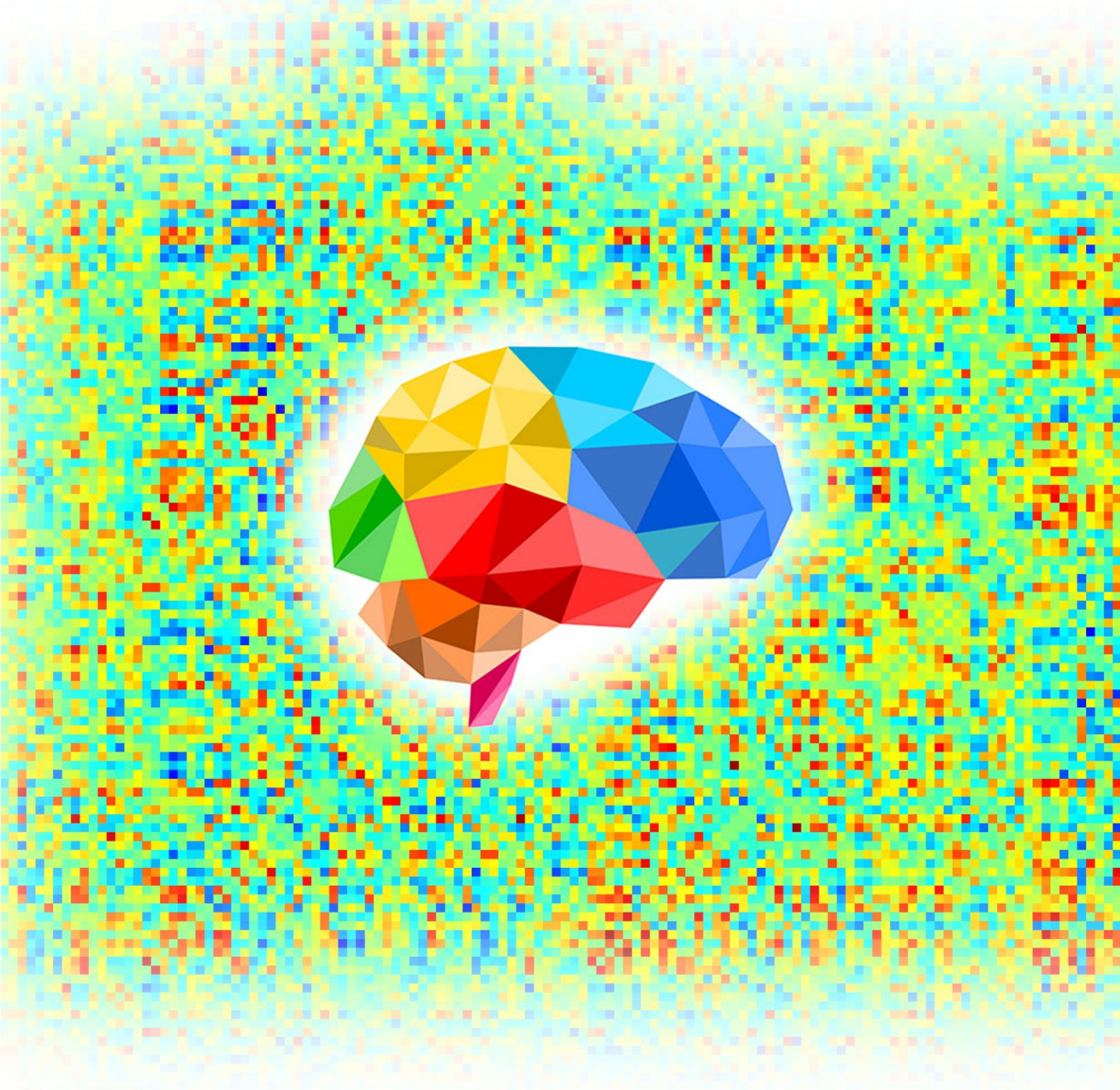
Master thesis

Improving the reproducibility of BOLD rs-fMRI signals by selective data elimination

Theo Driever



ImPhys
Image to insight
Insight in imaging
Department of Imaging Physics

To be defended on
15 Februari 2018

BIGR
Biomedical Imaging Group Rotterdam

Abstract

Connectivity mapping with resting state functional magnetic resonance imaging (rs-fMRI) is rapidly developing and has shown great promise for clinical applications. Before successful implementation in clinical setting, it is key to evaluate the long-term reproducibility of the functional connectivity profiles. To this end, the reproducibility of rs-fMRI data is studied in this work. The main research question revolves around the improvement of the overall reproducibility by selectively omitting single components (either nodes or elements) from BOLD rs-fMRI connectivity matrices (CM's). The scans of the subjects are parcellated using four different schemes, which are all analysed throughout this work. A reproducibility study is carried out on a dataset of 37 subjects that are scanned twice within 2 weeks on average. The inter-subject intraclass correlation coefficient (ICC) is used to quantify component reproducibility within the dataset. An algorithm is designed to quantify which component has the lowest inter-subject ICC, which is then eliminated from all CM's in the dataset. After every single component elimination, the intra-subject ICC is computed for every subject to quantify the reproducibility, and a matching accuracy (MA) test is performed on the set to quantify the distinctive power of the CM's.

The order in which components are eliminated and its effect on the overall reproducibility is tested by applying this to a larger test set of longitudinal data. To this end, a dataset of 521 subjects is used to quantify the reproducibility of the CM's after iteratively removing components in the order that is found in the reproducibility study. This larger dataset of 521 subjects is analysed, along with 4 subsets, namely: sex based, age based, interscan time based and based on the grounds for exclusion. The latter is a subset where the quality of the rs-fMRI scans could not be assured due to pathologies or excessive motion during image acquisition. No significant difference is found within the sex-based subsets, and no relation between the reproducibility and the interscan time (within the range that is assessed in this work, namely 5 years) is found. Significantly lower intra-subject ICC's are found for the subjects whose scan quality was subpar, due to excessive motion or pathology. For the age-based subset analysis, it is reported that reproducibility decreases with age.

The node removal algorithm clearly outperforms the element removal algorithm when looking at the intra-subject ICC. As the element removal algorithm can increase the intra-subject ICC by roughly 0.1, whereas the node removal algorithm manages to increase the intra-subject ICC of roughly 0.3. The MA, which is used as to quantify the distinguishing power between various CM's, is seen to increase from 82.4% to the maximum of 98.7% correctly matched subjects for the RSS100 parcellation scheme within the reproducibility study. Aside from the element removal within the reproducibility study, the matching accuracy is not improved for any of the other analyses. The component elimination algorithm can increase the intra-subject ICC's of the subjects of the longitudinal set. The MA is not found to increase with the component elimination algorithm.

Contents

Preface

In the past 20 years, the field of neuroimaging has become a mayor area of research. This is mostly due to increased imaging capabilities, pushing the boundaries of what was thought possible. This is especially the case with respect to imaging function of the brain. For a lot of people strongly identify with the brain. (i.e. 'I am my brain') This public interest appears to be interested in the concept of 'consciousness' which is widely described in main-stream 'popular' science outlets. As everybody seems to experience it daily, but what is exactly is remains extremely hard to explain.

There is also societal necessity to deepen the knowledge of the brain to address the numerous pathologies associated with it. With the senior population steadily growing through the years, and the general life expectancy increasing as well, it is essential to develop better treatment for age-related neurodegenerative pathologies. Understanding the brain is crucial to do so.

In this thesis an effort will be made to broaden the knowledge of the functional connectome, as to help improve future neurodegenerative treatment options. This work, and the master thesis are performed at the faculty of Applied Physics at the Technical University Delft as part of the master education Biomedical Engineering.

1. Introduction

In an adult human male of 30 years, the brain accounts for roughly 2% of the body mass, while consuming 20% of the energy intake. In children the brain accounts for 50% of total energy intake [1]! Increased neuronal signalling due to cognitive task paradigms is not more than 5% of the total energy intake [2], which is surprisingly small. So how is this energy used by the brain?

One of the first people to get an insight into this energy expenditure of the human brain is Hans Berger, who in 1924 used his newly invented electroencephalogram (EEG) device to make arguably the first ever recording of brain activity. Importantly, the readout of the electrical activity appeared to be ever ongoing, so not only related to the processing of external stimuli, but also responsible for making sure the body works appropriately. The understanding of these signals remained very crude as the only means of interpretation was attained from cognitive task paradigms. It was only in the last 20 years or so, that further investigation of the brain revealed more about the workings of the brain, mainly brought about by new imaging technologies. The CT scan, invented by sir Godfrey Hounsfield was revolutionary, as it was now possible to look inside the brain without the need for risky surgeries. Both magnetic resonance imaging (MRI) and positron emission tomography (PET) were introduced a few years later. With PET it became possible to image function, by introducing radioactively labelled sugar molecules into a subject and see where the biggest uptake of the sugar arose within the brain. While MRI was initially used to make remarkably detailed anatomical images of the brain, it also harboured the potential to image function of the brain. A promising tool to investigate this function more in-depth with, is by means of resting state functional MRI (rs-fMRI), which is a type of MRI scan that gives insight into the functional behaviour of the brain. The use of rs-fMRI has great potential for clinical applications pertaining to various neuropathology's.

2. Theory

2.1. In vivo neuroimaging

There are two main types of tissue that can be distinguished in the brain: Grey matter and white matter. The grey matter largely forms the outer shell of the brain, and the white matter is predominantly located on the inside of the brain. Both tissue types contain parts of brain cells or neurons, the difference being that the grey matter consists of the cell bodies, while the white matter consists of the axons covered by the myelin sheath.

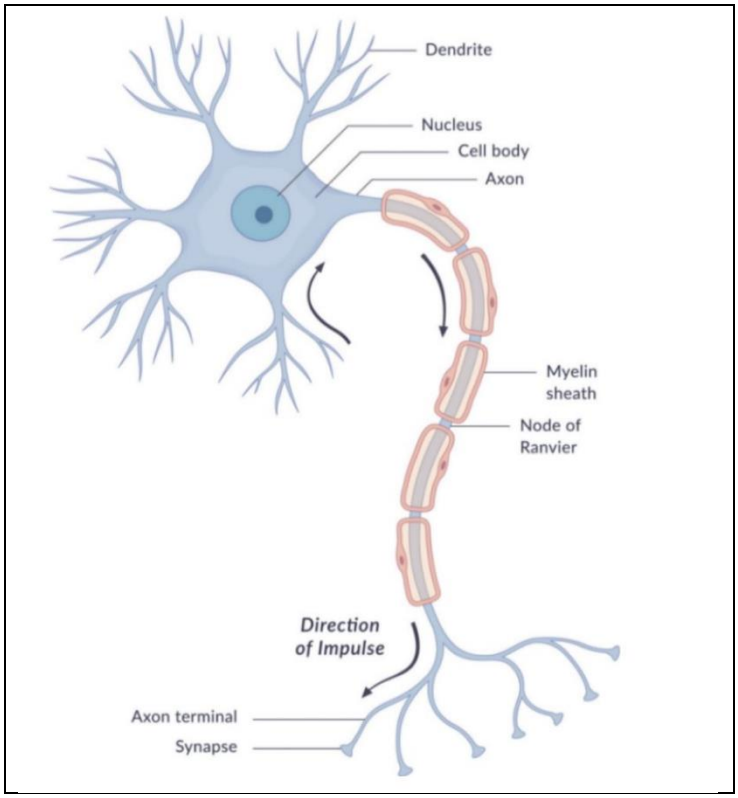


Figure 1 A diagram of a typical neuron with several parts and the direction of impulse labelled [3].

The distinction in colour is brought about by the myelin sheath covering the axons of the neurons, which is a fatty layer that speeds up the action potential. The signal exchange between various neurons brings about the computational power of the brain, which is mainly done in the grey matter. The white matter is mainly concerned with connecting various parts of the brain to itself and to the rest of the body. Using modern MRI techniques, it is not only possible to make extremely detailed anatomical images of the brain, but also allow for the investigation of brain function. Diffusion tensor imaging (DTI) shows the white matter structure, and complementary is fMRI which images the oxygen consumption of the grey matter, thus indirectly imaging the activity of the grey matter.

DTI measures the diffusivity of water in the brain in three directions [4]. As the water molecules move around the axons by means of diffusion, their movement is restricted by the presence of the white matter tracts' cell walls. Diffusion through the white matter tracts is obstructed by the tissue, while movement along the white matter tracts is not. It is this anisotropic diffusivity within white matter tissue that forms the basis of DT imaging [5]. By connecting the orientations of the largest diffusivity in voxels the structure of the white matter tracts is visualised. This type of imaging is called tractography, and it is used to study the brain's structural architecture.

With fMRI, the focus is placed on imaging the activity of the grey matter. The primary technique of fMRI used, is blood-oxygen-level dependant (BOLD) fMRI which reflects the cerebral blood flow (CBF) [6]. Because a neuron does not have an energy reserve in the form of oxygen or sugar, it is solely dependent on the CBF for all its metabolic needs required for appropriate functioning. Local neuronal activity evokes a haemodynamic response to supply the tissue with increased quantities of oxygen and nutrients. The haemodynamic response basically consists of local vasodilation, enlarging the surface area between brain tissue and vascular tissue. An increased amount of oxygenated haemoglobin (oxy-Hb) is delivered to the neuron, leading to a decrease of the ratio between deoxygenated haemoglobin (deoxy-Hb) and oxy-Hb in the respective voxel [7][8]. Deoxy-Hb is paramagnetic and oxy-Hb is diamagnetic [9][10][11]. The resulting relatively lowered concentration of deoxy-Hb produces magnetic inhomogeneities that increase the dephasing of the hydrogen atoms in a region which is approximately two times the radius of the capillary [12]. This disturbance leads to a reduced MR signal, which indirectly reflects brain activity.

Interpretation of BOLD data has, for the most part, been similar to that of other neurological data. By administering a stimulus to a subject, or letting a subject undergo a cognitive task paradigm, a neurological response is evoked which is recorded by the measuring device. This approach has been very successful, so much so that a large portion of what is currently known about brain function comes from this type of research [13]. An example of such a study is shown in Figure 2. A subject was given a task to open and close the eyes at specific intervals in time. By subtracting the average signals of the 'eyes-open' state from the 'eyes-closed' state, it can clearly be seen that the biggest neurological response is evoked in the visual cortex. Averaging of the signal is performed because of relatively large fluctuations that are present in the BOLD signal. At the time, these so-called spontaneous fluctuations were considered to be noise.

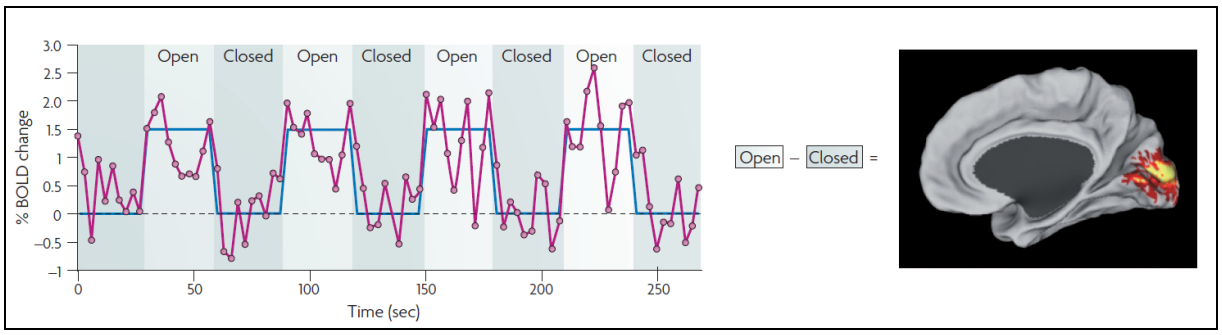


Figure 2 Shown in magenta is the unaveraged change in BOLD signal as taken from a region in the primary visual cortex. The subject was given the cognitive task paradigm to open and close the eyes at fixed intervals in time. In blue is the average change of the BOLD signal within the specific time interval. By subtracting the two states a clear activation can be seen in the primary visual cortex, as shown in the right-hand side of the image [13].

2.2. Acquisition and analysis of BOLD data

In the first paper studying the connectivity of the motor cortex in humans using BOLD fMRI, it was observed that there was a specific correlation between brain activity arising in the left and right motor cortex, even in the absence of a specific task [14]. One of the conclusions of this work was that these spontaneous fluctuations are not just random noise, but harbour function. This paper implicitly initiated a new way of interpreting neurological data: by means of analysing the correlation of BOLD signals. But before the BOLD signal can be analysed, pre-processing must be performed.

The raw BOLD data acquired with fMRI consists of a 4-dimensional dataset, namely a time series for each voxel in the brain. Pre-processing entails the sequential application of spatio-temporal image processing steps applied to the data. These pre-processing steps address issues that arise from (among others): Head motion, magnetic field inhomogeneities and within-volume slice-timing

differences [15]. Additionally, to facilitate comparison between fMRI data of successive scans of the same subject and scans of different subjects, the data is often mapped to an atlas. There is an abundance of algorithms and/or techniques employed to address these issues, all of which have certain benefits and drawbacks [15].

Next should be determined what type of the frequency spectrum from the BOLD fMRI data should be used. It is known that the power spectral density function of a BOLD fMRI data exhibits a 1/f relation in the frequency domain, meaning there is more activity in the lower frequencies [16][17]. The higher frequencies of the BOLD fMRI data are related to cardiac and respiratory signals [18], Therefore a cut-off frequency of <0.1 Hz or 0.08 Hz tends to be used. There appears to be no decisive literature about a lower limit to the frequency spectrum that contributes to the spatially specific correlations [13]. This is in-part because the acquisition of very low frequency fMRI data would take longer to acquire than a regular scan.

A wide-spread method to process the BOLD data is a seed-based analysis. In this technique a region of interest is selected a priori, after which the correlation to all other voxels is determined. The technique was inspired by the work of Biswal [14] and enabled researchers to investigate the relation of the seed region to all other regions in the brain [19]. Apart from the ease of implementation, an apparent drawback of seed-based analysis is the restriction to only investigate the relation of a single region. It is possible to generalize the seed-based analysis to the entire brain, by extending the technique. This is made possible by applying a hierarchical clustering on the extracted time series, and thus producing a hierarchical tree or topological map that reflect the relations of multiple brain regions with one another [13]. In any case, both techniques still rely on the a priori definition of seed regions and is thus heavily influenced by the user.

Another means of analysing the BOLD fMRI data is by means of independent component analysis (ICA). This technique requires no a priori definition of seed regions, but instead aims to decompose the data into independent components via maximization of mutual independence among these components [19]. Each component is accompanied with a spatial map, that can either reflect a noise component or a neurological system. It is up to the discretion of the researcher to interpret the components. Additionally, the outcome is influenced by the number of components, which is a parameter set by the researcher [13].

2.3. Grasping intrinsic activity

With a means of analysing the raw data as outlined in the previous section, the next step is representing the data in a manner that is both comprehensive and informative. Graph theory delivers promising tools to study the functional architecture of the human brain, and how it effects emotion and cognition [20][21][22]. Graph theory governs the study of network clustering and hierarchies, the study of network hubs, (the nodes of modular groups that are interconnected to other modular groups) studying general network efficiency and deriving network summary statistical measures such as the degree or shortest path length [23]. The most general implementation of graph theory on fMRI data is by means of cross-correlation matrices, or connectivity matrices (CM's). An example of a CM is given in Figure 3.

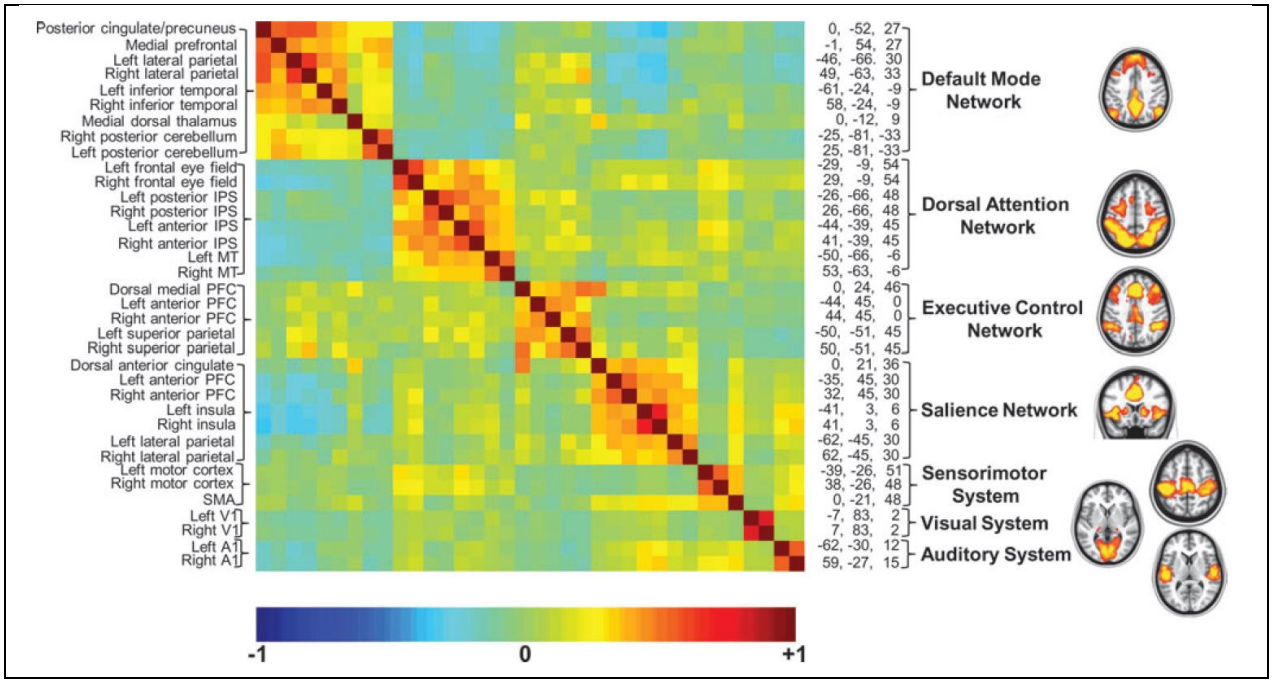


Figure 3 A connectivity matrix displaying Pearson's correlation coefficients between various nodes shown as a heat-map, along with the spatial coordinates on the right of the matrix and the name of the regions on the left. The matrix is rearranged in order to group regions with similar spatial correlations. On the far right-hand side are the names of 7 ICN's along with an image of the brain regions that they respectively consist of [25].

Figure 3 shows a connectivity matrix with a heatmap overlay, along with the location of the nodes and the names of a few intrinsic connectivity networks (ICN's). Every row and column represent a node, and every element represents the correlation between the respective nodes and is generally called an edge. As two identical timeseries have a correlation of 1, the diagonal of the CM only contains ones.

