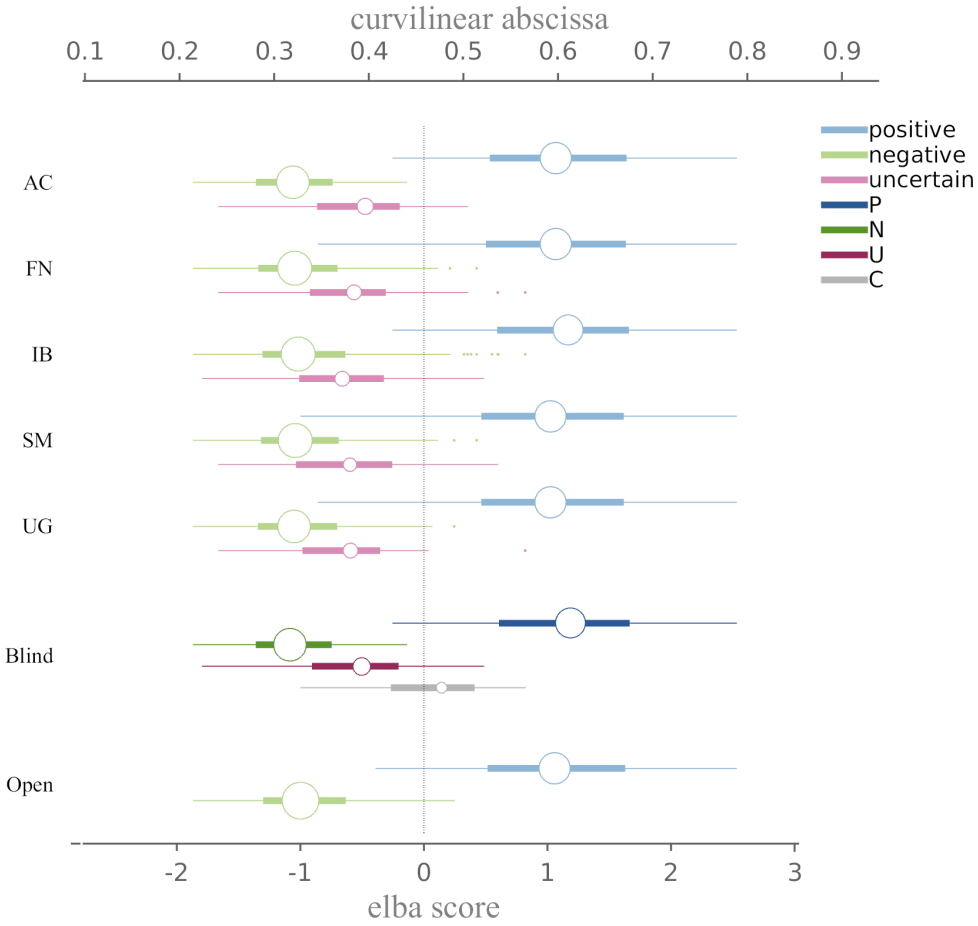

Figure 7.1: Visual evaluation versus the ELBA score (or equivalently the curvilinear abscissa) before and after the blind phase session. *negative*, *positive* and *uncertain* labels are given by each reader (AC, FN, IB, SM and UP). In the blind session summary ('Blind'), P and N refers to scans consistently (i.e. by all readers) tagged *positive* and *negative* respectively. U refers to scans which received at least one *uncertain* tag but no contrasting assessment; C refers to scans which received contrasting assessment (even when together with *uncertain* tags). The open session summary ('Open') shows the dichotomic consensus. Circles are centred on the median value of the respective cohort and their areas are proportional to the sample size. The vertical line marks the cut-off. Thick lines mark the 25% and 75% percentile, thin lines extend up to to 1.5 times the interquartile range.

7.1.2 Comparison with SUVr-based methods

Figure 7.11 and table 7.2 present semi-quantification values and binary summary separately for the different diagnosis groups, using the visual assessment after the blind session. Figure 7.4 shows the comprehensive ELBA score vs. SUVr semi-quantification on all scans labelled according to the visual evaluation after the open phase session, the confusion matrix is provided in table 7.1a and 7.1b. In these figures and tables the optimal SUVr cut-off is 1.23 and it was computed maximizing the accuracy using the open session results, in the same way as with the ELBA score.

The SUVr AUC= 0.978 [0.964 - 0.985] (accuracy= 0.94). The Pearson correlation between ELBA and SUVr scores is $r = 0.86$ ($p < 10^{-4}$).

We matched our SUVr computation on 111 baseline scans which were provided with

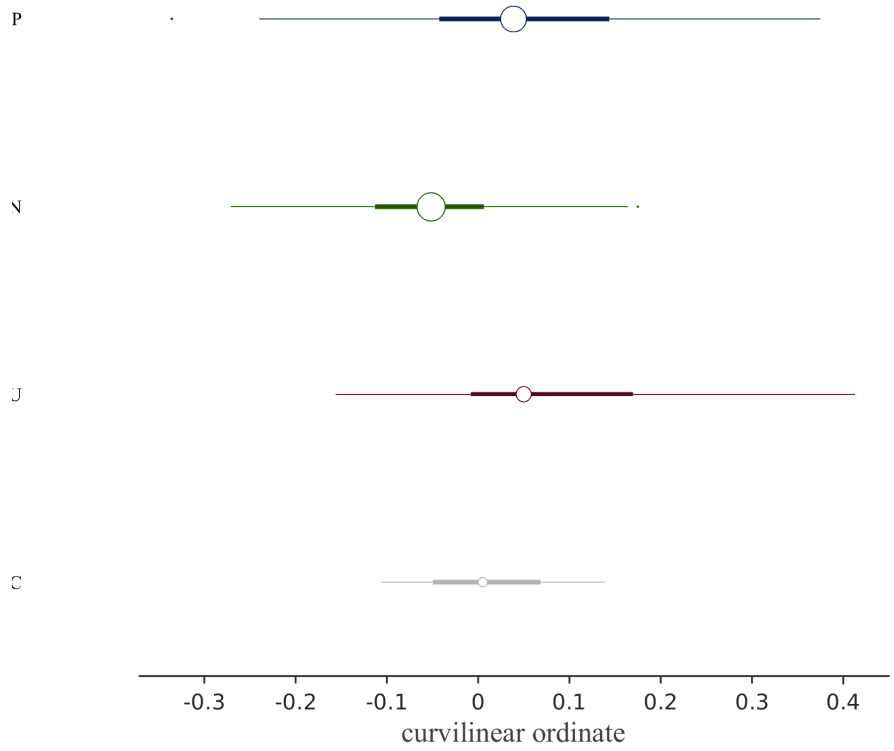

Figure 7.2: Set distribution over the curvilinear ordinate y_c after the blind phase session. P and N refers to scans consistently (i.e. by all readers) tagged *positive* and *negative* respectively. U refers to scans which received at least one *uncertain* tag but no contrasting assessment; C refers to scans which received contrasting assessment (even when together with *uncertain* tags). Circles are centred on the median value of the respective cohort and their areas are proportional to the sample size. Thick lines mark the 25% and 75% percentile, thin lines extends up to to 1.5 times the interquartile range.

independent SUVr values by ADNI (sPAP and AVID methods). Pearson correlation is: $r = 0.98$ ($p < 10^{-4}$) this work vs. sPAP; $r = 0.99$ ($p < 10^{-4}$) this work vs. AVID. The least square line $y = ax + b$ between our SUVr and the AVID one has a slope [CL=0.95] $a = 0.90$ $[0.87, 0.92]$ and intercept $b = 0.07$ $[0.03, 0.10]$ which translates in an equivalent optimized cut-off = 1.18 on the AVID values.

Using the cross-validation procedure to establish cut-off values and estimate the generalized performance (section 6.1.6) we found that the combined accuracy, sensitivity and specificity for ELBA are 0.96, 0.97 and 0.95 respectively. Similarly, SUVr results are: 0.95, 0.96 and 0.93. The ELBA cut-off range was found within the interval [-0.14 – 0.18] (95% CL).

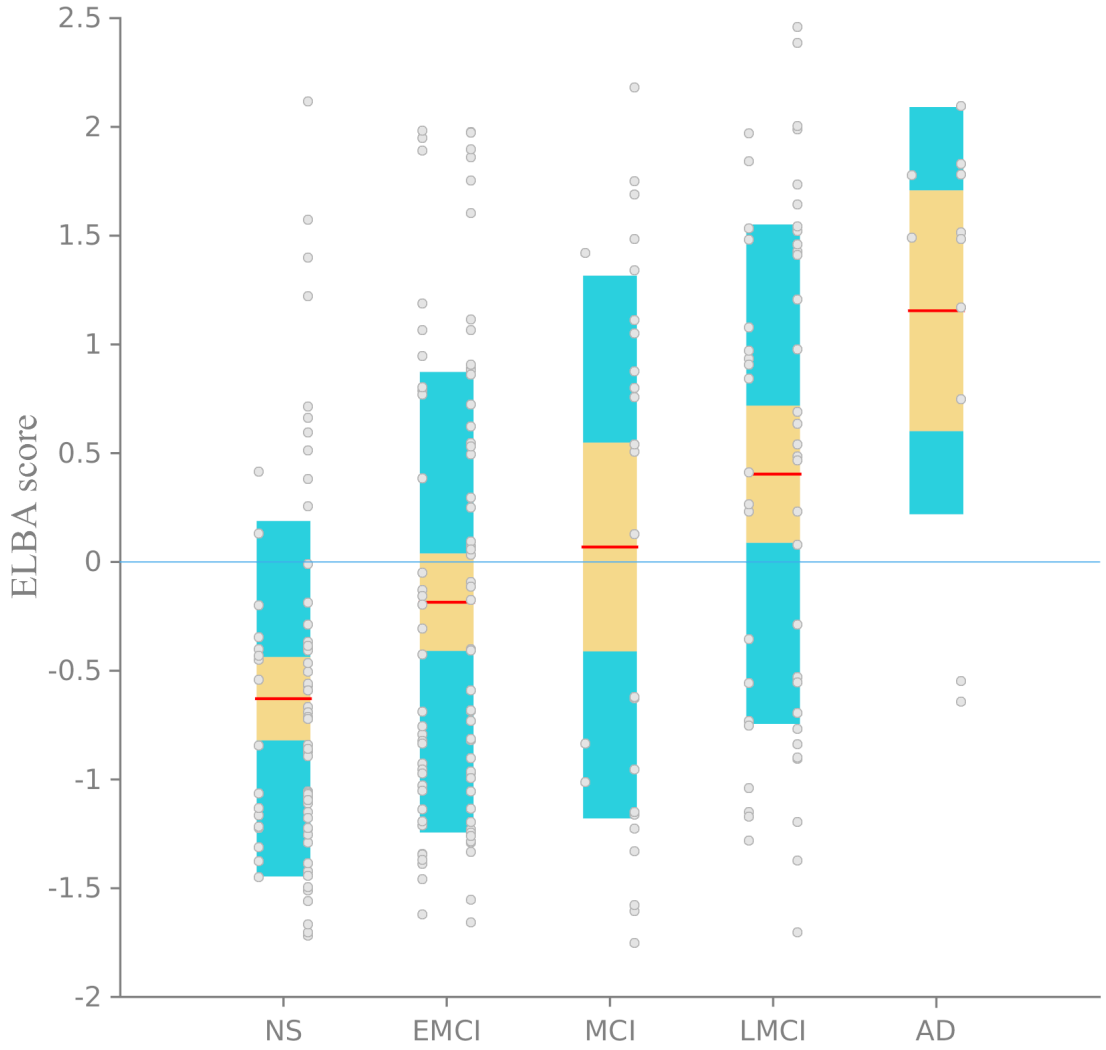

Figure 7.3: Mean (baseline and follow-up) ELBA score distribution vs. clinical cohort. Dots (subjects) are shown with the median (red line), the 95% conf. level on the median (yellow band) and the interquartile range (azure band). In each cohort, the leftmost [rightmost] dots represent age class < 70 [> 70] years old. NS = cognitively normal subjects, EMCI/LMCI/MCI = early- / late- / mild cognitive impairment, AD = Alzheimer's Disease.

7.1.3 Comparison with CSF Aβ_{42}

The scatterplot between baseline ELBA score and CSF Aβ_{42} concentration is provided in figure 7.5, where the open-phase visual assessment was used to group data.

Concordance between ELBA score and Aβ_{42} score classification was achieved in 184 out of 203 (90.6%) instances; in 7.5% of patients an altered Aβ_{42} score was found with normal ELBA score and, on opposite, 1.5% of patients had a (slightly) increased ELBA score with normal Aβ_{42} levels.

The related SUVr representation is in figure 7.6. The confusion matrix for CSF versus visual assessment (open phase), ELBA and SUVr scores is provided in table 7.3, where the accuracy is found to be 0.90, 0.91 and 0.89 respectively.

The number of subjects with confirmed NS and AD clinical status at follow-up and with CSF analysis is 108. The area under the ROC curve (auc, CL=0.95) for NS (57) versus

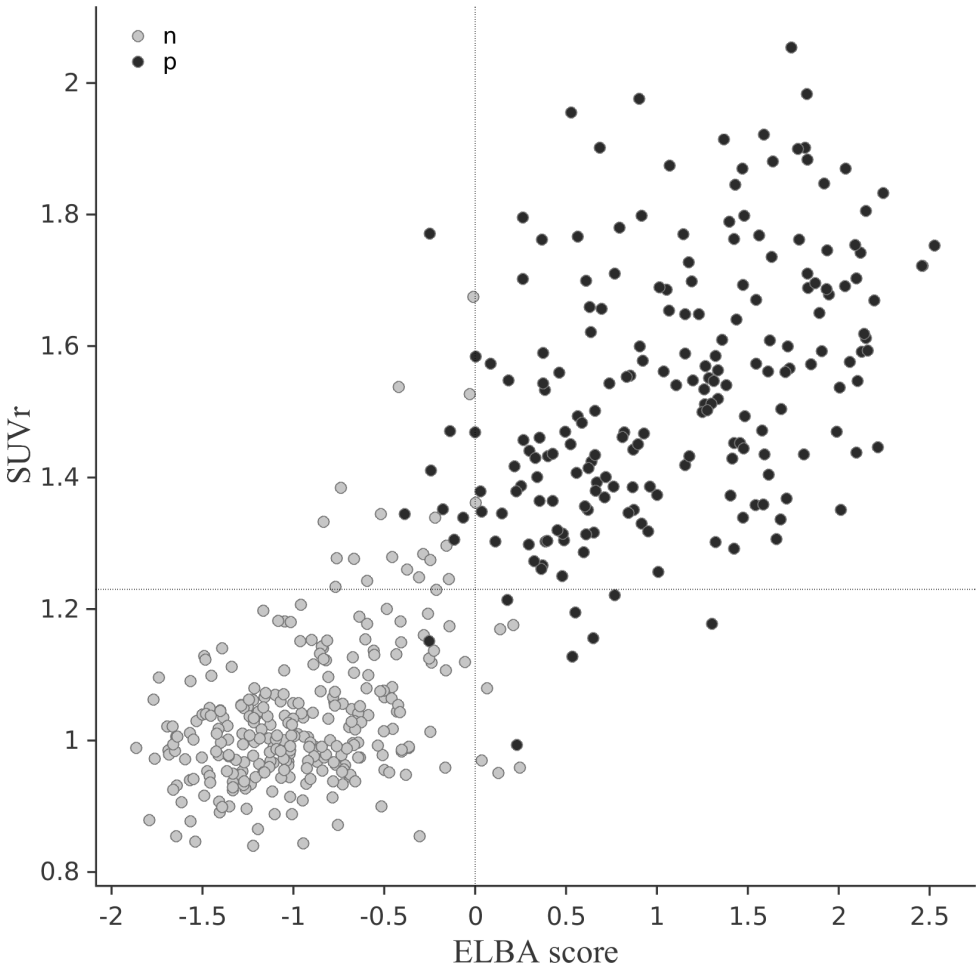

Figure 7.4: ELBA and SUVr scores scatter plot for all scans. 'n' and 'p' tag the consensus evaluation after the open phase session. Dotted lines mark the cut-off values.

AD (51) is found to be auc= $0.88[0.78 - 0.95]$ (CSF), auc= $0.96[0.90 - 0.98]$ (ELBA) and auc= $0.92[0.84 - 0.97]$ (SUVr). A graphical representation is in figure 7.7.

7.1.4 Clinical follow-up

Figure 7.10a shows the scores for those subjects whose clinical evaluation changed over time. The agreement between ELBA and SUVr dichotomized scores is 97.4%. Most MCI→AD (89%) do fall into the SUVr positive / ELBA positive quadrant, as well as most MCI→NS (87.5%) fall into the SUVr negative / ELBA negative quadrant.

When considering CSF Aβ_{42} values – which were available for 63 out of 78 converters – the agreement with ELBA was slightly better than with SUVr (93.7% vs. 90.5%, figure 7.10b and 7.10c).

Very limited discrepancies with the clinical evaluation are apparent. For instance, there are two subjects (1 MCI→AD and 1 NS→AD) who exhibit consistently negative markers (ELBA, SUVr and CSF) despite their final clinical assessment, another NS→AD is borderline negative for CSF and ELBA and borderline positive for SUVr.

Table 7.1: Confusion matrix for ELBA score / SUVr (a) and visual assessment (open session, b)

a)

N=488		ELBA	
		positive	*negative*
SUVr	*positive*	190	25
	negative	13	260

b)

SUVr	ELBA	Visual assessment	
		positive	*negative*
positive	*positive*	189	1
positive	*negative*	7	18
negative	*positive*	7	6
negative	*negative*	1	259

7.2 TDr

TDr was validated on the Multicentric Pilot Dataset, which has significantly less scans and centers than the ADNI dataset used in the ELBA validation. While this dataset has enough subject and variety to serve as validation, we lack proper information on clinical follow-up, CSF and other auxiliary steps like the consensus-open session. For this reason we limit the results to a descriptive statistics.

First we show the TDr statistics per center. As seen in figure 7.8 all centers but Geneva share compatible statistics. This discrepancy though cannot be judged at face value because it can depend on several factors such as the patient selection and – most importantly – on the acquisition protocol.

Performancewise, TDr shows excellent results both when grouped by each center and globally. The AUC with respect to the visual assessment is: AUC=1.00 (Brescia, Ginevra, Pavia) and AUC=0.99 (Genova). Overall, TDr scores AUC=0.99. For comparison, the respective SUVr and ELBA performances are summarized in table 7.4.

In terms of correlation with other semi-quantification methods, results are summarized in figure 7.9 and table 7.4. We find that TDr significantly correlates with the two validated methods, although it clearly relates better with ELBA, particularly when we consider the negative and positive classes separately. The cut-offs shown in figure 7.9 are: $c_{SUVr} = 1.12$, $c_{ELBA} = 0.90$, $c_{TDr} = 0.62$.

ELBA/SUVr quadrant	Visual assess. (blind)	Baseline clinical eval. NS	EMCI	MCI	LMCI	AD	*Scan tot.*
-/-	N	81	71	21	31	4	208
	P	0	0	0	0	0	0
	U	20	18	3	6	0	47
	C	3	2	0	0	0	5
-/+	N	4	5	0	0	0	9
	P	1	2	0	1	0	4
	U	2	4	1	1	0	8
	C	2	2	0	0	0	4
+/-	N	0	0	0	0	0	0
	P	3	1	0	2	0	6
	U	1	2	0	0	0	3
	C	0	4	0	0	0	4
+/+	N	0	0	0	0	0	0
	P	19	57	25	57	18	176
	U	2	0	1	2	0	5
	C	2	4	1	2	0	9
Scan tot.		140	172	52	102	22	488

Table 7.2: Binary semi-quantification versus visual assessment and clinical evaluation. Classification summary of all scans using the visual assessment after the blind session (see figure 7.11). P and N refers to scans consistently (i.e. by all readers) tagged *positive* and *negative* respectively. U refers to scans which received at least one *uncertain* tag but no contrasting assessment; C refers to scans which received contrasting assessment (even when together with *uncertain* tags). NS = cognitively normal subjects, EMCI/LMCI/MCI = early- / late- / mild cognitive impairment, AD = Alzheimer's Disease. Values represent the number of scans in each class.

Table 7.3: Confusion matrix for CSF Aβ_{42} concentration versus the consensus visual assessment (a), ELBA (b) and SUVr scores (c).

a)

N=203		Visual evaluation *negative*	*positive*
Aβ_{42} [ng/L]	> 174	100	3
	< 174	18	82

b)

N=203		ELBA score < 0	> 0
Aβ_{42} [ng/L]	> 174	100	3
	< 174	16	84

c)

N=203		SUVr score < 1.23	> 1.23
Aβ_{42} [ng/L]	> 174	96	7
	< 174	15	85

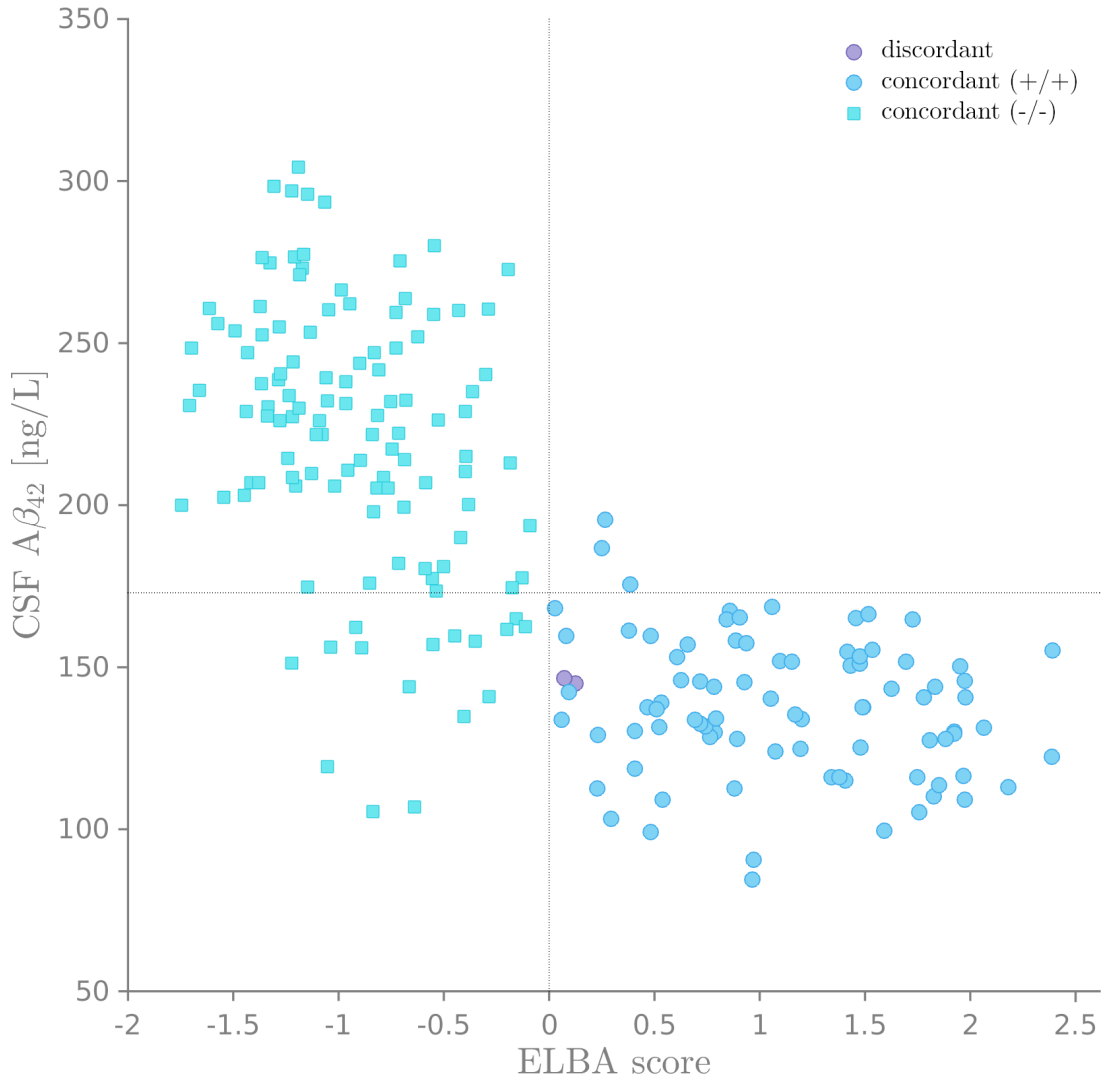

Figure 7.5: CSF Aβ_{42} level versus baseline ELBA. Markers are grouped by binarized ELBA score being either concordant or discordant with the consensual visual evaluation (open phase). Cut-off are marked with thin dotted lines.

Table 7.4: Performance (AUC) versus visual assessment on the Multicentric Pilot Dataset (left). Pearson r (p-value) between methods (right).

Site	TDr	SUVr	ELBA
Brescia	1.00	0.99	1.00
Ginevra	1.00	0.94	0.99
Pavia	1.00	0.94	0.98
Genova	0.99	0.76	0.99
Whole set	0.99	0.92	0.99

Methods	Pearson r (p-value)		
	whole set	negative	positive
TDr/SUVr	0.65 ($< 10^{-3}$)	0.21 (0.068)	0.09 (0.429)
TDr/ELBA	0.85 ($< 10^{-3}$)	0.57 ($< 10^{-3}$)	0.32 (0.006)
SUVr/ELBA	0.67 ($< 10^{-3}$)	0.35 (0.002)	0.03 (0.786)

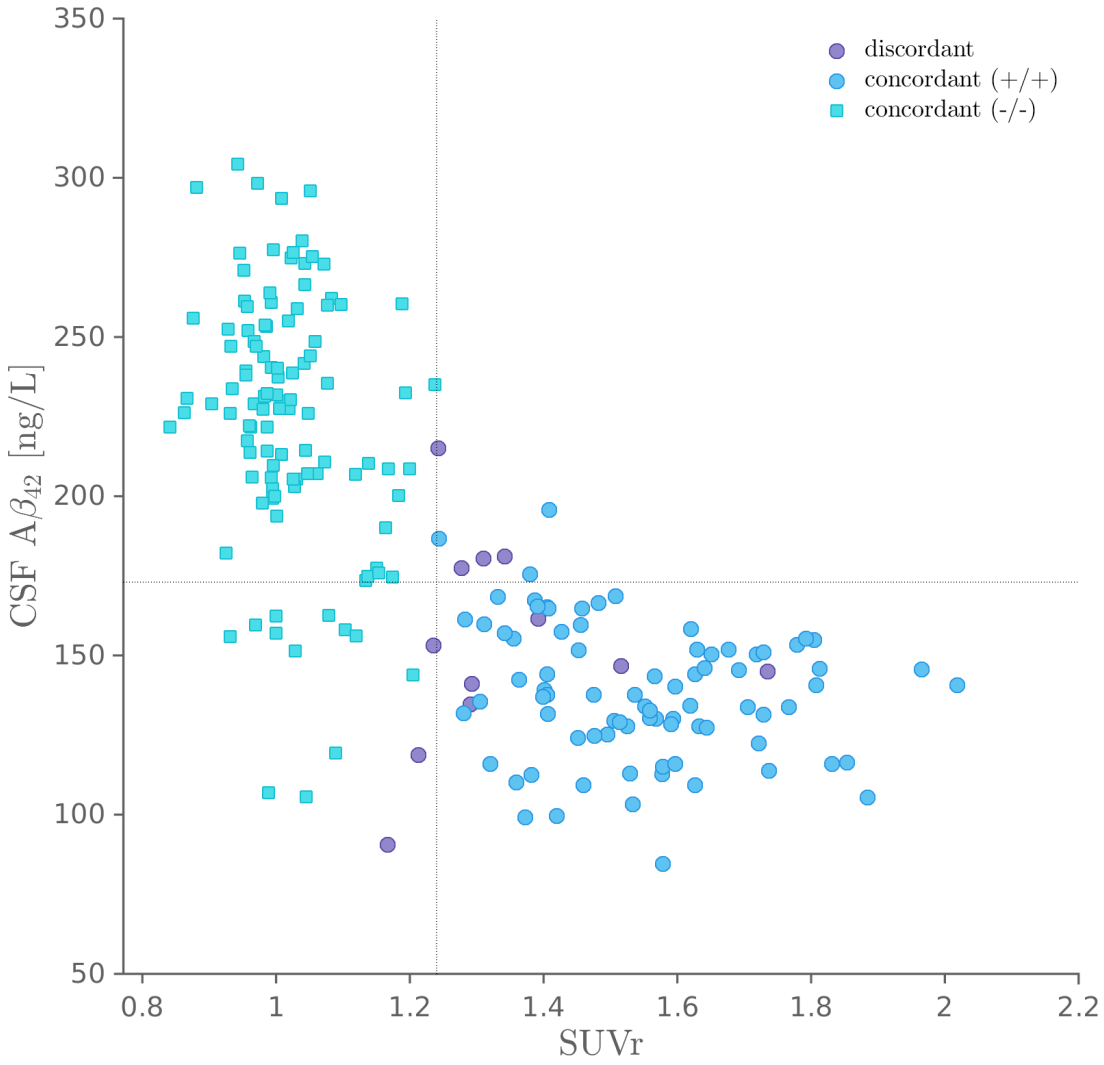

Figure 7.6: CSF Aβ_{42} concentration versus baseline SUVr score. Markers are grouped by binarized SUVr score being either concordant or discordant with the visual evaluation. Cut-off are marked with thin dotted lines.

