Evaluation of Synthetic MRI for Clinical Use

Teresa Helmersson

LIU-IMH/RV-A--10/002--SE

Linköping 2010
Evaluation of Synthetic MRI for Clinical Use

Master’s Thesis
Biological Engineering

Teresa Helmersson
Linköping 2010

Supervisor: Marcel Warntjes
CMIV, Linköpings Universitet

Examiner: Peter Lundberg
IMH, Linköpings Universitet
Abstract
Conventional Magnetic Resonance Imaging (MRI) is a qualitative method for obtaining images of soft tissues in patients. Conventional MRI is the standard method used today and it results in gray-scale images in which the different magnetic properties of biological tissues determine the image contrast. However, the magnitude of the measured signal is only relative and therefore not directly comparable between images. Synthetic MRI is a relatively new technique which can be used to post-synthesize different images based on absolute measurement of several magnetic properties of tissues. Synthetic MRI can therefore provide quantitative information together with the contrast images.

In order to use synthetic MRI clinically an evaluation of the image quality and diagnostic ability is required. The purpose of this thesis is to evaluate if synthetic MRI and conventional MRI produce images with equal contrast.

A study was designed and conducted for statistical evaluation of contrast and Contrast-to-Noise Ratio (CNR) generated with different imaging methods. A total of 22 patients were examined using both conventional MRI and synthetic MRI and the results were pairwise analyzed.

The contrast and CNR could not be stated as equal for the imaging methods. Typically the contrast was higher in the synthetic images for the T1 and T2 weighted images. This was not observed with CNR which suggests that the noise is higher in the synthetic images. The higher contrast obtained in synthetic images resulted in a better separation of different tissues using synthetic MRI. The synthetic T2 FLAIR images contained artifacts that are not good for clinical use. However the fact that the different imaging methods produce different image quality is not proven to be clinically decisive.
Acknowledgement
This thesis has been conducted at the Department of Medical and Health Sciences, Linköpings University on behalf of SyntheticMR and CMIV (Center for Medical Image Science and Visualization). I want to thank everybody that have helped and believed in me during this work. A special thanks to my supervisor Marcel Warntjes for giving me this opportunity and guiding me, Erika Peterson and Janne West at SyntheticMR for all support, Ida Blystad for medical information and support, Martin Olsson at the Department of Mathematics and Olle Eriksson at the Department of Computer Science for statistical support.
## Contents

1 Introduction .......................................................................................................................... 1  
   1.1 The Principle of Magnetic Resonance Imaging ................................................................. 1  
      1.1.1 MRI Physics ............................................................................................................. 1  
      1.1.2 Examination Settings ............................................................................................ 2  
   1.2 Project Background ....................................................................................................... 3  
   1.3 Objectives ...................................................................................................................... 4  
   1.4 Resources .................................................................................................................... 4  

2 Statistical Theory .................................................................................................................. 5  
   2.1 Descriptive statistics ..................................................................................................... 5  
      2.1.1 Averages and Spread ............................................................................................. 5  
      2.1.2 Visualization of Distributions .............................................................................. 6  
      2.1.3 Plots for Evaluating Agreement between Two Variables ..................................... 7  
   2.2 Analytic Statistics ........................................................................................................ 8  
      2.2.1 Hypotheses and Test Statistic .............................................................................. 8  
      2.2.2 Significance and Strength ..................................................................................... 8  

3 Materials and Methods ....................................................................................................... 11  
   3.1 Subjects ....................................................................................................................... 11  
   3.2 Evaluated Parameters .................................................................................................. 11  
   3.3 Sampling ..................................................................................................................... 12  
   3.4 Blinding ...................................................................................................................... 14  
   3.5 Statistical methods ..................................................................................................... 15  
      3.5.1 Alternative Statistical Methods ........................................................................... 16  
   3.6 Hypothesis .................................................................................................................. 17  
   3.7 Software .................................................................................................................... 17  

4 Results from the Statistical Analysis .................................................................................... 19  
   4.1 Hypothesis .................................................................................................................. 19  
   4.2 T1 Weighted Images .................................................................................................... 19  
   4.3 T2 Weighted Images .................................................................................................... 23  
      4.3.1 Pathology Measurements ..................................................................................... 25  
   4.4 T2 FLAIR Images ....................................................................................................... 27  