The group of nodes that exhibit integrated and coherent behaviour with one another is referred to as an ICN [19]. The BOLD signals of these ICN's fluctuate in a synchronous manner [24], while the relation between ICN's can be positively or negatively correlated. What should also be noted is that these networks even persist in the absence of a specific task paradigm, continuing to behave in an integrated manner, both within the ICN as well as between various ICN's [25]. Further studies revolving around the finer architecture of the human brain network have found that the brain exhibits a high degree of modularity [26][27], which can be altered by mental disease, (e.g. schizophrenia, Alzheimer's disease, ADD/ADHD) [28][29][30][31] is affected by age and sex [32][33][34] and can undergo rapid plastic changes [35][36].

One way to try and understand this type of spontaneous, yet coherent, activity is by using an orchestra as a metaphor. Generally, orchestras contain a brass, strings, percussion and woodwind section. Even though these groups of instruments play a different tune, the sum of all their interactions form a piece of music. The brain works similarly, with a degree of modularity within the architecture giving rise to emergent function [37].

2.4. Reproducibility

A key issue with fMRI data is reproducibility. Due to the highly dynamic nature of the brain, it is important to be able to quantify the reproducibility of the data. A common metric used to quantify the reproducibility of BOLD fMRI data is the intra-class correlation coefficient (ICC), which informs on the ability of fMRI to assess the differences in brain activation between subjects [38].

There are three types of ICC, each with a two-fold choice about the type of data that can be used [39]. Apart from the widespread adaptation of the ICC for fMRI reproducibility studies, there appears to be no consensus on what type of ICC should be used to quantify fMRI reproducibility [40]. Overall, the studies that employ the two-way random effect model report lower overall reproducibility [41][42] than studies employing a one-way random effect model [43][44][45]. The main reason for this apparent effect, is that the one-way random effects model assumes that the only source of variance arises from the systematic variances brought forth by the subjects, whereas the two-way model also assumes that variance arises both from systematic variance of the subjects and the session [46]. The ICC that will be used throughout this work is ICC_3, which is the version that quantifies the consistency of data. Due to the dynamic nature of the brain no two fMRI scans of the same subject will be identical. Additionally, due to errors in data acquisition and data parcellation, no two nodes will ever encompass the exact same region of the brain. For these reasons, quantifying the consistency of the data is more meaningful than quantifying its absolute agreement0 [46]. ICC_3 is shown in equation 1. Throughout the rest of this work, ICC_3 will be referred to as the ICC.

$$ICC_3 = \frac{MS_p - MS_e}{MS_p + (k-1)MS_E} \tag{1}$$

The way the ICC works is by having two sets of data, referred to as targets. Here, MS_p denotes between target (participant) mean sum of squares, MS_e denotes the error mean sum of squares and k denotes the amount of numbers in a target. See Appendix A for a detailed derivation of the ICC and its components as shown in equation 1. The way the output of the ICC is interpreted is shown in Table 1.

Table 1 Interpretation of ICC values [47].

ICC value	**Interpretation of the reproducibility**
1 > ICC ≥ 0.80	Outstanding
0.80 > ICC ≥ 0.60	Substantial
0.60 > ICC ≥ 0.40	Moderate
0.40 > ICC ≥ 0	Poor

2.5. The resting state

Throughout this work fMRI data is acquired from subjects without a task paradigm, also known as the resting state. Even though requiring subjects to perform a task during image acquisition has been the obvious approach for the longest time. Biswal and colleagues were the first to observe that during rest, the left and right hemispheric regions of the primary motor cortex exhibit a high degree of high correlation of the BOLD fluctuations [14]. Aside from establishing a new way of interpreting BOLD fMRI data by means of correlating (groups of) voxels, the research also laid the foundation of analysing the resting state. As this result suggested that even during rest, the brain continues process information and exhibits functional connectivity [18][48]. Further support in favour of a neuronal basis of resting state fMRI comes from observing energy consumption in the brain. As mentioned in the introduction: the brain requires 20% of all energy uptake to function in the resting state, and an additional 5% is necessary to perform certain cognitive task paradigms. [1][2] If only the effects of the cognitive task paradigm are of interest, you effectively consider 80% of your fMRI signal to be noise. As the resting state accounts for the biggest energy consumption, analysing the function is of the resting state is essential for understanding neuroanatomy.

One ICN specific to the resting state is the default mode network (DMN). The network originally referred to as a group of brain regions that include the hippocampus, posterior cingulate, medial prefrontal and inferior parietal cortex [49]. The DMN is specific to the resting state because the network experiences reduced activity during cognitive task activation [50], while experiencing enhanced metabolism during internal cognitive processes [51]. The function of the DMN has been related to mind wandering [52], episodic memory [53], personality [54] and conceiving the perspective of others (i.e. mentalizing) [55]. In addition, changes in the DMN have been observed in ageing subjects [56] and in neurodegenerative diseases such as Alzheimer's disease [57].

Aside from determining whether CM's are reproducible on the group level, it is also important to determine if inferences of the individual subject can be formed based on their CM. In a study by Finn [58] this was the central question. In the study, a group of 126 subjects were scanned twice with a two-day interval. The Pearson's correlation coefficient was computed of each CM of one day to all other CM's of the other day. The CM that gave the highest correlation was labelled as a 'match'. Among the results of the study, the highest matching accuracies (MA's) were found for the matching tests of the resting state scans (as there were also a handful of task paradigms) with an accuracy of 93.7%. The conclusion of the study was that it is plausible to draw inferences about single subjects based on their functional connectivity profile. An additional consideration postulated in the work was that the MA test was performed on scans of subjects with only two days apart, but it remained unclear is to what degree these individual connectivity profiles are consistent over bigger periods of time, as it is known that the functional connectivity decreases with age [59].

2.6. Research motivation

Numerous studies have been performed on how the functional connectivity profile changes with varying pathologies or conditions, an overview of which is shown in Table 2. Mapping functional connectivity through resting state fMRI is rapidly developing, and the technique has already shown potential for clinical applications. Before successful implementation of rs-fMRI in the clinical practice, it is essential to assess the long-term reproducibility of the functional connectivity profiles. Therefore, the reproducibility of different components of the CM's need to be quantified, as well as the overall reproducibility between successive scans. The robustness of the reproducibility study is assessed by applying the results of a reproducibility study on a test set. Aside from looking at the reproducibility of the data, the distinguishing power of the CM's is also of interest. To this end a MA test is performed along with the reproducibility study. The resulting research question is two-fold, namely:

Can the reproducibility of functional connectivity profile be improved by selectively eliminating spatially dependant components from the connectivity matrices?

Can the order in which the selective component elimination is performed lead to improved generalized reproducibility for a longitudinal test set?

Table 2 Group differences in rs-fMRI patterns observed in various brain diseases or conditions. Salience network: includes regions in the dorsal anterior cingulate and bilateral fronto/insular cortices; dACC = dorsal anterior cingulated cortex; PIB = PIB = Pittsburg compound B, a marker of amyloid plaque accumulation in the brain. PTSD = post-traumatic stress disorder; ALS = amyotrophic lateral sclerosis; ADHD = attention deficit hyperactivity disorder.

Pathology/condition	**Findings**
Alzheimer's	Decreased correlations within the DMN including hippocampi, decreased anticorrelations with the DMN, and reduced local connectivity as reflected in clustering coefficients
PIB positive	Decreased correlations within the DMN
Mild cognitive impairment	Decreased correlations within the DMN and decreased anticorrelations with the DMN.
Fronto-temporal dementia	Decreased correlations within the salience network
Healthy aging	Decreased correlations within the DMN
Multiple sclerosis	Decreased correlations within the somatomotor network
ALS	Decreased connectivity within the DMN and within the somatomotor network (esp. premotor cortex)
Depression	Variable: Decreased corticolimbic connectivity (esp. with dorsal anterior cingulate), increased connectivity within the DMN (esp. subgenual prefrontal cortex), decreased connectivity between DMN and caudate
Bipolar	Decreased corticolimbic connectivity
PTSD	Decreased connectivity within the DMN
Schizophrenia	Variable: Decreased or increased correlations within the DMN. Decreased, increased or unchanged correlations and anticorrelations between the DMN and other systems.
Schizophrenia 1° relatives	Increased connectivity within the DMN
ADHD	Variable: reduced connectivity within the DMN, reduced anticorrelations with the DMN, increased connectivity in the salience network
Autism	Decreased connectivity within the DMN (although hippocampus is variable, and connectivity may be increased in younger patients)
Tourette syndrome	Delayed maturation of task-control and cingulo-opercular networks
Epilepsy	Variable: decreased connectivity in multiple networks including the medial temporal lobe, decreased connectivity within the DMN (esp. in patients with generalized seizures)
Blindness	Decreased connectivity within the visual cortices and between visual cortices and other sensory and multimodal regions
Chronic pain	Variable: Increased/decreased connectivity within the salience network, decreased connectivity in attention networks
Neglect	Decreased connectivity within the dorsal and ventral attention networks
Coma/vegetative state	Progressively decreased DMN connectivity with progressive states of impaired consciousness
Generalized anxiety disorder	increased connectivity between amygdala and frontoparietal control network and decreased connectivity between amygdala and salience network

3. Methods & Materials

CM's will be used as the representation of the functional connectivity profiles and are used in all quantitative analyses. Firstly, the data acquisition will be discussed and what programs are used for data analyses. Next, the different data sets that are used are briefly discussed as well as the metrics. A reproducibility study is carried out to answer the first research question, and next the longitudinal set is analysed, using parts of the component order as found in the reproducibility study to answer the second research question.

3.1. Data acquisition and means of analysis

All computation required for this work has been carried out on MATLAB 2015b and 2017b. The CM's that are required have been obtained through the Erasmus Medical Centre in Rotterdam, which is leading the Rotterdam Scan Study (RSS). The RSS is initiated as a part of the Rotterdam study with the goal to gain a greater insight into neurological diseases by obtaining data in a prospective population study [61]. Because of the sensitive nature of this data, it is required to do all computation with the CM's in the Erasmus MC. The data was accessed on the cluster of the Erasmus MC through PuTTY and WinSCP.

A 1.5T MRI unit (General Electric Healthcare, Milwaukee, USA, software version 11x) is used as part of the ongoing Rotterdam Scan Study. The unit is fitted with an 8-channel head coil. BOLD rs-fMRI data is required for this work. An overview of the MRI protocol used is given in Table 3. All studies make use of the echo planar imaging (EPI) scanning protocol. EPI is a type of gradient echo sequence, where multiple spin echoes are acquired by using rephasing gradients. Here, the frequency encoding gradient is rapidly reversed, causing multiple gradient echo's drastically reducing imaging time compared to traditional MR imaging sequences. This technique allows for rapid imaging which is required to image the BOLD fluctuations.

Table 3 rs-fMRI acquisition protocol used in the Rotterdam Scan Study.

Readout module	**Time (min:sec)**	**TR/TE (ms)**	**BW (kHz)**	**Flip Angle (degrees)**	**Number of slices**	**Slice thickness (mm)**	**FOV (cm^2)**	**Matrix**
EPI	7:44	2900/66	7.81	90	31	3.3	21	64 x 64

Pre-processing is performed at the Erasmus MC by use of the FMRIB software library (FSL). FSL is a software library used for the image analysis and statistical analysis of functional, structural and diffusion MRI data. Registration of the rs-fMRI volumes to the subjects structural and standard space is carried out by FNIRT. ICA is used to decompose the fMRI data into various ICN's. To distinguish noise components from actual components, Xnoisefier (FIX) is used. MCFLIRT and temporal filtering is used to account for low frequency drift and motion components.

Each subject that had taken part in the Rotterdam Scan Study was scanned twice with varying interscan times. Four different parcellation schemes are applied to the scans: The RSS100 scheme, the UKBiobank scheme, the HCP820 d100 scheme and the HCP820 d200 scheme. The RSS scheme was developed in-house at the Erasmus MC, as part of the Rotterdam scan study. The two HCP820 schemes are provided by the Human Connectome Project, an international initiative which aims to gain greater insight into the workings of the human brain. The UKBiobank scheme has been obtained from the UK Biobank, which is a large biobank based in the UK that is involved in various epidemiologic studies.

3.2. Data sets and metrics

The pre-processing and parcellation procedures ultimately result in a CM for every single scan. The CM's form the data on which all further analyses are performed on. Datasets from 2 cohorts of the RSS are used: The Reproducibility (Repro) set and the longitudinal set. Both sets consist of 2 scans for every subject. The Repro set is a relatively small set, consisting of 37 subjects with an average interscan time of 2 weeks. This set is ideal for reproducibility studies due to the small interscan times. The longitudinal set is a larger set of 521 subjects, whose subjects have more varied interscan times. These interscan times vary from 3 months to slightly over 5 years.

All the elements in the CM's contain Pearson's correlation coefficients between the two respective nodes. The correlation of two identical time signal is 1, so the diagonal of the CM will therefore always consist solely of ones. As the correlation between two nodes is the same in both directions, it follows that the CM's are symmetrical. This entails that the number of unique elements is described by Equation (2).

$$E_u = \frac{N(N-1)}{2} \tag{2}$$

With E_u the number of unique elements and N the number of nodes of the respective parcellation scheme. When performing computations involving elements, only the unique elements are used. This reduces computation times significantly. The number of elements, nodes and unique elements for all parcellation scheme are given in Table 4.

Table 4 Overview of the dimensions of the used parcellation schemes. The number of elements is equal to the square of the number of nodes. The number of unique elements is equal to half the number of nodes multiplied to the number of nodes minus 1.

Parcellation scheme	Number of nodes	Number of elements	Number of unique elements
RSS100	50	2500	1225
UKBiobank	53	2809	1378
HCP820 d100	58	3364	1653
HCP820 d200	97	9409	4656

The components of the CM as defined here, can either entail whole nodes or individual elements. Quantifying the reproducibility of these two requires a small reformatting procedure on the dataset. The ICC is computed for a set of data, generally called a target. The way the targets are formatted thus determines what ICC is computed. The quantification of the reproducibility of a single element of a CM is referred to as an inter-subject element ICC. The reformatting procedure required to compute the inter-subject element ICC is illustrated in Figure 4.

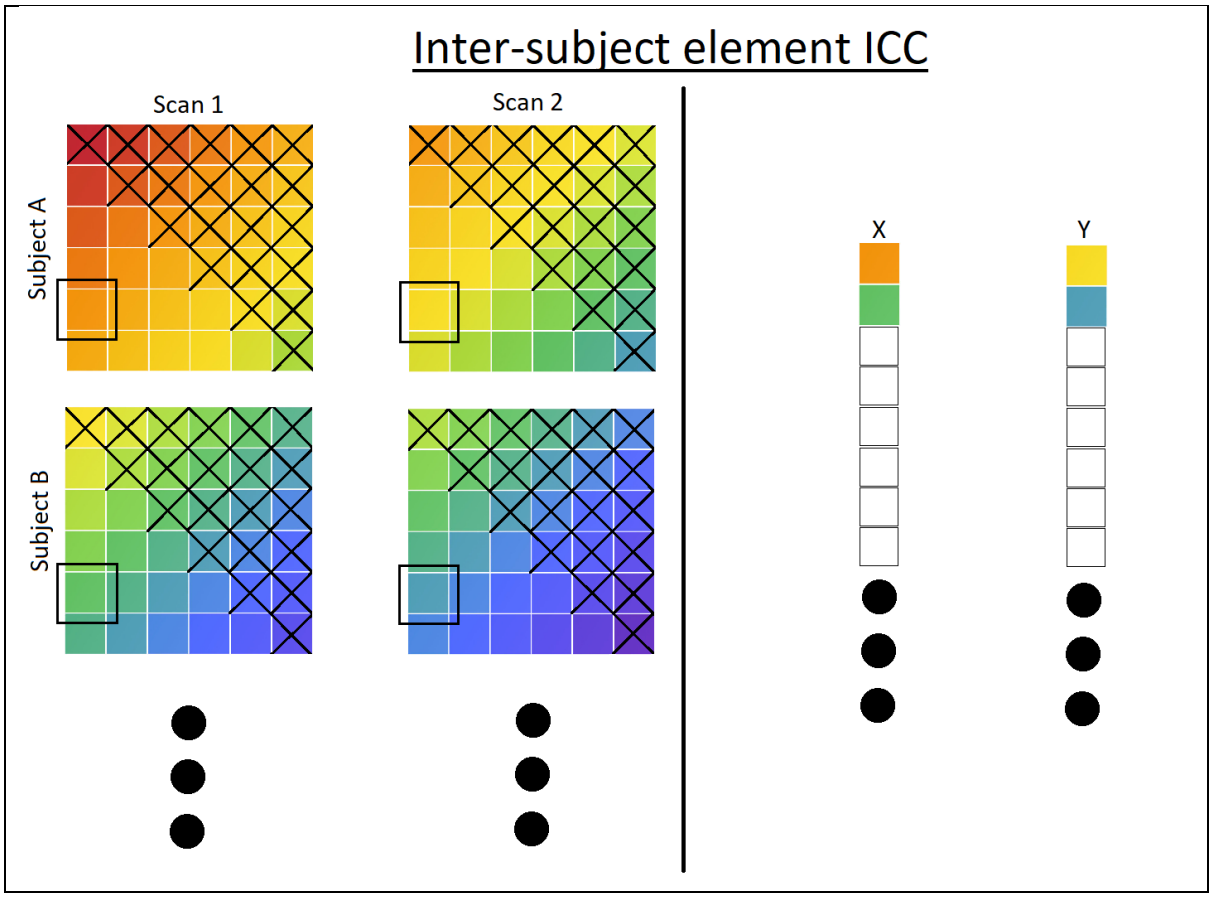


Figure 4 An illustration of the inter-subject element ICC. All elements 5,1 of the first scanning round are concatenated in target X, and the same elements of the second scanning round are concatenated in target Y. The ICC of targets X and Y quantifies the reproducibility of a single element as seen in the dataset seen in the left-hand side.

The left-hand side of Figure 4 shows a dataset of CM's with 2 scans per subject. As only the reproducibility of the unique elements are required, the top half of all CM's are omitted from the inter-subject element ICC calculation. The diagonal of the CM's contains only ones, therefore these are also not considered to be unique elements. Figure 4 only shows the concatenation of element 5,1, but when the inter-subject element ICC is computed for all elements in the set, the reproducibility of all the single elements is found.

To quantify the reproducibility of all nodes in a set of CM's, the inter-subject node ICC is performed. Again, targets X and Y need to be defined in a slightly different as seen in Figure 4. This time target X consists of a concatenation of a single node from the CM's of the first scan round, and target Y consists of a concatenation of a single node from all CM's of the second scan round. Calculating the ICC on targets X and Y as shown on the right-hand side of Figure 5, results in a quantification of the reproducibility of that specific node.

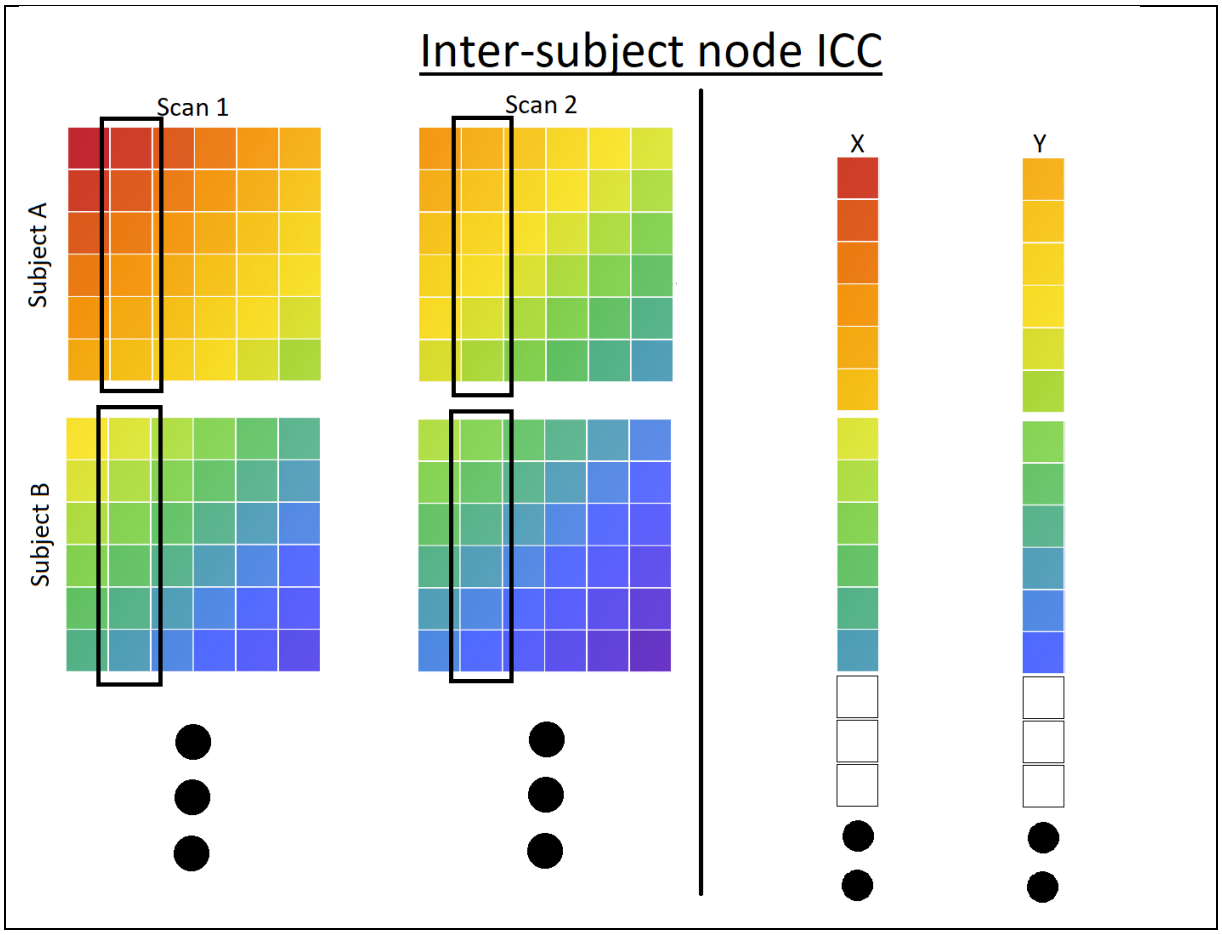


Figure 5 An illustration of the inter-subject node ICC. On the left-hand side of the picture, node 2 of all subjects is selected. All the nodes from the first scan round are concatenated in target X, and all nodes from the second scan round in target Y. Calculating the ICC on targets X and Y results in a quantification of the reproducibility of node 2 as seen in the dataset on the left-hand side.

As will be explained in Section 3.3, it will be necessary to remove nodes from the set of CM's. When a node is removed, this results in the deletion of a row and column. When a node is removed in this way, all other nodes will also be affected because they will lose the specific connection to that node. What follows, is that the inter-subject node ICC will also be different after that node is removed. Therefore, the inter-subject node ICC will have to be recalculated for every removed node in order to determine the least reproducible node in the set.

The main metrics that is used throughout this work to quantify the overall reproducibility is the intra-subject ICC. Calculating the overall reproducibility by means of the intra-subject ICC requires minimal reformatting of the data. With the intra-subject ICC, targets X and Y consist of successive scans of the same subject.