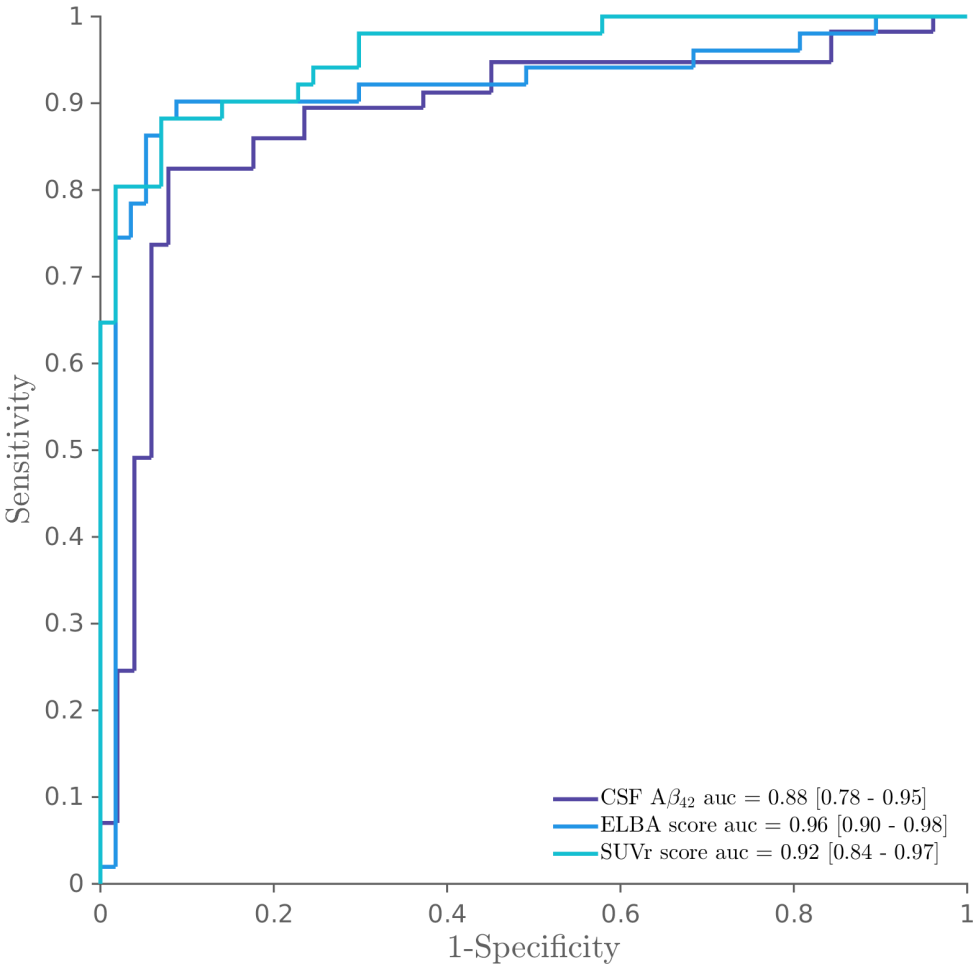

Figure 7.7: ROC curves for 57 cognitively normal (NS) and 51 with Alzheimer's disease (AD) subjects with confirmed diagnosis at follow-up visits. 95% CL are provided on the area under the ROC curves (auc).

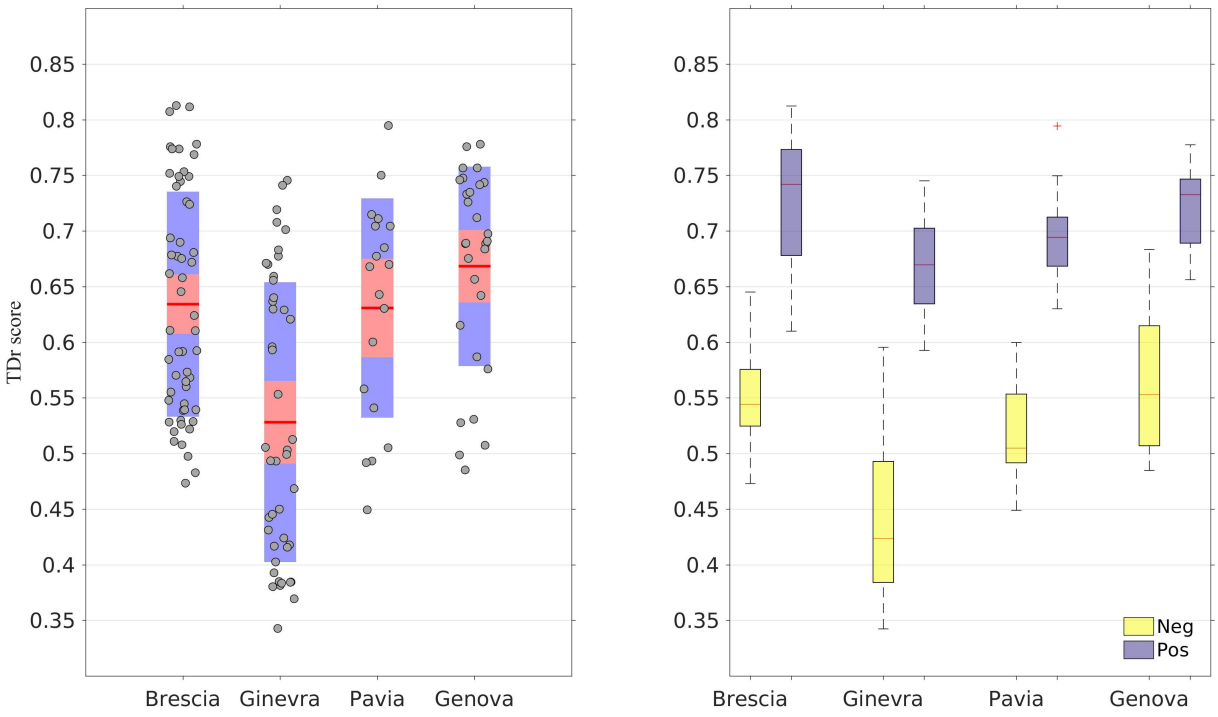

Figure 7.8: Raw TDr scores versus center in the Multicentric Pilot Dataset. Left: TDr distribution versus center; median (red line), median CL (0.95%, red area); ± one standard deviation (blue area); dots are the single data points. Right: TDr distribution versus visual assessment split by center; median (red line); 25% and 75% are within the colored boxes.

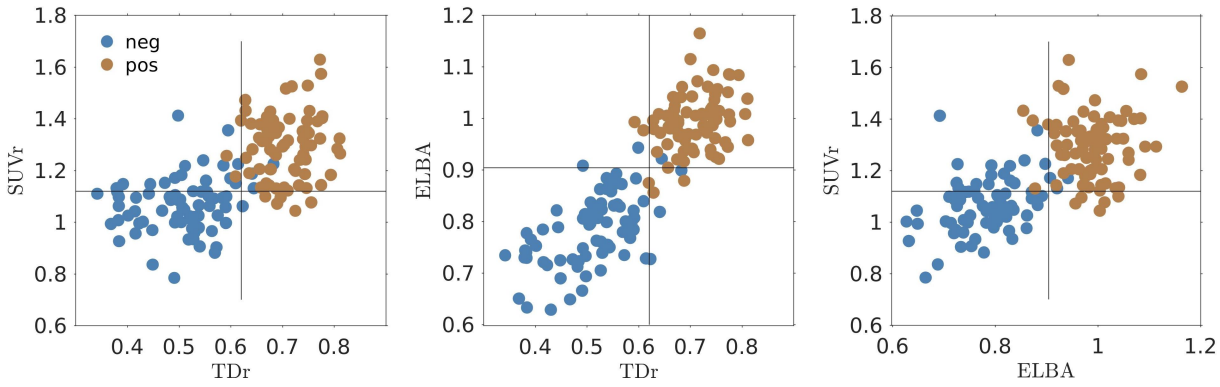

Figure 7.9: Scatter plot of all three semi-quantification methods versus the visual assessment. Black horizontal and vertical lines show the cut-off.

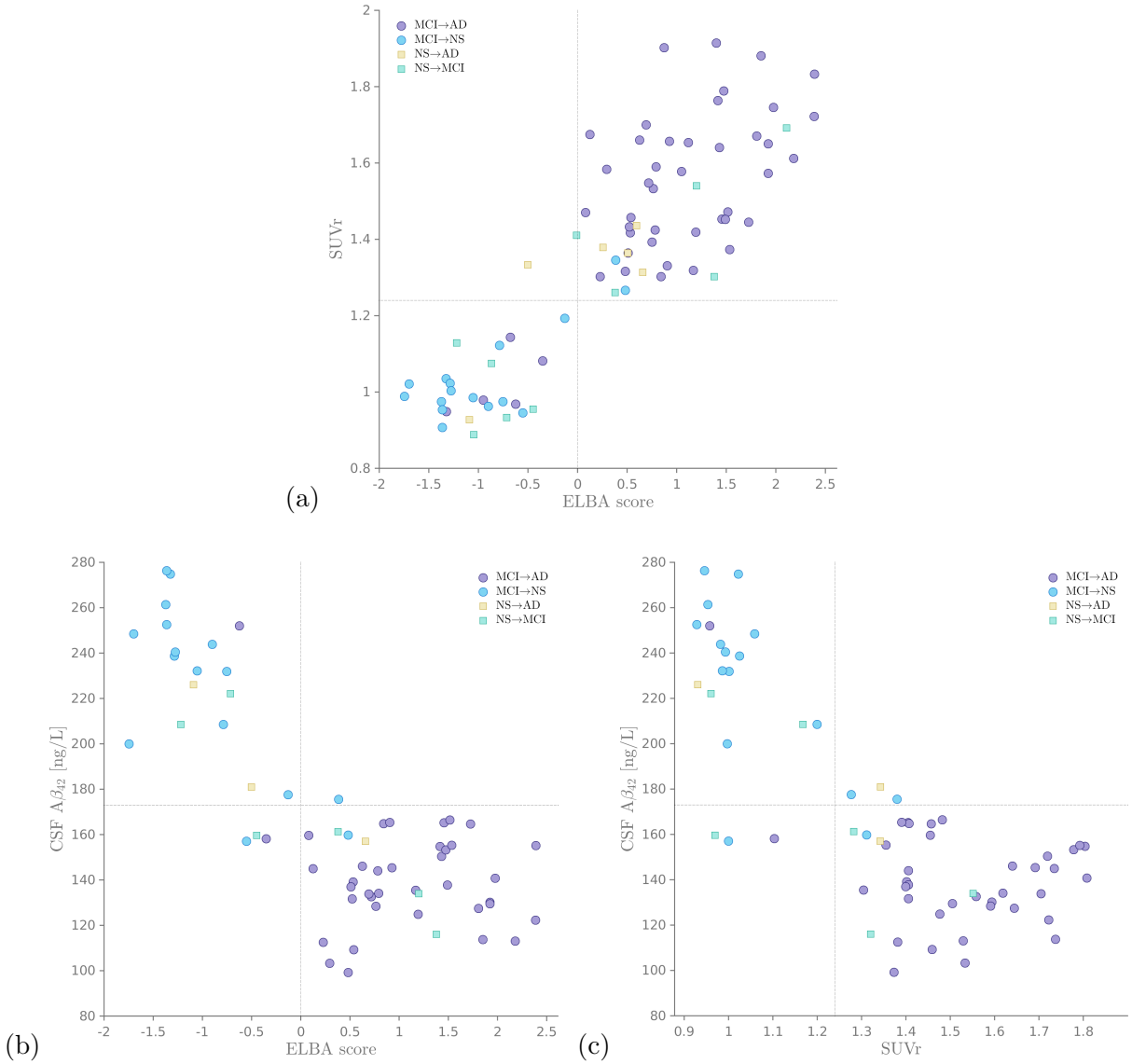

Figure 7.10: Baseline CSF Aβ_{42}, ELBA and SUVr scores grouped by clinical evaluation (baseline→follow-up) for subjects whose initial assessment changed at some later visit. Cut-off are marked with dotted lines. In plot (a) the number of subjects is 78. In plot (b) and (c) the number of subjects is 63 out of 78, that are those for which CSF data were available too. NS = cognitively normal subjects, MCI = mild cognitive impairment, AD = Alzheimer's Disease.

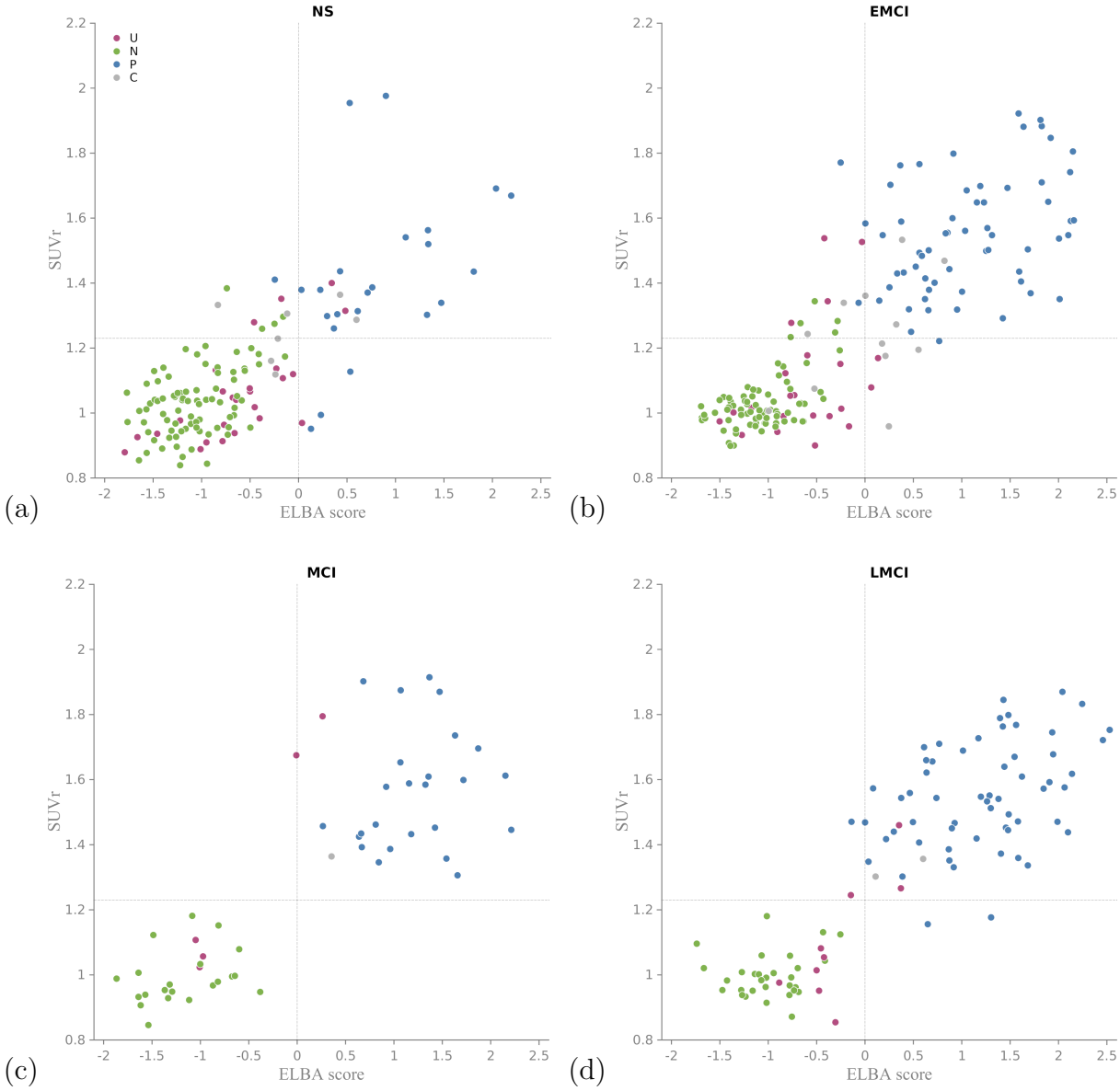

Figure 7.11: ELBA score / SUVr values versus visual assessment after the blind session, grouped by baseline clinical evaluation. P and N refers to scans consistently tagged *positive* and *negative* respectively. U refers to scans which received at least one *uncertain* tag but no contrasting assessment; C refers to scans which received contrasting assessment. NS = cognitively normal subjects, EMCI/LMCI/MCI = early- / late- / mild cognitive impairment, AD = Alzheimer's Disease.

Chapter 8

Advanced topics

8.1 Parcellation

So far, we have only considered a semi-quantitative analysis performed on a macro-region (b_i) that contains all cortical areas subject to accumulation of Aβ.

Several studies (Clark et al., 2012; Nelissen et al., 2009; Thurfjell et al., 2014) have assessed the amyloid load in specific brain regions (frontal, lateral, temporal, parietal, anterior cingulate and precune cortex), as these regions show greater tracer retention in positive subjects than in the normal control (Ziolko et al., 2006; Engler et al., 2006), consistently with pathological studies of accumulation of amyloid (Braak and Braak, 1991; Thal et al., 2002). In some cases, the analysis of specific regions can provide a more sensitive measure than the average over the entire cerebral cortex (Rodrigue et al., 2012; Yotter et al., 2013). For these reasons it is interesting to perform a semi-quantitative regional analysis.

From the MNI Cortical Atlas we selected the subdivision of the cortex and adjacent tissued into 25 contralateral regions P_k, $k = 1..50$ (figure 8.1). The P_k were added to the macro-region consisting of the whole brain, thus obtaining 51 indicators of the amyloid load per scan and per method.

The integration of a predefined parcellation with the specific ROIs relevant to each semi-quantification method is not straightforward. The proposed solution is not necessarily the best nor the only one. This adaptation though is useful to compare several methods

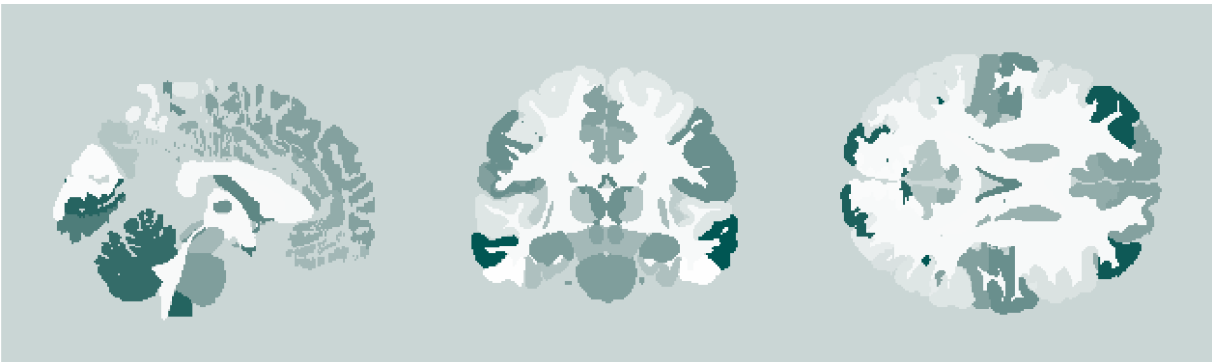

Figure 8.1: Sample cross-sections of the parcellation regions P_k.

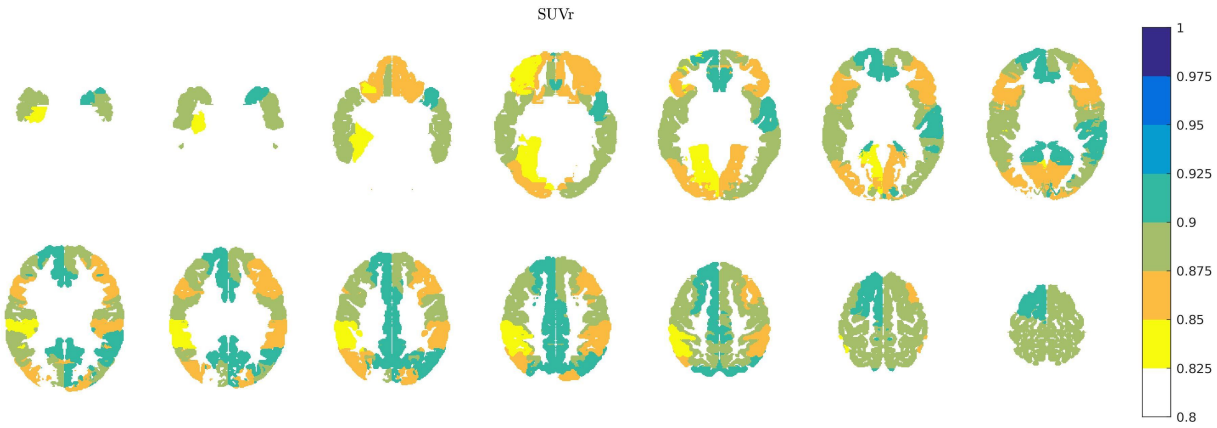

Figure 8.2: Regional AUC versus visual assessment on the parcelled atlas.

on the same ground.

The regional analysis was carried out with some slight differences with respect to the three methods (SUVr, ELBA and TDr) due to their peculiar approach to ROIs.

8.1.1 SUVr regionalization

The normative dataset used to compute SUVr ROIs (see paragrah 4.4) was reworked to give a whole-image, voxel-based value that can be used to weight the uptake information on each voxel. This whole image map expresses a value w proportional to the AUC between negative and positive scans at the voxel level: $w = 2(AUC - 0.5)$, see figure 4.5.

The regional SUVr value S on the region P_k is therefore

$$S_k = \frac{\sum_{i \in P_k} w_i I_i}{\sum_{i \in \{\cup_k P_k\}} w_i}$$

where i is the index of all voxels, P_k is the parcellation under consideration, $\{\cup_k P_k\}$ is the union of all parcellation (i.e. the atlas), I_i is the normalized intensity in the voxel i (intensity normalized to the whole cerebellum).

Obviously, SUVr values do not exhibit the same discrimination ability on all P_k. An approximate performance indicator is shown in figure 8.2, where the AUC is color-encoded on the parcellations.

8.1.2 ELBA regionalization

The peculiar treatment of intensity patterns in ELBA do not allow the direct parcellation. For instance, ELBA cannot work on strict cortical P_k as it measures the intensity distribution when passing between the white and the gray matter. In addition, the definition of iso-intensity surfaces require a relatively big volume (i.e. one with a sufficient number of voxels) so we expect that ELBA would perform poorly on small P_k.

To partially overcome these issues, each P_k undergoes a dilation with a radius of 4 mm ($P_k \rightarrow \hat{P}_k$). This provides an enlargement of small parcels and ensures that the white

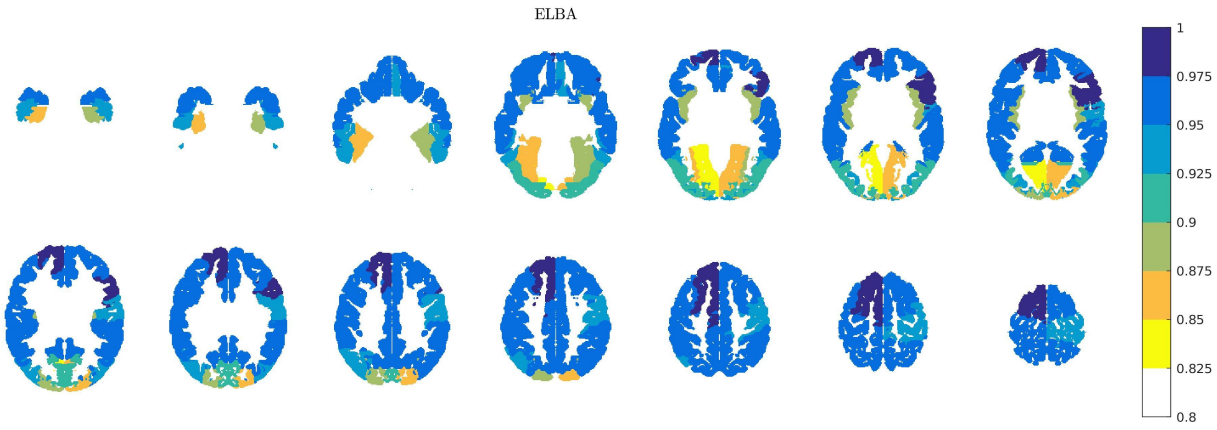

Figure 8.3: Regional AUC versus visual assessment on the parcelled atlas.

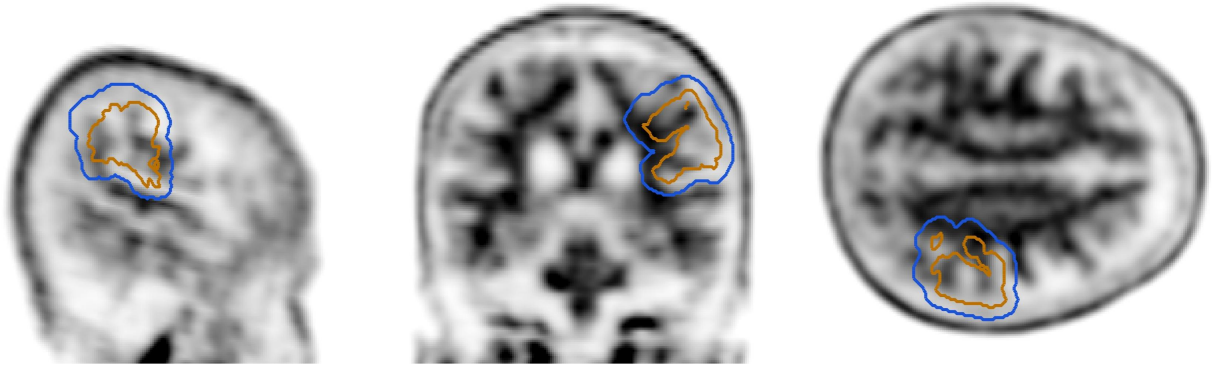

Figure 8.4: Sample parcellation region P_k (orange) and its extension $\hat{P}_k$ (blue) superimposed onto a sample negative scan.

matter adjacent to the gray cortical one is included in the $\hat{P}_k$. Figure 8.4 shows an example of the enlarged parcellation $\hat{P}_k$.