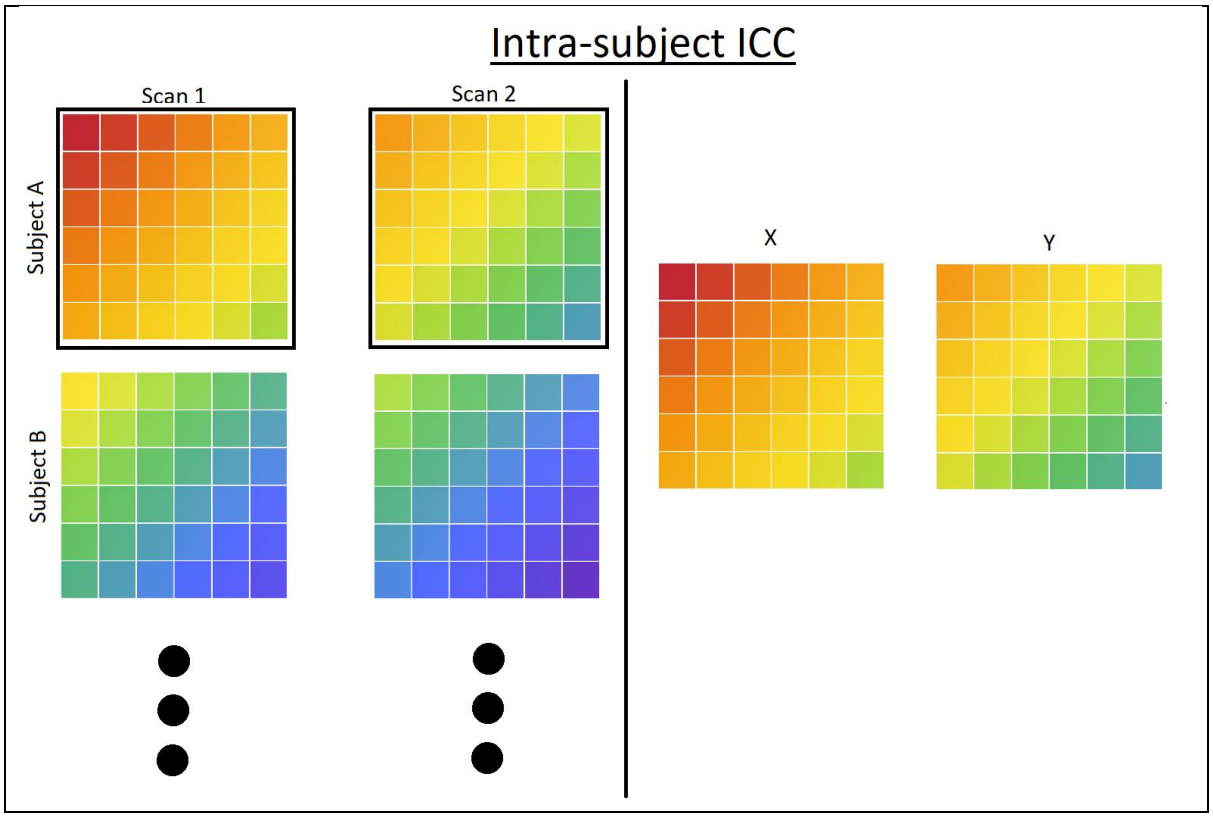


Figure 6 An illustration of the intra-subject ICC. 2 successive scans of the same subject are taken, one scan is labelled target X and the other target Y. The ICC is computed on targets X and Y, thus producing the reproducibility of the FC profile of a single subject.

The intra-subject ICC is thus determined for every subject in the respective dataset. The intra-subject ICC's as shown in the result section are the means of the intra-subject ICC's. The error bars give the associated standard error of the mean, as shown in Equation (3).

$$\sigma_{ICC} = \frac{s_{ICC}}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(ICC_i - \overline{ICC})^2} \tag{3}$$

Where n represents the number of subjects in the respective set, ICC_i the intra-subject ICC of the subjects and $\overline{ICC}$ the average intra-subject ICC.

Another metric used is the MA. When a MA test is performed the subject ID's are removed from one of the two scan sessions (either from all subjects from scan 1 or from scan 2). One of the unknown subjects' CM is taken, and the Pearson's correlation coefficient between the unknown CM and the CM's of the other scan session are calculated. Pearson's correlation coefficient is shown in Equation (4).

$$r = \frac{\sum_{i=1}^{L}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{L}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{L}(x_i - \bar{x})^2}} \quad (4)$$

Where x_i are all elements of the CM of the first scan round, and y_i all elements of the CM of the second scan round. $\bar{x}$ and $\bar{y}$ are the averages of all elements in their respective CM. The summation goes to L, the number of elements in the CM's. After quantifying the correlation between all CM's, the unknown subject is matched to the subject ID whose CM gave the highest correlation coefficient.

The result of the MA test is a percentage of subjects whose scans were rightfully matched. The test is performed twice, where during the second test the subject ID's from the other scan session is removed. The average matching accuracy is then the average of the two rightfully matched subjects. In this way, only a single MA percentage is found for an entire set.

3.3. Reproducibility study

The first research question requires selective data elimination from the CM's to see how the overall (intra-subject) reproducibility is affected. This is where the difference between the intra-subject ICC and the inter-subject node and element ICC becomes apparent: The two versions of the inter-subject ICC allows for the quantification of the components of the CM's of the set, whereas the overall reproducibility is quantified by the intra-subject ICC. A distinction is made between a node removal approach and an element removal approach.

In the node removal algorithm, the inter-subject node ICC is computed and the node with the lowest inter-subject ICC score is eliminated from the CM's of the dataset. Elimination of the node entails eliminating the specific row and column of that node from all CM's in the dataset.

After the node is successfully removed from the dataset, both the intra-subject ICCs are calculated, and a MA test is performed on the set. As the elimination of the node also removed an element of that node in all other nodes, the inter-subject node ICC needs to be computed after every node elimination. This process is repeated until the CM's only consist of a single node, after which the intra-subject ICC can't be computed anymore, because there is no within target variation to be quantified on a single value. The order of the nodes that have been removed is saved in vector Q_n.

The element removal algorithm works similarly to the node removal algorithm. The clear distinction being, that single elements are removed as opposed to whole nodes. Quantification of the reproducibility of the elements is performed by calculating the inter-subject element ICC for all elements of the CM's in the set. As the removal of a single element from the CM's in the set does not affect the inter-subject element ICC of the other remaining elements, the inter-subject element ICC only has to be computed once. The order of the elements to be removed in determined from the inter-subject element ICC's, and after the removal of an element the intra-subject ICC is computed, and the MA test is performed. This process continues until there is only a single element left in the CM's, after which the intra-subject ICC becomes undefined, just like in the node removal algorithm. The order in which the elements of the CM's is removed is saved to vector Q_e. An overview of the component removal reproducibility study algorithm is shown in Figure 7.

The number of iteration that need to be performed before the CM's are reduced till the point that the intra-subject ICC can no longer be performed, is determined by the number of unique elements or nodes in that parcellation scheme (as shown in Table 2). To facilitate the comparison between the different parcellation schemes, the number of elements or nodes removed is represented in the result section as a fraction of the total number of elements or nodes in that parcellation scheme.

The outcome of the element or node removal algorithm is thus a function of the fraction of elements or nodes removed, which is always a value between 0 and 1.

This reproducibility study is carried out on the Repro set. The Repro set is the most obvious choice, as the subjects are scanned with the shortest interscan time. Therefore, this set should give the most apparent overview of the reproducibility of the nodes and elements that form the CM's.

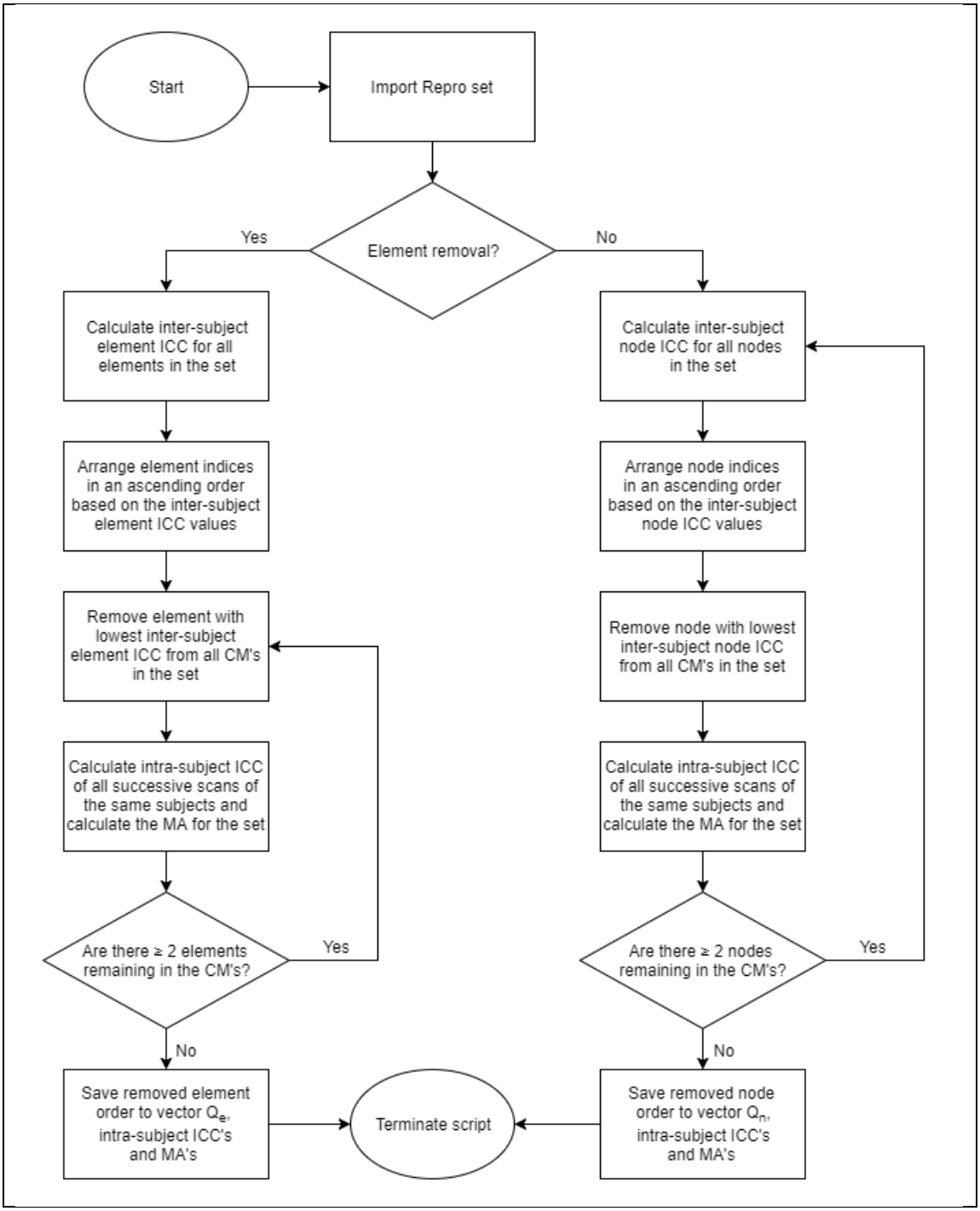


Figure 7 Flowcharts of the algorithm used in the component removal reproducibility study

3.4. Scaling and error estimation of the matching accuracy

There are two issues with the MA that need to be addressed before starting the longitudinal set analyses.

The first issue has to do with error quantification: as the MA is calculated for the entire set, only a single value is found for every iteration of the node- or element removal algorithm. The node- or element removal order will not change after completing the reproducibility study, therefore repetitions of the script will give identical MA outcomes. Therefore, it is required to make an estimate of the errors found in with the MA test.

The second issue that becomes apparent has to do with comparing MA's of subsets with varied sizes. Because the MA is based on the correlation between different CM's, a larger subset will have more potential matches whose difference in correlation coefficients become smaller with increasing subset size. The MA's of a smaller subset will therefore give higher outcomes overall. Therefore, scaling matrices are employed to compare MA's of varied subset sizes. A resampling technique is employed to address both issues.

A random training sample (with overlap) of the scans of 37 subjects from the longitudinal set are taken, on which the inter-subject element ICC is performed. The element order Q_e is derived from the inter-subject element ICC and is applied to a random test sample (with overlap) from the longitudinal set. The number of subjects in the test set is varies, consisting of 16, 22, 31, 37, 59, 78, 118 and 470 subjects. The number of subjects in the test set is based on subset sizes used in the longitudinal set analyses. 11 fixed fractions of the elements were removed from the test set, at every iteration the MA test was performed on the set. These 11 removed fractions of elements were 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1 of the total amount of unique elements in the parcellation scheme. This resampling procedure is repeated 250 times. By resampling both the training- and test sets, the element order Q_e will vary with every iteration. The means and the standard deviation is computed for the resulting MA's.

To find a scaling factor to compare the MA's of different subset sizes, all mean MA's as found in the resampling technique are divided by one another for every fraction of elements removed. The resulting 11 scaling factor are averaged again, as to form a single scaling factor that can be used to scale an entire MA curve. The scaling factors are computed for every subset size combination and are displayed in a scaling matrix. Every element in the scaling matrix is used as a scaling factor to go from the subset size as shown in the rows, to the subset size as represented in the column.

The standard deviation is used as an estimate for the variance of the population of the various subset sizes. As every other analysis takes a single sample from the population, the standard deviation as found in the method as described here serves as an estimation of the standard error of the mean, see Equation (3). These estimated errors are used as indications of the variation throughout the longitudinal set analyses.

3.5. longitudinal set analyses

The second research question was concerned whether it is possible to improve the overall reproducibility by removing the components whose inter-subject ICC is anticipated to be minimal from the reproducibility study. The anticipated minimally reproducible nodes or elements are ordered in the vectors Q_n and Q_e, respectively. The order of Q is applied to the longitudinal dataset, computing the intra-subject ICC of the subjects in the set and the MA for the entire set at every iteration.

Several subgroups were formed from the longitudinal set on which the order of Q is applied to see whether the reproducibility of different subgroups is affected differently. To this end, four different subgroup categories were analysed, based on: grounds for exclusion, sex, interscan time and age.

After fMRI image acquisition, the data is pre-processed as described earlier. The quality of the fMRI scans is also determined at this stage. During the quality control stage, certain factors may come to light that impair image quality. Head motion during image acquisition is a commonly occurring problem. A threshold of maximum head motion is defined, and when this threshold is exceeded the subject's scans given the 'Motion' label. When a type of pathology is discovered during quality control stage, such as mild cognitive impairment or early symptoms of Alzheimer's disease, the subject's scans are given the 'Pathology' label. These pathologies could hamper the quality of the analyses performed on the CM's of the set, in the same way that the subjects with the 'Motion' labels do. When the scan is of low quality without a clear cause, the scans are given the label 'Unclear'. If none of these three grounds for exclusion are found on a scan, they are given the 'Included' label, as these scans can are of sufficient quality and taken from healthy subjects. The collection of these labels is referred to as the exclusion parameters. The total number of subjects with each of the exclusion parameters that are contained in the longitudinal set is given in Table 5.

Table 5 Overview of exclusion parameters as found in the subjects of the longitudinal set.

Parameter	Number of subjects
Included	470
Pathology	16
Motion	22
Unclear	13
Total	521

Both the Pathology subgroup and the Motion subgroup are compared to the included group, to quantify the extent of the inclusion parameters on the reproducibility of the CM's. In all following analyses of the longitudinal set, only the scans of the subjects that have the Included label are used.

The second subset to be analysed is based on the sex of the subjects: The scans with the Included label are split in an all-male subgroup and an all-female subgroup. The all-male subgroup consists of 221 subjects, and the all-female subgroup consists of 249 subjects.

The third subset of the longitudinal set to be analysed is based on the interscan time of the scans. The included subjects are divided into four subgroups of nearly equal size: two of 117 subjects and 2 of 118 subjects. By constructing the IST subgroups in such a way that they almost have the same number of subject, the comparison of the MA's of the subgroups with one another becomes more reliable. A consequence of constructing the subgroups in this way, is that the IST range of the subsets are not equal with one another.

The fourth and final subset of the longitudinal set is based on the age of the subjects during the first scan. Just like with the IST based subgroup, 4 subgroups are constructed: two of 117 subjects and 2 of 118 subgroups.

All subsets were analysed for the four parcellation schemes, for both the node removal approach and the element removal approach, the order of which determined by the vector Q derived from the reproducibility study. An overview of the longitudinal set analysis is shown in Figure 8.

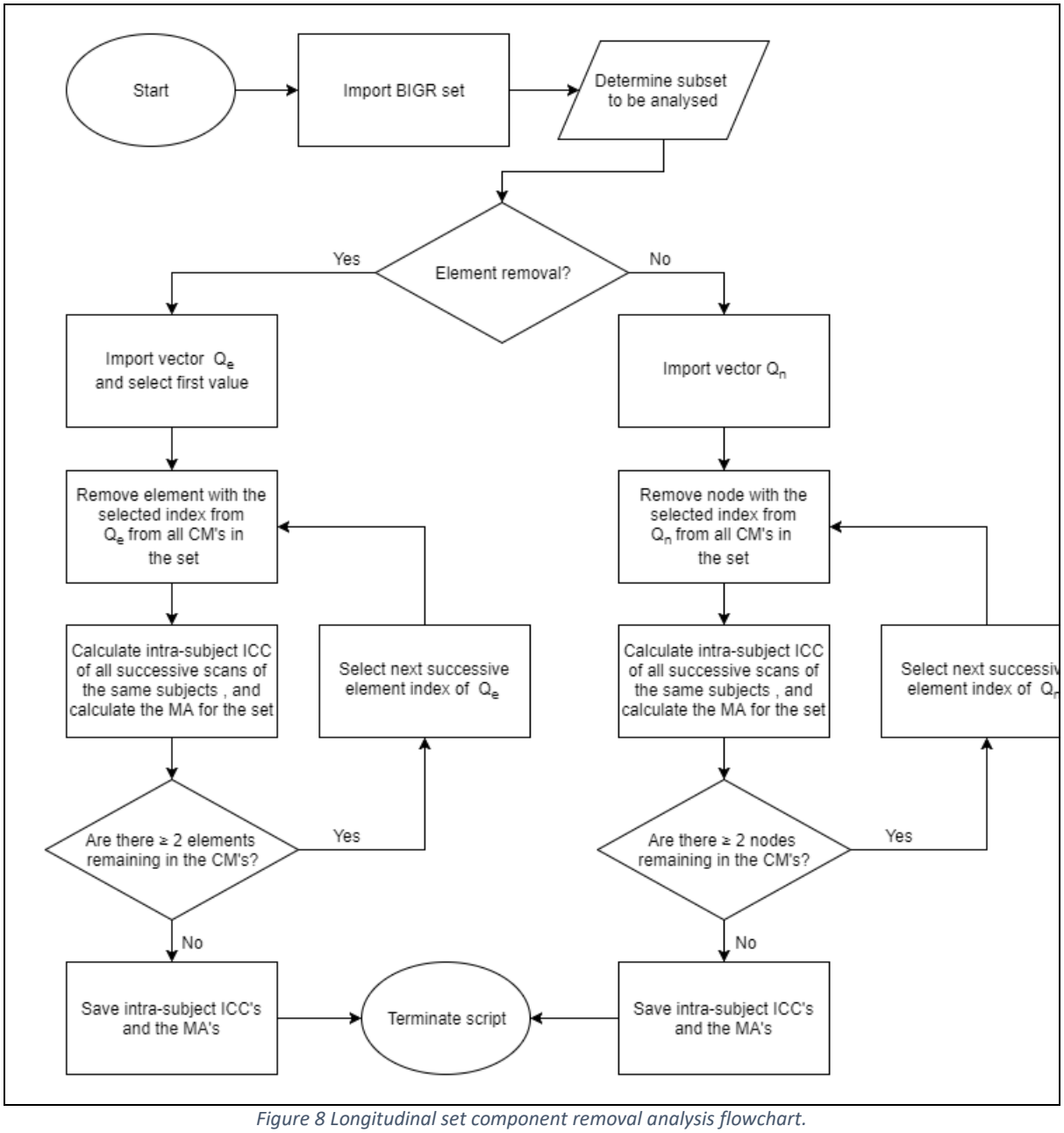


Figure 8 Longitudinal set component removal analysis flowchart.

4. Results

The average intra-subject ICC and MA of the whole CM's are computed and serve as a base of comparison, see Table 6.

Table 6 Average Intra-subject ICC and MA test of the unaltered repro set of four different parcellation schemes

Parcellation scheme	Average ICC ± σ_d	$MA_{1\rightarrow2}$ (%)	$MA_{2\rightarrow1}$ (%)	MA_{avg} (%)
RSS100	0.59 ± 0.12	81.1	83.8	82.4
UKBiobank	0.59 ± 0.11	89.2	75.7	82.4
HCP820 d100	0.53 ± 0.11	73.0	73.0	73.0
HCP820 d200	0.49 ± 0.10	78.4	86.5	82.4

The average MA is the mean of the MA of scan 1 to 2, and scan 2 to 1. When talking about the MA, the average MA is meant. Taking Table 6 as a starting point, the following section will show how the intra-subject ICC and MA will be affected when different components of the CM will be removed from the data.

4.1. Reproducibility study – Element removal

For the reproducibility study, the algorithm as seen in figure 7 is used. Starting off the with element removal approach, the inter-subject element ICC's are computed first. A histogram of the inter-subject element ICC is shown of parcellation scheme RSS100 in Figure 9.

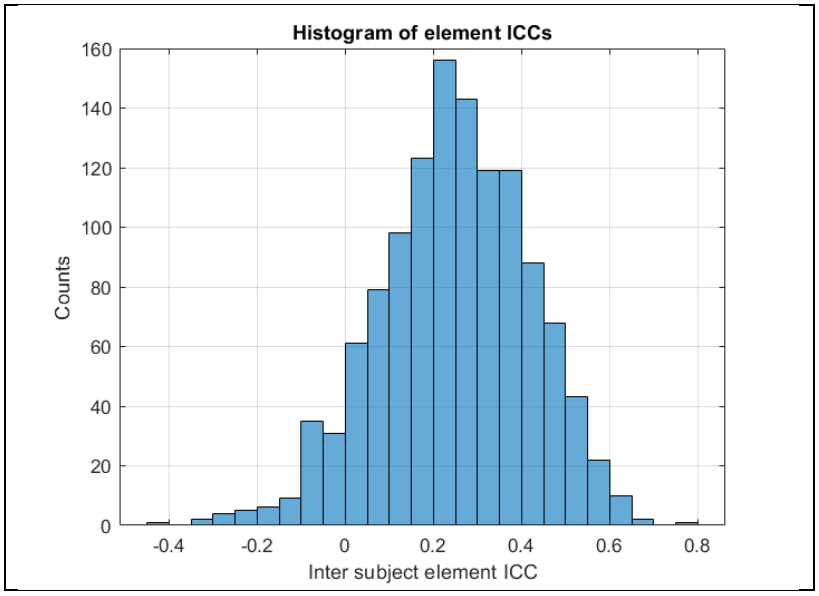


Figure 9 Histogram of the inter-subject element ICC for parcellation scheme RSS100

Note that the average inter-subject element ICC of 0.25 is smaller than the average intra-subject ICC of RSS100 as found in table 2. Additionally, multiple negative inter-subject element ICC's are found. The inter-subject elements ICC's are sorted in an ascending order, and the indices of the elements are saved in vector Q_e. Elements are removed in the order dictated by Q_e, at each iteration the average intra-subject ICC is calculated and the MA of the set. Figure 10 shows these results for all 4 parcellation schemes.

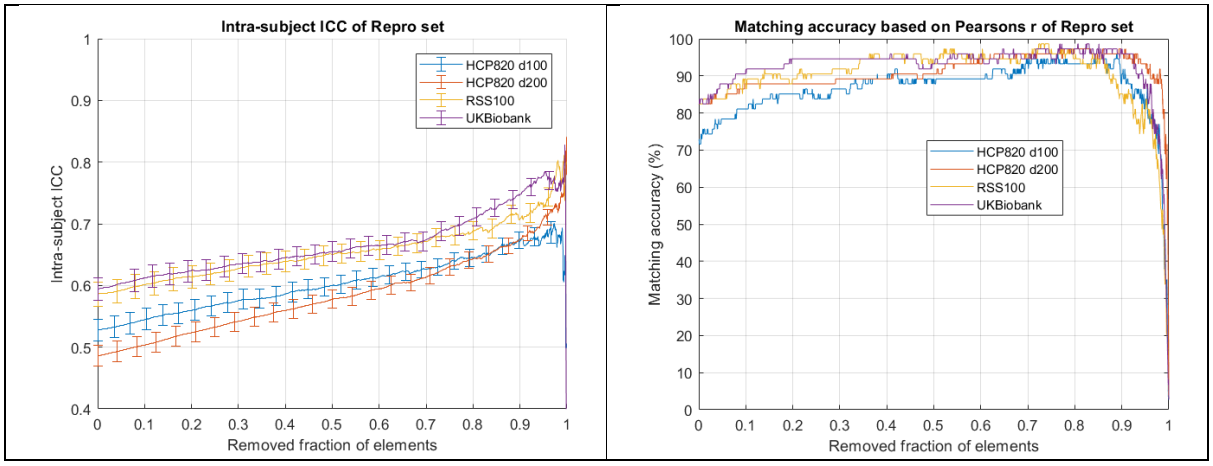


Figure 10 (left) Intra-subject ICC of the repro set for the four different parcellation schemes along with the standard error of the mean. (right) MA of the repro set for the four different parcellation schemes as a function of the fraction of removed elements.

Figure 10 shows that both the MA and the intra-subject ICC increase for the range of 0 to 0.8 removed fraction of elements. For all parcellation schemes, the ICC seems to increase at least 0.1 until a removed fraction of 0.9 of the elements. Between a removed fraction of elements of 0.9 to 1 it is seen that the ICC graph becomes rather erratic, before dropping to zero. As more elements are removed, the influence of single elements on the reproducibility of the remaining data increases, which explains the erratic behaviour.

With the MA curves, the overall matching accuracy of all parcellation schemes increases compared to the baseline as seen in Table 6. The highest MA's found for every parcellation scheme are: 98.7 % for the RSS100 and UKBiobank parcellation, 97.3% for HCP820 d200 and 96.0% for HCP820 d100. Just like with the ICC graphs, it is seen that the MA goes to zero when the removed fraction of nodes goes to 1. When the CM's are completely empty it becomes impossible to match corresponding scans to one another.

4.2. Reproducibility study – Node removal

The reproducibility study built upon node removal is comparable to the element removal procedure. The baseline values as found in table 2 are also the same. The clear difference being that nodes are removed iteratively, as opposed to single elements. A difference is that the inter-subject ICCs have to be computed at every step when a node is removed from the CM's, as the elements that are part of the removed node are still part of the remaining nodes. This can be seen when making a histogram of the inter-subject node ICC at different numbers of removed nodes: not only are the lowest nodes removed from the histogram, it can also be seen that the number of counts of nodes with a high inter-subject node ICC, is increased, as shown in Figure 11.

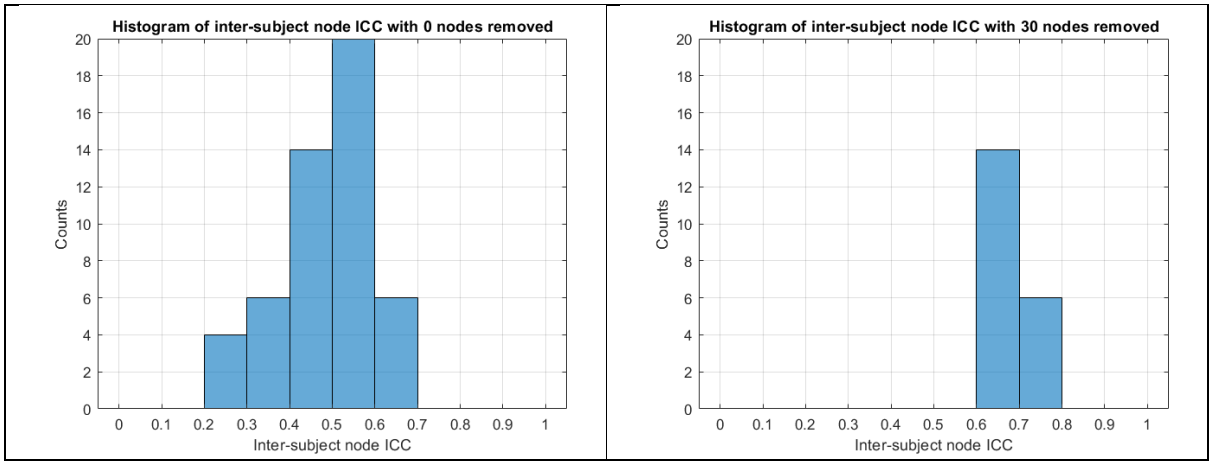


Figure 11 Histogram of inter-subject ICCs of the RSS100 scheme in the Repro set. Left shows the histogram when zero nodes are removed, and the right shows the histogram when 30 nodes are removed. The number of nodes in the bin of 0.65 has increased from 6 to 14 counts.

The order of the node indices that are removed from the Repro set is saved in vector Q_n, for the analysis of the longitudinal dataset. When removing the nodes from the CM's whose inter-subject node ICC values were the lowest of the remaining nodes, an intra-subject ICC and a MA test was performed at each iteration. The results are shown in Figure 12.

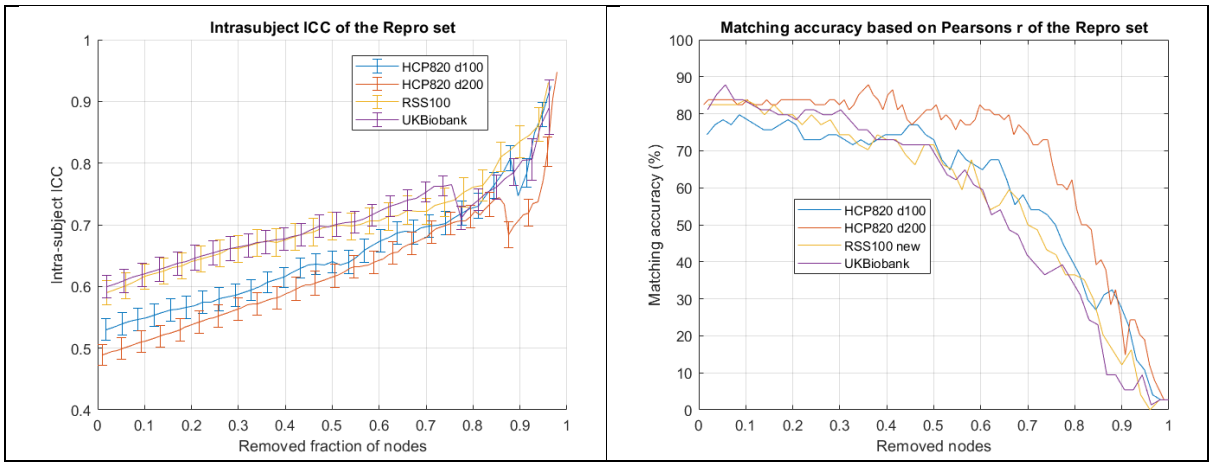


Figure 12 (left) Intra-subject ICC of the repro set for the four different parcellation schemes along with the standard error of the mean. (right) MA of the repro set for the four different parcellation schemes as a function of the fraction of removed nodes.

The left graph of Figure 12 shows the intra-subject ICC curves. All these curves follow an upward trend similar to its element removal counterpart as shown in Figure 10. The clear difference being that these curves are steeper than those of figure 10 for all parcellation schemes, thus a similar fraction of the total data removed yields higher reproducibility when whole nodes are removed.

The MA curves as depicted in the right of Figure 12 vary from that of its element removal counterpart: Whereas there is a clear upward trend to be seen in Figure 10, this is absent in Figure 12. One might argue that for the first fraction 0.25 nodes removed, the MA is somewhat stable. But a clear improvement as seen in the element removal MA curves is simply not present. It appears counter-intuitive that the intra-subject ICC increases faster while the MA goes down.

4.3. Matching accuracy scaling and error estimation

The error of the MA test is estimated by resampling the longitudinal set. To estimate the error of the MA for various subset sizes, a random training set of 37 subjects was sampled from the longitudinal set on which the inter-subject element ICC determined the element order to be removed. This order of elements to be removed is applied to a test set of sizes 16, 22, 31, 37, 59, 78, 118 and 470 subjects sampled from the longitudinal set, which was repeated for 250 times. Table 7 shows the standard deviation of RSS100 that serve as an estimate for the error for the MA's for the various subset sizes and for a certain fraction of elements removed. The error estimates of the other three parcellation schemes are shown in Appendix B.

Table 7 standard deviation found on the resampled test set of RSS100. The rows represent different set sizes, whereas the columns represent different fraction of elements removed

RSS100	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
16	8.76	8.75	8.73	8.73	8.71	8.86	8.72	8.82	8.88	9.74
22	7.55	7.52	7.32	7.53	7.51	7.61	7.69	7.68	7.91	8.83
31	6.32	6.30	6.17	6.23	6.32	6.45	6.32	6.43	6.67	7.02
37	6.27	6.26	6.26	6.31	6.36	6.35	6.48	6.43	6.64	6.65
59	4.96	4.90	4.92	4.91	4.96	5.07	5.07	4.92	5.08	5.36
78	4.33	4.34	4.15	4.18	4.23	4.30	4.36	4.38	4.36	4.58
118	3.56	3.63	3.64	3.70	3.69	3.66	3.69	3.72	3.8	3.99
470	1.54	1.53	1.50	1.44	1.50	1.39	1.48	1.46	1.56	1.48

For the construction of the scaling matrices, the means of the resampled MA's were used. The MA curves for various set sizes are divided by one another for every fraction of elements removed, after which an average is taken of all 10 fractions of elements removed. This result is used as a scaling factor used to compare MA's of different set sizes. Table 8 shows the scaling matrix used to compare the MA's of different subset sizes for various sub set sizes of the RSS100 parcellation scheme. The scaling matrices of the other three parcellation schemes are shown in Appendix C.

Table 8 MA scaling matrix of parcellation scheme RSS100 from the longitudinal set

Set size	16	22	31	37	59	78	118	470
16	1	0.96	0.82	0.88	0.83	0.79	0.72	0.45
22	1.05	1	0.96	0.92	0.87	0.82	0.75	0.47
31	1.09	1.04	1	0.96	0.90	0.86	0.78	0.49
37	1.13	1.08	1.04	1	0.94	0.89	0.81	0.51
59	1.21	1.16	1.11	1.07	1	0.95	0.87	0.54
78	1.27	1.21	1.16	1.12	1.05	1	0.91	0.57
118	1.40	1.34	1.28	1.23	1.16	1.10	1	0.63
470	2.23	2.13	2.04	1.97	1.84	1.75	1.59	1

To give a small example for the usage of the scaling matrix: to compare the MA's of two subsets, one with 16 subjects, and the other with 470 subjects. Multiply the 470-subject curve with 2.23 to compare it to the 16-subject MA curve.

4.4. Longitudinal set subset analysis

For the analysis of the various subsets of the longitudinal set, the algorithm as described in Figure 8 was employed. Several subsets of the longitudinal set are explored in this section, for both the element removal and node removal processes. First, the subsets based on the grounds of exclusion are investigated, where both the Pathology subset and the Motion subset will be compared to the Included subset. Figure 13 shows the intra-subject ICC and the MA's of Included subset and the Motion subset. A corrected motion subset curve is also shown for the MA curves to compare it to the Included subset. Additional results of the Motion subset analysis of the other three parcellation schemes are found in appendix D.

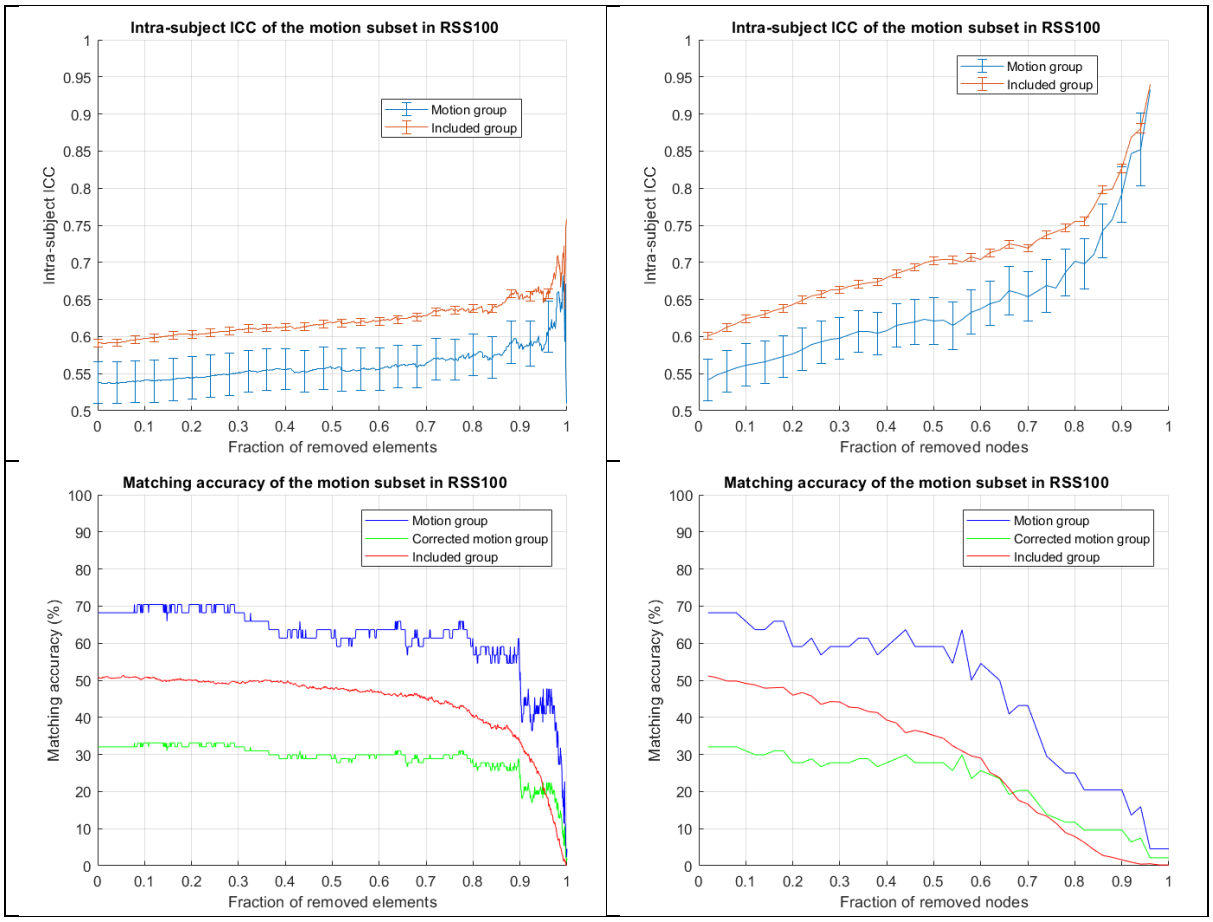


Figure 13 Motion subset analysis of element removal on RSS100 parcellation scheme. The errors as shown in the intra-subject ICC curves are the standard errors of the mean. (Top right) Intra-subject ICC of element removal (Top right) Intra-subject ICC as a function of node removal. (Bottom left) MA as a function of element removal. (Bottom right) MA as a function of node removal.

In the top two graphs of Figure 13, it is seen that higher intra-subject ICCs are found for the included group. As the subjects move during image acquisition, the reproducibility between the two successive scans of that subject is expected to be less compared to the subject that does not exceed the motion threshold. It can also be seen that for the last fraction of elements or nodes remaining, the difference between the two subgroups becomes less. As more elements and nodes are removed from the dataset, the more reproducible elements and nodes remain part of the CM's.

In the bottom two graphs the MA curves are shown of the motion subgroup, the included group and the adjusted motion group. The adjusted motion group is constructed by using the corresponding scaling factor as found in Table 8.

Figure 14 shows the included subgroup along with the Pathology subgroup for the RSS100 parcellation scheme. In the graphs with the MA curves, a corrected pathology curve is also shown, which allows for comparison of the Pathology subset with the Included subset. Additional results of the Pathology subset analysis of the other three parcellation schemes are found in appendix E.

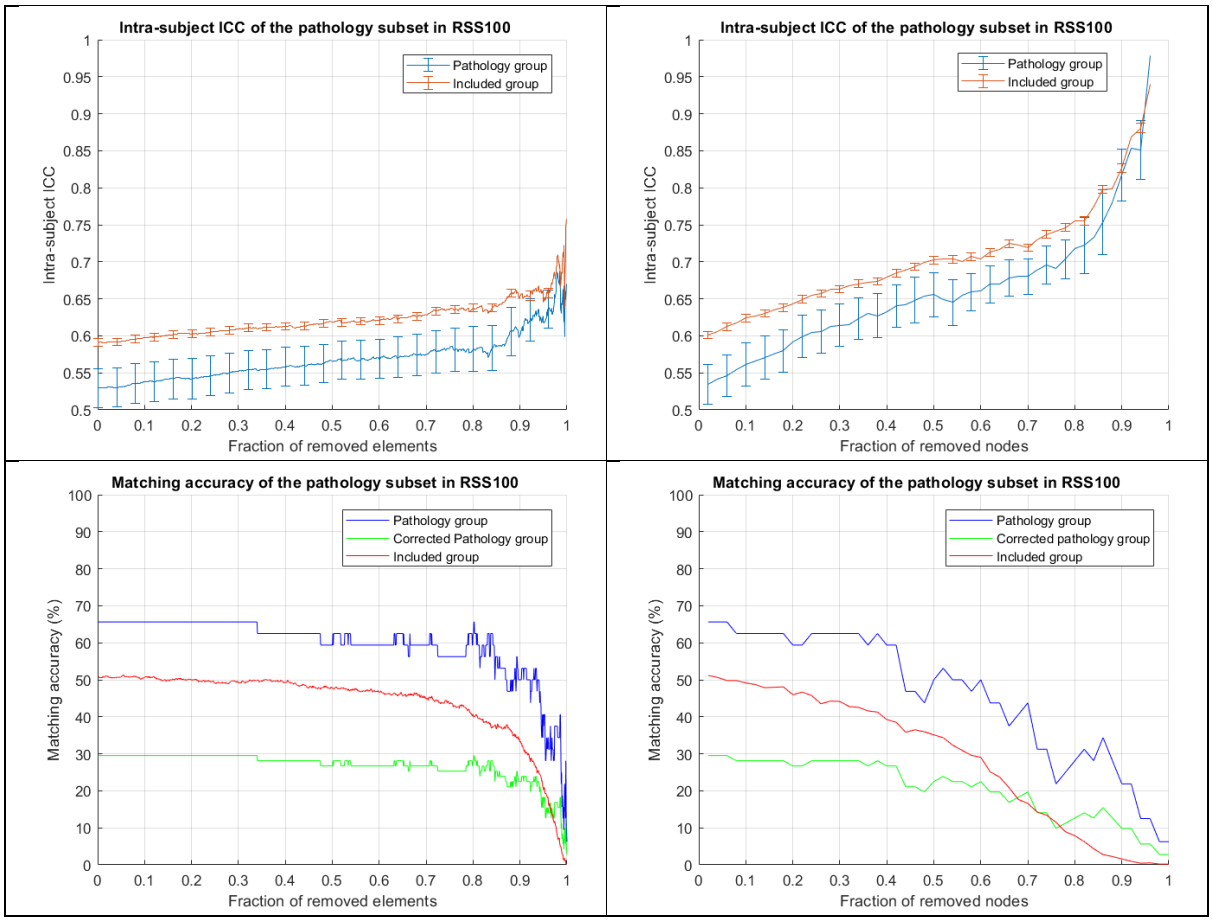


Figure 14 Pathology subset analysis on RSS100. The error shown in the intra-subject ICC curves represent the standard errors of the mean (Top left) intra-subject ICC of element removal. (Top right) Intra-subject ICC as a function of node removal. (Bottom left) MA as a function of element removal. (Bottom right) MA as a function of node removal.

Just like with the motion subgroup, the ICC of the included group is higher than the pathology group, but as more elements or nodes are removed, the difference between the two becomes smaller. The MA curves are also similar to the motion group in the sense that the included group should have higher MA values, which is the case after a correction with the scaling factor.

The next subset is constructed by dividing the included group in two, based on the sex of the subjects. This gives an all-male and an all-female subset. Results of the RSS100 parcellation scheme are shown in Figure 15. Additional results of the subset based on sex of the other three parcellation schemes are found in appendix F.