The drawback is that $\hat{P}_k$ are now overlapping among themselves. This though does not constitutes a real problem as the regional ELBA values are loosely related with the global one. Indeed, the sphericity G (paragrah 5.1.3, geometric feature) of an iso-intensity surface computed – say – on two adjacent regions $\hat{P}_k \cup \hat{P}_h$ is not the sum of those computed on the single regions. The same reasoning holds for the intensity feature I (paragrah 5.1.3, intensity feature) Therefore, in ELBA, the regional values might be completely different from the whole brain assessment. Figure 8.3 shows the color-encoded AUC.

8.1.3 TDr regionalization

This method is the most adaptable to parcellation. The relatively high spatial density computed from the early scan ensures that in each region P_k there is a sufficient number of TDr ROIs to be relevant for the analysis. The regional TDr values comes therefore from the intersection between each P_k and the TDr ROIs from the early scan. The related AUC plot is in figure 8.5.

We can see that regional performance is rather scattered, particularly when SUVr is

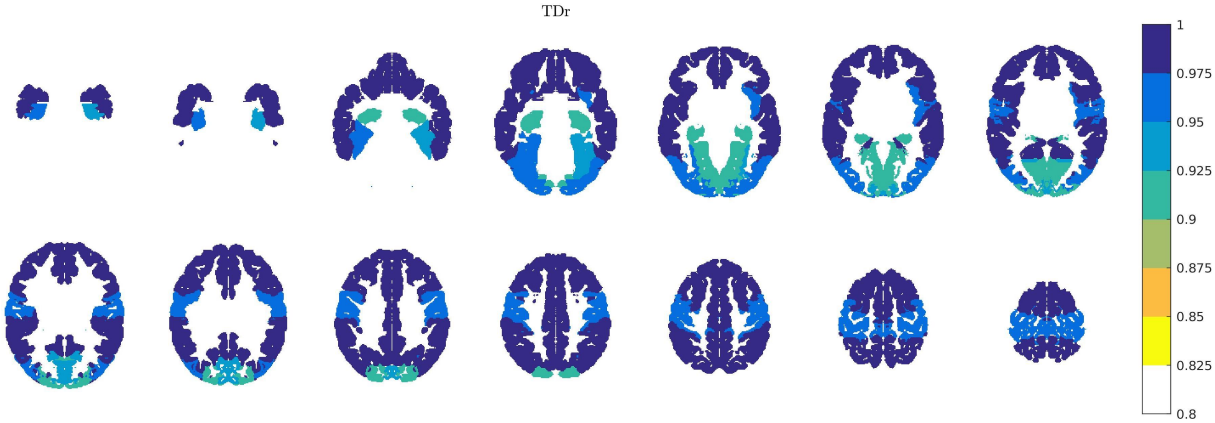

Figure 8.5: Regional AUC versus visual assessment on the parcelled atlas.

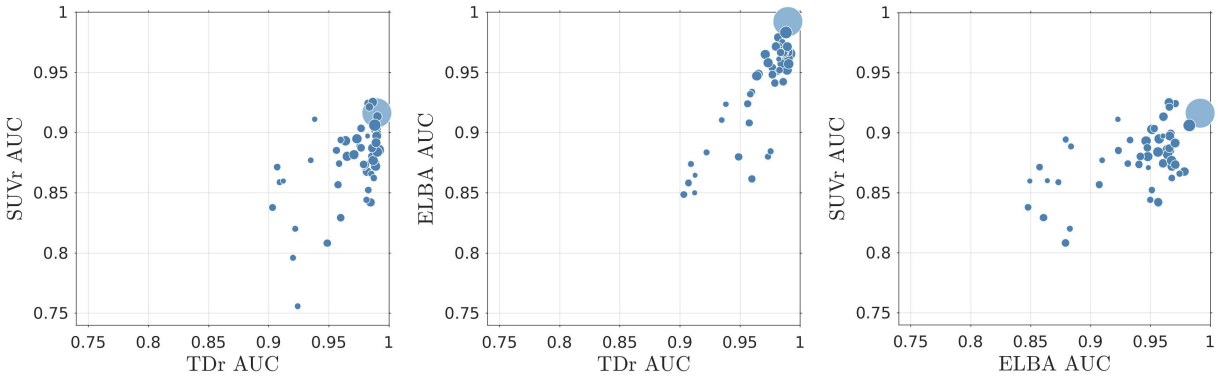

Figure 8.6: Compared AUC on the parcelled atlas. Marker area is proportional to the parcellation volume. The biggest marker (light blue) is the whole brain.

involved (figure 8.6).

8.2 Longitudinal studies

Amyloid load quantification is especially important in the context of longitudinal analysis. With the foreseen availability of disease-modifying drugs that will likely induce anti-amyloid effects, it is important to evaluate a change in amyloid load with great accuracy.

Unfortunately, current semi-quantification techniques are still rather lacking in terms of reproducibility and test-retest error. In this chapter I will show a detailed comparison of ELBA and SUVr in terms of measure accuracy in a longitudinal setting. TDr will also be investigated, albeit with a different approach.

8.2.1 ADNI dataset analysis with ELBA and SUVr

We computed the annualizes score change (ASC) using follow-up scans. For the 228 subjects with a baseline and a follow-up scan we used the formula

$$\mathrm{ASC} = \frac{s_f - s_b}{t_f - t_b}$$

where f and b label the follow-up and baseline score and subject's age. For the 16 subjects with two follow-up we computed the least square linear regression

$$s + \epsilon = mt + b$$

of the score s versus time t (b is the intercept and ϵ the residuals) and the ASC is simply

$$\mathrm{ASC} = m$$

To compare the numbers with other indexes (such as those computed with SUVr quantification) we considered the normalized quantity

$$\delta = \frac{\mathrm{ASC}}{\mathrm{iqr}(s)} \times 100$$

where the interquartile range (iqr) of the score s is used as normalization factor to estimate the relative score change over the observed population variability.

Using the 16 subjects with three scans each we attempted to estimate the analysis stability and robustness. With the present data we could not evaluate a test/retest paradigm on the same subject (i.e. two repeated scans with subject reposition) so we used the residuals ϵ on the linear regression of the score versus time as a proxy. This works under the hypothesis that the amyloid burden piles up slowly and linearly at least within the follow-up time (approx. 4 years) – an assumption that agrees with the accepted neuropathological models – and assuming that the technical errors on repeated scans are independent from the subject's amyloid burden.

Deviation from the linear behaviour is then used as a surrogate to the test/retest error and treated as analysis uncertainty - due to protocol, image acquisition, reconstruction and processing - and used to estimate the error on the single examination. It can be thus compared to literature works on the various Aβ ligands ((Joshi et al., 2012; Vandenberghe et al., 2013; Klinger et al., 2013) and (Lopresti et al., 2005)), which show an average test/retest relative error using the global SUVr measurement in the range $3\% - 7\%$.

Obviously, the measured uncertainty would not be due to the feature processing only, but also to the different acquisition conditions, scanners and reconstruction parameters. For instance, we remark that among the 16 subjects with three scans, 11 were acquired on a single scanner on baseline and follow-up, and 5 subjects were acquired with two different scanners at some follow-up.

Finally, we selected a subset of subjects with either negative or borderline ELBA score (i.e. ELBA¡0.5) which were also evaluated by CSF analysis. These subjects were divided into three groups, of progressively lower average Aβ_{42} concentration (ng/L): group A, A$\beta_{42} > 230$; group B, $174 <$ A$\beta_{42} < 230$; group C, A$\beta_{42} < 174$.

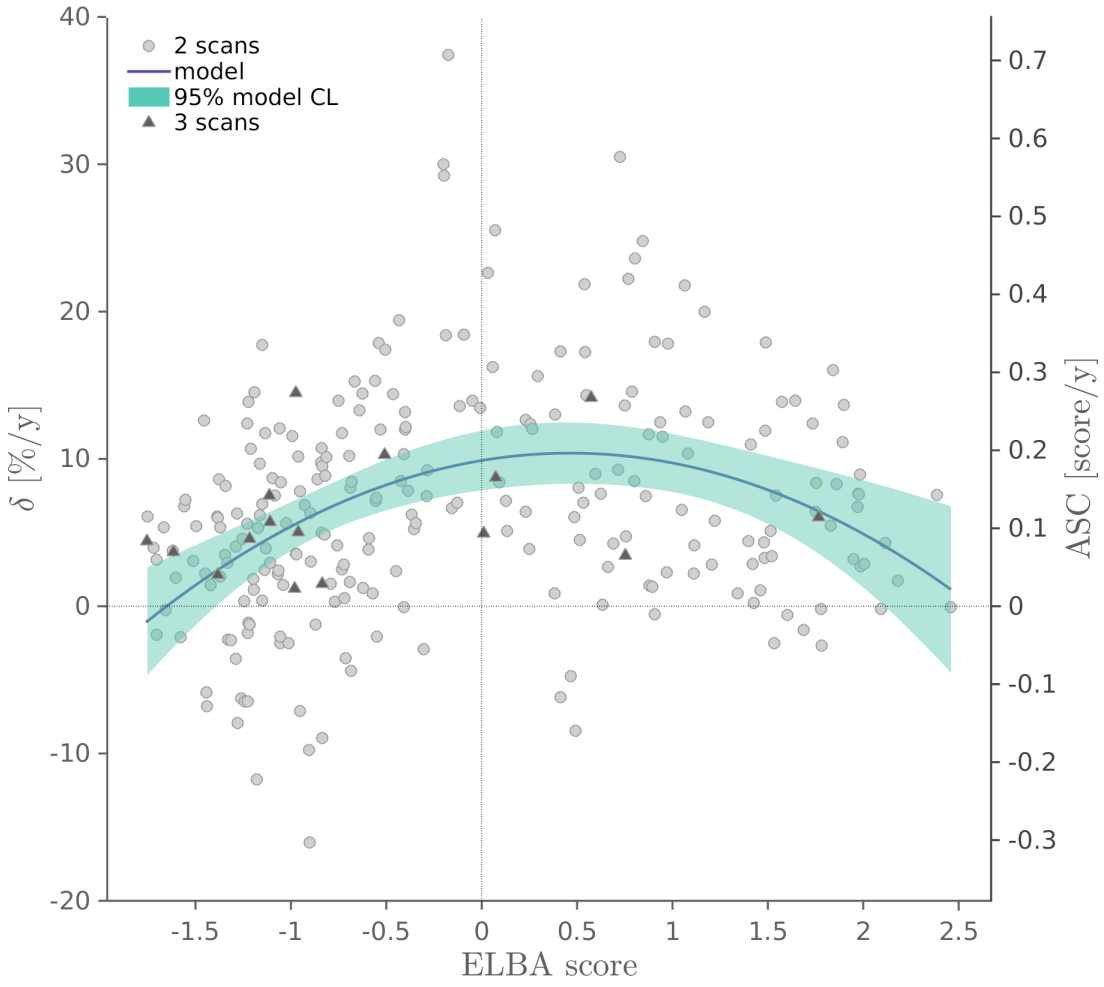

Figure 8.7: Relative (δ, interquartile-range normalized) and absolute annualized score change (ASC) for ELBA vs baseline score. A quadratic model and the CL band on the model are superimposed. Subjects with three scans (triangles) are marked separately from subjects with only two scans (dots).

These three groups contained an approximately equal number of subjects (52, 48 and 34 respectively) where the above-cutoff ensemble was divided into A and B to better reflect the possible trend in longitudinal behaviour with respect to CSF outcome. An analysis of variance was applied to check for significant differences among groups.

Results

Longitudinal analysis on the subjects with a baseline and one follow-up scans showed rather scattered values, although a pattern could be clearly discerned. Figure 8.7 shows the distribution of δ vs. the average ELBA score together with a 2^{nd}-order polynomial model, used to fit the data. The interquartile value used for normalization of the ASC is $\mathrm{iqr}(s) = 1.89$. An equivalent graph using SUVr values is reported in figure 8.8.

Subjects with two follow-up scans are also plotted on the same graph. Cohortwise, these subjects belonged to NS (5), EMCI (2) and MCI (9). Using these latter 16 subjects we computed the residuals ϵ from the linear fit, to be used as proxy of the test/retest error. The standard deviation of the residuals is approximately $\sigma = 0.084$, which amounts to

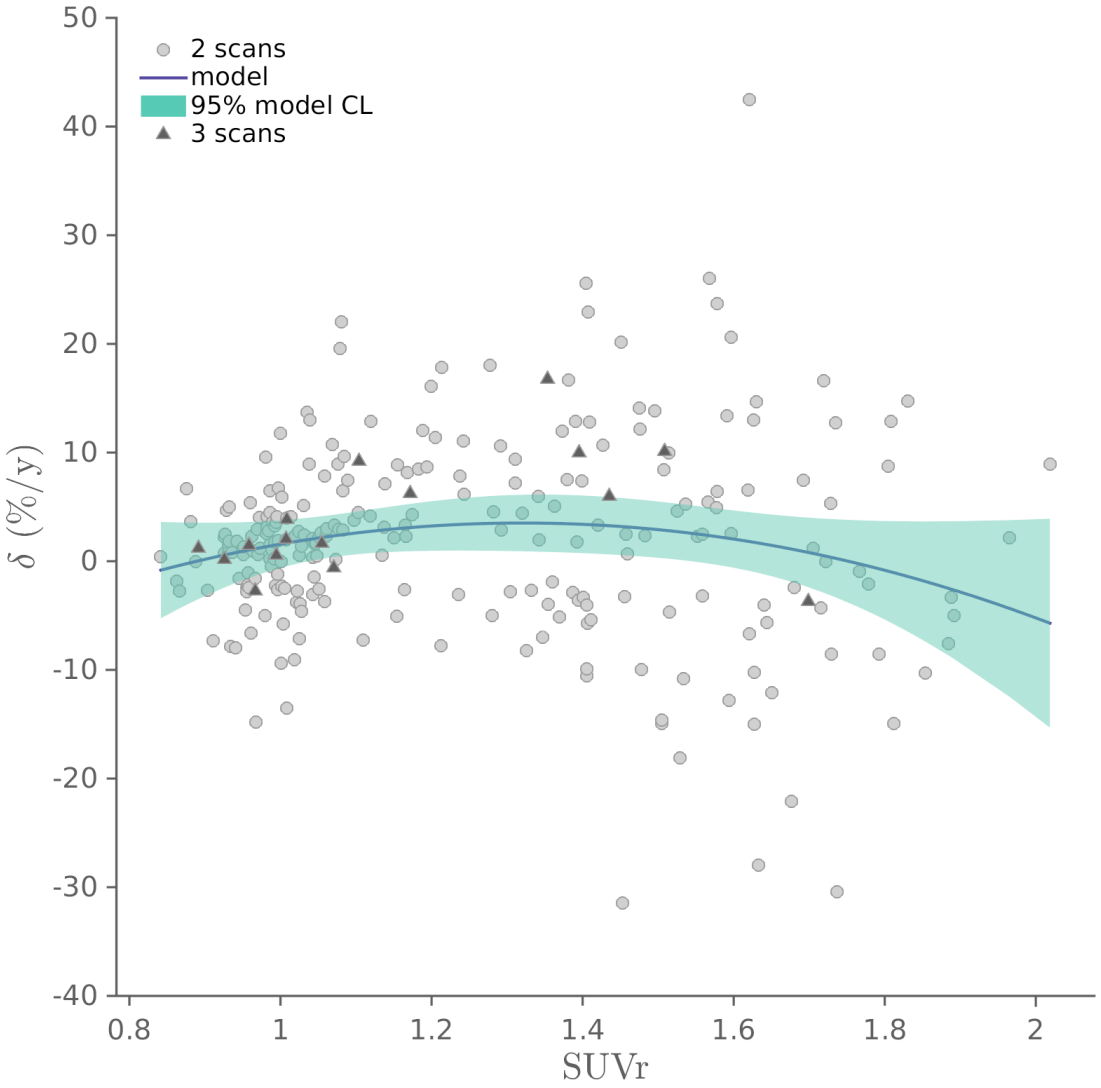

Figure 8.8: Relative score change (δ, interquartile-range normalized) for SUVr vs baseline score. A quadratic model and the CL band on the model are superimposed. Subjects with three scans (triangles) are marked separately from subjects with only two scans (dots).

an estimated relative error $\sigma_r = \sigma/\text{iqr}(s) = 4.4\%$ ($\sigma_r = 2.3\%$ if we normalize on the 95% percentile of the score range). The equivalent SUVr values are $\sigma_{SUVr} = 0.023$ and $\sigma_r = \sigma_{SUVr}/\text{iqr}(SUVr) = 4.6\%$

We computed also the relative annualized score change δ grouped by cohorts. Reported values (expressed in $\%/year$) are the median and its confidence interval (at CL=95%): $\delta = 4.6$ [3.1, 6.1] (NS), $\delta = 8.1$ [6.6, 9.6] (EMCI), $\delta = 3.4$ [0.6, 6.2] (MCI), $\delta = 6.0$ [3.9, 8.2] (LMCI), $\delta = 5.6$ [−0.7, 11.9] (AD). A t-test found the values to be significantly different between EMCI and NS cohorts ($p = 0.02$) and between EMCI and MCI cohorts ($p = 0.02$).

Finally, we analized the ELBA ASC against the CSF Aβ_{42} concentration (figure 8.9 and figure 8.10 for SUVr ASC) for 134 subjects with negative or borderline ELBA score (¡0.5). The Pearson correlation between the two quantities is small but significant ($r = -0.30$, $p = 0.0004$). The grouping by CSF intervals allows to see a trend beyond the noisy ASC data. From the highest mean Aβ_{42} concentrations to the lowest, the ASC ranges from group A = 0.07 [0.03 – 0.10] to group B = 0.14 [0.11 – 0.17], to group C = 0.20 [0.14 – 0.27] (mean and 95% CL on the mean). The analysis of variance indicates that group

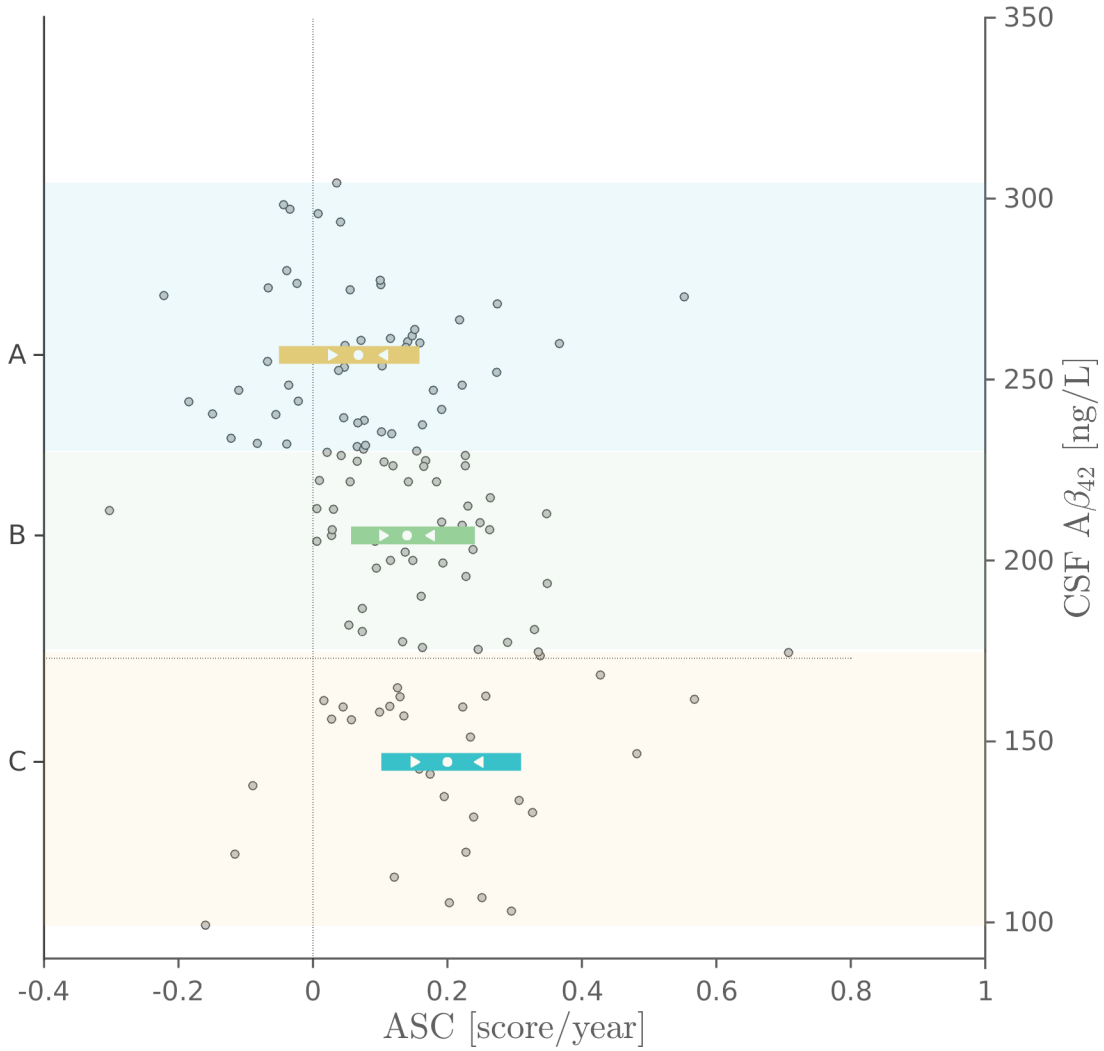

Figure 8.9: CSF Aβ_{42} versus ELBA annualized score change (ASC) for 134 subjects whose baseline ELBA score was < 0.5 (i.e. negative or borderline according to ELBA). Subjects are divided by CSF range (ng/L) into three groups: A ($230 <$ Aβ_{42}), B ($174 <$ A$\beta_{42} < 230$) and C(A$\beta_{42} < 174$). The corresponding ensemble statistics is summarized in the boxplot, where the box length spans the 25% to the 75% ASC percentile, the white dot and the white triangles are the mean and the 95% CL on the mean respectively. Groups A/C, and B/C are significantly different.

A and C, as well as group B and C are significantly different ($p < 10^{-4}$ and $p = 0.02$ respectively). The corresponding analysis on SUVr ASC gives: group A = 0.005 [-0.00 – 0.01], group B = 0.01 [0.00 – 0.02] and group C = 0.03 [0.01 – 0.04]. The between-groups comparison is $p = 0.001$ (A vs. C), and $p = 0.03$ (B vs. C).

8.3 Test-retest: a cross-sectional approach

The trend in amyloid load that can be estimated by longitudinal analysis depends on the test-retest error of the measure. In paragraph 8.2.1 we have seen that the ASC has a rather large uncertainty, and that uncertainty is severely limiting the effectiveness of a longitudinal measure and its clinical potential.

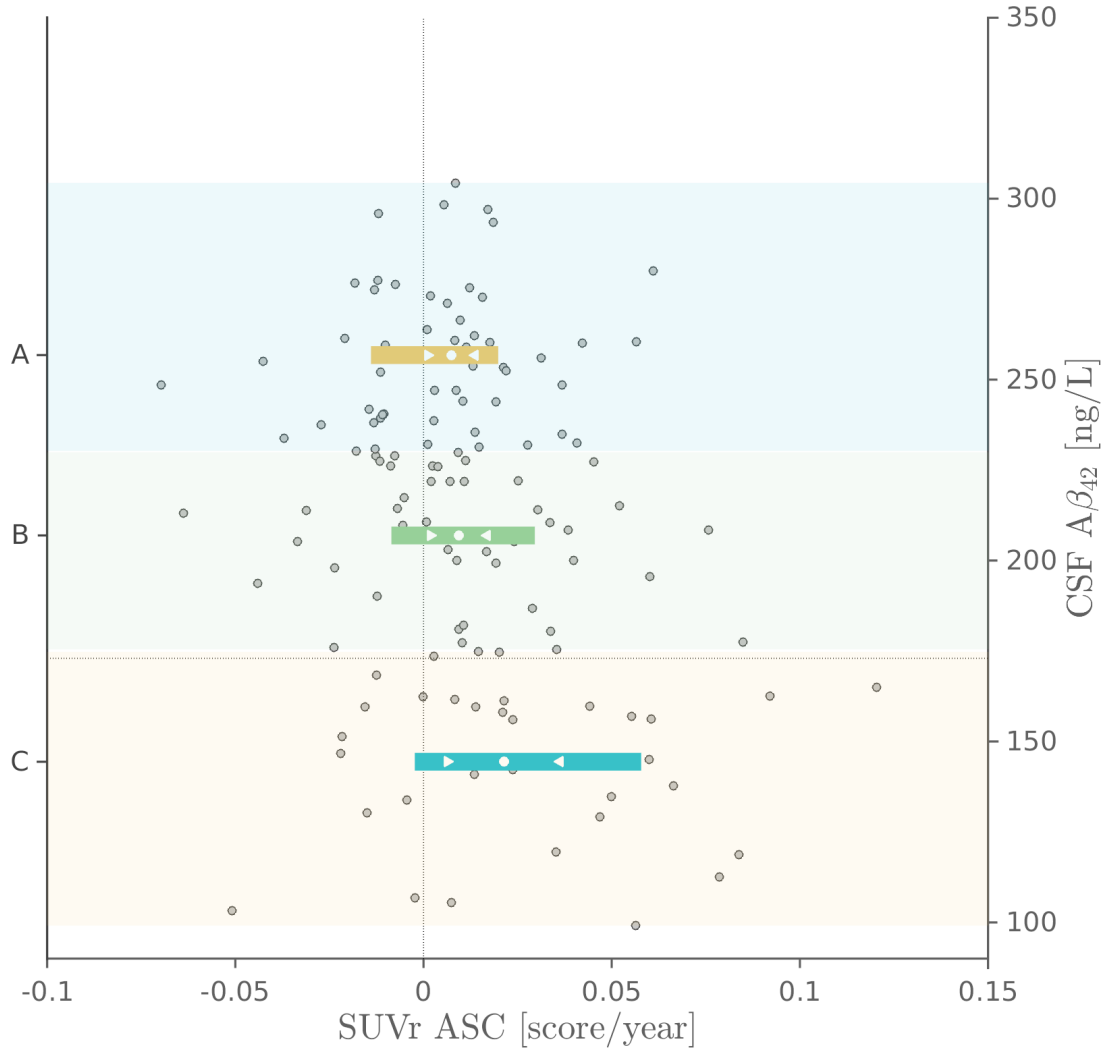

Figure 8.10: CSF Aβ_{42} versus SUVr annualized score change (ASC) for 134 subjects whose baseline score negative or borderline. Subjects are divided by CSF range (ng/L) into three groups: A (230 $<$ Aβ_{42}), B (174 $<$ Aβ_{42} $<$ 230) and C(Aβ_{42} $<$ 174). The corresponding ensemble statistics is summarized in the boxplot, where the box length spans the 25% to the 75% ASC percentile, the white dot and the white triangles are the mean and the 95% CL on the mean respectively. Groups A/C, and B/C are significantly different.

A way to model the actual semi-quantification measure m is to think of it as the sum of two contributions: the "true" amyloid load due to the physiological variability a, and the measure error (both due to the acquisition and to the semi-quantification) δa. So we can write

$$m = a + \delta a$$

If the measure error is unbiased and with zero mean

$$< m > = < a >$$

A cohort-wise measure then approximates the true mean physiological value.

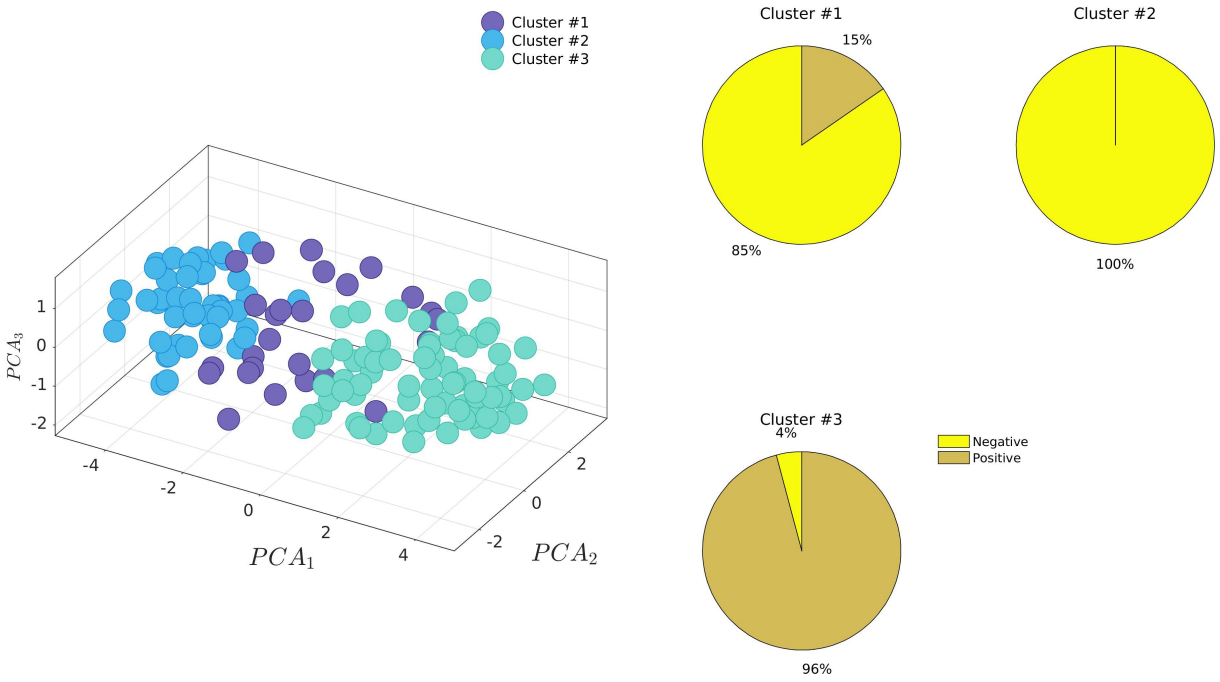

Figure 8.11: Data clustering over the first three PCA components. Pie charts show the percentage of negative and positive subject in each cluster.

In the simplest longitudinal scheme

$$\Delta m = (a_{t_1} - a_{t_0}) + \delta(a_{t_1} - a_{t_0})$$

which becomes:[1]

$$\Delta m = \Delta a + \sqrt{(\delta a_{t_1})^2 + (\delta a_{t_0})^2} \approx \Delta a + \sqrt{2}\delta a$$

In general, δa constitutes the limiting factor in all longitudinal analyses[2].

Typically, one can estimate the measure error δa with a test-retest approach, where the physiological change is assumed to be zero. Ethical considerations though do not allow test-retest protocols with amyloid-PET. While the test-retest error cannot be assessed without a dedicated experiment, we can still use a cross-sectional approach to rank different semi-quantification measures. In this case we can use a cohort-wise estimate of the variance.

The idea is rather simple: suppose you have a set of identical objects and two different tools to measure them, but you do not know the tools' intrinsic error. You want to find out which tool has the lowest error.

Under the hypothesis of independence among measures and tools versus objects, and if you know that the objects are indeed identical, then the variance of all the measures made with one tool is a proxy of the tools intrinsic error. This approach is best achieved in a true test-retest experiment but we can still approximate the "identical objects" condition by population partitioning.

[1] Propagation of errors in a sum.

[2] In longitudinal analyses with multiple time points one can use a model (i.e. linear) to estimate the rate of change. In this case the uncertainty on the rate can be smaller than δa

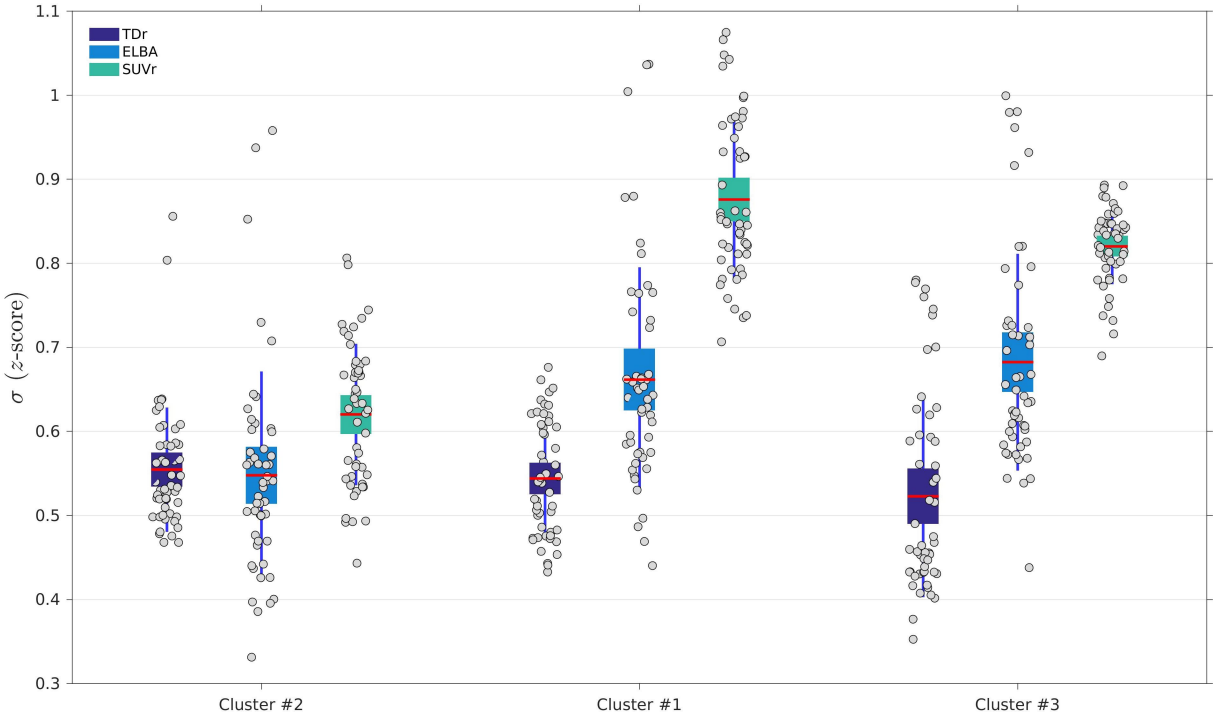

Figure 8.12: Standard deviation of the normalized data (z-score) grouped by clusters and semi-quantification method. Red line is the median, colored boxes are 1 standard deviation of the points, lines are the 95% CL. Clusters are ordered by increasing amyloid positivity.

Let us consider a set of amyloid quantifiers q_j, $j = 1..n$ and apply them to a population P of m subjects. We get a matrix Q_{ij} $i = 1..m$, $j = 1..n$ of measures. This matrix can also be described in terms of its PCA components s_k and let us take enough s_k so that the measure matrix is described up to a high explained variance (for instance 90%). We can now cluster data using the s_k principle components and divide subjects into C number of clusters. By definition, cluster are more homogeneous than the whole population, that is, subjects belonging to one cluster are more similar among themselves than they are to subjects belonging to different clusters.

We can now compute the standard deviations of all measures q on the partitioned measure matrix Q. We get a matrix of standard deviations S_{aj} $a = 1 : C$, $j = 1..n$, that is S_{aj} is the standard deviation computed over the value set $\{Q_{ij},\, i \in C_a\}$, that is computed over the column j of Q where the rows belongs to the cluster C_a.

For two semi-quantification techniques ($i = 1..2$) applied to *the same* scan group and under the hypothesis that the physiological variability and the semi-quantification are independent, we can write

$$\sigma^2_{m_i} = \sigma^2_a + \sigma^2_{\delta a_i}$$

We should now compare the two variances $\sigma^2_{m_1}$ and $\sigma^2_{m_2}$. This time though we have limited σ_a because of the clustering (although we cannot quantify it yet). Actually, if we have a set of values for each technique (such as the regional information or paragraph 8.1), we have to compare a set of σ too. In addition, as we assumed independence of physiological variability and the semi-quantification, we can take all homogeneous cohorts and consider

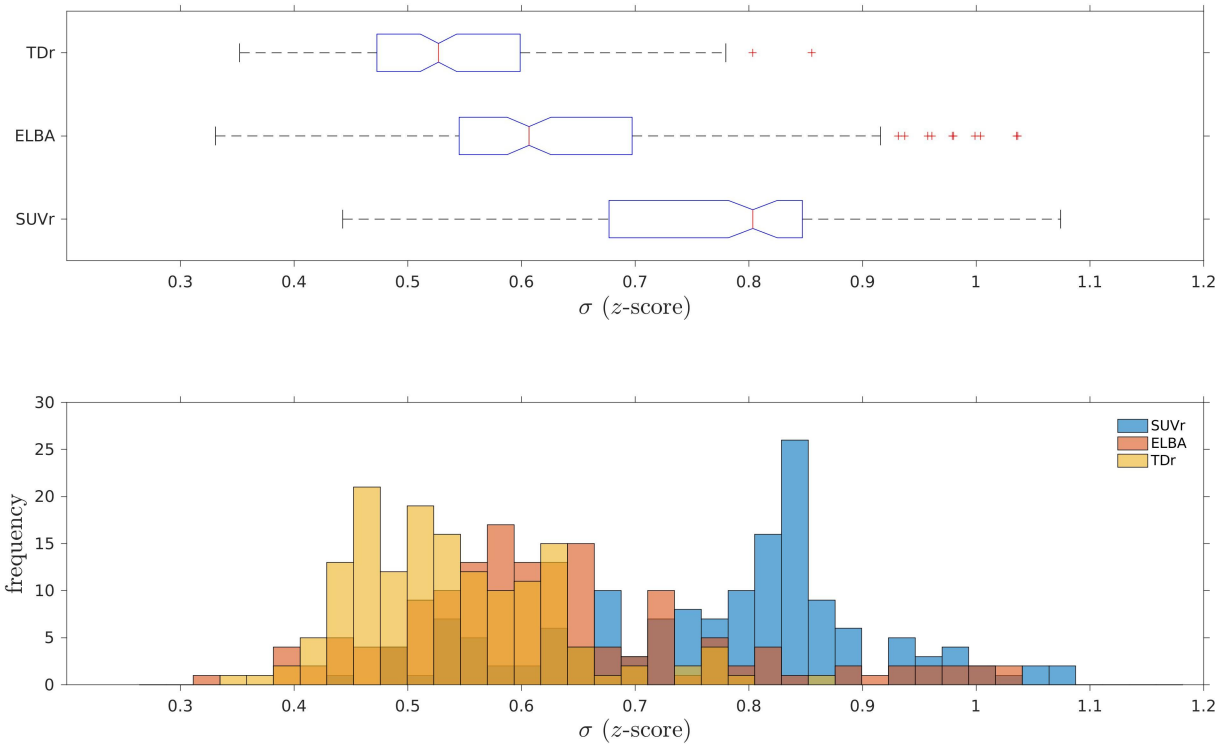

Figure 8.13: Overall distribution of the standard deviation (on normalized data). The distribution merges all regional quantifications and clusters.

their variances together. In the end therefore, we can compare a whole set of variances from one technique with another set of variances from another technique. The one which score lower is the "better" technique in terms of test/retest capabilities as it shows that it has a significantly lower intrinsic variance when measured over the whole population.

We applied this reasoning to the Multicentric Pilot Dataset (MPD) to test the performances of SUVr, ELBA and TDr. We remind that this is test does not imply a direct measure of the test-retest error, rather, it can only rank methods on the smallest variance[3]. Still, as we could not apply the TDr to the ADNI dataset (as in paragraph 8.2.1) we can at least assess the TDr potential in terms of intrinsic error by comparing it to the other two methods on the same dataset.

8.3.1 Implementation and results

Data were clustered into three partitions (of 26, 48 and 73 elements respectively) using the first three PCA components, which explain $\approx 85\%$ of the total variance (figure 8.11). We applied hierarchical clustering with Ward's linkage, and these clusters roughly correspond to negative, positive and borderline subjects (cluster 2, 1 and 3 respectively).

We constructed a measure matrix by juxtaposing the 51 quantifiers from each method, resulting in a matrix of size 147×153 . As each method has its own value range, we normalized the measure matrix into z-score in order to compare standard deviations. For each cluster and for each method, we computed the standard deviation of the regional

[3]although it can be considered as a superior limit of the test-retest error.

and global z-scored quantifiers, which amounts to 51 values per method and per cluster. In total, each method is characterized by 51 (quantifiers per method) $\times$ 3 (clusters) = 153 standard deviation values.

Results are shown in figure 8.12 and 8.13. The overall distributions (shown in figure 8.13 both as histograms and box plots) clearly indicate the TDr as the method with the lowest intrinsic error ($p < 10^{-3}$) with respect to both ELBA and SUVr (ANOVA). The test between ELBA and SUVr shows that ELBA is also significantly lower than SUVr ($p < 10^{-3}$). Incidentally, the results comparing ELBA and SUVr confirm the findings in paragraph 8.2.1, where ELBA had a better consistency in a longitudinal analysis.

Data grouped by cluster and method show that TDr is consistent on all data while ELBA and SUVr are both significantly different in clusters 1 and 3 versus cluster 2. This indicates that the uncertainties in these two methods are biased by the amyloid load, so that our assumption of independence of the error from the data is not verified.

Chapter 9

Beyond dichotomy

All semi-quantification methods and their variants proved sufficiently accurate when compared to the negative / positive assessment, either provided by a single reader or in the more complex multi-reader approach. Yet, if we limit the quantification information to the simple dichotomic answer we find that it is hardly worth the analysis.

In a recent paper (Pontecorvo et al., 2017) authors complemented both expert and novice readers with a SUVr-based semi-quantification software. The goal was to examine the feasibility of quantitation to augment interpretation of PET amyloid imaging. They found that augmentation of visual interpretation of PET amyloid images with quantitative information (obtained using commercially available software packages) was significant, raising the overall accuracy from 90.1% to 93.1% ($p < 0.0001$). Curiously though, they also found that the use of quantitation did not reduce the accuracy of readers who were already performing with above average accuracy on the visual read (a meagre result...), and could improve the accuracy and confidence of some readers in clinically relevant cases (a much better result!).

That is to say that, on one side, easy-to-read cases are finely assessed both by novice and by expert readers, limiting the quantification tool to a confirmatory role. On the other side, difficult-to-read scans cannot easily be trusted to a single implementation of a semi-quantification software, due to all the known uncertainties in image treatment, normative dataset and physiological peculiarity. Hence, there is little to gain with semi-quantification when used only for confirmation of the binary assessment[1].

We propose a paradigm shift that abandons the binary assessment in favor of a continuous scale (rank), ranking scans form the most negative to the most positive. We understand that this rank is but the tip of the iceberg, as a single value cannot possibly entail all the subtle nuances and complex information found in a 3D image.

9.1 From binary reading to ranking

A dichotomous evaluation is appropriate in the assessment of amyloid accumulation in cases where the presence or absence of Aβ is evident: in these patients this process has either not yet begun or it has already reached a plateau. Due to the current use of

[1] An obvious exception though is in the analysis of large dataset, where the statistical approach is central to the study.

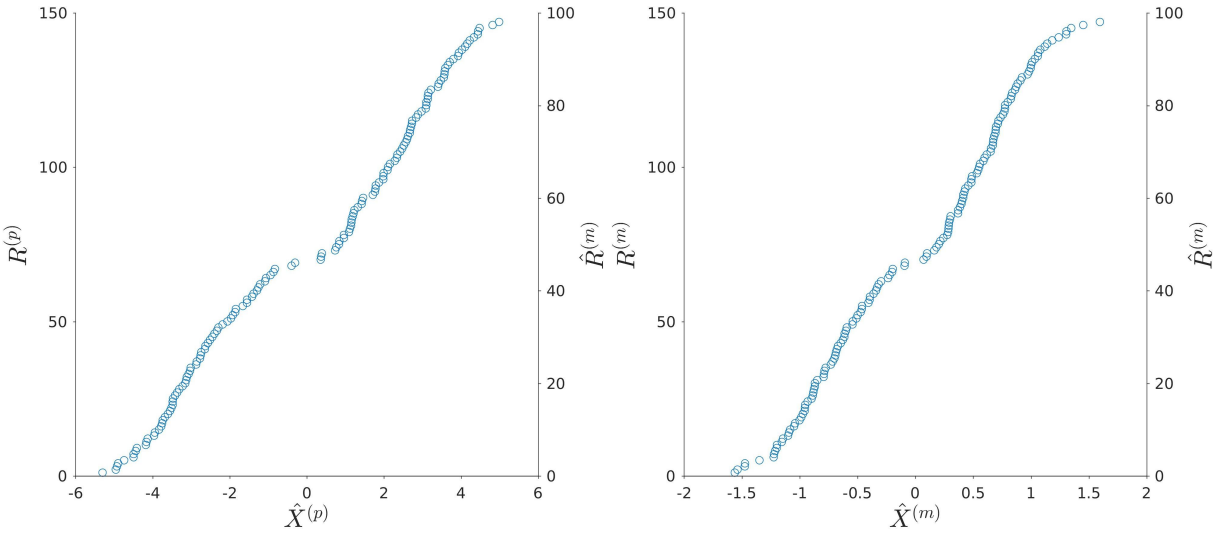

Figure 9.1: Examples of the link between the estimator $\hat{X}$ and the rank R. Left: PCA-based weights with quantile normalization ($\hat{X}^{(p)}$ and $R^{(p)}$); right: constant weights with z-score normalization ($\hat{X}^{(m)}$ and $R^{(m)}$). Each rank can also be normalized ($R \to \hat{R}$, right axis of each plot) on a standard range $[0 - 100]$ to get a number independent from the normative dataset size.

amyloid-PET in the diagnostic flowchart (see section 3.2), most patients should easily fall into these categories.

Yet, clinical experience show that the number of difficult classifications is not negligible, with an approximate figure of merit of $\approx 20\%$ of all scans. These scans can be difficult to interpret for a number of reasons: they can belong to subjects whose accumulation process is still in progress; they can present confusing elements (i.e. bad quality); or there can be structural and physiological reasons (such as the presence of significant local atrophy).

Whatever the reason, these scans are almost always very close to the cut-off value regardless ot the semi-quantification method. Therefore the dichotomous evaluation is "forced" on them.

If we assume that technical issues and atrophy can be ruled out by appropriate protocols and clinical investigation, we can concentrate onto those subjects which may indeed present a general borderline amyloid load. Because future therapies are likely aimed at counteracting the accumulation process, these cases are also most interesting from a clinical point of view.

9.2 Integrating estimators

Overcoming the distinction into two categories can be achieved by finding criteria for scans subdivision into several classes, or by a method for ordering them. The strategy is the integration of all semi-quantitative values.

The idea here is that each method (SUVr, ELBA and TDr) is a proxy measure of the "true" amyloid load. As each method is based on a different approach – reference-tissue intensity for SUVr, geometric properties for ELBA and delayed-ROI intensity ratio for

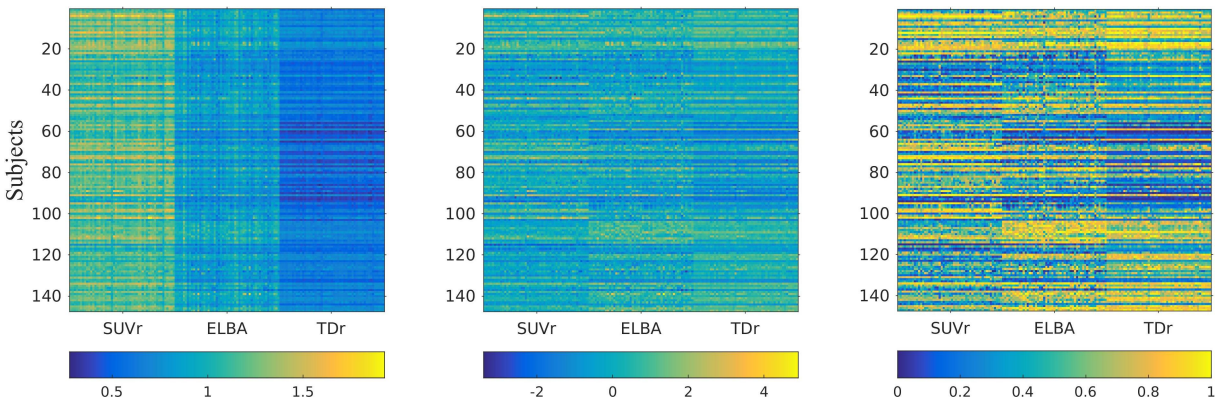

Figure 9.2: Data matrices: raw (x_{ij}, left), z-score normalized (z_{ij}, center) and quantiled-normalized (q_{ij}, right).

TDr – we can assume that they behave coherently with the amyloid load (see the general correlation results in section 7.2) but they are uncorrelated with respect to the intrinsic / data treatment errors.

As a rough proof of this, let us briefly analyze the pattern of the residuals from section 7.2, figure 7.9. We tested whether residuals correlate with any of the quantifiers and whether residuals are normally distributed and we found no significant correlation and no significant deviation from normality on all residuals and on all methods comparison. Thus is it reasonable to use all three semi-quantification methods to provide a single and more robust estimate.

According to the sample parcellation introduced in paragraph 8.1, each of the three methods now delivers 51 variables (from the 25 contralateral regions plus the whole brain). Obviously, each 51×3 variables could be used as a ranking value but we are now looking to obtain a more robust result.

Out of the infinite ways to combine several observations, we chose the linear combination (i.e.weighted average). The general functional form is therefore

$$\hat{X}_i^{(k)} = \sum_{j=1}^{N} x_{ij}\, w_j^{(k)} \tag{9.1}$$

Where N is the number of variables (153 in our case) and $i = 1..S$ the number of subjects. The subscript (k) indicates possible ways of choosing the weights. If we order $\hat{X}^{(k)}$ from the minimum to maximum value, we define the rank $R_i^{(k)}$ as the ordered position of subject i in $\hat{X}^{(k)}$ (where the weights are chosen with criteria (k)). Therefore while $\hat{X}_i^{(k)}$ assume real values, $R_i^{(k)}$ assume integer values (figure 9.1).

Some typical choices for the weights can range from a constant value (i.e. the simple mean) to the maximum variance methods (i.e. PCA).

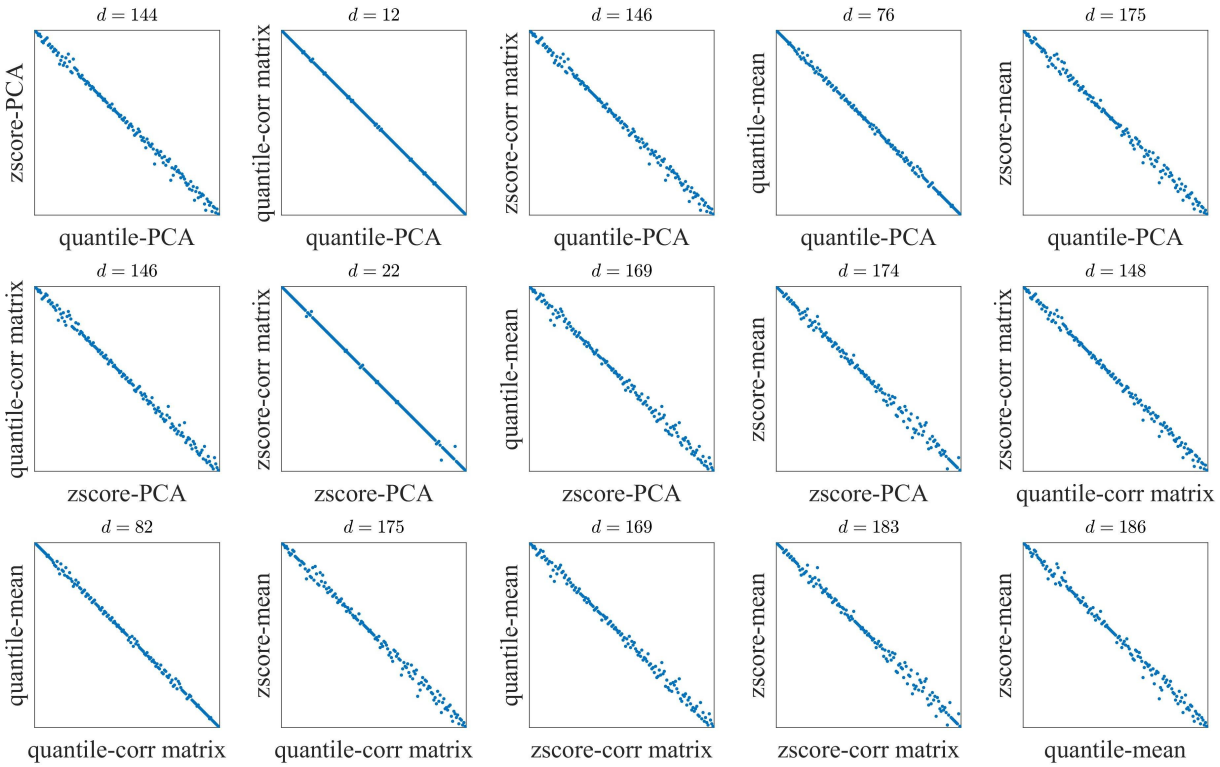

Figure 9.3: Comparison among ranks versus weight choice method (constant, correlation, PCA) and normalization (z-score, quantile). d is the rank distance (equation 9.4) with respect to the "quantile-PCA" rank.

9.2.1 Weights determination

The first and easiest choice for $w_j^{(k)}$ is the constant, that is $w_j^{(m)} = 1/N$, $j = 1, \ldots, N$, hence $\hat{X}^{(m)}$ is the aritmetic mean.

Other possible choices can take into consideration the AUC on the normative dataset, or, if we wish to be independent from it, we could exploit the correlation among quantifiers (columns of the matrix x_{ij}). In facts, it is reasonable to assume that high correlation among quantifiers is a good indication of their ability to estimate the amyloid load.

So the second tentative weight selection is proportional to the correlation among quantifiers. The intent is that correlated quantifiers are assigned a higher weight than those poorly correlated. This is achieved by averaging the correlation matrix ρ_{ij}

$$\rho_{jk} = \frac{\sum_{i=1}^{S} \left(x_{ij} - \frac{1}{S} \sum_{i'=1}^{S} x_{i'j} \right) \left(x_{ik} - \frac{1}{S} \sum_{i'=1}^{S} x_{i'k} \right)}{\sqrt{\sum_{i=1}^{S} \left(x_{ij} - \frac{1}{S} \sum_{i'=1}^{S} x_{i'j} \right)^2} \sqrt{\sum_{i=1}^{S} \left(x_{ik} - \frac{1}{S} \sum_{i'=1}^{S} x_{i'k} \right)^2}} \tag{9.2}$$

over its columns, so that

$$w_j^{(c)} = \bar{\rho}_j = \frac{1}{N} \sum_{k=1}^{N} \rho_{jk} \tag{9.3}$$

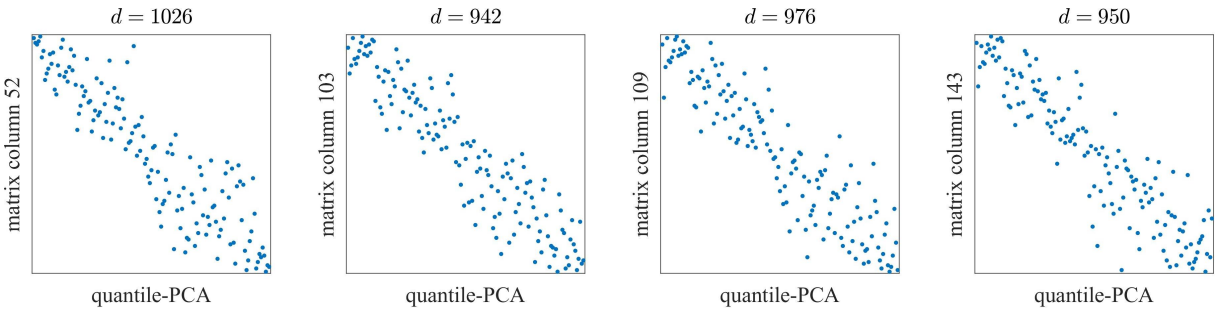

Figure 9.4: Comparison between the ranks due to four quantifiers (columns of the measure matrix x_{ij}) versus the reference rank (PCA over the quantile-normalized matrix). d is the rank distance (equation 9.4) with respect to the reference rank.