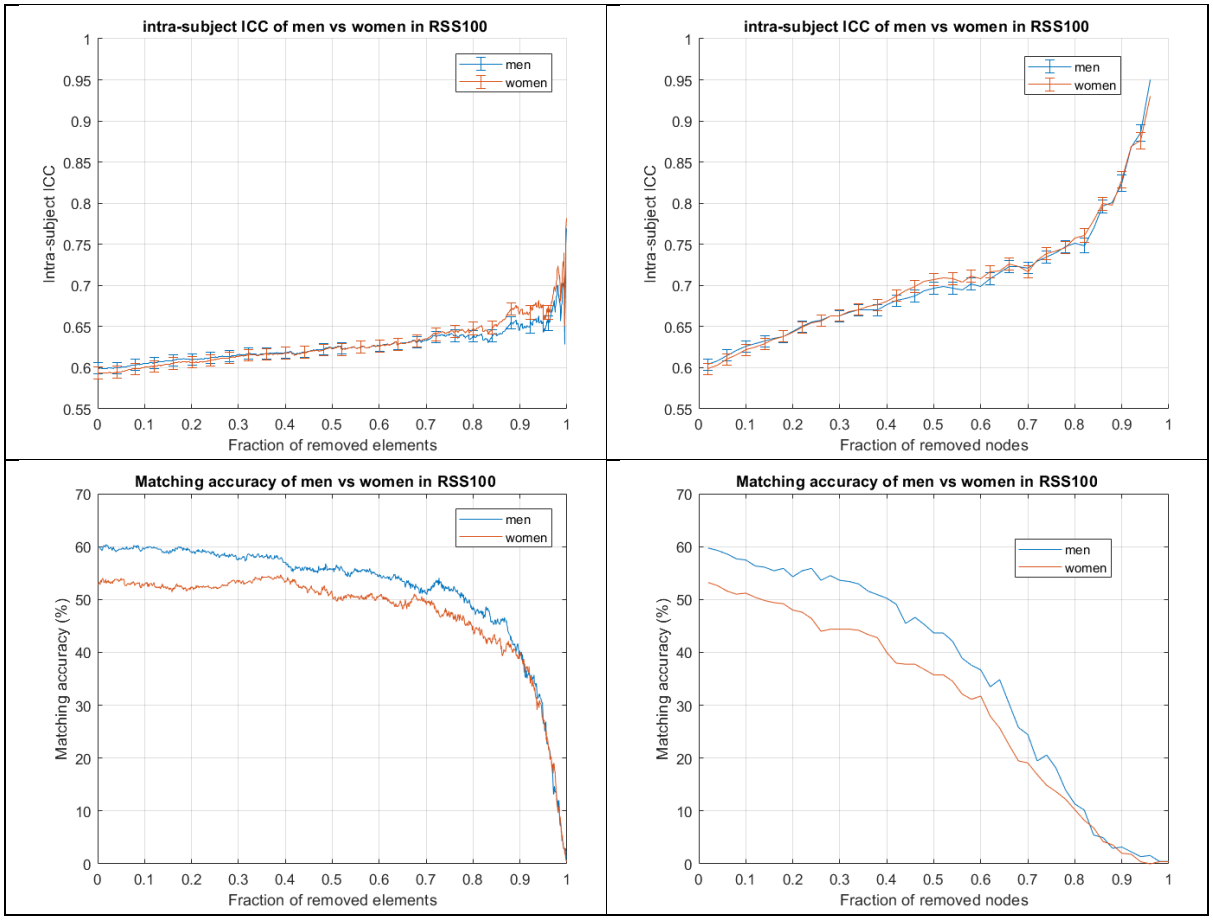


Figure 15 Men vs women longitudinal subset analysis of RSS100. The errors shown in the intra-subject ICC curves represent the standard errors of the mean. (Top left) intra-subject ICC as a function of removed elements, men vs women. (Top right) intra-subject ICC as a function of removed nodes, men vs women. (Bottom left) MA as function of elements. (Bottom right) MA as a function of removed nodes.

The intra-subject ICC curves of both the node and the element removal nearly overlap. The matching accuracy is slightly higher for the men in both the node removal and element removal curves.

The next subset to be analysed is that of the interscan times. To get an impression for the range of the interscan times that are present in the subsets, see Figure 16.

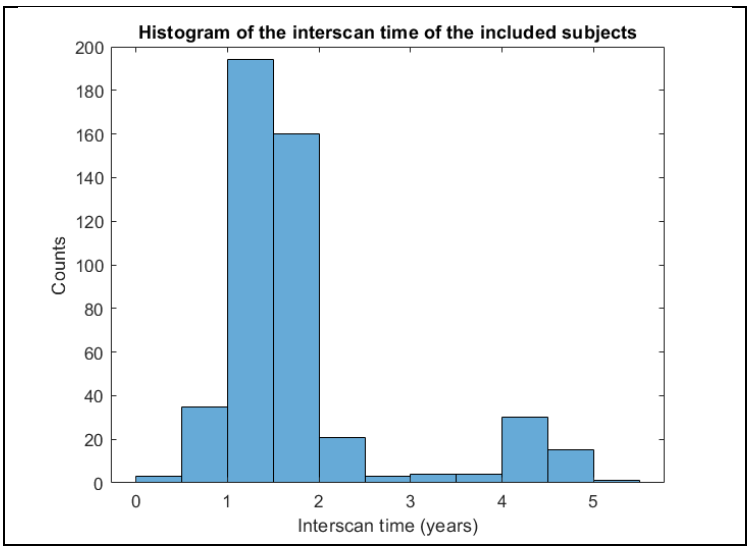


Figure 16 Histogram of the interscan times in the longitudinal set

The interscan time of the subjects of the longitudinal group will be divided into subsets of equal size. As the scaling matrices only give an indication of the difference in MA between sets of different sizes, it is of course better to have sets that are equally sized. Each of the IST subsets consist of either 117 or 118 subjects. These equally sized subsets do cause the range of the bins of the interscan time to vary as seen in the legend of Figure 17. Figure 17 shows the intra-subject ICC and MA for the different subsets based on IST of the RSS100 parcellation scheme. The results on the other three parcellation schemes are shown in Appendix G.

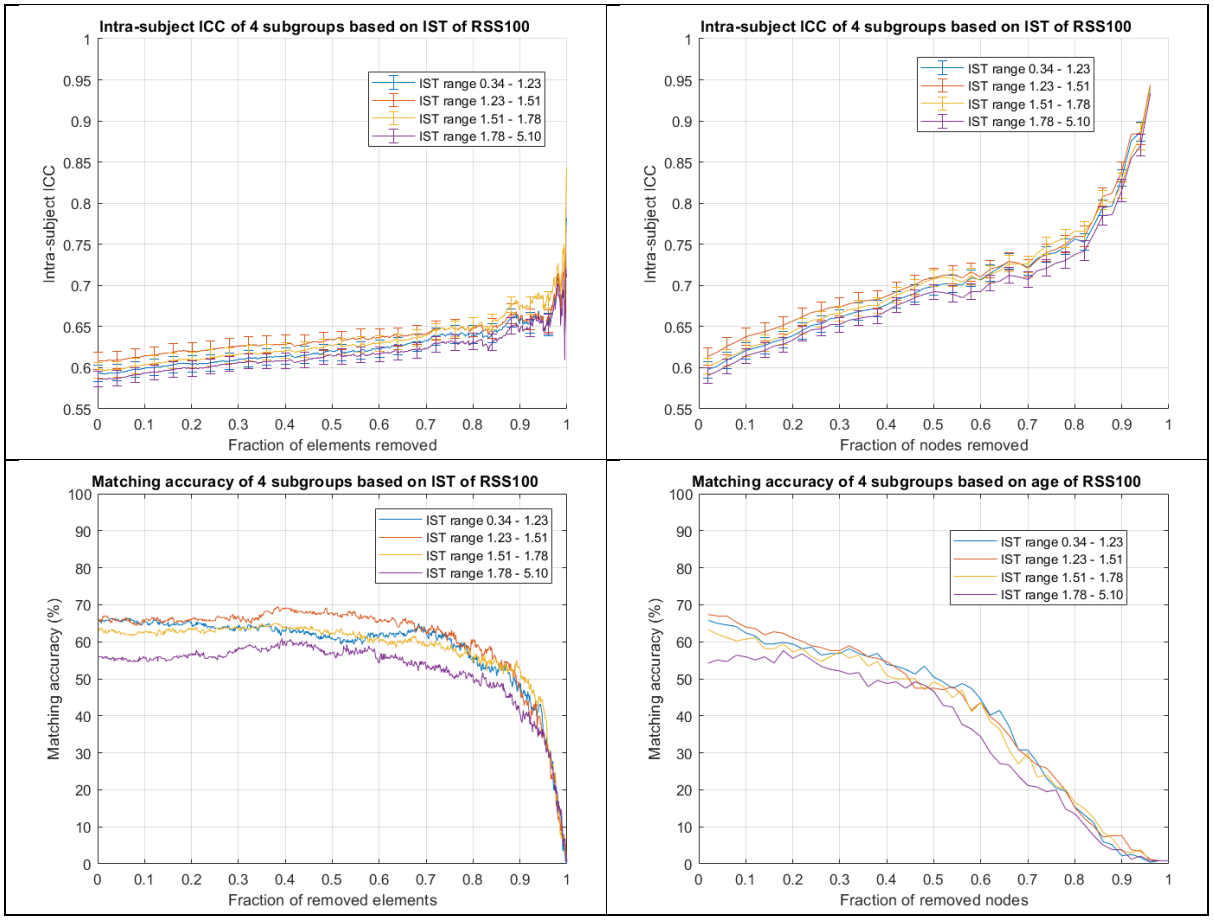


Figure 17 Four interscan time subset of the longitudinal set of the parcellation scheme RSS100. The errors as shown in the intra-subject ICC curves are the standard errors of the mean. (Top left) intra-subject ICC as a function of removed elements. (Top right) intra-subject ICC as a function of removed nodes. (Bottom left) MA for element removal. (Bottom right) MA for node removal.

There appears to be no direct relation between interscan times (IST) and reproducibility, as the highest ICCs are found for the IST range of 1.23 to 1.51 years. Additionally, this smaller time window also does not give a smaller error, as can be seen from the standard deviation. What is of note, is that the longest IST range, of 1.78 to 5.1 years, gives both the lowest intra-subject ICC scores, as well as the lowest MA results.

The last subset that is to be analysed is that of the age subgroups. To start, a histogram is given in Figure 18 to show the spread of the ages of the subjects at the time of the first scan.

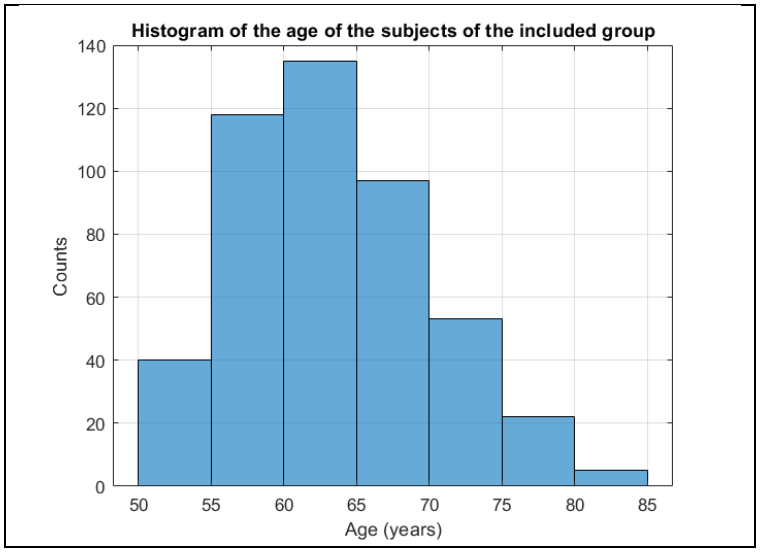


Figure 18 Histogram of the age of the subjects in the longitudinal set at the time of the first scan

Just like with the IST, the histogram does not give an equally distributed set. This entails that when making subgroups that have the same number of subjects, the size age range of that set varies. See Figure 19. Additional results of the age subset analyses of the remaining three parcellation schemes are found in Appendix H.

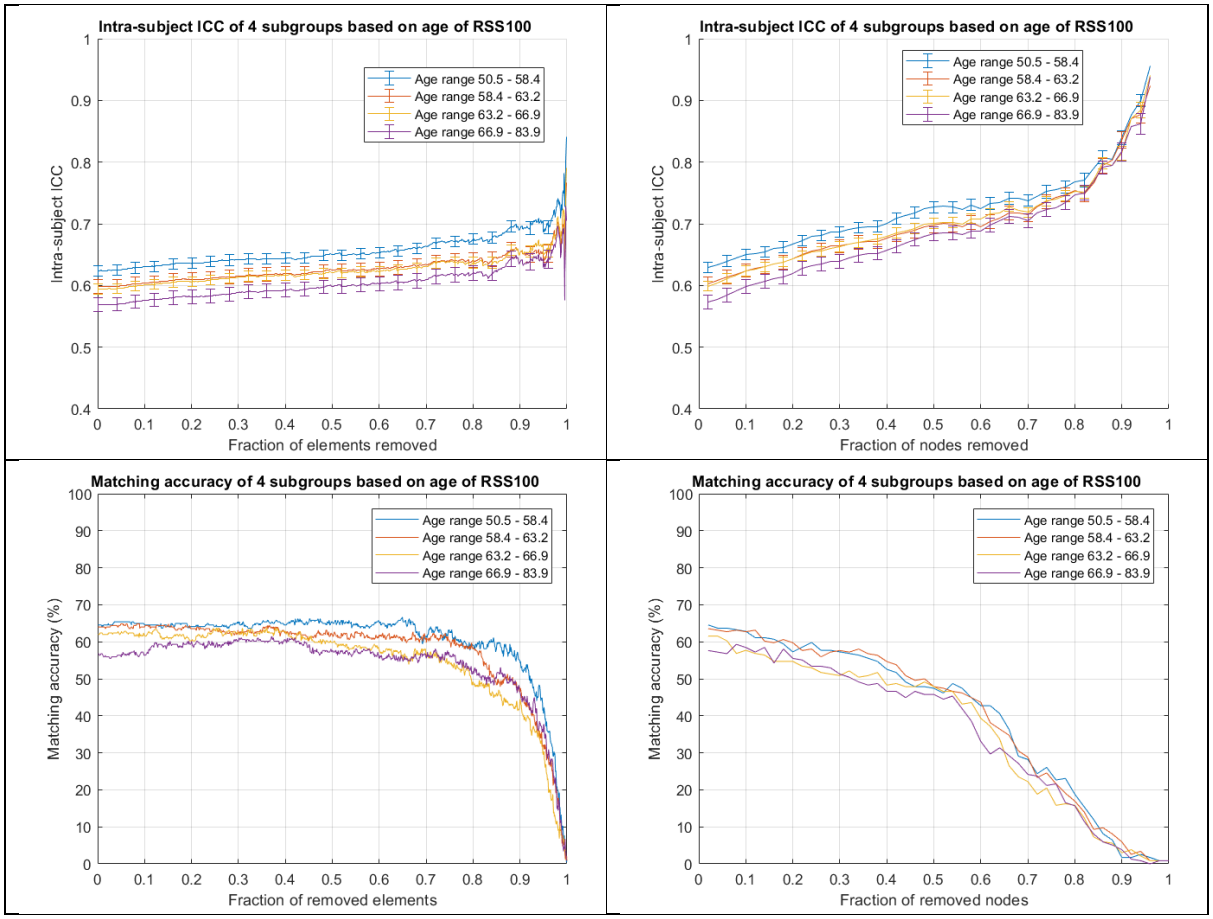


Figure 19 Four age-based subsets of the longitudinal set parcellation scheme RSS100. The errors as shown in the intra-subject ICC curves are the standard errors of the mean. (Top left) intra-subject ICC as a function of removed elements. (Top right) intra-subject ICC as a function of removed nodes. (Bottom left) MA for element removal. (Bottom right) MA for node removal.

The results of Figure 19 do suggest a relation between intra-subject ICC and age. It can be seen that the reproducibility decreases for aging subjects. Moreover, the component removal algorithm appears to enhance the reproducibility for all ages. This relation is not found as clearly in the MA curves though. To make the relation between the intra-subject ICC and the age of the subjects in the longitudinal set clearer, the top left image of Figure 19 is represented in another way in Figure 20.

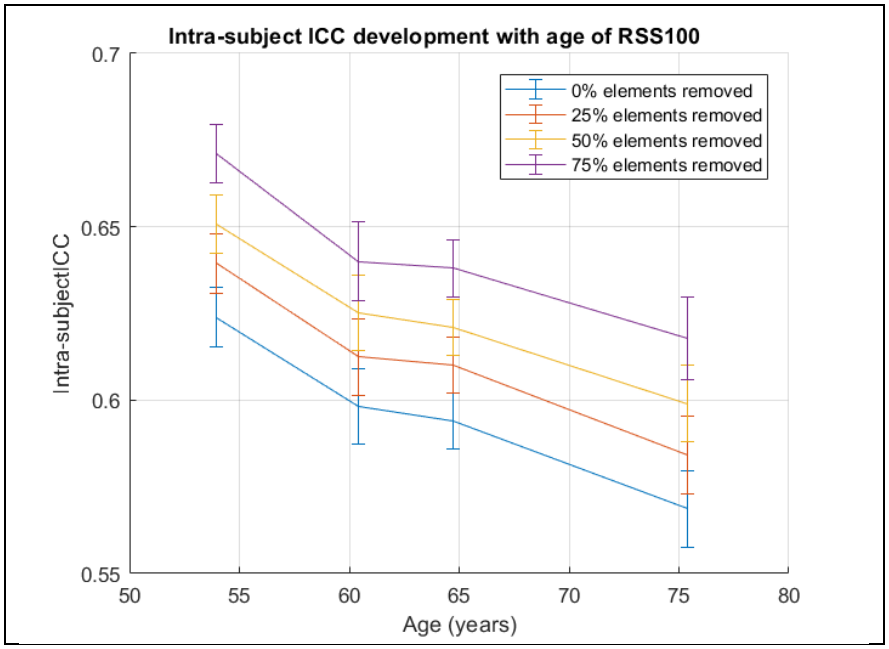


Figure 20 Intra-subject ICC development with age for RSS100. 4 curves depict 4 different fractions of the elements removed.

Figure 20 illustrates the relation between intra-subject ICC and age: as the subjects grow older, the intra-subject ICC goes down. Furthermore, for an increasing number of elements removed, the intra-subject ICC increases.

4.5. ICN Analysis of node and element order

Retaining the predicted optimally reproducible components, in order to enhance the overall reproducibility, is one of the main focusses of this work. This entails the isolation of certain elements/nodes and eliminating others. Although the method presented here is data-driven, another obvious way of partitioning the CM's is by isolating certain ICN's. To check the effectiveness of the method presented in this work, the reproducibility's of the different ICN's of the longitudinal set are also quantified. Table 9 shows the intra-subject ICC's along with the MA's of the included subject group of the longitudinal set. ICN analyses of other parcellation schemes is shown in Appendix I.

Table 9 Intra-subject ICC along with standard error of the mean (SEM) and MA of the included subjects of the RSS100 parcellation scheme. The CM's were divided based on their ICN.

Intrinsic connectivity network (ICN)	Number of nodes in ICN	ICC ± SEM	MA (%)
DAN	6	0.65 ± 0.0083	3.3
DMN	12	0.65 ± 0.0061	21
FPN	5	0.62 ± 0.0096	1.9
SMN	8	0.52 ± 0.0091	5.4
Subcort	2	0.50 ± 0.0187	.11
Temp	5	0.44 ± 0.0113	2.7
VAN	5	0.66 ± 0.0089	.64
Visual	7	0.66 ± 0.0079	4.6

The highest intra-subject ICC's are found for the visual and VAN network, closely followed by the DAN and DMN networks of values 0.66 and 0.65 respectively. Higher intra-subject ICC's of included group of the longitudinal set are found using both the element- and node removal approaches. Figure 21 shows to which ICN the elements and nodes belong whose inter-subject ICC is the highest. Histogram of the other parcellation schemes is shown in Appendix J. The average number of nodes in each ICN as shown in Table 6 is rounded to 6, therefore the origin of the 6 nodes with the highest inter-subject node ICC is shown. 6 nodes correspond to 15 unique elements according to Equation (2).

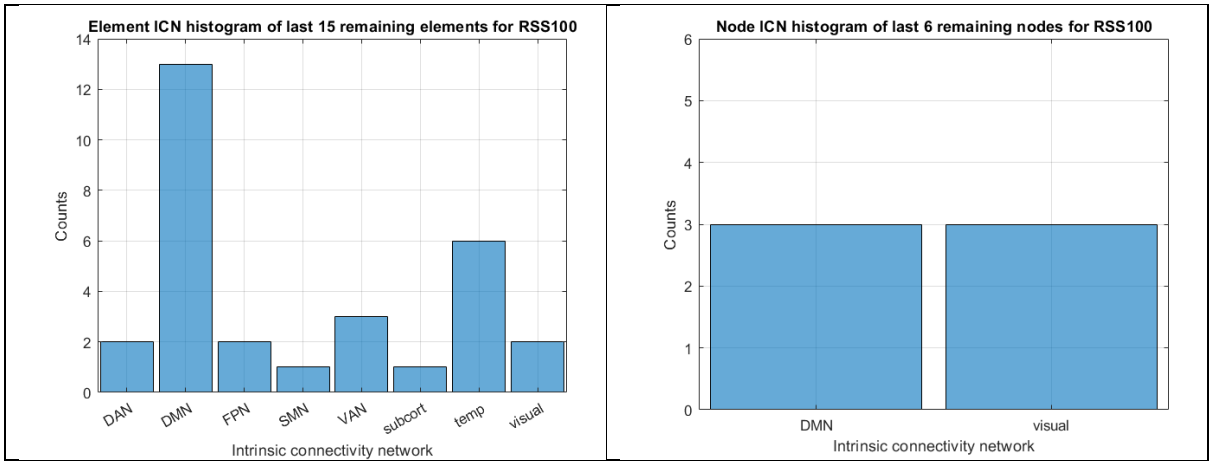


Figure 21 ICN origin of the 15 elements (left side) and 6 nodes (right side) with the highest expected inter-subject ICC for parcellation scheme RSS100. As every element is the connection between two nodes, the 15 best elements harbour the connection between two nodes, and thus ICN's. The total number of counts on the left figure should thus be 2 times 15 which equals 30 ICN's.

The node removal approach determined that the 6 most reproducible nodes originated from the DMN and the visual networks. The reproducibility is partially reflected in Table 6, as these networks produce the highest intra-subject ICCs. In the data-driven method presented in this work in appears that when considering the connections between these networks, higher intra-subject ICC's can be found compared to looking at each network separately. In the element removal approach, the results are more varied, as all ICN's are represented. The DMN proved to have the highest number of counts.

Table 9 also shows the MA's from the different ICN's. These results vary significantly, with the lowest MA of 0.11% in the subcort network. The highest MA is found in the DMN where 21% of the scans were matched to the correct counterpart. The MA as found in the longitudinal analysis for the included group starts off slightly higher than 50%, which is significantly higher than any of the MA's found for the ICN based node groupings.

5. Discussion

For the inter-subject element ICC, several negative values were found. In theory the ICC can't be negative [62]. A negative ICC can be found when the numerator of equation (1) is smaller than 1, so MS_p is smaller than MS_e. MS_e scales with the difference of the within target sum of squares and the between rater sum of squares. This can only lead to negative inter-subject element ICC's when the variance of an element taken within successive scans of the same subject is bigger than the variance between elements of different subjects. In other words: The variance between subjects' elements is smaller than the variance within a subject's element. If there is no connection between the two targets, the variance between the target and within the targets should have the same size which should give a reproducibility of zero. The negative ICC's should thus be interpreted as being poorly reproducible. Negative ICC's were only encountered when calculating the inter-subject element ICC's, and since the calculation was only performed to compute the order of the elements, it did not have a significant impact on the rest of the work.