(the subscript (c) here stands for *correlation*, as (m) stands for *mean*).

Finally, we can use the principle components analysis. In this case we only consider the first component (largest variance explained and projection onto the first autovector α_{1j}) and $w_j^{(p)} = \alpha_{1j}$.

As quantifiers come from three different methods, it can be necessary to normalize them before summation. Range normalization can be achieved in several ways. We studies the effect of a linear (z-score, z_{ij})

$$z_{ij} = \frac{x_{ij} - X_j}{\sigma_j}$$

and a non-linear normalization (quantiles, q_{ij}), implicitly defined as.

$$x_{ij} = \int_{-\infty}^{q_{ij}} pdf_j(n)dn$$

where $pdf(n)$ is the probability density function of the entire normative dataset[2]. The latter transformation assumes that the normative subjects be the same for all methods. See figure 9.2 for a graphical representation of the measure matrix x_{ij} and its normalizations z_{ij} and q_{ij}.

The three weights choices are applied on the normalized quantifiers (z_{ij} or q_{ij}), giving out 6 possible outcomes.

Now the question rises on how to tell which weight choice and quantifier normalization is preferrable. An obvious solution would be to use $\hat{X}$ as a classifier and look for a rank that maximizes a metric over the ground truth of confirmed negative/positive labels (i.e. AUC, accuracy, etc). This approach though is not very informative as any random permutation of subjects **within** the same class does not alter the classification metric.

To find the "best" rank therefore we need to resort to another kind of metric. To help us we define a distance $d(R^{(1)}, R^{(2)})$ between two ranks $R^{(1)}$ and $R^{(2)}$ as the number of permutations that we need to make $R^{(1)} = R^{(2)}$, multiplied by the distance between each

[2]Alternatively we can write

$$x_{ij} = cdf_j(q_{ij}) \rightarrow q_{ij} = icdf_j(x_{ij})$$

that is with the cumulative distribution function (cdf) and its inverse ($icdf$)

two elements that have been permuted (we remind that $R^{(k)}$ assume integer values):

$$d(R^{(1)}, R^{(2)}) = \frac{1}{2}\sum_{1}^{S} |R^{(1)} - R^{(2)}| \tag{9.4}$$

where S is the number of subjects in our normative dataset. For instance, if $R^{(1)} = 1..n..n+l..S$ and $R^{(2)} = 1..n+l..n..S$, then $d(R^{(1)}, R^{(2)}) = 1/2(|-l| + |l|) = l$. We can visually compare ranks as in figure 9.5, where we plot $R^{(1)} = 1..3..5..10$ versus $R^{(2)} = 1..5..3..10$ with dots. We can clearly see the exchange in the position of elements #3 and #5, as well as the rank distance $d = 2$.

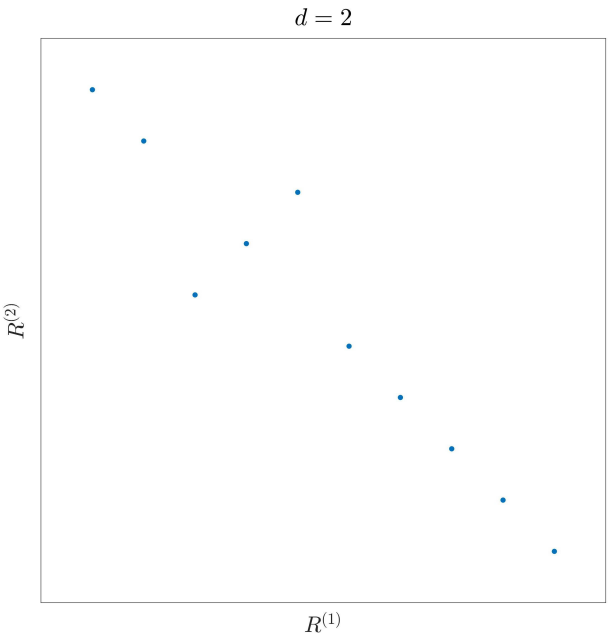

Figure 9.5: Sample visualization comparing two ranks.

We now examine the six ranks (three weight methods by two normalizations) in figure 9.3, where the PCA/quantile-based rank was used as reference for the plot. We note that the choice of the weights and the quantifier normalization have a rather small influence on the rank, influence which is mostly limited to elements permutation in the extreme positions (d is rather small).

We remark that rankings do not use any ground truth label (neg / pos), they are entirely determined by the raw quantifier values (save for an intensity normalization based on the overall normative dataset).

We now need to determine whether there is one (or more) rank that is preferrable to others, given the fact that all ranking (and even some single quantifiers) score an AUC ≈ 1 (and an accuracy $= 0.97$) on the binary reading. In absence of a finer visual grading, there is virtually no way to search for and validate a rank so that it represents a scale from the most negative to the most positive subject.

For instance, let us visually compare a rank of choice and those coming from some of the single quantifiers which score the same accuracy over the visual assessment. The rank of choice is $R^{(p)}$, which comes from the ordered indexes of $\hat{X}^{(p)}$ (first component of the PCA with quantile-based normalization). The other ranks come from the ordered indexes of some quantifiers such as $x_{i\,52}$ and $x_{i\,109}$, figure 9.4.

We can see that the distance among the single quantifiers and the reference rank is larger than that between all the other rankings in figure 9.3, and yet all these score AUC ≈ 1. This is due to the fact that all these selections do not mix scans between classes (neg / pos) so that there is no way to tell which one is "better" than the other from a simple dichotomic assessment.

While it is reasonable to assume that single quantifiers on small ROIs are more prone to errors than a value built on all available information, we need a more robust criteria to decide it.

9.3 One-dimensional transition model

To help identifying a possible alternative criteria for ranks, we built a simple one-dimensional, 2-classes model (low / high) having S values taken from a sigmoid function plus some added Gaussian noise. A pictorial representation is in figure 9.6. In this simple construct, noiseless elements below 0.5 are "negative" and above 0.5 are "positive" (the Gaussian noise then mixes them up a little). We suppose to know the "true" ordering $R^{(1)}$ of these S elements, which is the one represented in the top left plot of figure 9.6. Following the same reasoning of paragraph 9.2.1 we can: (a) randomly permute the elements in each class (without mixing between classes, ordering $R^{(2)}$) and (b) randomly permute all elements, with class mixing (ordering $R^{(3)}$). The corresponding permutation plots are shown in figure 9.6, rightmost plots.

Now we compute the moving standard deviation (msd), that is the standard deviation of a moving subset of s contiguous elements, having chosen s to be comparable to the transition width of the sigmoid.

In the ordered system ($R^{(1)}$), the standard deviation for each moving subset s is dominated by the Gaussian noise. As soon as the moving subset approaches the transition, the "true" ordering guarantees the smoothest transition (low values of the msd).

With the $R^{(2)}$ rank instead, the msd assumes approximately the same values as before when s is at the extreme of the rank. When approaching the transition though (now sharper) it has much higher values, resulting in an overall distribution with a higher mean (figure 9.6, center). Finally, With the $R^{(3)}$ rank, the msd is at the highest, as each moving subset s contain both low and high values.

The distribution of the msd values, computed over moving subsets of the whole population, reflects the permutation of the rank and, in this simple 1-dimensional model, its average is minimum for the "true" rank.

9.4 Degree of order

We can apply the same concept to our data, where an approximate equivalent of the standard deviation of the one-dimensional model is the *degree of order* (DO).

The DO is a value that is used to characterize a system dynamic and it is usually defined as a measure of the internal correlation of a system (Chandler, 1987). In general, when a system undergoes a change either due to internal reordering or due to external stresses, this often reflects into the DO. During a transition, for instance, the number of significant components is reduced (Gorban et al., 2010).

A synthetic measure can be built using correlation among the variables that describe the system. A typical implementation uses principle component analysis (Gorban et al., 2010) and its derivative. For instance, some authors use the variance explained by the first n autovectors of the PCA (Gorban et al., 2010; Jolliffe, 1986).

An exhaustive discussion of the DO and its implementation is out of the scope of this work. Here we shall use an operative definition of the DO (denoted by the symbol Ω) as

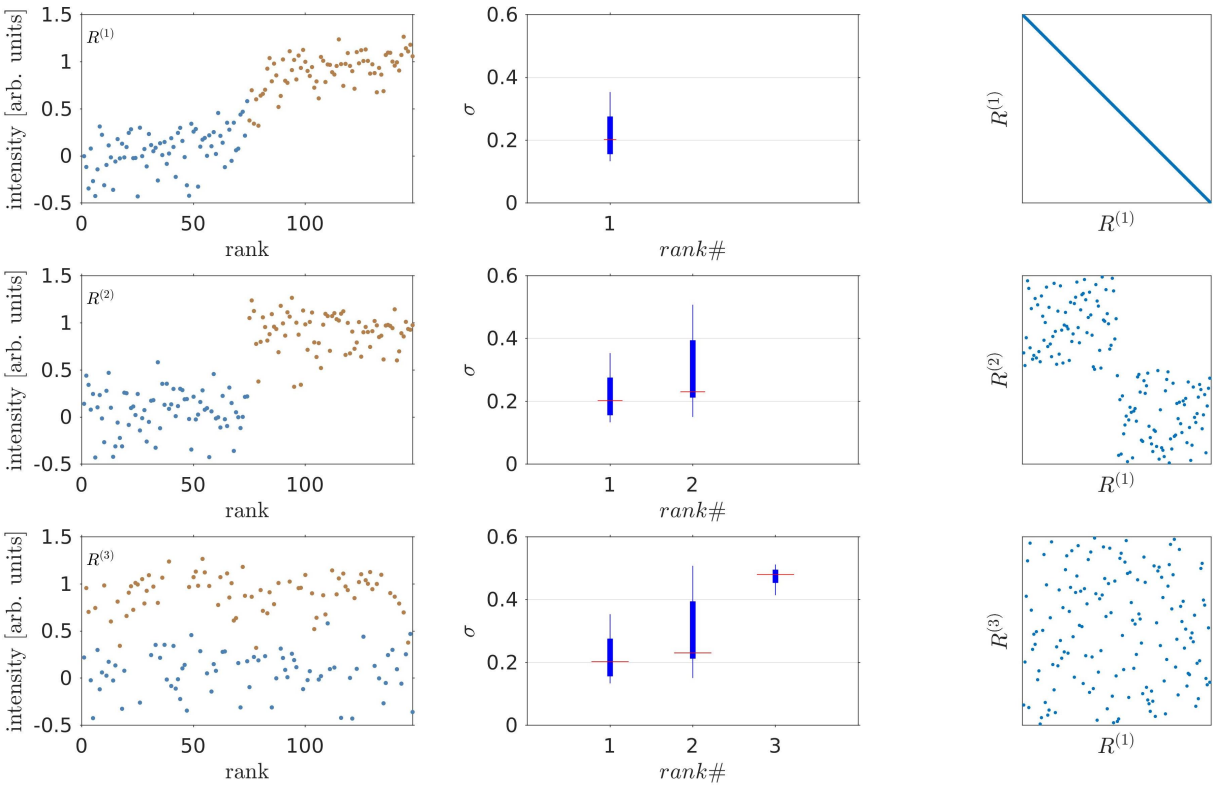

Figure 9.6: Simple 1-dimensional transition model. Colors show the two classes. Top plots: data in the "true" order $R^{(1)}$ (left), distribution of the moving standard deviation (center), rank comparison with respect to $R^{(1)}$ (right, in this case it is trivial). Center plots: data mixed within classes only (order $R^{(2)}$) (left), distribution of the moving standard deviation due to $R^{(2)}$ now compared with that of $R^{(1)}$ (center), rank comparison with respect to $R^{(1)}$ (right, no between-classes mix). Bottom plots: data mixed with a random order ($R^{(3)}$, left), comparison of the three distributions of the moving standard deviation (center), rank comparison with respect to $R^{(1)}$ (right).

follows:

$$\Omega = \frac{\tilde{\alpha_1}}{\sum_{j=2}^{S} \tilde{\alpha_j}} \tag{9.5}$$

where $\tilde{\alpha_j}$ is the variance explained by the j-th PCA component ($\sum_{j=1}^{S} \tilde{\alpha_j} = 1$).

In a completely random system with N degrees of freedom, the variance explained is approximately shared by all components so that $\Omega \approx {}^1/_N$. Viceversa, a system where one or more degrees of freedom exhibit large variances, the first component will tend to capture this range and $\Omega \gg 1$. (see figure 9.7 for a simple example).

In a recent work by Pagani et al. (2016), the DO was used to quantify the loss of coherence within clinical groups measured by FDG-PET. Inspired by this work, we apply the DO to the measure matrix x_{ij} and we study the ordering effect of the various ranks.

Similarly to the simple approach in paragraph 9.3, we compare different ranks: one completely random, some based on a single quantifier (see figure 9.4) and the $R^{(p)}$. As we do not know the real transition width, the moving average is carried out over several sets ranging from 8 subjects to 20 subjects.

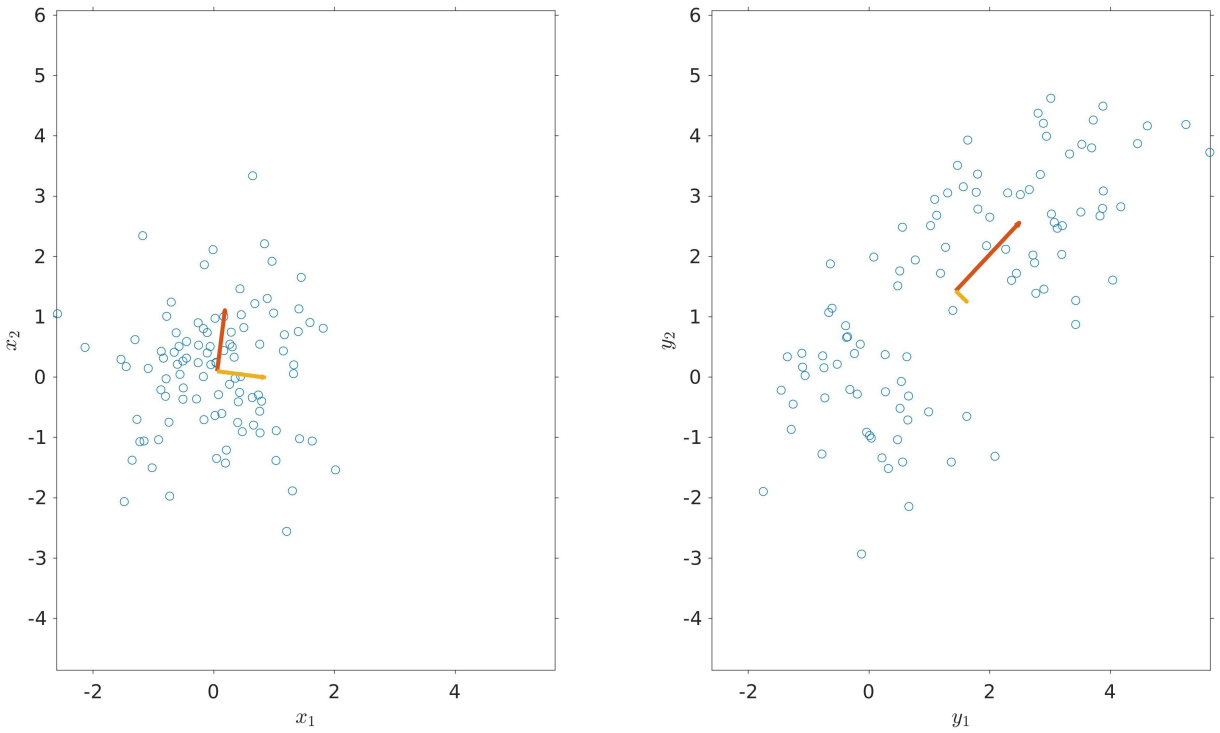

Figure 9.7: Simple comparison of two systems with two degrees of freedoms. System x (left) is a random system, system y (right) has some correlation. The two PCA vectors length (colored arrows) are proportional to the variance explained by the respective components. The degrees of order are $\Omega = 1.3$ and $\Omega = 6.1$ respectively.

Results are shown in figure 9.8, where the dots are the Ω values for all moving sampling (that is a sliding subset of s subjects that covers the entire range, and $s = 8..20$). Here too, we find that the random rank has the maximum variance because in each subset s we can find subjects from different classes. More interestingly though, the ranking due to a single matrix column behaves similarly to the intra-class permutation rank in the simple 1-dimensional model. We conclude therefore that the PCA-quantile-based rank $R^{(p)}$ is the **relative** best choice. At this point we cannot conclude that $R^{(p)}$ is the **absolute** best rank, but only that it performs best (in terms of Ω distribution) with respect to all other tested choices.

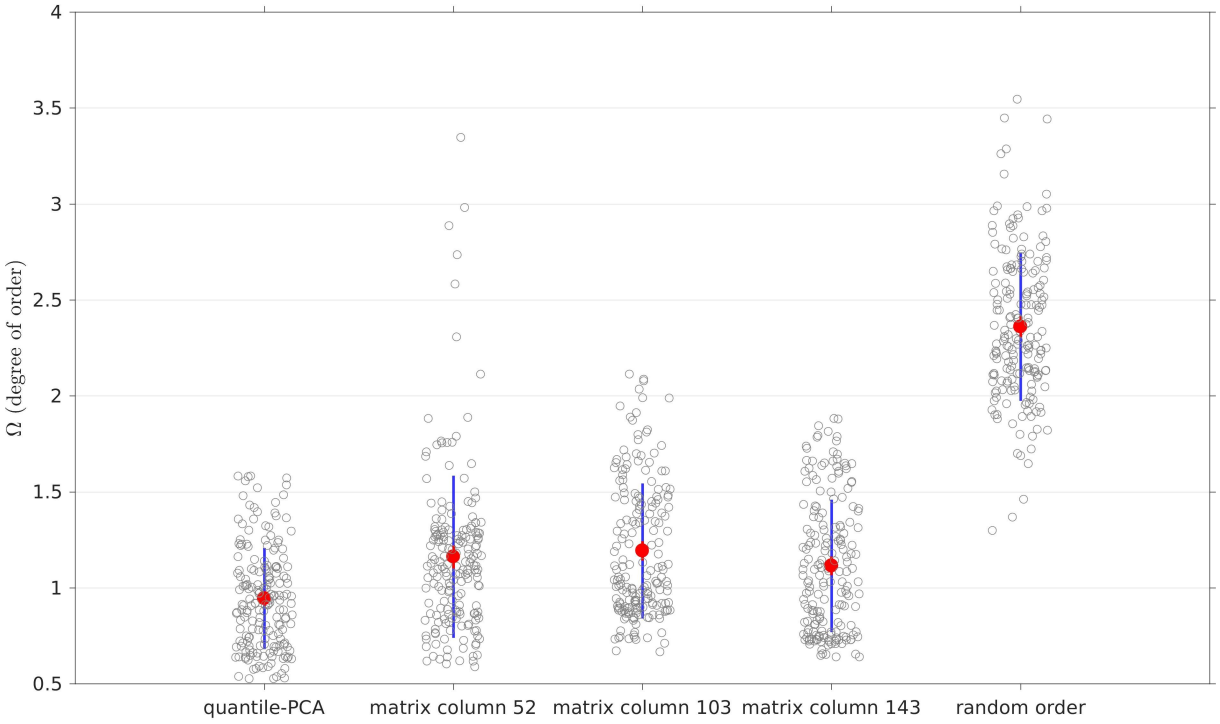

Figure 9.8: Degree of Order distribution for different rankings of the measure matrix.

Chapter 10

Clinical impact

The analyses and methods presented in this work were developed to fulfill the lack of reliable tools for quantitative analyses in the everyday clinical evaluations of amyloid PET. We believe that a more sophisticated approach to amyloid PET evaluation can benefits clinicians in many ways, both in standard practice and in research.

The application shown here was prepared on the ^{18}F-florbetapir tracer (Amivid), although a completely analogous tool can be given on all tracers, provided we get a proper normative dataset.

10.1 Summary plots

We have now the opportunity to summarize all the analyses into a single report that can be used in clinics.

The idea is to present each semi-quantification method and their combination into a single rank within a unified report that can be easily accessed by clinicians.

We designed a simple report that is delivered for each patient and includes the amyloid rank plot and all the three semi-quantification assessment with both regional and whole brain values. Numbers and the respective cut-offs are provided both raw and within the normative dataset cumulative probabilities.

Rank is computed with the PCA and quantile normalization. First we introduce the rank plot: this plot simply shows the normalized rank of a new subject onto the normative $\hat{R} - \hat{X}$ curve (figure 10.1). The analysis procedure first computes the 51 quantifiers for all the three methods (SUVr, ELBA and TDr) and combine them into a single quantifier $\hat{X}$, which is then mapped into the rank $\hat{R}$ by means of the normative data curve. This plot then is the synthesis of all previous analyses and delivers the most robust quantifier.

Then we have the single methods' local probability plot, which needs a bit more explanation. The main feature of this plot is the cumulative probability curve: this curve consists of the sum of two cumulative probability functions, one for the negative and one for the positive normative data. The curve formula is:

$$y = \int_{-\infty}^{x} pdf^{(-)} + \int_{-\infty}^{x} pdf^{(+)} - 1 = \int_{-\infty}^{x} pdf^{(+)} - \int_{x}^{\infty} pdf^{(-)} \tag{10.1}$$

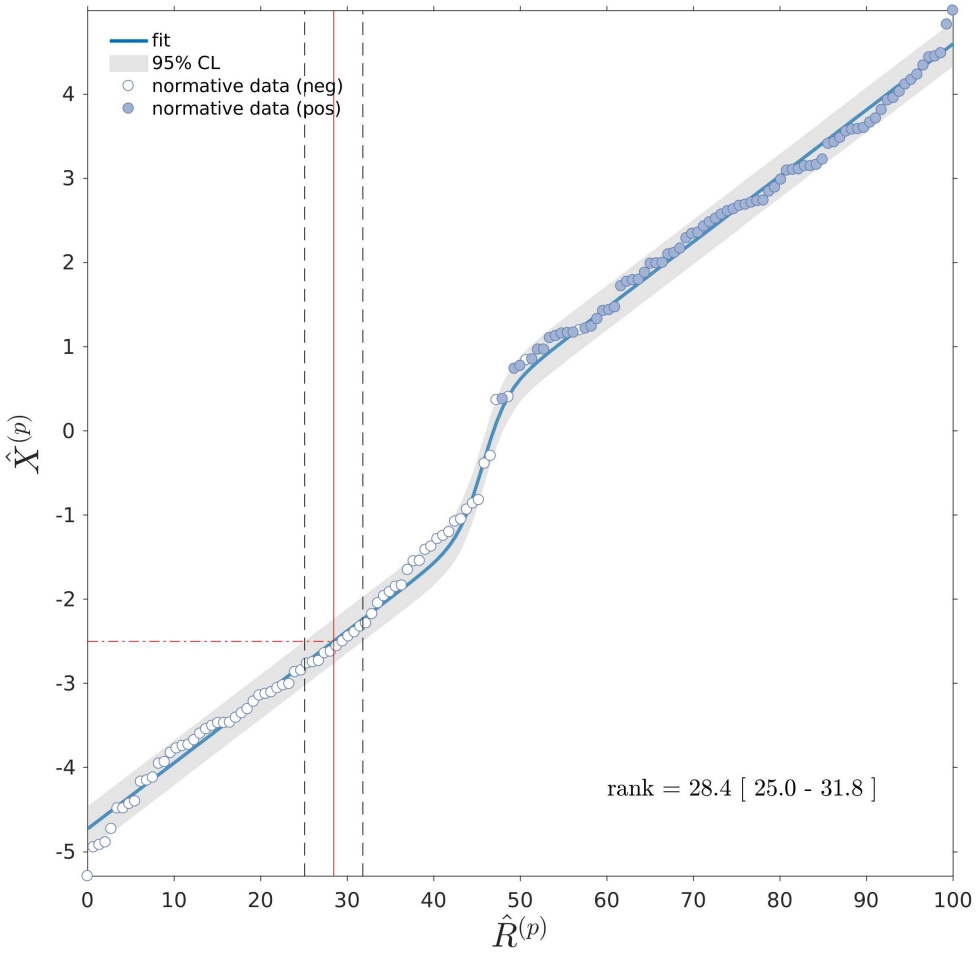

Figure 10.1: Rank plot linking the quantifier weighted sum $\hat{X}^{(p)}$ with the normalized rank $\hat{R}^{(p)}$. White dots are the position of the normative data, blue thick line is the non-linear fit, shadowed area is the 95% CL on the fit. The new data at $\hat{X}^{(p)} = -2.5$ is mapped into a normalized rank $\hat{R}^{(p)} = 28.4$. Rank CL (within square brackets) are estimated by intersecting the new data with the confidence limits from the normative data fit.

where $pdf^{(-)}$ and $pdf^{(+)}$ are the probability distribution functions of the negative and positive subjects in the normative dataset. The pdfs are separately computed for each quantifier of the measure matrix x_{ij}.

Figure 10.3 graphically shows a sample cumulative probability curve. This visualization allows to immediately grasp the probability of belonging to the positive and negative group. The curve crosses zero at a value $\hat{x}$ when the probability of finding a negative normative subjects with quantifier value $x_n > \hat{x}$ ($\int_{\hat{x}}^{\infty} pdf^{(-)}$) is the same as finding a positive normative subject with quantifier $x_p < \hat{x}$ ($\int_{-\infty}^{\hat{x}} pdf^{(+)}$). In other words, the negative portion of the curve indicates quantifier values for which the probability of finding a negative subject is higher than that of finding a positive subject.

Each quantifier in the measure matrix x_{ij} is used to draw a cumulative probability curve and this is the basis for the probability plot, which shows the detailed analysis for each method and each parcellation P_k (figure 10.2).

This plot also shows the position of a subject on the normative probability curves. The subject quantifiers are places on each curve, which have been shifted so that the all their

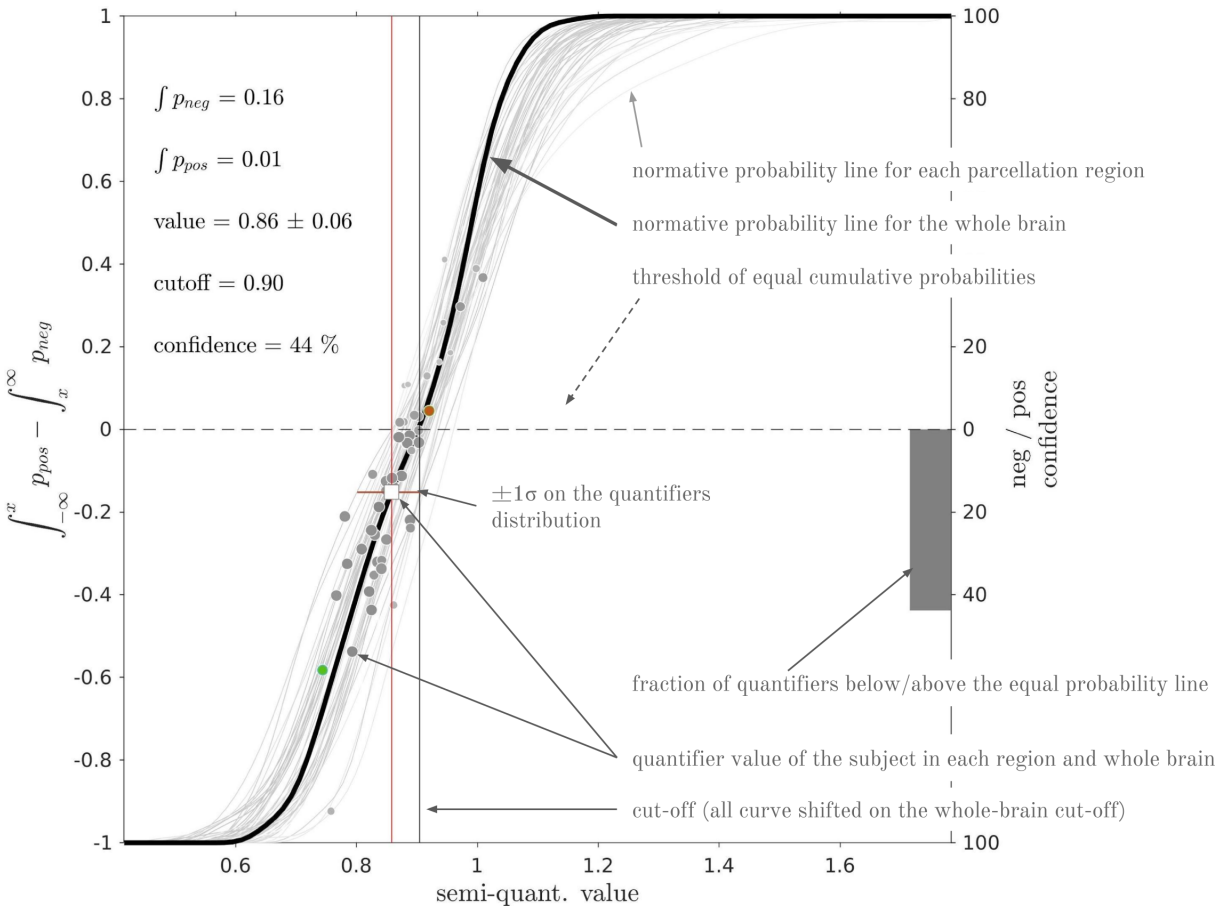

Figure 10.2: Placement of a subject's quantifiers (single method) on the normative cumulative probability curves.

cut-offs are at the same value than the whole-brain one. On the y-left axix we have the difference between the cdf of the negative and the positive scans in the normative dataset (equation 10.1).

The subject's quantifiers are plotted as gray dots on the curves, according to each own's value. The dot size is proportional to the AUC of the parcellation region P_k (see paragraph 8.1 and figures 8.2, 8.3 and 8.5) so that parcellation with less significance are visually unobtrusive. The regional markers are also complemented with a green and a red dot, showing the most negative and positive regions among those which perform with AUC> 0.95 on the normative dataset.

The whole brain quantifier is drawn on a thicker probability line with a white square marker and it is complemented by a red vertical and horizontal lines. The vertical line visually enhance the whole-brain quantifier value on the x axis, while the horizontal line shows ±1 standard deviation computed on all P_k quantifiers with respect to the whole brain.

On the y-right axis we plot a bar representing the fraction of regional quantifiers which lye below / above the equal cumulative probability line (depending on the position of the whole brain marker). It is labeled "neg / pos confidence" because it measures the consistency within regional quantifiers.

10.2 The amyloid-PET quantification report

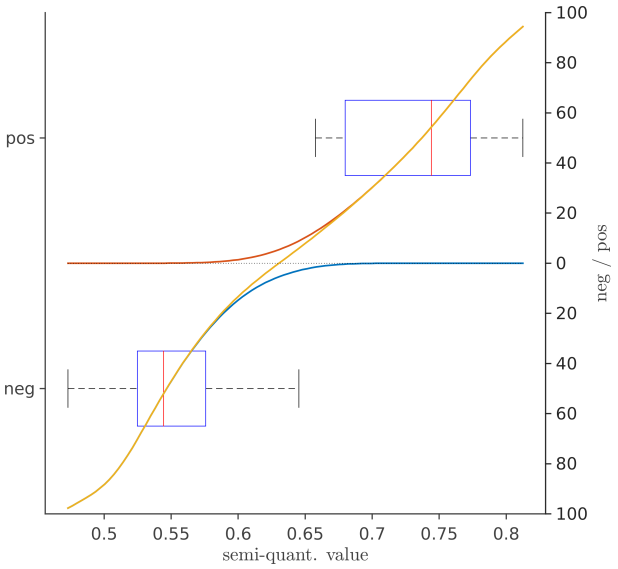

Figure 10.3: Cumulative probability curve example. Boxplots shows the negative and positive normative data distribution over a given quantifier. Cumulative probability function of the negative and positive distributions (blue and red curve, the blue curve is shifted by -1). The yellow curve is described by equation 10.1

A typical report is shown in figure 10.4. It consists of several parts:

- patient name and age, taken from the DICOM files;
- the rank plot (figure 10.1), showing the patient rank with respect to the normative dataset;
- the regional semi-quantification versus probability curve for each method (figure 10.2)
- the most negative and most positive regions among those with high performance (AUC on the normative dataset > 0.95)

To clarify the reading, each curves plot also shows two probabilities: the probability $\int_{\tilde{x}}^{\infty} pdf^{(-)}$ of finding a negative normative subject with whole-brain score greater then the subject score $\tilde{x}$; the probability $\int_{-\infty}^{\tilde{x}} pdf^{(+)}$ of finding a positive normative subject with whole-brain score lower then the subject score $\tilde{x}$. We can also read the subject's whole brain score with the stardard deviation computed on the regional scores, the cut-off and the numerical value of the confidence (gray bar on the right y axis in figure 10.2).

Next to each curve probability plot we also show the method's take on the most negative and most positive region (green and red region), which correspond to the green and red dot in the probability plot. Sagittal, coronal and axial sections are visualized on the subject's scan in the MNI space, and centered on the relevant parcellation centroid.

The report is complemented with a disclaimer, a brief legend and a summary table with the normative dataset demographics.

We have made the analysis completely automatic. Currently, we provide a web site (`https://nextmr01.ge.infn.it/myPET/`) where selected and registered users (typically nuclear medicine physicians) can access the analysis and run it remotely. This was necessary because the average computer which is usually present in hospitals cannot run the whole analysis in a reasonable time. The analysis pipeline is therefore hosted on a dedicated computing infrastructure at the National Institute for Nuclear Physics (INFN), which provides computing power adequate for the task.

The typical processing time is ≈ 20 minutes. At the end of the process all relevant and confidential data are deleted from the system, while the registered clinician receives the amyloid report vie email.

10.3 Clinical research and borderline cases

Outside the general clinical goal to tell negative and positive scans, we enter the clinical research framework where we are interested in less obvious evaluations. Borderline scans can occur for several reasons: "technical" reasons include patients with marked cerebral atrophy, scan of not optimal technical quality (movement, low counting statistics, noisy images or too smoothed), reader experience (inter-observer agreement), modality of analysis (visual, semiquantitative ...).

Still, there can be "true" borderline cases, that is patients with borderline deposits of amyloid plaques. There are relatively few works in literature on these cases although we know that borderline deposits of amyloid plaques have indeed been observed (Sojkova et al., 2011).

A search was carried out on PubMed to look for explanatory works on amyloid borderline cases, using a variable combination of keywords: "limitation Amyloid PET"; "Borderline SUVR amyloid imaging"; "Unclear cases / uncertain results amyloid PET"; "Unclear visual assessment amyloid PET"; "Difficult amyloid scan"; "Uncertain florbetapir scan"; "Prognostic value amyloid PET / imaging"; "Stratification of risk amyloid PET Very few relevant articles were found, except perhape for for Mormino et al. (2012), although his work places the accent on the presence of dubious scans, without providing hypotheses or solutions on what could be their meaning or definitions.

These borderline subjects are likely at the beginning of the amyloid accumulation process, and will probably continue the neurodegenerative path; they are therefore the ideal target for future therapies. The dichotomical assessment is inadequate in these cases and it is not trivial to find operative definitions of borderline cases (Baker et al., 2017; Mormino et al., 2012; Fleisher et al., 2011). In addition, one cannot rely on simple numerical estimates either, as SUVr cut-offs on the same tracer vary among studies on a considerable range (from 1,06 to 1,80).

Our ranking approach was then used for a **very exploratory** study on a small cohort of mildly negative and borderline patients, where the uptake quantification (and hence the inclusion criteria) was not defines by SUVr values, but by a range in the normalized rank.

We selected outpatients from a naturalistic population who were examined at the S. Martino Hospital in Genoa, Italy (HSM, the same center which also contributed to the Multicentri pilot Dataset). All patients had an amyloid PET scan following the diagnostic pathway outlined in paragraph 3.2.

Out of these patients, we selected 12 who were found to be intermediate, that is within a normalized rank $\hat{R} \in [30 - 70]$. We than selected a set of clinical, neuropsychological and neuroimaging medical report on MRI, FDG-PET and amyloid-PET. The idea is to try profile these patients, looking for possible clues to their etiology and – perhaps – their prognosis.

The high heterogeneity in clinical data collection and assessment did not allow a thorough dataset collection and we had to select and sift through several possible choices, in order to keep the dataset as complete as possible. Out of all clinical examinations we selected a subset that was common to all patients and we operated a categorization of all clinical variable in order to enhance internal normalization and to provide a solid base for the rank analysis. A sample of the clinical data gathered is in table 10.1. This is only a small

subset of the data that were gathered; for brevity, not all clinical information are shown.

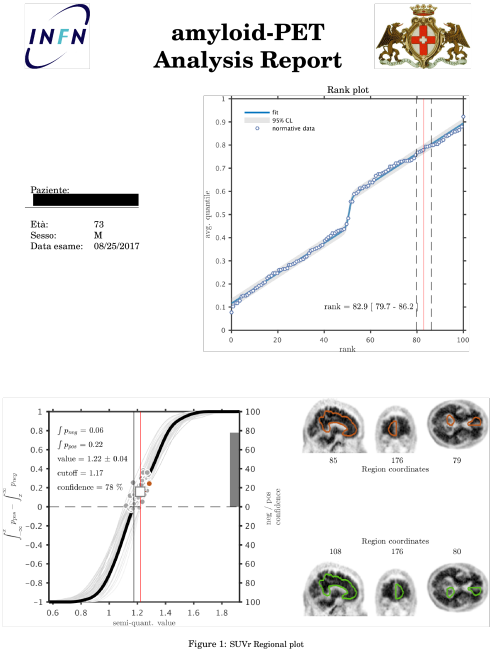

Figure 1: SUVr Regional plot

(a)

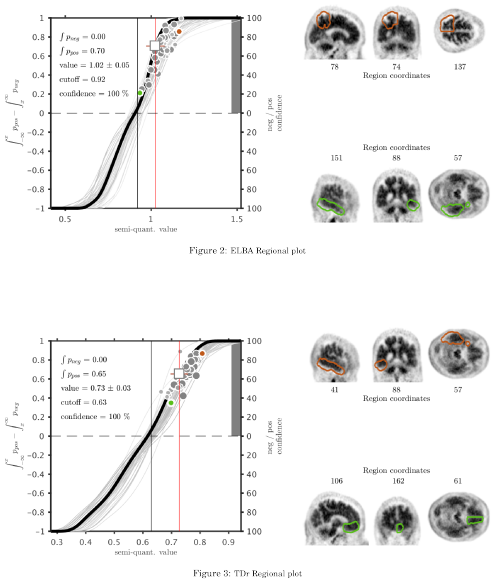

Figure 2: ELBA Regional plot

Figure 3: TDr Regional plot

(b)

Notes

Methods and cut-off values were obtained on a normative dataset consisting of ^{18}F-florbetapir scans. Analysis of scans acquired with a different tracer may yield inaccurate and unreliable results.

Legend

- Nomenclature

SUVr	Standardized Uptake Value ratio
ELBA	Evaluation of Brain Amyloidosis [1,2]
TDr	Time-Delayed ratio [2]

- Rank plot

x-axis	Rank (linearized mean quantile)
y-axis	Weighted mean of all quantile-transformed scores (both whole brain and regional)
dots	Normative data
blue line	Model fit
gray-ish area	95% conf. interval of the model on the normative data
red line	Subject rank
dotted lines	95% conf. interval on the rank
text	Rank [95% conf. interval]

- Scores plot (SUVr, ELBA, TDr), regional probability plot (left)

x-axis	Score value
y-axis	Difference between cumulative probabilities of neg. and pos. distributions on normative data
dots	Regional (ROI) scores. Dot size is proportional to AUC between neg./pos. distributions
square marker	Whole-brain score
green dot	Most negative ROI (selected ROI with an AUC on norm. data > 0.9)
red dot	Most positive ROI (selected ROI with an AUC on norm. data > 0.9)
thick black line	Cum. prob. difference on whole-brain neg./pos. distributions. Given the scores s_{neg} and s_{pos} of the negative and positive normative subjects, for each score s^* on the x-axis it weights the probabilities difference $p(s_{pos} < s^*) - p(s_{neg} > s^*)$
thin gray lines	Cum. prob. difference on regional neg./pos. distributions
vert. red line	Subject score (whole-brain)
horz. short red line	$\pm 1\,\sigma$, std. estimated from ROI scores values
vert. black line	Score cut-off (maximum accuracy) evaluated for the whole brain
horz. black dotted line	Null difference between cumulative probability distribution $p(s_{pos} < s^*) = p(s_{neg} > s^*)$
vert. gray band to the right	Percent of ROI scores below/above the null line
$\int p_{neg}$	$p(s_{neg} > s^*)$ probability to find *negative* normative subjects *above* the score s^* (red line)
$\int p_{pos}$	$p(s_{pos} < s^*)$ probability to find *positive* normative subjects *below* the score s^* (red line)
value	Subject score $\pm 1\,\sigma$ (standard deviation also shown as short red horizontal line crossing the square marker)
cutoff	Score cutoff value based on maximum accuracy of whole-brain normative data
confidence	Percent of ROI scores below/above the null line (also shown as vert. gray band to the right)

- Scores plot (SUVr, ELBA, TDr), significant ROIs (right)

upper plot	Sagittal, coronal and axial cross-sectional plot of the MNI-oriented subject scan, together with the most significantly positive ROI (red line and red dot in the scores probability plot)
lower plot	Sagittal, coronal and axial cross-sectional plot of the MNI-oriented subject scan, together with the most significantly negative ROI (green line and green dot in the scores probability plot)
regional coordinates	Slice coordinates of significant ROIs are given in the MNI-oriented space

(c)

Normative dataset:

Provenance	N subj.	M/F	Age [95% CI]	neg/pos
FPA-BS[1]	55	24/31	71 [69 72]	29/26
HUG-GE[2]	44	20/24	72 [69 74]	27/17
FSM-PV[3]	19	9/10	76 [73 79]	7/12

[1]Department of Nuclear Medicine, Fondazione Poliambulanza, Brescia Italy
[2]University Hospitals and University of Geneva, Geneva, Switzerland
[3]Nuclear Medicine Unit, IRCCS Fondazione Salvatore Maugeri, Pavia, Italy

Radioligand used on all normative subjects: ^{18}F-florbetapir.
Cut-off value for each method was computed on the normative dataset by accuracy maximization method.
Positive/negative evaluation was provided by visual assessment of two independent readers.

References:

[1] *SUVr-independent evaluation of brain amyloidosis*, Journal of Alzheimer's Disease, Vol. 54-4 2016, doi:10.3233/JAD-160232
[2] *Approaches to semi-quantification: beyond SUVr in amyloid imaging*, European Conference on Clinical Neuroimaging, Roma, March 14-15, 2016

DISCLAIMER

This software is not intended for diagnosis or for clinical applications. Measurements and statistics are provided for research purposes only. The application of this software outside this scope is a responsibility of the end-user.

Questo software ha uno scopo puramente esplorativo e quindi non diagnostico. La responsabilità del suo uso è totalmente dell'utilizzatore.

(d)

Figure 10.4: The amyloid-PET report. (a) patient data, rank plot, SUVr probability curve and its most significant regions; (b) ELBA and TDr probability curves and most significant regions; (c) legend; (d) disclaimer, references and normative dataset demographics.

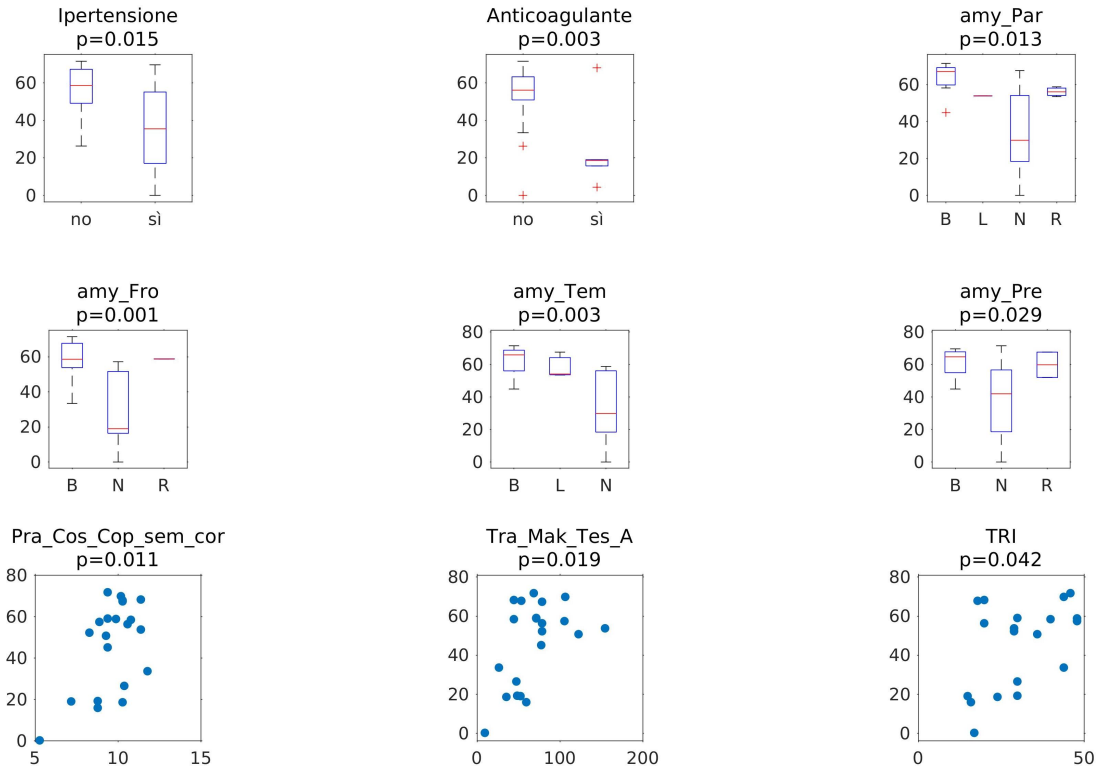

Figure 10.5: Preliminary results on a small subset of borderline patient. Medical report on amyloid PET included regional (Parietal, Frontal Temporal and Precuneus) with the qualifiers B,L,R,N (bilateral,left,right,normal); Neuropsychological assessment included Trail making test (A and B), and Constructive Praxia – simple copy with semantic correction.

Table 10.1: Sample demographics from borderline cases. B=bilateral; R=right; L=left; N=negative; aMCI=amnestic mild cognotive impairment (AD=Alzheimer's Disease); SNAP=Suspected Non Alzheimer's Pathology; APPsem=Semantic primary progressive aphasia.

SubjectID	Sex	Age	Diagnosis	Amyloid PET						Rank	Physiological Anamnesis			
				pos/neg	Parietal	Frontal	Temporal	Occipital	Precuneus	Posterior cingulate		Sleep cycle	Alveus	Diuresis	Menopause
barpier	m	66	aMCIAD	pos	B	B	B	B	B	B	58.2	Impaired	Regular	Regular	
quaaur	f	75	aMCI	pos	N	B	N	N	N	N	58.6	Regular	Regular	Regular	Regular
dabcar	f	76	aMCIAD	pos	R	R	N	R	N	N	58.8	Regular	Regular	Regular	Regular
picben	m	80	aMCIAD	pos	B	B	B	B	B	N	67.1	Regular	Regular	Regular	
fiamic	f	76	SNAP	neg	N	B	N	N	N	N	33.4	Impaired	Regular	Regular	Regular
firbar	m	81	aMCIAD	pos	B	B	B	N	B	N	68	Regular	Impaired	Regular	
marmar	f	73	AD	pos	B	B	B	B	B	B	69.6	Regular	Regular	Regular	Regular
rizter	f	76	aMCIAD	pos	B	B	B	B	N	B	67.5	Regular	Impaired	Regular	Regular
rosmat	f	77	aMCIAD	pos	N	B	L	N	R	N	67.6	Impaired	Regular	Regular	Precoce
sanrit	f	75	FTD	neg	N	N	L	N	N	N	54.1	Regular	Regular	Regular	Regular
belros	f	78	APPsem	pos	B	B	B	B	B	B	64.7	Impaired	Regular	Regular	Regular
cargia	m	84	AD	pos	R	B	L	N	N	N	53.5	Impaired	Regular	Regular	

SubjectID	Education	Familial risk	Smoke	Alcohol	Diabetis	Hypertension	Hypercholesterolemy	Concussions	Therapies							EEG
									Anticoagulant	Antiplatelet	Antidepressant	Cholinesterase inhibitor	Antiepileptic	Antidiabetic	Regular	Base rhythm [Hz]
barpier	13	no	no	no	no	yes	no	no	no	no	no	no	no	no	no	8
quaaur	13	no	no	no	no	no	yes	no	no	no	yes	no	no	no	no	11.5
dabcar	11	no	no	no	no	no	yes	no	no	no	yes	no	no	no	yes	8
picben	5	no	no	no	no	no	yes	yes	no	yes	no	yes	no	no	no	9.5
fiamic	8	no	no	no	no	no	no	no	no	no	no	yes	no	no	no	9
firbar	13	yes	no	no	no	no	yes	no	yes	no	no	yes	no	no	yes	9.5
marmar	5	no	no	no	no	yes	yes	no	no	no	yes	no	no	no	yes	9
rizter	13	no	no	no	no	no	no	no	no	no	yes	no	no	no	no	9
rosmat	5	yes	no	no	no	no	no	yes	no	no	yes	no	no	no	yes	9.5
sanrit	5	yes	no	no	no	yes	no	yes	no	yes	yes	no	no	no	no	9
belros	5	no	no	no	no	no	no	no	no	no	no	yes	no	no		
cargia	13	no	no	no	no	yes	yes	yes	no	no	yes	yes	no	no		

Table 10.2: Linear model on borderline and negative cases.

	Estimate	Standard Error	t-Statistics	p-Value
(Intercept)	-14.087	25.46	-0.55	0.587
Hypertension	-13.587	7.22	-1.88	0.077
Trail Making Test (A)	0.263	0.10	2.42	0.026
Constr. praxia (simple copy) corrected	5.024	2.55	1.96	0.065

These patients were compared to another small set of 13 decidedly negative patients (i.e. $\hat{R} \in [0-30]$) that were enrolled in the same center and under the same inclusion/exclusion conditions (that is symptomatic outpatients with uncertain diagnosis but with AD as possible outcome). Therefore these are not controls, nor normal aging subjects. Rather, they are totally similar to the borderline patients from the pre-amyloid-scan point of view, save for the fact that they tested negative to the amyloid PET (and therefore AD was excluded as diagnosis).

Once the target and reference population were defined, and the clinical data gathered, we proceeded with two simple analysis: (a) each covariate is separately checked against the rank: if the covariate is a continuous variable we used Pearson correlation, for categorical variables we used the ANOVA test; (b) we built a general linear model to predict the rank: the idea is to use a mixed regression for categorical data and continuous variables, looking for predictors with high significance.

Preliminary results are in figure 10.5. We fixed the p value for the ANOVA (categorical covariates) and correlation (continuous covariates) to $p = 0.05$. The only significant neuroimaging covariates come from the Nuclear Medicine report of the amyloid PET (an expected result). No MRI or FDG-PET related information from the clinical report had any significant correlation; this obviously do not exclude more subtle relationships (i.e. by direct analysis of the neuroimages), it only states at – at this level of details – the amyloid-PET quantification is not related to other neuroimaging techniques, (it provides independent information).

The second preliminary result is the linear model (table 10.2). Due to the dataset lack of completeness, the model statistics are the following: 21 observations, 17 error degrees of freedom, estimated Dispersion: 248, F-statistic vs. constant model: 6.72, (p-value = 0.00343). The Anticoagulant feature is not present because it correlates with the Hypertension.

These preliminary analyses indicate that further studies on borderline patients are warranted, and that we can probably profile them both with thorough neuropsychological tests and with detailed anamnesis.

Due to the very limited number of patients, the single-center provenance and the lack of a proper common protocol in the clinical data, we remark once again that the borderline patient analysis is exploratory at best, and that there is no clinical conclusion to draw from this results as yet.

Chapter 11

Discussion

This work deals with the clinical and methodological use of amyloid-PET. We have studied in particular the problem of quantification and its relationship with clinical assessment, visual evaluation and longitudinal analysis. We have also developed a simple and robust quantifier derived from all available and validated methods.

This discussion will focus primarily on the ELBA method, which was studied more thoroughly and validated on an extensive dataset. Then we shall discuss the TDr, its relationship with SUVr and ELBa and the advanced topics of regional and longitudinal analysis. Finally we discuss the ranking and its potential.

11.1 ELBA validation

The proposed analysis shows that it is feasible to construct a semi-quantification method on amyloid-PET images without relying on counts-ratio approaches. The ELBA method shows good performance versus the dichotomic visual assessment and has ranking characteristics, proven both on cohort-based and longitudinal analyses.

From a methodological point of view, both SUVr-based and ELBA approaches require image registration techniques (spatial normalization), although the lack of small cortical ROIs in ELBA renders the registration process and the template choice less demanding. ELBA tries to mimic the human visual process, in that it captures information on global contrast and intensity distribution rather than weighing intensity in predefined regions. While further tests are necessary – particularly on all other major PET tracers and with histologically validated scans – it is worth noting that this process delivers comparable (if not slightly better) information in terms of semi-quantification, classification and ranking, with respect of the widely used SUVr methods.

11.1.1 Blind and open phase performance

Expert readers were concordant and confident in reaching a diagnosis (*positive* or *negative*) on 403 scans (82.6%), without the influence of clinical information and other imaging data. This finding confirms that a trained reader can safely rely on his/her experience on over 80% of images, when evaluating by means of visual analysis.

As a second point, the five expert readers were not able to give a concordant diagnosis (i.e., amyloidosis present/absent) in 22 (4.5%) images and in 63 (12.9%) more images at least some of them declared that the initial assessment was unclear (i.e., doubtful). Interestingly, *uncertain*-tagged scans were mostly located in the negative domain whereas conflicting scans were evenly placed around the cut-off. As a consequence, looking at the fraction of uncertain and contrasting labels vs clinical cohort (blind session, table 7.2) we find that for cognitively normal subjects and EMCI it is 23% and 21% respectively, versus a 11% found in MCI and LMCI (AD has 0%).

Overall, these 85 (17.4%) scans may represent the borderline scans where a quantification approach can help in reaching a definite diagnosis. This fraction of scans is polled from 61 subjects and could derive both from healthy subjects or mainly early MCI patients in a stage when Aβ is accumulating in the brain but still in a limited amount (Villemagne et al., 2013; Jack et al., 2013), thus leading to difficult reading. In facts, they belonged to NS (22), EMCI (27), MCI (4) and LMCI (8).

When the experts were aided by the closest negative and positive images - as yielded by the ELBA output - the 22 conflicting cases and 63 uncertain were solved, which is a non trivial aid this automatic system can provide even to expert readers. We have not tested the potential aid ELBA can give to less experienced readers but it should be intuitively higher.