When examining the MA's of the repro study, it can be seen that the MA test outcome appears to make 'jumps' from one percentage to the next. These jumps can be explained by the limited number of subject that the Repro set consists of. As there are only 37 subjects that can be matched in two directions (from scan 1 to scan 2 and vice versa) this effectively gives MA's as fractions of 2 times 37 is 74. 72 out of 74 is 97.3%, 73 out of 74 is 98.6% and 71 out of 74 is 96.0%.

Also note the apparent disjoint between Figures 10 and 12: These figures show the results of the reproducibility study of both the node and element removal algorithms. It appears counter-intuitive that for the element removal a moderate increase is found for the intra-subject ICC, while the MA climbs to percentages of over 90%. While in the node removal algorithm, the ICC goes beyond values of 0.9, and the MA does not appear to increase at all. As the order in which the elements or nodes are removed are determined based on their inter-subject ICC, it is to be expected that the intra-subject ICC is increased for both sets. The suspected cause for the larger intra-subject ICC increase for the node removal approach is because the within target variation for the intra-subject ICC of the node removal approach is smaller, as nodes act in a more synchronous matter compared to individual elements. It is this line of reasoning that also causes the MA to be better for the element removal compared to the node removal: As more and more elements are removed with larger variation, it leads to a greater distinguishing power, because in the node removal algorithm the CM's are to similar.

The longitudinal set analysis is mainly concerned with the second research question, which is focussed around the intra-subject reproducibility of the predicted optimal removal of nodes and elements as found in the reproducibility study.

The first thing that should be noted is that with all the subsets the intra-subject ICC development is similar to that of its corresponding counterpart with the reproducibility study. It appears the predicted element- and node orders as found in the reproducibility study are able to increase the overall reproducibility of the longitudinal group, which served as the test set.

Perhaps the biggest discrepancy when going from the reproducibility study to the longitudinal set analysis, is the difference found in the MA of the element removal algorithm. As MA's upward of 90% are found in the reproducibility study for the element removal algorithm, results that have not been approximated in the longitudinal study. It appears that the MA test based on the inter-subject element ICC of the same set is highly successful, but the algorithm lacks predictive power. Finn [58] found an average MA of 93.7% for a MA consisting of 126 subjects. When the scaling matrix is used to compare the results of this work with Finn', it turns out that the results of Finn are still better than those presented in this work. Though a different parcellation

scheme is used in his work (one with 268 nodes). Moreover, a 3T MRI scanner was used for the research of Finn et all. This increase in magnetic field strength results in MRI images with a higher resolution. Therefore, it can be expected that Finn's results are better than those presented here.

Moving on to the longitudinal set analysis, the use of the scaling matrices immediately proved useful when analysing the Motion and Pathology subsets: As the quality of the fMRI scans is impaired by the motion of the subjects, or any apparent pathologies, it can be expected that the MA of these sets to be lower than the included subset. This proved to be the case. It is also interesting to see to what extent the ICC's are affected in these subsets. The use of the scaling matrices allowed for a comparison of MA's of subsets of varied sizes, but it remains an approximation. Ideally, MA's should be compared to subsets of the same size, because then the subset size will not influence the outcome. A recommended approach would be to take a sample of the larger subset, equal to the size of the smaller subset and perform a MA test on both.

The test that compared the reproducibility of functional connectivity profiles of men to women, proved to be highly similar. There is nearly a complete overlap of the intra-subject ICC's of both the node and element removal algorithms, and a slightly higher MA for men. Using the MA estimate, it can be seen that these results are significant (assuming both sets have roughly the same size). It appear that the functional connectivity profiles of men are slightly more distinctive than women's.

With the subgroups based on their interscan time, no relation between the interscan time and the intra-subject ICC development or the MA was found. As the included subgroup of the longitudinal set consisted of healthy subjects, the reproducibility study should not result in extreme deviations in intra-subject ICC outcome. On the other hand, 5 years is a relatively long time. To see no (degree of) change in reproducibility over this period of time is somewhat surprising. It is suggested to research even longer periods of interscan time to see if longer periods of time do result in changes in reproducibility.

As for the subgroups based on the age of the subjects: It has been reported the functional connectivity is diminished with aging [59]. This diminished functional connectivity is reflected in the decreasing reproducibility as found in this study. The element and node removal algorithms can increase the reproducibility of the scans by selectively omitting data as predicted from the reproducibility set.

Lastly, a comparison is made between the performance of the node removal algorithm, and node selection based on ICN grouping. The MA of the subgroup based on ICN turned out higher for all ICN's. This is to be expected, as one of the requirements is that the node exhibit high correlation with the other nodes of the ICN. The reproducibility was considerably higher for the node removal algorithm. A notable result, as the reproducibility was predicted based on the Repro set.

In the end a comparison is made between the longitudinal set component removal algorithm and a division of the CM's based on their ICN. When comparing all parcellation schemes, there proved to be no single network that consistently appeared to be among the most reproducible fraction of nodes or elements. The intra-subject ICC's found using the longitudinal set component removal algorithm (for roughly the same number of nodes and elements as the average ICN size) are higher than any of the intra-subject ICC's found using the ICN's. It appears that the component order as determined in the reproducibility study is able to obtain higher a higher reproducibility compared to the ICN partitioning. The MA's of the ICN's turned out to be higher in the ICN partitioning compared to the longitudinal set component removal algorithm. This is to be expected, because the internal variation within an ICN is required to be small in order to be grouped together in an ICN.

6. Conclusions

The two-fold research question as posted in subsection 2.6. are as follows:

Can the overall reproducibility of brain connectivity profiles be improved by selectively omitting parts of the FC profiles?

Can this selective data reduction improve the predicted reproducibility for a population study?

The reproducibility study forms the basis of the answer of the first research question, as it showed that it is indeed possible to enhance the MA by selectively omitting elements that have the lowest inter-subject element ICC. With the RSS100 parcellation scheme, the matching accuracy is increased from an unaltered CM of 82.4% to 98.7%. All other parcellation schemes also saw increased MA for the element removal algorithm. The overall reproducibility of the repro set is enhanced in both the node removal and element removal algorithms. The node removal approach proved to be better at enhancing the intra-subject reproducibility, as the intra-subject ICC was higher than 0.8 which is indicative of 'outstanding' reproducibility. The element removal algorithm was able to enhance the reproducibility of all parcellation schemes from 'moderate' to 'substantial'.

With the second research question, the robustness of the predicted optimal element/node order removal was explored. The elements/nodes of the longitudinal set were iteratively removed in the order as found in the repro set. At every iteration the intra-subject ICC and the MA were computed to see how the overall reproducibility is affected. This algorithm was applied to various subsets of the longitudinal set based on: sex, age, interscan time and taking the grounds for exclusion into account. The development of the reproducibility is not affected by the sex of the subject, or the interscan time (within the range explored in this work, namely 5 years). When the subjects crossed a predetermined motion threshold, a clear decrease in reproducibility is seen. With the subjects who possessed a certain pathology a similar decrease in reproducibility is seen. The algorithm did accurately predict that the intra-subject ICC of the age subsets did reduce as the subjects grew older. Furthermore, the algorithms is able to enhance the reproducibility for all age groups as more nodes and elements were removed from the set.

Lastly, what followed from the comparison between the node removal algorithm and the comparison with the ICN MA's, was that the MA's of the ICN's turned out to be higher than the node removal algorithm. But the node removal algorithm is able to reduce the datasets in a way that they retain a greater degree of the reproducibility compared to the division based on ICN's.

In the end a comparison drawn between the longitudinal set component removal algorithm and a partitioning of the CM's based on their ICN. It is reported that the longitudinal set component removal algorithm is able to produce a reduced CM that has a higher reproducibility compared to the ICN partitioning intra-subject ICC outcomes. The MA's of the ICN partitioning are higher than the longitudinal set component removal algorithm.

Acknowledgements

I would first like to thank my thesis supervisor dr. ir. Frans Vos of the department of Imaging Physics. For his introduction into the world of fMRI research allowed me to cultivate my own interest into the workings of the human brain. At the same time, Frans always allowed me to do my own research, only steering me in the right direction where needed. The freedom I had while doing my master thesis was greatly appreciated.

Secondly, I would like to thank my direct supervisor Willem van Valenberg MSc. It was thanks to his patience and critical feedback, that allowed this work to be elevated above what I dared think was achievable. Moreover, Willem was always available to answer any of my questions regarding the work. I am gratefully indebted to his contributions to the thesis.

Lastly, I want to thank the research group of Imaging Physics at the TU Delft, and my fellow master students over at the Biomedical Imaging Group Rotterdam. All these people provided me with an academic environment that nurtured the work I was doing. They were always up for a laugh as well, which provided a particularly nice working environment.

Theo Driever

References

[1] Lennie, P. (2003). The cost of cortical computation. *Current biology*, *13*(6), 493-497.

[2] Raichle, M. E., & Mintun, M. A. (2006). Brain work and brain imaging. *Annu. Rev. Neurosci.*, *29*, 449-476.

[3] [Diagram of neuron biology]. Retrieved on January 13, 2018, from: https://www.gettyimages.nl/detail/illustratie/neuron-diagram-royalty-free-illustraties/531825569

[4] Assaf, Y., & Pasternak, O. (2008). Diffusion tensor imaging (DTI)-based white matter mapping in brain research: a review. *Journal of molecular neuroscience*, *34*(1), 51-61.

[5] Basser, P. J., Mattiello, J., & LeBihan, D. (1994). MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, *66*(1), 259-267.

[6] Huettel, S. A., Song, A. W., & McCarthy, G. (2009). Functional magnetic resonance imaging (2nd ed). Massachusetts: Sinauer. ISBN 978-0-87893-286-3.

[7] Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic resonance in medicine*, *25*(2), 390-397.

[8] Kwong, K. K., Belliveau, J. W., Stern, C. E., Chesler, D. A., Goldberg, I. E., Poncelet, B. P., ... & Cheng, H. M. (1992). Functional MR imaging of primary visual and motor cortex. *JMRI*, *2*(P), 76.

[9] Pauling, L., & Coryell, C. D. (1936). The magnetic properties and structure of hemoglobin, oxyhemoglobin and carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences*, *22*(4), 210-216.

[10] Weisskoff, R. M., & Kiihne, S. (1992). MRI susceptometry: Image-based measurement of absolute susceptibility of MR contrast agents and human blood. *Magnetic Resonance in Medicine*, *24*(2), 375-383.

[11] Thulborn, K. R., Waterton, J. C., Matthews, P. M., & Radda, G. K. (1982). Oxygenation dependence of the transverse relaxation time of water protons in whole blood at high field. *Biochimica et Biophysica Acta (BBA)-General Subjects*, *714*(2), 265-270.

[12] Ogawa, S., & Lee, T. M. (1992). Blood oxygenation level dependent MRI of the brain: effects of seizure induced by kainic acid in rat. In *Proc Int Soc Magn Reson Med* (Vol. 1, p. 501).

[13] Fox, M. D., & Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nature Reviews Neuroscience*, *8*(9), 700-711.

[14] Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic resonance in medicine*, *34*(4), 537-541.

[15] Strother, S. C. (2006). Evaluating fMRI preprocessing pipelines. *IEEE Engineering in Medicine and Biology Magazine*, *25*(2), 27-41.

[16] Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. *Neuroimage*, *5*(3), 179-197.

[17] Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological review*, *108*(1), 33.

[18] Cordes, D., Haughton, V. M., Arfanakis, K., Carew, J. D., Turski, P. A., Moritz, C. H., ... & Meyerand, M. E. (2001). Frequencies contributing to functional connectivity in the cerebral cortex in “resting-state” data. *American Journal of Neuroradiology*, *22*(7), 1326-1333.

[19] Zuo, X. N., Kelly, C., Adelstein, J. S., Klein, D. F., Castellanos, F. X., & Milham, M. P. (2010). Reliable intrinsic connectivity networks: test–retest evaluation using ICA and dual regression approach. *Neuroimage*, *49*(3), 2163-2177.

[20] Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186-198.

[21] Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

[22] Eryilmaz, H., Van De Ville, D., Schwartz, S., & Vuilleumier, P. (2011). Impact of transient emotions on functional connectivity during subsequent resting state: a wavelet correlation approach. *Neuroimage*, *54*(3), 2481-2491.

[23] Smith, S. M. (2012). The future of FMRI connectivity. *Neuroimage*, *62*(2), 1257-1266.

[24] Park, H. J., & Friston, K. (2013). Structural and functional brain networks: from connections to cognition. *Science*, *342*(6158), 1238411.

[25] Raichle, M. E. (2011). The restless brain. *Brain connectivity*, *1*(1), 3-12.

[26] Ferrarini, L., Veer, I. M., Baerends, E., van Tol, M. J., Renken, R. J., van der Wee, N. J., ... & van Buchem, M. A. (2009). Hierarchical functional modularity in the resting-state human brain. *Human brain mapping*, *30*(7), 2220-2231.

[27] Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. D. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral cortex*, *15*(9), 1332-1342.

[28] Bassett, D. S., Bullmore, E., Verchinski, B. A., Mattay, V. S., Weinberger, D. R., & Meyer-Lindenberg, A. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, *28*(37), 9239-9248.

[29] Rubinov, M., Knock, S. A., Stam, C. J., Micheloyannis, S., Harris, A. W., Williams, L. M., & Breakspear, M. (2009). Small-world properties of nonlinear brain activity in schizophrenia. *Human brain mapping*, *30*(2), 403-416.

[30] Supekar, K., Menon, V., Rubin, D., Musen, M., & Greicius, M. D. (2008). Network analysis of intrinsic functional brain connectivity in Alzheimer's disease. *PLoS computational biology*, *4*(6), e1000100.

[31] Wang, L., Zhu, C., He, Y., Zang, Y., Cao, Q., Zhang, H., ... & Wang, Y. (2009). Altered small-world brain functional networks in children with attention-deficit/hyperactivity disorder. *Human brain mapping*, *30*(2), 638-649.

[32] Gong, G., Rosa-Neto, P., Carbonell, F., Chen, Z. J., He, Y., & Evans, A. C. (2009). Age-and gender-related differences in the cortical anatomical network. *Journal of Neuroscience*, *29*(50), 15684-15693.

[33] Micheloyannis, S., Vourkas, M., Tsirka, V., Karakonstantaki, E., Kanatsouli, K., & Stam, C. J. (2009). The influence of ageing on complex brain networks: a graph theoretical analysis. *Human brain mapping*, *30*(1), 200-208.

[34] Tian, L., Wang, J., Yan, C., & He, Y. (2011). Hemisphere-and gender-related differences in small-world brain networks: a resting-state functional MRI study. *Neuroimage*, *54*(1), 191-202.

[35] Achard, S., & Bullmore, E. (2007). Efficiency and cost of economical brain functional networks. *PLoS computational biology*, *3*(2), e17.

[36] Polanía, R., Paulus, W., Antal, A., & Nitsche, M. A. (2011). Introducing graph theory to track for neuroplastic alterations in the resting human brain: a transcranial direct current stimulation study. *Neuroimage*, *54*(3), 2287-2296.

[37] Grinvald, A., Arieli, A., Tsodyks, M., & Kenet, T. (2003). Neuronal assemblies: single cortical neurons are obedient members of a huge orchestra. *Biopolymers*, *68*(3), 422-436.

[38] Caceres, A., Hall, D. L., Zelaya, F. O., Williams, S. C., & Mehta, M. A. (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *Neuroimage*, *45*(3), 758-768.

[39] Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, *86*(2), 420.

[40] Zandbelt, B. B., Gladwin, T. E., Raemaekers, M., van Buuren, M., Neggers, S. F., Kahn, R. S., ... & Vink, M. (2008). Within-subject variation in BOLD-fMRI signal changes across repeated measurements: quantification and implications for sample size. *Neuroimage*, *42*(1), 196-206.

[41] Raemaekers, M., Vink, M., Zandbelt, B., Van Wezel, R. J. A., Kahn, R. S., & Ramsey, N. F. (2007). Test–retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage*, *36*(3), 532-542.

[42] Wei, X., Yoo, S. S., Dickey, C. C., Zou, K. H., Guttmann, C. R., & Panych, L. P. (2004). Functional MRI of auditory verbal working memory: long-term reproducibility analysis. *Neuroimage*, *21*(3), 1000-1008.

[43] Aron, A. R., Gluck, M. A., & Poldrack, R. A. (2006). Long-term test–retest reliability of functional MRI in a classification learning task. *Neuroimage*, *29*(3), 1000-1006.

[44] Fernandez, G., Specht, K., Weis, S., Tendolkar, I., Reuber, M., Fell, J., ... & Elger, C. E. (2003). Intrasubject reproducibility of presurgical language lateralization and mapping using fMRI. *Neurology*, *60*(6), 969-975.

[45] Specht, K., Willmes, K., Shah, N. J., & Jäncke, L. (2003). Assessment of reliability in functional imaging studies. *Journal of Magnetic Resonance Imaging*, *17*(4), 463-471.

[46] McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological methods*, *1*(1), 30.

[47] Chou, Y. H., Panych, L. P., Dickey, C. C., Petrella, J. R., & Chen, N. K. (2012). Investigation of long-term reproducibility of intrinsic connectivity network mapping: a resting-state fMRI study. *American Journal of Neuroradiology*, *33*(5), 833-838.

[48] Damoiseaux, J. S., Rombouts, S. A. R. B., Barkhof, F., Scheltens, P., Stam, C. J., Smith, S. M., & Beckmann, C. F. (2006). Consistent resting-state networks across healthy subjects. *Proceedings of the national academy of sciences*, *103*(37), 13848-13853.

[49] Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253-258.

[50] Shulman, G. L., Fiez, J. A., Corbetta, M., Buckner, R. L., Miezin, F. M., Raichle, M. E., & Petersen, S. E. (1997). Common blood flow changes across visual tasks: II. Decreases in cerebral cortex. *Journal of cognitive neuroscience*, *9*(5), 648-663.

[51] Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences*, *98*(2), 676-682.

[52] Mason, M. F., Norton, M. I., Van Horn, J. D., Wegner, D. M., Grafton, S. T., & Macrae, C. N. (2007). Wandering minds: the default network and stimulus-independent thought. *Science*, *315*(5810), 393-395.

[53] Buckner, R. L., Snyder, A. Z., Shannon, B. J., LaRossa, G., Sachs, R., Fotenos, A. F., ... & Mintun, M. A. (2005). Molecular, structural, and functional characterization of Alzheimer's

disease: evidence for a relationship between default activity, amyloid, and memory. *Journal of Neuroscience*, *25*(34), 7709-7717.

[54] Adelstein, J. S., Shehzad, Z., Mennes, M., DeYoung, C. G., Zuo, X. N., Kelly, C., ... & Milham, M. P. (2011). Personality is reflected in the brain's intrinsic functional architecture. *PloS one*, *6*(11), e27633.

[55] Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1-38.

[56] Sambataro, F., Murty, V. P., Callicott, J. H., Tan, H. Y., Das, S., Weinberger, D. R., & Mattay, V. S. (2010). Age-related alterations in default mode network: impact on working memory performance. *Neurobiology of aging*, *31*(5), 839-852.

[57] Greicius, M. D., Srivastava, G., Reiss, A. L., & Menon, V. (2004). Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(13), 4637-4642.

[58] Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., ... & Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature neuroscience*, *18*(11), 1664-1671.

[59] Hampson, M., Tokoglu, F., Shen, X., Scheinost, D., Papademetris, X., & Constable, R. T. (2012). Intrinsic brain connectivity related to age in young and middle aged adults. *PloS one*, *7*(9), e44067.

[60] Fox, M. D., & Greicius, M. (2010). Clinical applications of resting state functional connectivity. *Frontiers in systems neuroscience*, *4*, 19.

[61] Hofman, A., Brusselle, G. G., Murad, S. D., van Duijn, C. M., Franco, O. H., Goedegebure, A., ... & Stricker, B. H. C. (2015). The Rotterdam Study: 2016 objectives and design update. *European journal of epidemiology*, *30*(8), 661-708.

[62] Giraudeau, B. (1996). Negative values of the intraclass correlation coefficient are not theoretically possible. *Journal of clinical epidemiology*, *49*(10), 1205.

Appendix A – ICC derivation

Acquisition of a BOLD fMRI scan can be modelled by Equation (5).

$$r = \ Y_{ij} = \mu + p_i + t_j + e_{ij}, \quad 1 \ll i \ll n, \ 1 \ll j \ll k \tag{5}$$

Where Y_{ij} represents the acquired time series of subject number i, as taken during scan number j. Therefore, n represents the total number of subjects, and k the total number of successive scans. μ is a fixed variable, p represents the participants effect, t is the systematic error and e is the measurement error. All sources of error in this model are independent random effects with mean zero. Definition of the ICC is then given in Equation (6).

$$ICC = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_t^2 + \sigma_e^2} \tag{6}$$

For ICC_3 the standard deviation in the systematic error t (or the scanning occasion effect) is assumed to be zero, giving the specific form of ICC_3 in Equation 7.

$$ICC_3 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2} \tag{7}$$

Now two sums of squares can be computed from the components of equation 2, see Equations 8 to 10.

$$SS_p = k \sum_{i=1}^{n} (\bar{Y}_{i.} - \bar{Y}_{..})^2 \tag{8}$$

$$SS_e = \sum_{j=1}^{k} \sum_{i=1}^{n} (\bar{Y}_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})^2 \tag{9}$$

$$\bar{Y}_{i.} = \frac{\sum_{j=1}^{k} Y_{ij}}{k}, \quad \bar{Y}_{.j} = \frac{\sum_{i=1}^{n} Y_{ij}}{n}, \quad \bar{Y}_{..} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} Y_{ij}}{nk} \tag{10}$$

The expectation values of $MS_p = SS_p/(n-1)$ and $MS_e = SS_e/(n-1)(k-1)$ are $d\sigma_p^2 + \sigma_e^2$ and σ_e^2 respectively. The estimators of the standard deviation are thus given by Equations 11 and 12.

$$\hat{\sigma}_p^2 = \frac{MS_p - MS_E}{k} \tag{11}$$

$$\hat{\sigma}_e^2 = MS_e \tag{12}$$

When combining Equations (11) and (12) with Equation (7), an expression for ICC_3 is found. See Equation 13.

$$ICC_3 = \frac{\hat{\sigma}_p^2}{\hat{\sigma}_p^2 + \hat{\sigma}_e^2} = \frac{MS_p - MS_e}{MS_p + (k-1)MS_e} \tag{13}$$

This derivation is constructed using [18][38][45].