A substantial number (about 19%) of healthy subjects had a positive score, although the positivity fraction is mainly ascribed to the elderly ($>$ 70 years) ones, on a par with literature findings. Also, a few patients with AD had a normal score, which is in keeping with literature works (Johnson et al., 2013b) and raising the possibility of wrong clinical diagnosis (ADNI subjects do not have pathological confirmation), patients with discrepant normal amyloid-PET scan but abnormal Aβ_{1-42} levels in cerebrospinal fluid (Landau et al., 2013) or, alternatively, of patients with suspected non Alzheimer pathology (SNAP, Jack (2014)).

In the middle, MCI subjects in progressive stages of cognitive impairment were roughly halved between positive and negative scans which highlights the presence of causes of MCI (such as frontotemporal lobar degeneration or vascular cognitive impairment) other than AD, and possibly of SNAP. Ranking based on the overall amyloid burden is also apparent and it is a benefit of the method, as the average ELBA score progressively increases from healthy subjects to patients with AD.

11.1.2 Comparison with SUVr-based methods

The performance of both approaches are very similar, with a modest but statistically significant edge in favour of ELBA. The optimized SUVr cut-off (=1.23 and based on the visual assessment) is higher than the values typically considered in literature for this ligand (1.10 - 1.14, Johnson et al. (2013b); Hutton et al. (2013); Camus et al. (2012)) but it agrees with other works (e.g. Mattsson et al. (2014)), where their cut-off value (1.24) was obtained by AUC optimization.

Table 7.1b and 7.2 summarizes the binary classification (ELBA and SUVr) versus the visual assessment in both blind and open session. The comparison to cortico-cerebellar SUVr in the cases where both methods agree shows that SUVr values are not alone in providing good classification relative to visual assessment, and that and alternative and

independent approach can enrich the information obtainable from the PET scan.

Although the ensemble on which the two methods disagree is limited to a small number of subjects (quadrants for which ELBA / SUVr are pos/neg and neg/pos), the apparent trend is that subjects negative to the visual assessment are likely to be considered negative by ELBA (although not far from its cut-off) but are considered positive for SUVr. This suggests that whole-brain amyloid burden evaluation is more concordant with respect to a visual read than (small) ROI-based quantification on borderline cases. When keeping the blind visual assessment as reference (that is using only the N and P set, table 7.2) this distinction is even more pronounced: 15 out of 19 cases agree with ELBA versus 4 out of 19 for SUVr.

11.1.3 Comparison with CSF Aβ_{42}

The comparison between ELBA score and CSF Aβ_{42} assay yielded a satisfactory agreement with 90.6% concordance, in the same range (Landau et al., 2016) or even a bit higher (Landau et al., 2013; Zwan et al., 2016) than those achieved in previous works using the SUVr approach. Discrepancies (7.5%) were mostly found in patients with abnormal Aβ_{42} levels and a normal ELBA score. Such a limited discrepancy may be explained with the notion that decreased Aβ_{42} reduction in CSF can precede brain amyloid deposition (Palmqvist et al., 2016).

11.1.4 Clinical follow-up

Results shown in figure 7.10 suggest a cautionary attitude when using clinical evaluation (even at follow-up visits) as gold standard. For instance, the lack of amyloid markers – as evidenced by both CSF and PET analysis – is virtually incompatible with the diagnosis /conversion to AD. It has recently been shown that a small but non-trivial part of AD patients of the ADNI population shares a negative Florbetapir scan (Landau et al., 2016). These Florbetapir negative patients would have a variety of clinical and neurodegeneration biomarker features distinct from Florbetapir positive patients, suggesting that one or more non-AD etiologies – such as cerebrovascular disease and SNAP – may be the main cause of their cognitive deficit, mimicking AD.

On the other hand, MCI who reverted to NS condition may still have a negative or borderline amyloid PET burden; among them an occasional patient may show abnormally low levels of CSF Aβ_{42} and might theoretically become Florbetapir positive in the future due to the earlier CSF positivity (Palmqvist et al., 2016).

11.1.5 Image quality issues

A non negligible fraction of scans (60 out of 504) were flagged as "sub-optimal quality" by at least one reader, and 37 were flagged so by all readers (a sample image is provided in figure 6.1). This flag did not imply the impossibility of visual assessment, it meant though that - within the boundary of the readers clinical experience - they deemed that their evaluation was made more difficult by the image quality.

Indeed, the apparent higher performance of the intensity feature C versus the geometric feature G is actually incidental for this study, as scans provenance and quality are so heterogeneous. In another study (unpublished) where data came from a single center, we observed no significant performance difference between the intensity and geometric feature.

The number of flagged scans is rather relevant, given the cost of the ligand and radiation exposure to the patient. The peculiar relevance of image quality and resolution in these investigations is most notable when dealing with difficult cases, as it is with those subjects affected by significant gray matter atrophy. In addition, the relationship between flagged scans and the blind phase *uncertain* label suggest that diagnostic errors are more likely to be made on amyloid-negative subjects.

This highlights the need of procedural guidelines which underline the importance of spatial resolution, partial volume effect correction, and the general minimization of acquisition nuisances; all to be addressed in the official nuclear medicine societies guidelines on acquisition protocols.

The general call to quality and consistency in nuclear medicine imaging peaks in longitudinal studies. Considering the absolute value of the residuals $|\epsilon|$ on the 16 subjects with three scans and dividing them into two batches – the first one consisting of those subjects acquired on the same site and scanner ($n_{same} = 11$) and a second one consisting of those who had at least one scan taken with a different system ($n_{diff} = 5$) – we find $|\epsilon_{same}| = 0.04 \pm 0.02$ and $|\epsilon_{diff}| = 0.14 \pm 0.07$. Although the modest number of samples excludes a definite statement, the influence of scan consistency (both within and among different sites) should be at least considered as a nuisance in the rather large spread found in the evaluation of the ASC. For the same reason, we suggest a minimum of three scans to attempt a reasonable estimation of the annualized score change, unless image quality and consistency can be guaranteed (and possibly quantified).

11.2 TDr validation

We have shown that it is possible to define a SUVr-like analysis with no predefined uptake and reference ROIs. The TDr method shows excellent performance versus the dichotomic visual assessment and - similarly to ELBA - it has ranking characteristics.

From a mehodological point of view, TDr requires an additional spatial normalization steps (to match the early scan onto the late) although the lack of predefined ROIs relax the requirements on the registration onto the MNI space.

TDr takes the SUV ratio to a more refined level while maintaining the mathematical simplicity of a ratio between average counts, thus avoiding the non trivial computing steps required by ELBA.

On the same line of TDr, we should also mention the work of Cecchin et al. (2017) as an excellent example of SUVr refinement. In order to improve the stability and therefore the clinical applicability of the SUVr, the authors have developed a method based on the integration of early late scan, plus the volumetric MRI of the subject. This multimodal approach was able to correct the atrophy and dependence on regional CBF, both typical bias for standard semi-quantification methods. Once more, this example indicates that the problem of semi-quantification is far from being resolved and that new and improved

procedures are needed.

11.2.1 Dataset issues

The mere fact that two scans per patient needs to be acquired, makes the TDr more difficult to validate on existing and public dataset.

TDr was validated on a multicentric dataset and on a single tracer. All limitations and discussions due to these conditions are the same as those just stated for the ELBA method. The Multicentric Pilot Dataset though is much less refined than the ADNI, even when limiting the considerations to the amyloid information alone. For instance, the MPD did not provide any CSF measurements, and not enough clinical data nor follow-up information to be used in validation as the ADNI dataset was used with ELBA.

We should remark though that the introduction of early scans is becoming more and more popular for it is believed to be a reliable proxy to CBF and even FDG-PET (Ottoy et al., 2017; Rostomian et al., 2011; Chen et al., 2015). Besides, the acquisition of an early scan poses little inconvenience on the patient and on the scanner management. We have therefore hope that more data will be available with two scans per subject, that we shall use to consolidate the validation procedure.

The excellent agreement with the visual assessment is a necessary but not sufficient condition to the proper TDr validation: first, the binary evaluations were not discussed in an open session (as in ELBA) and they were given by unevenly trained personnel; second, the inclusion/exclusion criteria were not stated or agreed before the creation and utilization of this dataset.

The rather large heterogeneity in the patients demographics and the lack of a pre-study agreement on scan quality related issues are probably responsible for the uneven response of TDr on the Geneva patients shown in figure 7.8.

11.2.2 Comparison with SUVr and ELBA

The main result for TDr is the good agreement and correlation with the other two validated methods. Results in figure 7.9 clearly indicate that the TDr measures the amyloid load at least with the same accuracy as SUVr and ELBA. When also considering the lack of correlation in the residual analysis (paragraph 9.2), we conclude that TDr is indeed another independent method.

As TDr is mathematically similar to SUVr (a ratio between two uptakes, averaged over some ROI), one would expect a closer relationship with it. Instead, we find TDr to be much more related to ELBA, both as correlation on the whole population and on the separate negative and positive subjects. This latter consideration is very important as it shows the ranking potential of TDr (and ELBA).

In addition, the intra-class correlation between TDr and ELBA impacts not only the TDR validation, but also the pathophysiological implications. It confirms that the transition between the negative and positive state is neither abrupt nor untraceable and that the concept of borderline scans is actually not a mere technical nuisance.

11.3 Longitudinal evaluation

The actual longitudinal experiment could be carried out on the ADNI dataset and with ELBA and SUVr only.

The fitted model in paragraph 8.2.1 is qualitatively comparable to that shown by (Villemagne et al., 2013) albeit using a much shorter observation time (24 months), with a different ligand and on a larger pool of PET scanners. The shorter follow-up time could also explain why our findings on the SUVr longitudinal analysis exhibits a larger variability than that shown by (Villemagne et al., 2013). The ASC shows a rather sparse distribution when computed on two scans only (with a wide range of positive and negative values), a behaviour which is reduced in the 3 scans analysis. In all the 16 subjects with at least three repeated scans the ELBA score increased, showing the sensitivity of the method to amyloid deposition even in a relatively short time span (48 months) and opening potential applications to pharmacological studies with anti-amyloid compounds.

Even taking into account the relatively strong uncertainty due to the use of only two scans, ASC values peaked in the EMCI cohort, are mildly positive within normal subjects and are substantially compatible with zero in AD, a behaviour which is expected according to the latest models of amyloid load (Jack et al., 2013).

The comparison between ELBA ASC and CSF Aβ_{42} points to a discrete, inverse relation between CSF Aβ_{42} level and variations in brain amyloid load, taking into account that they reflect indirect measures of a biologically complex phenomenon.

According to our results which focused on patients with a negative or borderline positive ELBA score at baseline, brain amyloid accumulation appears to be faster in those with a pathological low CSF Aβ_{42} levels, in line with current knowledge.

The relatively high scattering of ASC value of both ELBA and SUVr scores tells us that we are still rather far from being able to use differential measures at the single subject level. This limitation alone should sponsor new methodological approaches to semi-quantification on amyloid-PET images.

Compared to longitudinal SUVr values though, ELBA shows lesser variability. This is likely due to the fact that reference (and uptake) region is not needed, the selection of which has recently been shown to impair the reproducibility and accuracy of longitudinal SUVr measurements (Landau et al., 2015).

In addition, the average ASC values found in our work are comparable to those proposed by (Jack et al., 2013) using SUVr measurements, suggesting that the important clinical and pharmacological implications of an accurate longitudinal evaluation at the cohort level are within our reach, particularly if the protocol involves three or more PET scans.

The high uncertainty in the longitudinal analysis – hence the presence of a significant number of subjects with negative amyloid trend – is mostly due to a measurement / methodological noise rather than a real variation in the data. Indeed, in the cohort subset with three scans, this phenomenon is greatly attenuated.

In the matter of measurement uncertainty (or equivalently the test-retest error) which impairs the typical 2-points longitudinal study, we attempted the cross-sectional estimation. While this is not a true test-retest measure, we can at least rank the three methods on the base of their variances on a common, homogeneous population.

Results in paragraph 8.3 show that TDr is indeed the most reliable of the three measures and suggest that it could have a decidedly lower error in a proper longitudinal study.

The cross-sectional approach relies on several assumptions which are not always satisfied. For instance, the distribution of the standard deviation is not always independent on the cohort. In addition, many more subjects should be included so that we can vary the clustering parameters and provide proper homogeneous cohorts (for instance by optimizing on the degree of order). We should then verify that the results are independent from the number of clusters. In our study we tried 3 to 8 clusters, although we limited the results to 3 because of the optimal balance of the cohort sizes.

Even with all these limitations, we are still confident that TDr ir perhaps the most robust of all three methods.

11.4 Regional analysis

Another open area of investigation is the brain ROI specialization. In this work we considered the brain parcellation to be a pre-defined atlas with 50 cortical regions (25 per hemisphere), with the intent to describe the complexity of amyloid accumulation and have a more stable final result.

Results however indicate that the relationship between uptake regions and method is not so straightforward and each method might perform best on its own regions (figure 8.6).

For instance, ELBA was developed in order to overcome the weaknesses inherent to the SUVr computation, and for this reason the method is not suitable to be applied on small ROIs such as those used in SUVr analysis (see figure 8.3).

In facts, ELBA performance on the strictly defined atlas regions was rather poor, and this problem was solved by extending each uptake region (figure 8.4). Still, ELBA could probably work best when specialized on brain macro regions such as the frontal or parietal lobes.

Another area of investigation is the conciliation between fixed uptake ROIs (like those found in SUVr and TDr) with arbitrary parcellation regions – in principle – which might divide, partially overlap of skip them entirely. In this work we used a probability map to be used as local weights for SUVr, so that it could be computed regardless of the parcellation. This solution however might have impaired the regional performance and further studies are needed to look for possible extensions of SUVr when dealing with an externally imposed parcellation (figure 8.3 and 8.6).

On the same note, TDr was left with its own ROIs as we found them to be rather dense on the cortical areas. Therefore we assumed that each parcellation region would include at least one TDr ROI. This assumption however was not thoroughly verified and there is no guarantee that performance is not impaired on regions with unbalanced number of uptake ROIs.

A possible solution to all these problems is under study. We are currently investigating an adaptive atlas (macro-parcellation) that take into consideration each method's uptake ROIs, while preserving the latitude needed for ELBA to be properly applied.

The macro-parcellation approach should be investigated more as alternative for all fixed parcellations, as it could also prove easier to integrate into the human experience in the

process of clinical assessment.

11.5 Ranking

During our tests the clinicians feedback on the rank were very positive. Indeed, from the mathematical point of view we know the rank to be a better choice than any single measure. If a set of independent measures have uncertainties that are of the same order of magnitude, combining them has always been the method of choice for delivering more robust results (the standard deviation on the mean of n measurements is $\sigma/\sqrt{n}$).

There still is the open question whether the ordering on the first component of the PCA applied to the quantile-normalized measure matrix is the best choice. One criteria is that it shows the least degree of order statistics (figure 9.8).

The presence of a transition in cerebral amyloidosis is known from many studies (Chincarini et al., 2012; Burgold et al., 2014) and has motivated us to use the degree of order as sensor to gauge subgroups coherence as the accumulation process progresses.

We defined the state vector of the system as the ensemble of the regional and global quantifiers on all three methods. The DO was built by exploiting the PCA, as the fraction of variance expressed by the first component with respect to all the others. This approach is a variant of similar definitions found in literature (Gorban et al., 2010; Pagani et al., 2016).

Unlike the work of Pagani et al. (2016) – whose cohorts were provided by an external evaluation of the cognitive state – in our analysis the cohorts were data driven (i.e. generated by a moving subset on the ordered values of the first PCA).

The approach of the DO of the system is allowed assuming the validity of the ergodic hypothesis, as discussed in Pagani et al. (2016). The dynamics of a system is considered ergodic if it has the same behavior when averaged over time or averaged over the space of all the states of the system (phase space). In our case, that amounts to assume that the average of the amyloid accumulation process over time and that of the statistical (sub)set are the same. In other words, that implies that the ranked subject are actually a realization of the accumulation process, lying on the mean trajectory from negativity to positivity.

If the ergodic hypothesis is satisfied, the degree of order that was measured among patients of the same class will give the same result as that shown by the other patient at different times, if they are within the interval where we can safely assume the invariance of the clinical status.

In retrospective, the ranking using the first PCA component was expected to be the best ranking according to the DO variance. In facts, this is true if **the only relevant information** is the amyloid load and if this information is **completely explained** by the fist component only. Then, by the definition of the DO (equation 9.5) all other variances do not carry information, i.e. they only encode noise (the physiological variability). This was actually verified by showing that the other PCA components on the quantile-normalized matrix (up to the fifth) did not carry any information in terms of amyloid load (AUC ≈ 0.5).

11.6 Study limitations

An unavoidable limitation of this study is the lack of a true gold standard (i.e. neuritic plaques at autopsy) that can be used to evaluate the accuracy of the imaging quantitation, and to set an absolute threshold for positivity. Unfortunately there are relatively few examples of such cases in literature (see for instance Clark et al. (2012) or Sabri et al. (2015)). As practical solution – also shown by Johnson et al. (2013a) – we used the consensus visual read as the reference standard, together with the CSF Aβ_{42} and SUVr measurements when available.

In practice, the cross-comparison of all methods should provide the ground for a reasonable method validation targeted at the clinical practice.

Because of the peculiar image treatment which evaluates the intensity distribution patterns, the main drawback for the ELBA analysis is the image quality and consistency. A similar remark can be made on TDr, where the small size and scattered spatial distribution of the uptake regions can be prone to large measure uncertainty in noisy images.

Although convincing evidence can be drawn even from multi-centric and blind analysis such as this one, the weight of the acquisition-related variables can be significant. To correctly estimate this effect we are planning a more detailed analysis on images coming from a single center but with different acquisition protocols and image reconstruction parameters.

Finally, each method has its own strength and weakness. For instance, the ELBA method mimics the visual process in that it captures global geometrical and intensity features, and it is therefore reasonably correlated to the reader's assessment. TDr and SUVr both rely on intensity ratios but based on fundamentally different approaches. It is therefore possible that we can find cases where one method excels and others fail. This is why we believe that any usefulness in a clinical setting is likely to be most informative if we use all of them combined.

Chapter 12

Conclusion

Future diagnostic pathways to the early assessment of neurodegenerative diseases will strongly rely on the in-vivo measurement of cerebral amyloid load.

To date, amyloid-PET ranks among the most reliable and less invasive technique available to clinicians. Yet, the relative novelty of amyloid tracers and the challenges in proper quantification are still an obstacle to all but the most experienced and qualified research centers.

The goal of quantification in amyloid-PET is to extract reliable measures to describe the progress of of accumulation of cerebral Aβ. Some implementation of the quantification process (i.e. those that go under the name of semi-quantification) are widely applicable and prone to the development of automatic analysis.

At present, semi-quantitative analysis is carried out with the SUVr procedure (and its variants), generally calculated on entire brain volume. SUVr has several limitations, which prompted us to look for alternative methods and independent approaches.

We have developed and validated two additional novel semi-quantification methods: ELBA (Chincarini et al., 2016) and TDr (paper under preparation). The three methods (SUVr, ELBA and TDr) are all independent as they are based on fundamentally different approaches. Yet they investigate the uptake of the tracer so that each provides a proxy measure of the cerebral amyloid load. The benefit of independent methods is that their errors are uncorrelated and they have possibly different sensitivities to various nuisances such as the presence of atrophic regions, the image reconstruction parameters and the presence of noise.

A further and consequential step was to integrate all possible measures into a single quantifier. We have researched in this topic and developed a simple and robust quantifier from the whole set of measures and methods. In doing this, we eviscerated the information present in the PET scans, both in terms of local and whole-brain quantitation and in terms of clinical and longitudinal relationships.

Finally, we implemented all the analyses into an easy-to-use web service, which registered Nuclear Medicine and Neurological departments can access, and exploit the detailed information we packed into the report.

Both scientific results and direct feedback from the clinical departments are very encouraging. We have demonstrated that an alternative approach to quantification is possible, one that evolves the current binary assessment into a proper continuous scale and that,

at the same time, provides fundamental and complementary information to the clinical evaluation.

Bibliography

Aisen, P. S., Petersen, R. C., Donohue, M. C., Gamst, A., Raman, R., Thomas, R. G., Walter, S., Trojanowski, J. Q., Shaw, L. M., Beckett, L. A., Jack, C. R., Jagust, W., Toga, A. W., Saykin, A. J., Morris, J. C., Green, R. C., and Weiner, M. W. (2010). Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans. *Alzheimer's & Dementia*, 6(3):239–46.

Alessio, A. and Kinahan, P. (2006). PET image reconstruction. *Nuclear Medicine*, 1(figure 1):1–22.

Alessio, A. M., Kinahan, P. E., and Lewellen, T. K. (2006). Modeling and incorporation of system response functions in 3-D whole body PET. *IEEE Transactions on Medical Imaging*, 25(7):828–837.

Bailey, D. L., Townsend, D. W., Valk, P. E., and Maisey, M. N. (2005). *Positron Emission Tomography*. Springer-Verlag, London.

Baker, J. E., Lim, Y. Y., Pietrzak, R. H., Hassenstab, J., Snyder, P. J., Masters, C. L., and Maruff, P. (2017). Cognitive impairment and decline in cognitively normal older adults with high amyloid-β: A meta-analysis. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 6:108–121.

Barrett, H. H. (1990). Objective assessment of image quality: effects of quantum noise and object variability. *Journal of the Optical Society of America. A, Optics and image science*, 7(7):1266–78.

Becker, G. a., Ichise, M., Barthel, H., Luthardt, J., Patt, M., Seese, A., Schultze-Mosgau, M., Rohde, B., Gertz, H.-J., Reininger, C., and Sabri, O. (2013a). PET Quantification of 18F-Florbetaben Binding to -Amyloid Deposits in Human Brains. *Journal of Nuclear Medicine*, 54(5):723–731.

Becker, G. A., Ichise, M., Barthel, H., Luthardt, J., Patt, M., Seese, A., Schultze-Mosgau, M., Rohde, B., Gertz, H.-J., Reininger, C., and Sabri, O. (2013b). PET quantification of 18F-florbetaben binding to β-amyloid deposits in human brains. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 54(5):723–31.

Bianca, V. D., Dusi, S., Bianchini, E., Prà, I. D., and Rossi, F. (1999). beta-Amyloid Activates the O-2Forming NADPH Oxidase in Microglia, Monocytes, and Neutrophils. *Journal of Biological Chemistry*, 274(22):15493–15499.

Blennow, K., Hampel, H., Weiner, M., and Zetterberg, H. (2010). Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, 6(3):131–144.

Blennow, K., Zetterberg, H., Haass, C., and Finucane, T. (2013). Semagacestat's fall: where next for AD therapies? *Nature Medicine*, 19(10):1214–1215.

Bourgeat, P., Villemagne, V. L., Dore, V., Brown, B., Macaulay, S. L., Martins, R., Masters, C. L., Ames, D., Ellis, K., Rowe, C. C., Salvado, O., and Fripp, J. (2015). Comparison of MR-less PiB SUVR quantification methods. *Neurobiology of Aging*, 36(S1):S159—-S166.

Braak, H. and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, 82(4):239–259.

Braak, H. and Braak, E. (1995). Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, 16(3):271–278.

Bruyant, P. P. (2002). Analytic and iterative reconstruction algorithms in SPECT. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 43(10):1343–1358.

Buerger, K., Ewers, M., Pirttila, T., Zinkowski, R., Alafuzoff, I., Teipel, S. J., DeBernardis, J., Kerkman, D., McCulloch, C., Soininen, H., and Hampel, H. (2006). CSF phosphorylated tau protein correlates with neocortical neurofibrillary pathology in Alzheimer's disease. *Brain*, 129(11):3035–3041.

Burgold, S., Filser, S., Dorostkar, M. M., Schmidt, B., and Herms, J. (2014). In vivo imaging reveals sigmoidal growth kinetic of beta-amyloid plaques. *Acta Neuropathologica Communications*, 2(1):30.

Camus, V., Payoux, P., Barré, L., Desgranges, B., Voisin, T., Tauber, C., La Joie, R., Tafani, M., Hommet, C., Chételat, G., Mondon, K., De La Sayette, V., Cottier, J. P., Beaufils, E., Ribeiro, M. J., Gissot, V., Vierron, E., Vercouillie, J., Vellas, B., Eustache, F., and Guilloteau, D. (2012). Using PET with 18F-AV-45 (florbetapir) to quantify brain amyloid load in a clinical environment. *European Journal of Nuclear Medicine and Molecular Imaging*, 39(4):621–631.

Catafau, A. M. and Bullich, S. (2015). Amyloid PET imaging: applications beyond Alzheimer's disease. *Clinical and Translational Imaging*, 3(1):39–55.

Cecchin, D., Barthel, H., Poggiali, D., Cagnin, A., Tiepolt, S., Zucchetta, P., Turco, P., Gallo, P., Frigo, A. C., Sabri, O., and Bui, F. (2017). A new integrated dual time-point amyloid PET/MRI data analysis method. *European journal of nuclear medicine and molecular imaging*, 44(12):2060–2072.

Chandler, D. (1987). Introduction to modern statistical mechanics.

Chen, Y. J., Rosario, B. L., Mowrey, W., Laymon, C. M., Lu, X., Lopez, O. L., Klunk, W. E., Lopresti, B. J., Mathis, C. A., and Price, J. C. (2015). Relative 11C-PiB Delivery as a Proxy of Relative CBF: Quantitative Evaluation Using Single-Session 15O-Water and 11C-PiB PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 56(8):1199–205.

Cherry, S. R., Sorenson, J., Phelps, M. E., and Methe, B. M. (2004). Physics in Nuclear Medicine. *Medical Physics*, 31(8):2370.

Chincarini, A., Bosco, P., Gemme, G., Morbelli, S., Arnaldi, D., Sensi, F., Solano, I., Amoroso, N., Tangaro, S., Longo, R., Squarcia, S., and Nobili, F. (2012). Alzheimer's disease markers from structural MRI and FDG-PET brain images. *The European Physical Journal Plus*, 127(11):135.

Chincarini, A., Sensi, F., Rei, L., Bossert, I., Morbelli, S., Guerra, U. P., Frisoni, G., Padovani, A., and Nobili, F. (2016). Standardized Uptake Value Ratio-Independent Evaluation of Brain Amyloidosis. *Journal of Alzheimer's Disease*, 54(4):1–21.

Clark, C. M., Pontecorvo, M. J., Beach, T. G., Bedell, B. J., Coleman, R. E., Doraiswamy, P. M., Fleisher, A. S., Reiman, E. M., Sabbagh, M. N., Sadowsky, C. H., Schneider, J. a., Arora, A., Carpenter, A. P., Flitter, M. L., Joshi, A. D., Krautkramer, M. J., Lu, M., Mintun, M. a., and Skovronsky, D. M. (2012). Cerebral PET with florbetapir compared with neuropathology at autopsy for detection of neuritic amyloid-beta plaques: A prospective cohort study. *The Lancet Neurology*, 11(8):669–678.

Clark, C. M., Schneider, J. A., Bedell, B. J., Beach, T. G., Bilker, W. B., Mintun, M. A., Pontecorvo, M. J., Hefti, F., Carpenter, A. P., Flitter, M. L., Krautkramer, M. J., Kung, H. F., Coleman, R. E., Doraiswamy, P. M., Fleisher, A. S., Sabbagh, M. N., Sadowsky, C. H., Reiman, E. M., Zehntner, S. P., and Skovronsky, D. M. (2011). Use of Florbetapir-PET for Imaging beta-Amyloid Pathology. *Jama*, 305(3):275–283.

Contractor, K. B., Kenny, L. M., Coombes, C. R., Turkheimer, F. E., Aboagye, E. O., and Rosso, L. (2012). Evaluation of limited blood sampling population input approaches for kinetic quantification of [18F]fluorothymidine PET data. *EJNMMI Research*, 2(1):11.