Appendix B – longitudinal set MA error estimates for varying set sizes of other parcellation schemes

~~UKBiobank~~	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
16	8.53	8.37	8.36	8.32	8.25	8.15	8.19	8.26	8.94	9.15
22	7.26	7.19	7.02	6.98	6.98	7.08	7.03	7.43	7.50	8.00
31	6.40	6.50	6.64	6.51	6.41	6.46	6.58	6.69	7.01	7.33
37	6.23	6.28	6.24	6.27	6.40	6.35	6.25	6.40	6.48	7.09
59	4.82	4.83	4.80	4.67	4.71	4.77	4.85	4.92	5.11	5.32
78	4.28	4.27	4.32	4.31	4.41	4.50	4.46	4.48	4.60	4.76
118	3.57	3.62	3.63	3.63	3.73	3.73	3.77	3.74	3.85	4.01
470	1.49	1.55	1.47	1.47	1.53	1.53	1.54	1.75	1.75	1.74

~~HCP820 d100~~	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
16	8.93	8.81	8.73	8.56	8.62	8.64	8.74	8.72	9.07	9.91
22	7.77	7.64	7.43	7.45	7.55	7.50	7.71	7.75	8.18	8.39
31	6.99	7.03	7.09	6.83	7.01	7.02	6.90	7.06	7.29	7.51
37	6.40	6.38	6.22	6.33	6.45	6.42	6.61	6.76	6.83	6.93
59	4.97	5.09	5.08	5.16	5.06	5.12	5.13	5.28	5.41	5.38
78	4.47	4.49	4.45	4.46	4.53	4.50	4.51	4.49	4.85	5.26
118	3.50	3.57	3.63	3.62	3.67	3.67	3.64	3.65	3.84	3.82
470	1.43	1.50	1.32	1.35	1.36	1.50	1.59	1.62	1.63	1.65

~~HCP820 d200~~	**0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
16	8.26	8.15	8.21	8.04	7.96	7.93	7.83	7.82	7.99	7.92
22	7.49	7.32	7.29	7.16	7.07	7.21	7.28	7.32	7.54	7.76
31	6.61	6.69	6.53	6.48	6.38	6.36	6.32	6.42	6.34	6.69
37	5.70	5.73	5.79	5.82	5.86	5.71	5.79	5.83	5.97	6.22
59	4.86	4.83	4.73	4.92	4.88	4.89	4.78	4.95	5.12	5.17
78	4.35	4.42	4.32	4.31	4.35	4.43	4.41	4.53	4.59	4.82
118	3.35	3.29	3.24	3.26	3.27	3.29	3.30	3.31	3.38	3.54
470	1.29	1.22	1.24	1.31	1.34	1.38	1.35	1.39	1.52	1.49

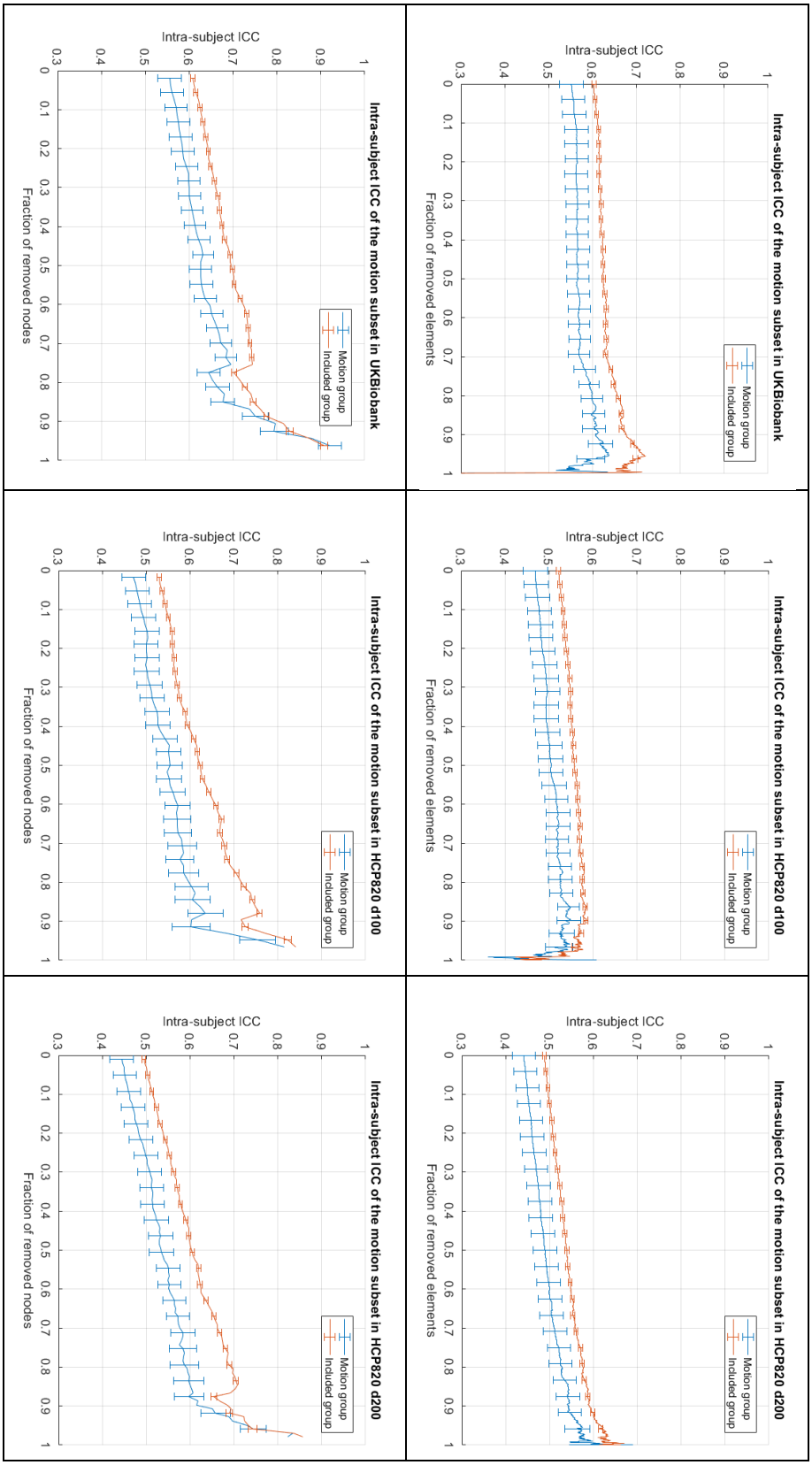
Appendix C – longitudinal set MA scaling matrices of other parcellation schemes

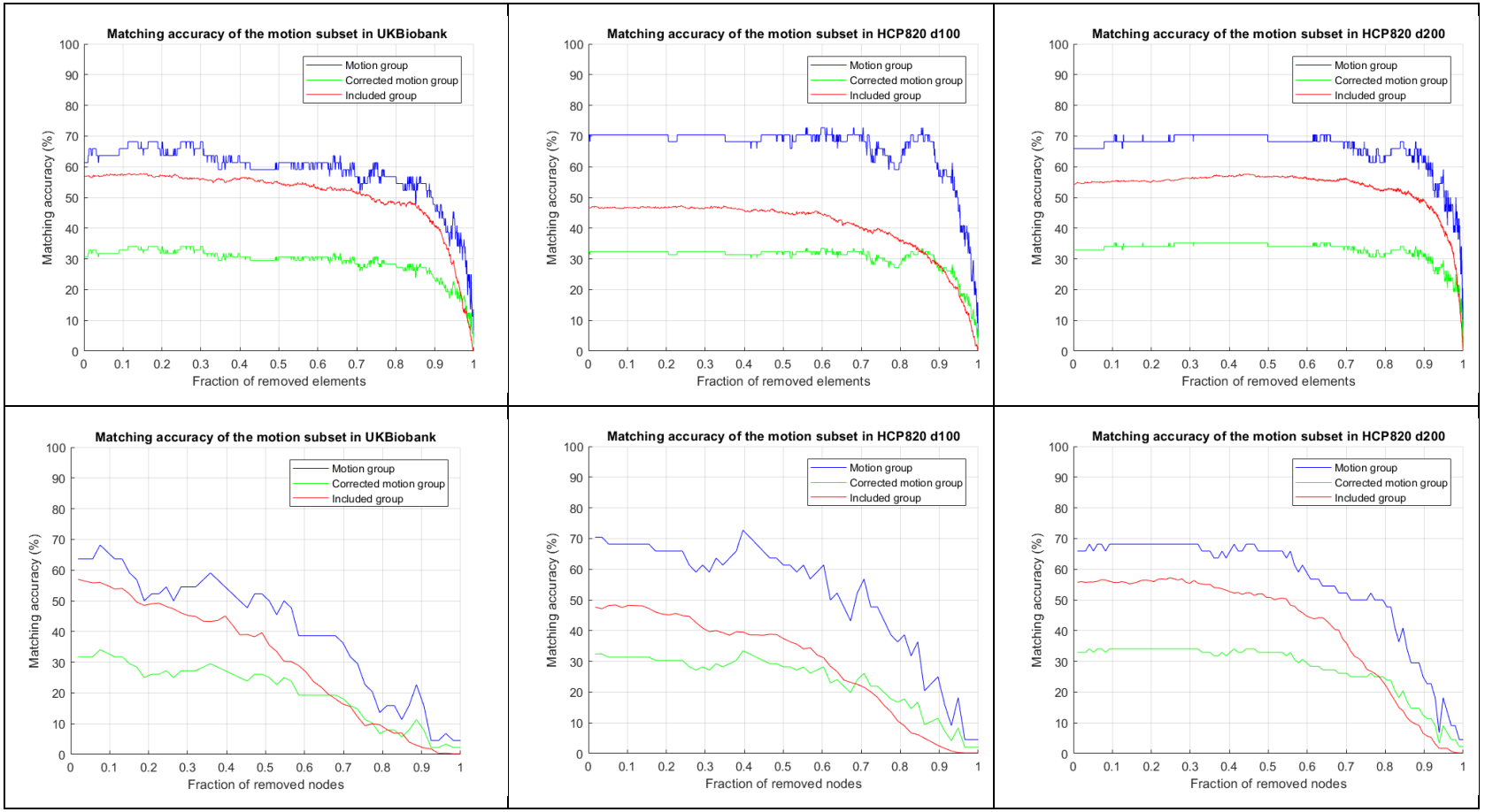
UKBiobank	**16**	**22**	**31**	**37**	**59**	**78**	**118**	**470**
16	1	0.96	0.93	0.91	0.85	0.82	0.75	0.49
22	1.04	1	0.97	0.94	0.88	0.85	0.77	0.50
31	1.07	1.03	1	0.98	0.81	0.87	0.80	0.52
37	1.10	1.06	1.03	1	0.94	0.90	0.82	0.53
59	1.17	1.13	1.09	1.07	1	0.96	0.87	0.57
78	1.23	1.18	1.14	1.12	1.05	1	0.91	0.60
118	1.34	1.30	1.25	1.22	1.14	1.09	1	0.65
470	2.07	1.99	1.92	1.88	1.76	1.68	1.54	1

HCP820_d100	**16**	**22**	**31**	**37**	**59**	**78**	**118**	**470**
16	1	0.95	0.91	0.89	0.81	0.77	0.70	0.43
22	1.05	1	0.95	0.93	0.85	0.81	0.74	0.46
31	1.11	1.05	1	0.98	0.90	0.85	0.77	0.48
37	1.13	1.08	1.02	1	0.92	0.87	0.79	0.49
59	1.23	1.17	1.12	1.09	1	0.95	0.86	0.53
78	1.30	1.24	1.18	1.15	1.06	1	0.91	0.56
118	1.43	1.36	1.29	1.27	1.16	1.10	1	0.62
470	2.31	2.20	2.09	2.04	1.87	1.77	1.62	1

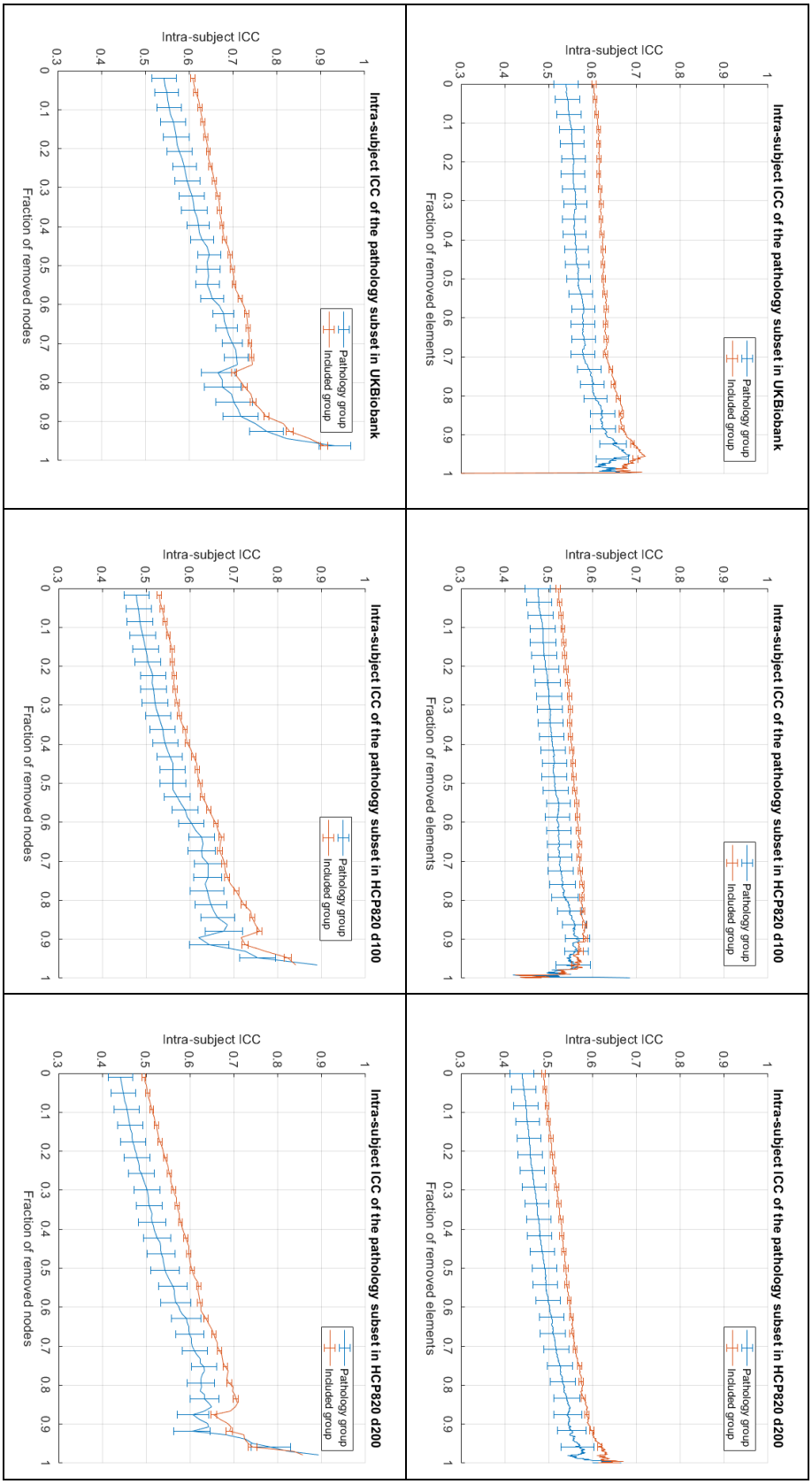
HCP820_d200	**16**	**22**	**31**	**37**	**59**	**78**	**118**	**470**
16	1	0.97	0.93	0.91	0.85	0.81	0.75	0.49
22	1.03	1	0.96	0.94	0.88	0.84	0.78	0.50
31	1.08	1.05	1	0.98	0.92	0.88	0.81	0.53
37	1.10	1.07	1.02	1	0.94	0.90	0.83	0.54
59	1.18	1.14	1.09	1.07	1	0.96	0.88	0.57
78	1.23	1.19	1.14	1.12	1.05	1	0.92	0.60
118	1.33	1.29	1.23	1.21	1.13	1.08	1	0.65
470	2.05	1.99	1.90	1.86	1.74	1.67	1.54	1

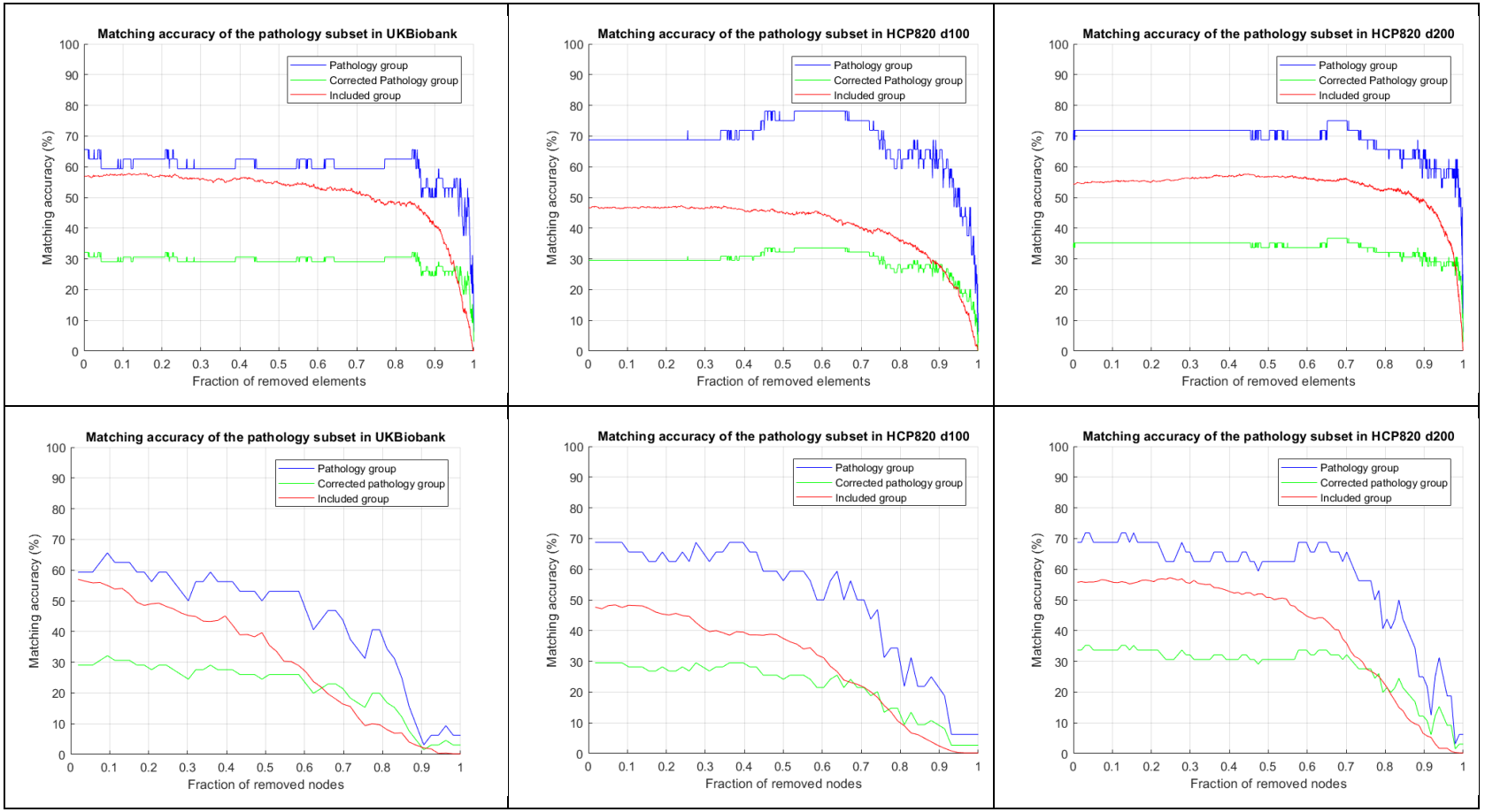
Appendix D – Motion subset figures of other parcellation schemes



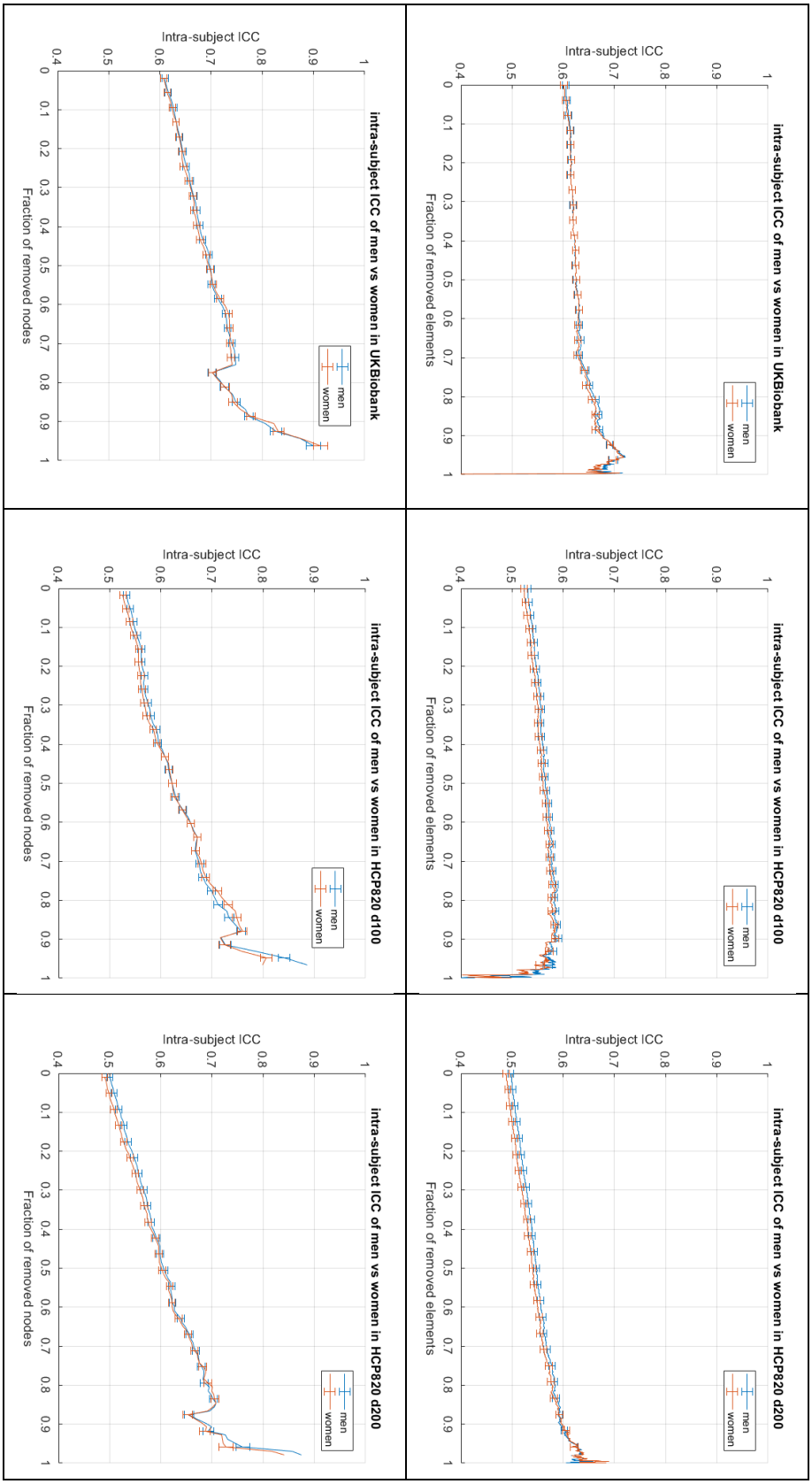

Matching accuracy of the motion subset in UKBiobank
Matching accuracy of the motion subset in HCP820 d100
Matching accuracy of the motion subset in HCP820 d200
Motion group
Corrected motion group
Included group
Matching accuracy (%)
Fraction of removed elements
Fraction of removed nodes

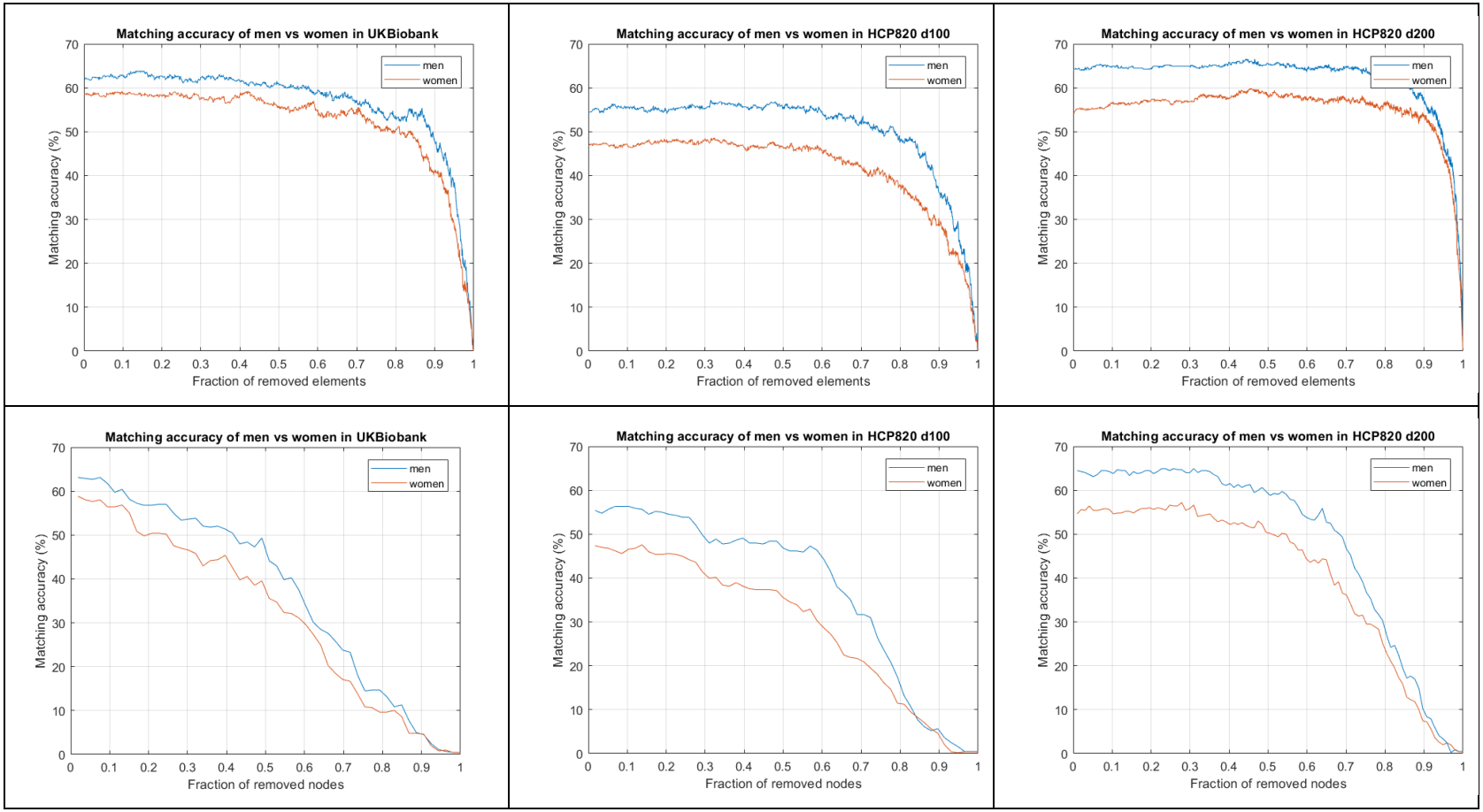
Appendix E – Pathology subset figures of other parcellation schemes



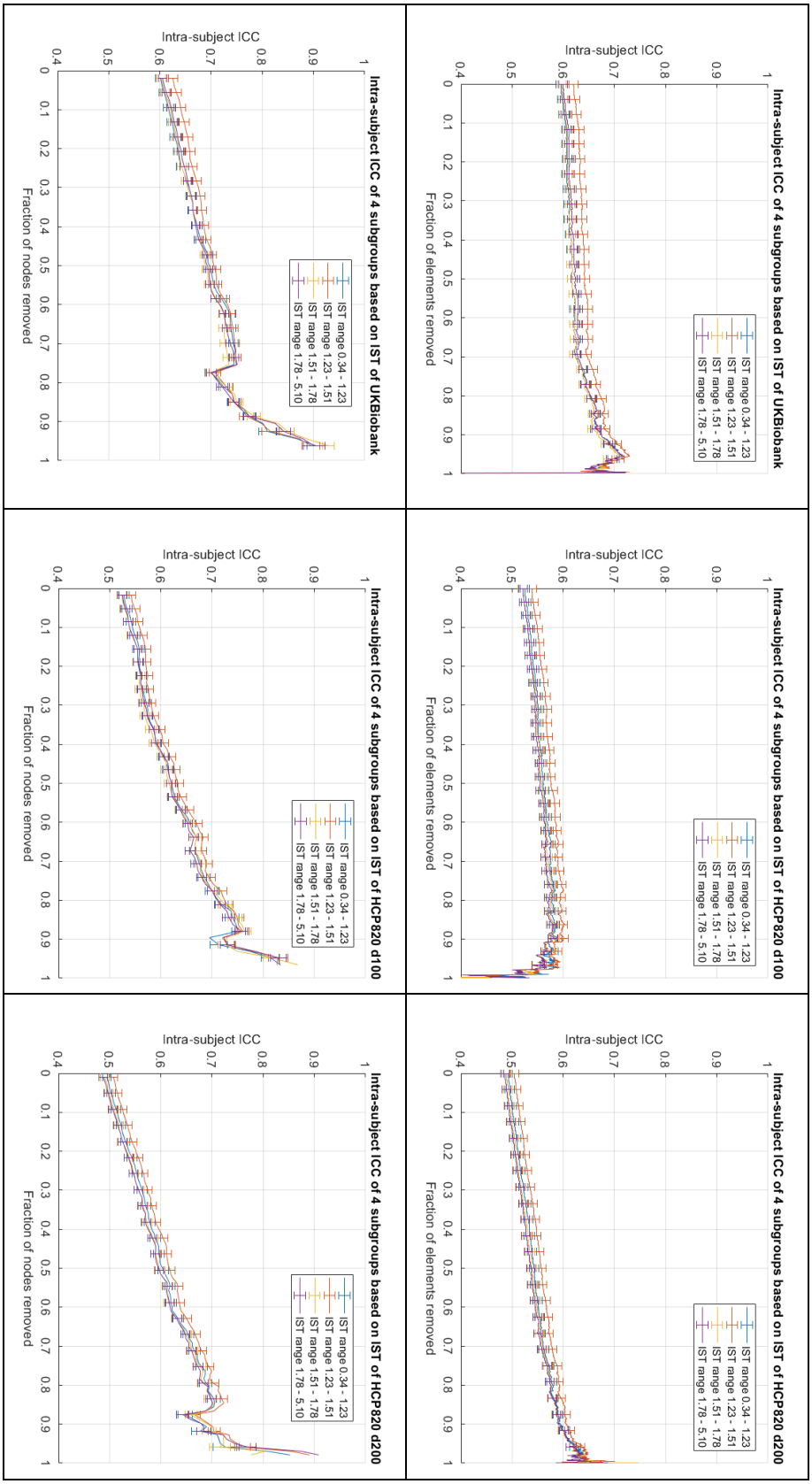

Matching accuracy of the pathology subset in UKBiobank
Pathology group
Corrected Pathology group
Included group
Matching accuracy (%)
Fraction of removed elements
Matching accuracy of the pathology subset in HCP820 d100
Pathology group
Corrected Pathology group
Included group
Matching accuracy (%)
Fraction of removed elements
Matching accuracy of the pathology subset in HCP820 d200
Pathology group
Corrected Pathology group
Included group
Matching accuracy (%)
Fraction of removed elements
Matching accuracy of the pathology subset in UKBiobank
Pathology group
Corrected pathology group
Included group
Matching accuracy (%)
Fraction of removed nodes
Matching accuracy of the pathology subset in HCP820 d100
Pathology group
Corrected pathology group
Included group
Matching accuracy (%)
Fraction of removed nodes
Matching accuracy of the pathology subset in HCP820 d200
Pathology group
Corrected pathology group
Included group
Matching accuracy (%)
Fraction of removed nodes

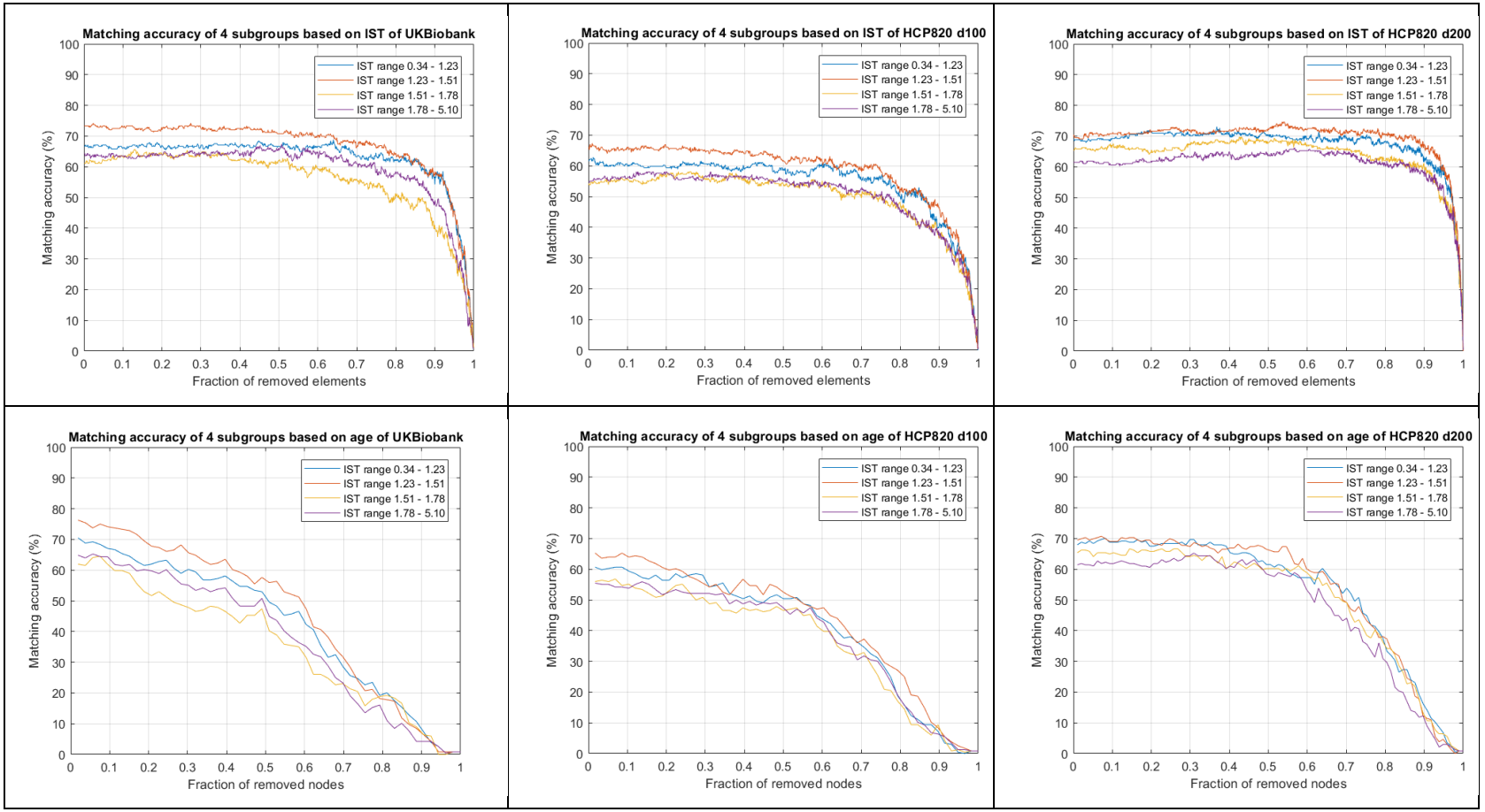
Appendix F – Figures of other parcellation schemes of men vs. women subset



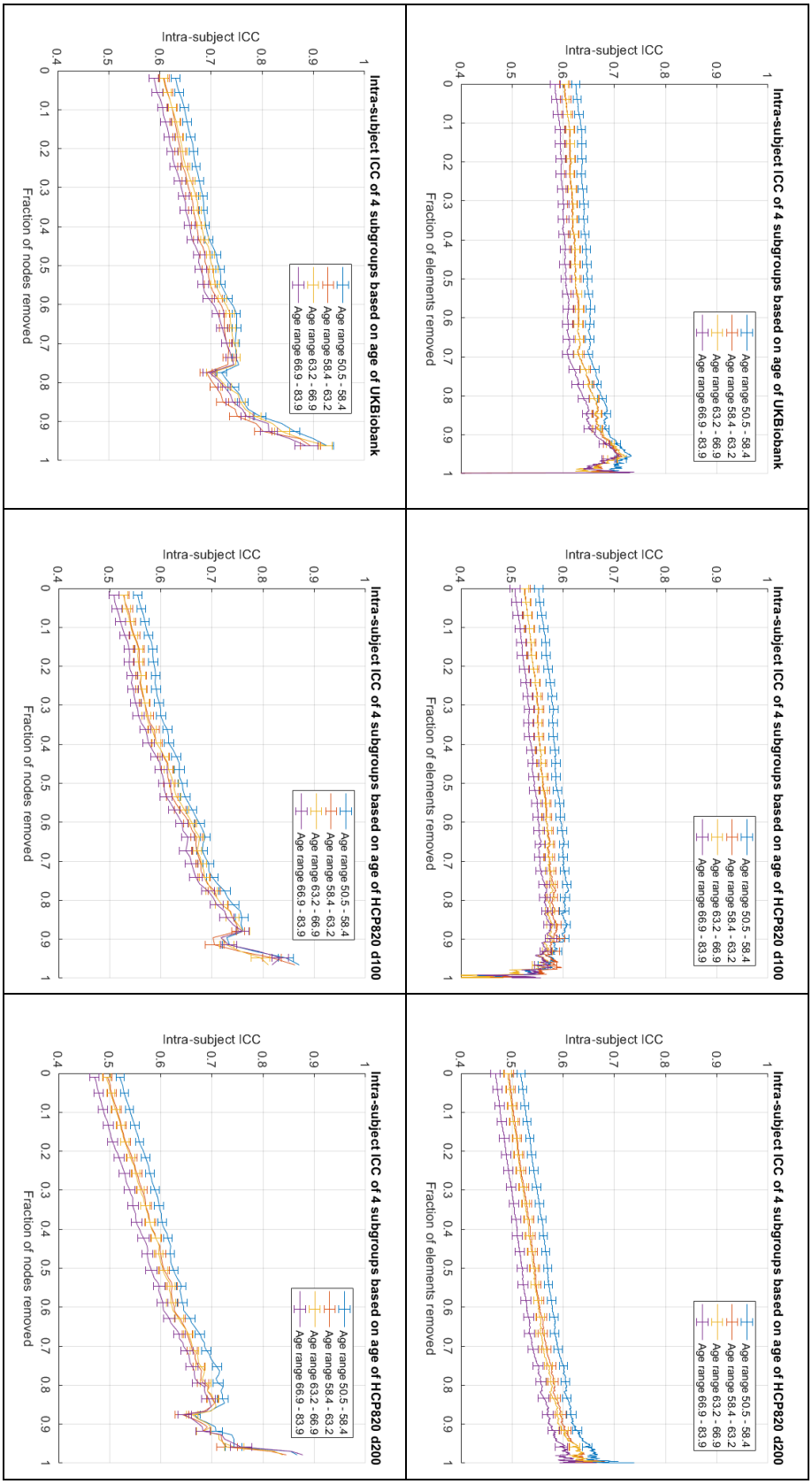

Matching accuracy of men vs women in UKBiobank
Matching accuracy of men vs women in HCP820 d100
Matching accuracy of men vs women in HCP820 d200
Matching accuracy of men vs women in UKBiobank
Matching accuracy of men vs women in HCP820 d100
Matching accuracy of men vs women in HCP820 d200
men
women
Matching accuracy (%)
Fraction of removed elements
Fraction of removed nodes

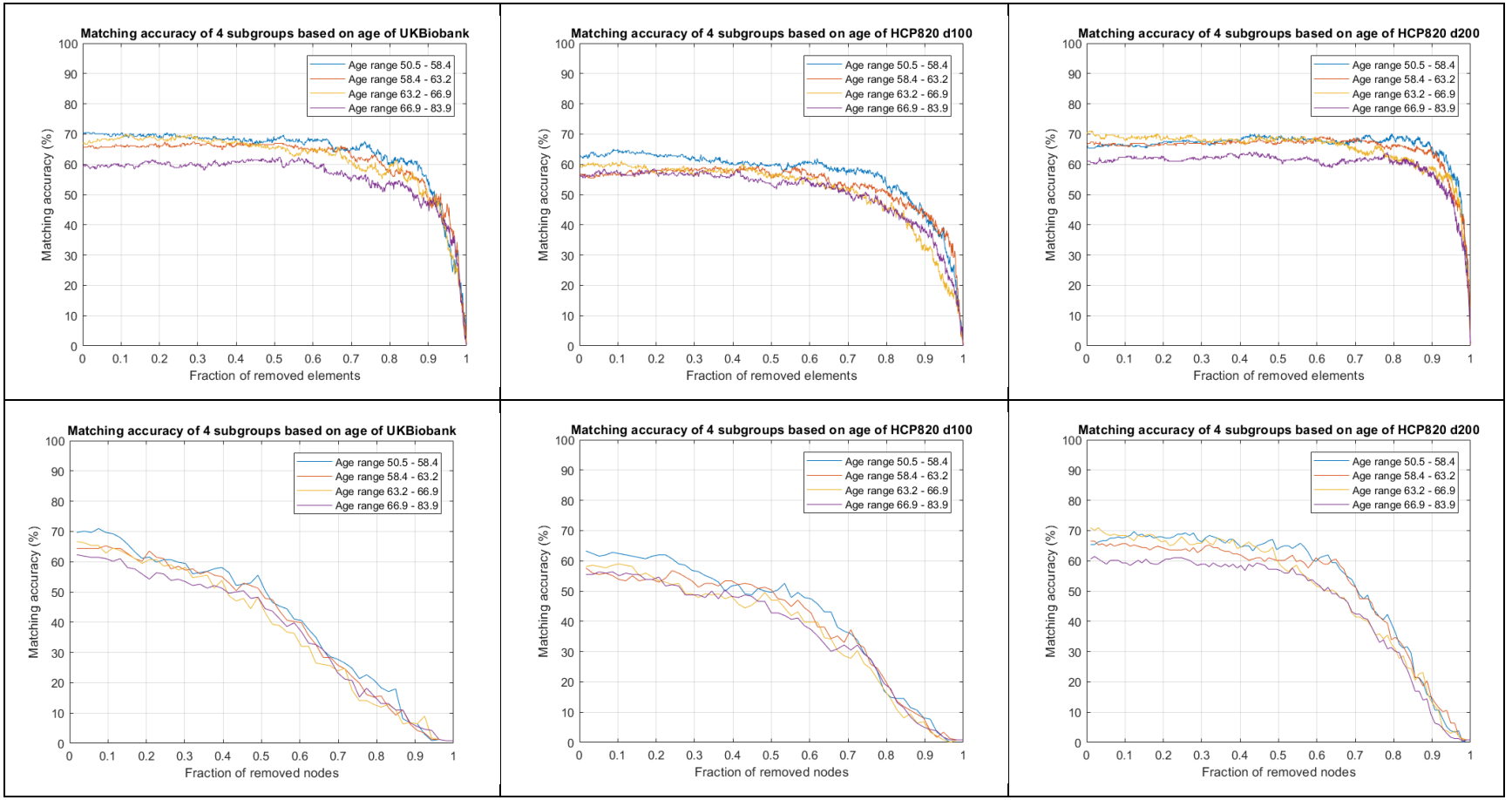
Appendix G – IST based subset figures of other parcellation schemes




Matching accuracy of 4 subgroups based on IST of UKBiobank
Matching accuracy of 4 subgroups based on IST of HCP820 d100
Matching accuracy of 4 subgroups based on IST of HCP820 d200
Matching accuracy of 4 subgroups based on age of UKBiobank
Matching accuracy of 4 subgroups based on age of HCP820 d100
Matching accuracy of 4 subgroups based on age of HCP820 d200
IST range 0.34 - 1.23
IST range 1.23 - 1.51
IST range 1.51 - 1.78
IST range 1.78 - 5.10
Matching accuracy (%)
Fraction of removed elements
Fraction of removed nodes

Appendix H – Age based subset figures of other parcellation schemes




Matching accuracy of 4 subgroups based on age of UKBiobank
Matching accuracy of 4 subgroups based on age of HCP820 d100
Matching accuracy of 4 subgroups based on age of HCP820 d200
Age range 50.5 - 58.4
Age range 58.4 - 63.2
Age range 63.2 - 66.9
Age range 66.9 - 83.9
Matching accuracy (%)
Fraction of removed elements
Fraction of removed nodes

Appendix I – ICN intra-subject ICC and MA of other parcellation schemes.

Table 3 MA and intra-subject ICC with standard error of the mean (SEM) for the different ICN's of the remaining three parcellation schemes only using the included 470 subjects of the longitudinal set

Intrinsic connectivity network	**UKBiobank**		**HCP820 d100**		**HCP820 d200**	
	ICC ± SEM	MA (%)	ICC ± SEM	MA (%)	ICC ± SEM	MA (%)
cogn	-	-	0.57 ± 0.0072	9.9	0.61 ± 0.0058	24
DAN	0.77 ± 0.0070	4.0	-	-	-	-
DMN	0.64 ± 0.0060	19	0.66 ± 0.0072	8.0	0.61 ± 0.0057	28
FPN	0.61 ± 0.0072	15	-	-	-	-
SMN	0.59 ± 0.0073	8.7	0.53 ± 0.0073	8.4	0.55 ± 0.0066	12
Subcort	-	-	0.49 ± 0.0068	3.5	0.44 ± 0.098	7.3
Temp	-	-	-	-	-	-
VAN	-	-	-	-	-	-
Visual	0.61 ± 0.0065	12	0.54 ± 0.0068	15	0.56 ± 0.0063	20

Appendix J – histogram ICN origin of best number of nodes and elements for other parcellation schemes