Crary, J. F., Trojanowski, J. Q., Schneider, J. A., Abisambra, J. F., Abner, E. L., Alafuzoff, I., Arnold, S. E., Attems, J., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., Gearing, M., Grinberg, L. T., Hof, P. R., Hyman, B. T., Jellinger, K., Jicha, G. A., Kovacs, G. G., Knopman, D. S., Kofler, J., Kukull, W. A., Mackenzie, I. R., Masliah, E., McKee, A., Montine, T. J., Murray, M. E., Neltner, J. H., Santa-Maria, I., Seeley, W. W., Serrano-Pozo, A., Shelanski, M. L., Stein, T., Takao, M., Thal, D. R., Toledo, J. B., Troncoso, J. C., Vonsattel, J. P., White, C. L., Wisniewski, T., Woltjer, R. L., Yamada, M., and Nelson, P. T. (2014). Primary age-related tauopathy ({PART}): a common pathology associated with human aging. *Acta Neuropathologica*, 128(6):755–766.

Dempster, A. P., Laird, N. M., and D.B. Rubin (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm.

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., DeKosky, S. T., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G. B., Fox, N. C., Galasko, D., Habert, M.-o., Jicha, G. A., Nordberg, A., Pasquier, F., Rabinovici, G., Robert, P., Rowe, C., Salloway, S., Sarazin, M., Epelbaum, S., de Souza, L. C., Vellas, B., Visser, P. J., Schneider, L., Stern, Y., Scheltens, P., and Cummings, J. L. (2014). Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria. *The Lancet Neurology*, 13(6):614–629.

Eli Lilly Inc., U. S. A. (2012). Amyvid (florbetapir F18) datasheet.

Engler, H., Forsberg, A., Almkvist, O., Blomquist, G., Larsson, E., Savitcheva, I., Wall, A., Ringheim, A., Langstrom, B., and Nordberg, A. (2006). Two-year follow-up of amyloid deposition in patients with Alzheimer's disease. *Brain*, 129(11):2856–2866.

Fagan, A. M., Mintun, M. A., Mach, R. H., Lee, S.-Y., Dence, C. S., Shah, A. R., LaRossa, G. N., Spinner, M. L., Klunk, W. E., Mathis, C. A., DeKosky, S. T., Morris, J. C., and Holtzman, D. M. (2006). Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid Abeta42 in humans. *Annals of neurology*, 59(3):512–9.

Fleisher, A. S., Chen, K., Liu, X., Roontiva, A., Thiyyagura, P., Ayutyanont, N., Joshi, A. D., Clark, C. M., Mintun, M. A., Pontecorvo, M. J., Doraiswamy, P. M., Johnson, K. A., Skovronsky, D. M., and Reiman, E. M. (2011). Using Positron Emission Tomography and Florbetapir F 18 to Image Cortical Amyloid in Patients With Mild Cognitive Impairment or Dementia Due to Alzheimer Disease. *Archives of Neurology*, 68(11):1404.

GEhealthcare (2012). Vizamyl (flutemetamol F18) datasheet.

Gjedde, A., Aanerud, J., Braendgaard, H., and Rodell, A. B. (2013). Blood-brain transfer of Pittsburgh compound B in humans. *Frontiers in Aging Neuroscience*, 5(NOV):1–9.

Glabe, C. (2001). Intracellular Mechanisms of Amyloid Accumulation and Pathogenesis in Alzheimer{\textquotesingle}s Disease. *Journal of Molecular Neuroscience*, 17(2):137–145.

Gorban, A. N., Smirnova, E. V., and Tyukina, T. A. (2010). Correlations, risk and crisis: From physiology to finance. *Physica A: Statistical Mechanics and its Applications*, 389(16):3193–3217.

Grimmer, T., Riemenschneider, M., Förstl, H., Henriksen, G., Klunk, W. E., Mathis, C. A., Shiga, T., Wester, H.-J., Kurz, A., and Drzezga, A. (2009). Beta Amyloid in Alzheimer's Disease: Increased Deposition in Brain Is Reflected in Reduced Concentration in Cerebrospinal Fluid. *Biological Psychiatry*, 65(11):927–934.

Guénette, S. Y. (2003). Mechanisms of Abeta Clearance and Catabolism. *Neuromolecular medicine*, 4(3):147–160.

Guerra, U. P., Nobili, F. M., Padovani, A., Perani, D., Pupi, A., Sorbi, S., and Trabucchi, M. (2015). Recommendations from the Italian Interdisciplinary Working Group (AIMN, AIP, SINDEM) for the utilization of amyloid imaging in clinical practice. *Neurological Sciences*, 36(6):1075–1081.

Gunn, R. N., Gunn, S. R., and Cunningham, V. J. (2001). Positron emission tomography compartmental models. *Journal of cerebral blood flow and metabolism*, 21(6):635–652.

Häggström, I. (2014). *Quantitative Methods for Tumor Imaging with Dynamic PET*.

Herman, G. T. (2009). *Fundamentals of Computerized Tomography*. Advances in Pattern Recognition. Springer London, London.

Heurling, K., Buckley, C., Van Laere, K., Vandenberghe, R., and Lubberink, M. (2015). Parametric imaging and quantitative analysis of the PET amyloid ligand 18F-flutemetamol. *NeuroImage*, 121:184–192.

Hudson, H. M. and Larkin, R. S. (1994). Accelerated Image Reconstruction Using Ordered Subsets of Projection Data. *IEEE Transactions on Medical Imaging*, 13(4):601–609.

Hutton, C., Declerck, J., Mintun, M. A., Michael, J., and Joshi, A. (2013). SPAP and Avid florbetapir Analysis Methods. Technical report, Siemens Molecular Imaging, Oxford, UK.

Ichise, M., Meyer, J. H., and Yonekura, Y. (2001). An introduction to PET and SPECT neuroreceptor quantification models. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 42(5):755–63.

Ikoma, Y., Watabe, H., Shidahara, M., Naganawa, M., and Kimura, Y. (2008). PET kinetic analysis: Error consideration of quantitative analysis in dynamic studies. *Annals of Nuclear Medicine*, 22(1):1–11.

Ikonomovic, M. D., Klunk, W. E., Abrahamson, E. E., Mathis, C. a., Price, J. C., Tsopelas, N. D., Lopresti, B. J., Ziolko, S., Bi, W., Paljug, W. R., Debnath, M. L., Hope, C. E., Isanski, B. a., Hamilton, R. L., and DeKosky, S. T. (2008). Post-mortem correlates of in vivo PiB-PET amyloid imaging in a typical case of Alzheimer's disease. *Brain*, 131(6):1630–1645.

Innis, R. B., Cunningham, V. J., Delforge, J., Fujita, M., Gjedde, A., Gunn, R. N., Holden, J., Houle, S., Huang, S. C., Ichise, M., Iida, H., Ito, H., Kimura, Y., Koeppe, R. A., Knudsen, G. M., Knuuti, J., Lammertsma, A. A., Laruelle, M., Logan, J., Maguire, R. P., Mintun, M. A., Morris, E. D., Parsey, R., Price, J. C., Slifstein, M., Sossi, V., Suhara, T., Votaw, J. R., Wong, D. F., and Carson, R. E. (2007). Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *J Cereb Blood Flow Metab*, 27(9):1533–1539.

Jack, C. R. (2014). PART and SNAP. *Acta neuropathologica*, 128(6):773–6.

Jack, C. R., Wiste, H. J., Lesnick, T. G., Weigand, S. D., Knopman, D. S., Vemuri, P., Pankratz, V. S., Senjem, M. L., Gunter, J. L., Mielke, M. M., Lowe, V. J., Boeve, B. F., and Petersen, R. C. (2013). Brain beta-amyloid load approaches a plateau. *Neurology*, 80(10):890–896.

Jagust, W. J., Bandy, D., Chen, K., Foster, N. L., Landau, S. M., Mathis, C. A., Price, J. C., Reiman, E. M., Skovronsky, D., and Koeppe, R. A. (2010). The Alzheimer's Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer's & Dementia*, 6(3):221–9.

Johnson, K. a., Minoshima, S., Bohnen, N. I., Donohoe, K. J., Foster, N. L., Herscovitch, P., Karlawish, J. H., Rowe, C. C., Carrillo, M. C., Hartley, D. M., Hedrick, S., Pappas, V., and Thies, W. H. (2013a). Appropriate use criteria for amyloid PET: a report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Alzheimer's & Dementia*, 9(1):e–1–16.

Johnson, K. A., Sperling, R. A., Gidicsin, C. M., Carmasin, J. S., Maye, J. E., Coleman, R. E., Reiman, E. M., Sabbagh, M. N., Sadowsky, C. H., Fleisher, A. S., Murali Doraiswamy, P., Carpenter, A. P., Clark, C. M., Joshi, A. D., Lu, M., Grundman, M.,

Mintun, M. a., Pontecorvo, M. J., and Skovronsky, D. M. (2013b). Florbetapir (F18-AV-45) PET to assess amyloid burden in Alzheimer's disease dementia, mild cognitive impairment, and normal aging. *Alzheimer's & Dementia*, 9(5 Suppl):S72–83.

Jolliffe, I. T. (1986). *Principal Component Analysis*, volume 2 of *Springer Series in Statistics*. Springer New York, New York, NY.

Joshi, A. D., Pontecorvo, M. J., Clark, C. M., Carpenter, A. P., Jennings, D. L., Sadowsky, C. H., Adler, L. P., Kovnat, K. D., Seibyl, J. P., Arora, A., Saha, K., Burns, J. D., Lowrey, M. J., Mintun, M. a., Skovronsky, D. M., and Florbetapir F 18 Study Investigators, t. F. F. . S. (2012). Performance characteristics of amyloid PET with florbetapir F 18 in patients with alzheimer's disease and cognitively normal subjects. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 53(3):378–84.

Karran, E. and Hardy, J. (2014). Antiamyloid Therapy for Alzheimer's Disease Are We on the Right Road? *New England Journal of Medicine*, 370(4):377–378.

Klein, W. L., Stine, W. B., and Teplow, D. B. (2004). Small assemblies of unmodified amyloid beta-protein are the proximate neurotoxin in Alzheimer's disease. *Neurobiology of Aging*, 25(5):569–580.

Klinger, R. Y., James, O. G., Wong, T. Z., Newman, M. F., Doraiswamy, P. M., and Mathew, J. P. (2013). Cortical β-amyloid levels and neurocognitive performance after cardiac surgery. *BMJ open*, 3(9):e003669.

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G.-f., Estrada, S., Ausén, B., Debnath, M. L., Barletta, J., Price, J. C., Sandell, J., Lopresti, B. J., Wall, A., Koivisto, P., Antoni, G., Mathis, C. A., and Långström, B. (2004). Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound-B. *Annals of neurology*, 55(3):306–19.

Klunk, W. E., Koeppe, R. a., Price, J. C., Benzinger, T. L., Devous, M. D., Jagust, W. J., Johnson, K. a., Mathis, C. a., Minhas, D., Pontecorvo, M. J., Rowe, C. C., Skovronsky, D. M., and Mintun, M. a. (2015). The Centiloid Project: Standardizing quantitative amyloid plaque estimation by PET. *Alzheimer's & Dementia*, 11(1):1—-15.e4.

Koh, J.-y., Yang, L. L., and Cotman, C. W. (1990). beta-Amyloid protein increases the vulnerability of cultured cortical neurons to excitotoxic damage. *Brain Research*, 533(2):315–320.

Lammertsma, A. A. (2017). Forward to the Past: The Case for Quantitative PET Imaging. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 58(7):1019–1024.

Lammertsma, A. A., Bench, C. J., Hume, S. P., Osman, S., Gunn, K., Brooks, D. J., and Frackowiak, R. S. (1996). Comparison of methods for analysis of clinical [11C]raclopride studies. *J Cereb Blood Flow Metab*, 16(1):42–52.

Landau, S. M., Fero, A., Baker, S. L., Koeppe, R., Mintun, M., Chen, K., Reiman, E. M., and Jagust, W. J. (2015). Measurement of Longitudinal -Amyloid Change with 18F-Florbetapir PET and Standardized Uptake Value Ratios. *Journal of Nuclear Medicine*, 56(4):567–574.

Landau, S. M., Horng, A., Fero, A., and Jagust, W. J. (2016). Amyloid negativity in patients with clinically diagnosed Alzheimer disease and MCI. *Neurology*, 86(15):1377–1385.

Landau, S. M., Lu, M., Joshi, A. D., Pontecorvo, M., Mintun, M. A., Trojanowski, J. Q., Shaw, L. M., and Jagust, W. J. (2013). Comparing positron emission tomography imaging and cerebrospinal fluid measurements of beta-amyloid. *Annals of Neurology*, 74(6):826–836.

Landau, S. M., Thomas, B. A., Thurfjell, L., Schmidt, M., Margolin, R., Mintun, M., Pontecorvo, M., Baker, S. L., and Jagust, W. J. (2014). Amyloid PET imaging in Alzheimer's disease: A comparison of three radiotracers. *European Journal of Nuclear Medicine and Molecular Imaging*, 41(7):1398–1407.

Logan, J., Fowler, J. S., Volkow, N. D., Wolf, A. P., Dewey, S. L., Schlyer, D. J., MacGregor, R. R., Hitzemann, R., Bendriem, B., Gatley, S. J., and Christman, D. R. (1990a). Graphical Analysis of Reversible Radioligand Binding from Time-Activity Measurements Applied to [N-11C-methyl]-(-)-Cocaine PET Studies in Human Subjects. *Journal of Cerebral Blood Flow & Metabolism*, 10(5):740–747.

Logan, J., Fowler, J. S., Volkow, N. D., Wolf, A. P., Dewey, S. L., Schlyer, D. J., MacGregor, R. R., Hitzemann, R., Bendriem, B., Gatley, S. J., and Christman, D. R. (1990b). Graphical Analysis of Reversible Radioligand Binding from TimeActivity Measurements Applied to [N - ¹¹ C-Methyl]-()-Cocaine PET Studies in Human Subjects. *Journal of Cerebral Blood Flow & Metabolism*, 10(5):740–747.

Lonsdale, M. N. and Beyer, T. (2010). Dual-modality PET/CT instrumentationToday and tomorrow. *European Journal of Radiology*, 73(3):452–460.

Lopresti, B. J., Klunk, W. E., Mathis, C. A., Hoge, J. A., Ziolko, S. K., Lu, X., Meltzer, C. C., Schimmel, K., Tsopelas, N. D., DeKosky, S. T., and Price, J. C. (2005). Simplified quantification of Pittsburgh Compound B amyloid imaging PET studies: a comparative analysis. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 46:1959–1972.

Lorenzo, A. and Yankner, B. A. (1994). Beta-amyloid neurotoxicity requires fibril formation and is inhibited by congo red. *Proceedings of the National Academy of Sciences*, 91(25):12243–12247.

Lucignani, G., Paganelli, G., and Bombardieri, E. (2004). The use of standardized uptake values for assessing FDG uptake with PET in oncology: a clinical perspective. *Nuclear Medicine Communications*, 25(7):651–656.

MacQueen, J. (1967). Some methods for classification and analysis. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, volume 233, pages 281–297. The Regents of the University of California.

Madsen, M. T. (2005). Emission Tomography: The Fundamentals of PET and SPECT. *Shock*, 23(4):390.

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters.

Matsubara, K., Ibaraki, M., Shimada, H., Ikoma, Y., Suhara, T., Kinoshita, T., and Ito, H. (2016). Impact of spillover from white matter by partial volume effect on quantification of amyloid deposition with [11C]PiB PET. *NeuroImage*.

Mattsson, N., Insel, P. S., Landau, S. M., Jagust, W. J., Donohue, M. C., Shaw, L. M., Trojanowski, J. Q., Zetterberg, H., Blennow, K., and Weiner, M. (2014). Diagnostic accuracy of CSF Ab42 and florbetapir PET for Alzheimer's disease. *Annals of clinical and translational neurology*, 1(8):534–43.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., and Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, 7(3):263–9.

Meda, L., Cassatella, M. A., Szendrei, G. I., Otvos, L., Baron, P., Villalba, M., Ferrari, D., and Rossi, F. (1995). Activation of microglial cells by beta-amyloid protein and interferon-gamma. *Nature*, 374(6523):647–650.

Montine, T. J., Phelps, C. H., Beach, T. G., Bigio, E. H., Cairns, N. J., Dickson, D. W., Duyckaerts, C., Frosch, M. P., Masliah, E., Mirra, S. S., Nelson, P. T., Schneider, J. A., Thal, D. R., Trojanowski, J. Q., Vinters, H. V., and Hyman, B. T. (2011). National Institute on Aging{\textendash}Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach. *Acta Neuropathologica*, 123(1):1–11.

Mormino, E. C., Brandel, M. G., Madison, C. M., Rabinovici, G. D., Marks, S., Baker, S. L., and Jagust, W. J. (2012). Not quite PIB-positive, not quite PIB-negative: slight PIB elevations in elderly normal control subjects are biologically relevant. *NeuroImage*, 59(2):1152–60.

Morris, E. D., Endres, C. J., Schmidt, K. C., Christian, B. T., Muzic, R. F., and Fisher, R. E. (2004). Kinetic Modeling in Positron Emission Tomography. *Emission Tomography: The Fundamentals of PET and SPECT*, 46(1):499–540.

Nelissen, N., Van Laere, K., Thurfjell, L., Owenius, R., Vandenbulcke, M., Koole, M., Bormans, G., Brooks, D. J., and Vandenberghe, R. (2009). Phase 1 study of the Pittsburgh compound B derivative 18F-flutemetamol in healthy volunteers and patients with probable Alzheimer disease. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 50(8):1251–1259.

Ottoy, J., Miedema, M., De Puydt, C., Verhaeghe, J., Deleye, S., Engelborghs, S., Stroobants, S., and Staelens, S. (2017). EARLY FRAME 18F-AV45 AND 18F-FDG-PET AS PROXIES OF CBF: COMPARISON TO 15O-H2O PET DATA. *Alzheimer's & Dementia*, 13(7):P763–P764.

Pagani, M., Giuliani, A., Öberg, J., Chincarini, A., Morbelli, S., Brugnolo, A., Arnaldi, D., Picco, A., Bauckneht, M., Buschiazzo, A., Sambuceti, G., and Nobili, F. (2016). Predicting the transition from normal aging to Alzheimer's disease: A statistical mechanistic evaluation of FDG-PET data. *NeuroImage*, 141:282–290.

Palmqvist, S., Mattsson, N., and Hansson, O. (2016). Cerebrospinal fluid analysis detects cerebral amyloid-β accumulation earlier than positron emission tomography. *Brain*, 139(4):1226–1236.

Palmqvist, S., Zetterberg, H., Mattsson, N., Johansson, P., Alzheimer's Disease Neuroimaging Initiative, L., Minthon, L., Blennow, K., Olsson, M., Hansson, O., and study Swedish BioFINDER Study Group, F. t. S. B. (2015). Detailed comparison of amyloid PET and CSF biomarkers for identifying early Alzheimer disease. *Neurology*, 85(14):1240–9.

Panin, V. Y., Kehren, F., Michel, C., and Casey, M. (2006). Fully 3-D PET reconstruction with system matrix derived from point source measurements. *IEEE transactions on medical imaging*, 25(7):907–921.

Patlak, C. S., Blasberg, R. G., and Fenstermacher, J. D. (1983). Graphical Evaluation of Blood-to-Brain Transfer Constants from Multiple-Time Uptake Data. *Journal of Cerebral Blood Flow & Metabolism*, 3(1):1–7.

Pontecorvo, M. J., Arora, A. K., Devine, M., Lu, M., Galante, N., Siderowf, A., Devadanam, C., Joshi, A. D., Heun, S. L., Teske, B. F., Truocchio, S. P., Krautkramer, M., Devous, M. D., and Mintun, M. A. (2017). Quantitation of PET signal as an adjunct to visual interpretation of florbetapir imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, 44(5):825–837.

Price, J. C., Klunk, W. E., Lopresti, B. J., Lu, X., Hoge, J. A., Ziolko, S. K., Holt, D. P., Meltzer, C. C., DeKosky, S. T., and Mathis, C. A. (2005). Kinetic modeling of amyloid binding in humans using PET imaging and Pittsburgh Compound-B. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*, 25(11):1528–1547.

Rodrigue, K. M., Kennedy, K. M., Devous, M. D., Rieck, J. R., Hebrank, A. C., Diaz-Arrastia, R., Mathews, D., and Park, D. C. (2012). Beta-Amyloid burden in healthy aging: Regional distribution and cognitive consequences. *Neurology*, 78(6):387–395.

Rostomian, A. H., Madison, C., Rabinovici, G. D., and Jagust, W. J. (2011). Early 11C-PIB frames and 18F-FDG PET measures are comparable: a study validated in a cohort of AD and FTLD patients. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 52(2):173–9.

Rowe, C. C., Ng, S., Ackermann, U., Gong, S. J., Pike, K., Savage, G., Cowie, T. F., Dickinson, K. L., Maruff, P., Darby, D., Smith, C., Woodward, M., Merory, J., Tochon-Danguy, H., O'Keefe, G., Klunk, W. E., Mathis, C. A., Price, J. C., Masters, C. L., and Villemagne, V. L. (2007). Imaging beta-amyloid burden in aging and dementia. *Neurology*, 68(20):1718–25.

Sabri, O., Sabbagh, M. N., Seibyl, J., Barthel, H., Akatsu, H., Ouchi, Y., Senda, K., Murayama, S., Ishii, K., Takao, M., Beach, T. G., Rowe, C. C., Leverenz, J. B., Ghetti, B., Ironside, J. W., Catafau, A. M., Stephens, A. W., Mueller, A., Koglin, N., Hoffmann, A., Roth, K., Reininger, C., and Schulz-Schaeffer, W. J. (2015). Florbetaben PET imaging to detect amyloid beta plaques in Alzheimer's disease: phase 3 study. *Alzheimer's & Dementia*, 11(8):964–74.

Saha, G. B. (2016). *Basics of PET Imaging*. Springer International Publishing, Cham.

Schmidt, M. E., Chiao, P., Klein, G., Matthews, D., Thurfjell, L., Cole, P. E., Margolin, R., Landau, S., Foster, N. L., Mason, N. S., De Santi, S., Suhy, J., Koeppe, R. A., and Jagust, W. (2015). The influence of biological and technical factors on quantitative analysis of amyloid PET: Points to consider and recommendations for controlling variability in longitudinal data. *Alzheimer's and Dementia*, 11(9):1050–1068.

Schmitt, D., Karuta, B., Carrier, C., and Lecomte, R. (1988). Fast point spread function computation from aperture functions in high-resolution positron emission tomography. *IEEE Transactions on Medical Imaging*, 7(1):2–12.

Seber, G. A. F., editor (1984). *Multivariate Observations*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Sojkova, J., Driscoll, I., Iacono, D., Zhou, Y., Codispoti, K.-E., Kraut, M. A., Ferrucci, L., Pletnikova, O., Mathis, C. A., Klunk, W. E., O'Brien, R. J., Wong, D. F., Troncoso, J. C., and Resnick, S. M. (2011). In Vivo Fibrillar β-Amyloid Detected Using [11C]PiB Positron Emission Tomography and Neuropathologic Assessment in Older Adults. *Archives of Neurology*, 68(2):232–40.

Tanzi, R. E., MOIR, R. D., and WAGNER, S. L. (2004). Clearance of Alzheimer's Abeta Peptide. The Many Roads to Perdition. *Neuron*, 43(5):605–608.

Thal, D. R., Rüb, U., Orantes, M., and Braak, H. (2002). Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology*, 58(12):1791–1800.

Thirion, J. P. (1998). Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical image analysis*, 2(3):243–60.

Thurfjell, L., Lilja, J., Lundqvist, R., Buckley, C., Smith, A., Vandenberghe, R., and Sherwin, P. (2014). Automated Quantification of 18F-Flutemetamol PET Activity for Categorizing Scans as Negative or Positive for Brain Amyloid: Concordance with Visual Image Reads. *Journal of Nuclear Medicine*, 55(10):1623–1628.

van Berckel, B. N. M., Ossenkoppele, R., Tolboom, N., Yaqub, M., Foster-Dingley, J. C., Windhorst, A. D., Scheltens, P., Lammertsma, A. A., and Boellaard, R. (2013). Longitudinal Amyloid Imaging Using 11C-PiB: Methodologic Considerations. *Journal of Nuclear Medicine*, 54(9):1570–1576.

Vandenberghe, R., Adamczuk, K., Dupont, P., Laere, K. V., and Chételat, G. (2013). Amyloid PET in clinical practice: Its place in the multidimensional space of Alzheimer's disease. *NeuroImage: Clinical*, 2(1):497–511.

Vandenberghe, S., Mikhaylova, E., D'Hoe, E., Mollet, P., and Karp, J. S. (2016). Recent developments in time-of-flight PET. *EJNMMI Physics*, 3(1):3.

Villemagne, V. L., Burnham, S., Bourgeat, P., Brown, B., Ellis, K. A., Salvado, O., Szoeke, C., Macaulay, S. L., Martins, R., Maruff, P., Ames, D., Rowe, C. C., and Masters, C. L. (2013). Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. *The Lancet. Neurology*, 12(4):357–67.

Villemagne, V. L., Ong, K., Mulligan, R. S., Holl, G., Pejoska, S., Jones, G., O'Keefe, G., Ackerman, U., Tochon-Danguy, H., Chan, J. G., Reininger, C. B., Fels, L., Putz, B., Rohde, B., Masters, C. L., and Rowe, C. C. (2011). Amyloid Imaging with 18F-Florbetaben in Alzheimer Disease and Other Dementias. *Journal of Nuclear Medicine*, 52(8):1210–1217.

Watabe, H., Ikoma, Y., Kimura, Y., Naganawa, M., and Shidahara, M. (2006). PET kinetic analysis - Compartmental model. *Annals of Nuclear Medicine*, 20(9):583–588.

Wong, D. F., Rosenberg, P. B., Zhou, Y., Kumar, A., Raymont, V., Ravert, H. T., Dannals, R. F., Nandi, A., Brašić, J. R., Ye, W., Hilton, J., Lyketsos, C., Kung, H. F., Joshi, A. D., Skovronsky, D. M., Pontecorvo, M. J., Brasić, J. R., Ye, W., Hilton, J., Lyketsos, C., Kung, H. F., Joshi, A. D., Skovronsky, D. M., and Pontecorvo, M. J. (2010). In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (florbetapir [corrected] F 18). *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 51(6):913–20.

Yotter, R. A., Doshi, J., Clark, V., Sojkova, J., Zhou, Y., Wong, D. F., Ferrucci, L., Resnick, S. M., and Davatzikos, C. (2013). Memory decline shows stronger associations with estimated spatial patterns of amyloid deposition progression than total amyloid burden. *Neurobiology of Aging*, 34(12):2835–2842.

Zhang, K., Herzog, H., Mauler, J., Filss, C., Okell, T. W., Kops, E. R., Tellmann, L., Fischer, T., Brocke, B., Sturm, W., Coenen, H. H., and Shah, N. J. (2014). Comparison of cerebral blood flow acquired by simultaneous [15O]water positron emission tomography and arterial spin labeling magnetic resonance imaging. *Journal of Cerebral Blood Flow & Metabolism*, 34(8):1373–1380.

Ziolko, S. K., Weissfeld, L. A., Klunk, W. E., Mathis, C. A., Hoge, J. A., Lopresti, B. J., DeKosky, S. T., and Price, J. C. (2006). Evaluation of voxel-based methods for the statistical analysis of PIB PET amyloid imaging studies in Alzheimer's disease. *NeuroImage*, 33(1):94–102.

Zlokovic, B. V. (2004). Clearing amyloid through the blood-brain barrier. *Journal of neurochemistry*, 89(4):807–11.

Zwan, M. D., Rinne, J. O., Hasselbalch, S. G., Nordberg, A., Lleó, A., Herukka, S.-K., Soininen, H., Law, I., Bahl, J. M., Carter, S. F., Fortea, J., Blesa, R., Teunissen, C. E., Bouwman, F. H., van Berckel, B. N., and Visser, P. J. (2016). Use of amyloid-PET to determine cutpoints for CSF markers. *Neurology*, 86(1):50–58.