5 Radiologists Experience .................................................................................................... 33  
   5.1 Image quality .............................................................................................................. 33  
   5.2 Problems with Clinical Use of Synthetic MRI .............................................................. 33  
   5.3 Benefits with Synthetic MRI in Clinical Use ............................................................... 33
6 Discussion........................................................................................................................................ 35
6.1 Overview of the Results............................................................................................................... 35
6.2 T1 Weighted Images.................................................................................................................. 35
   6.2.1 Contrast ............................................................................................................................ 35
   6.2.2 CNR ............................................................................................................................... 36
   6.2.3 Negative Contrast Measurements .................................................................................. 37
   6.2.4 Summary ....................................................................................................................... 37
6.3 T2 Weighted Images.................................................................................................................. 37
   6.3.1 Contrast ............................................................................................................................ 38
   6.3.2 CNR ............................................................................................................................... 38
   6.3.3 Summary ....................................................................................................................... 39
   6.3.4 Pathology Measurement ............................................................................................... 39
6.4 T2 FLAIR Images...................................................................................................................... 40
   6.4.1 Contrast ............................................................................................................................ 40
   6.4.2 CNR ............................................................................................................................... 40
   6.4.3 Negative Contrast Measurements .................................................................................. 42
   6.4.4 Summary ....................................................................................................................... 42
6.5 Resulting Scan Time................................................................................................................ 42
6.6 General Problems..................................................................................................................... 42
   6.6.1 Unexpected Relation of the Tissue Intensity ................................................................. 43
   6.6.2 Selected Structures and ROI Placement ....................................................................... 43
   6.6.3 Imaging Problems ......................................................................................................... 43
   6.6.4 Software Problems ....................................................................................................... 44
   6.6.5 Simultaneous Confidence Level .................................................................................... 44
6.7 Further Development and Research......................................................................................... 44
   6.7.1 Improvements for Similar Studies ............................................................................... 44
   6.7.2 Suggestions for Future Research ................................................................................... 45

7 Conclusions.................................................................................................................................. 47

Bibliography....................................................................................................................................... 49

Appendix A .......................................................................................................................................... 51
   A.1 Shapiro-Wilk test - Test for Normality.................................................................................. 51
   A.2 t-test – Significance Analysis for Normal Distributed Variables ......................................... 51
   A.3 Wilcoxon Signed Rank Test – Nonparametric Test of Symmetrical Distributions .......... 52
   A.4 Analysis of Variance – ANOVA ......................................................................................... 53
   A.5 F-test – Test of Equal Variances ......................................................................................... 55

Appendix B.......................................................................................................................................... 57
List of Figures

Figure 1-1. Proton precession and the resulting signal................................................................. 1

Figure 1-2. Relaxation curves ...................................................................................................... 2

Figure 1-3. Quantification map for the T1 parameter and a synthetic T1 weighted image. ....... 3

Figure 2-1. Histogram and Box-and-Whiskers plot................................................................. 7

Figure 2-2. Q-Q plot.................................................................................................................... 7

Figure 2-3. Bland-Altman plot and correlation plot ................................................................. 8

Figure 3-1. Different MRI weightings ....................................................................................... 12

Figure 3-2. Placement of all ROIs ........................................................................................... 14

Figure 4-1. Bland-Altman plot for the contrast and correlation plot for CNR in T1 weighted images .............................................................................................................. 19

Figure 4-2. Contrast in the T1 weighted images ......................................................................... 20

Figure 4-3. Contrast between gray matter and white matter in the T1 weighted images ....... 20

Figure 4-4. CNR in T1 weighted images ................................................................................... 21

Figure 4-5. CNR between white and gray brain matter in T1 weighted images ..................... 21

Figure 4-6. Bland-Altman plot and correlation plot for contrast in T2 weighted images ....... 23

Figure 4-7. Contrast in T2 weighted images ............................................................................. 23

Figure 4-8. Contrast between gray brain matter and white brain matter in T2 weighted images ...... 24

Figure 4-9. CNR in T2 weighted images .................................................................................. 24

Figure 4-10. CNR between gray brain matter and white brain matter in T2 weighted images .... 25

Figure 4-11. Bland-Altman plot and correlation plot of CNR in T2 weighted images ........... 25

Figure 4-12. Contrast and CNR between lesions and white matter in T2 weighted images .... 27

Figure 4-13. Bland-Altman plots of the CNR and contrast in T2 FLAIR images....................... 28

Figure 4-14. Contrast in the T2 FLAIR images .......................................................................... 28

Figure 4-15. Contrast measurements between gray and white brain matter in the T2 FLAIR images ............................................................................................................... 29

Figure 4-16. CNR in T2 FLAIR images ..................................................................................... 29

Figure 4-17. CNR between gray and white brain matter in T2 FLAIR images ....................... 30

Figure 5-1. Artifacts in T2 FLAIR weighted synthetic images ..................................................... 34

Figure 6-1. Noise in T2 FLAIR images ....................................................................................... 41
List of Tables
Table 4-1. Results from the analysis of T1 weighted images. ............................................................... 22
Table 4-2. Results from the analysis of T2 weighted images. ............................................................... 26
Table 4-3. Results from the analysis of lesions in T2 weighted images. ........................................... 27
Table 4-4. Results from the analysis of T2 FLAIR images. ................................................................. 31
Table A-1. Two-Way Analysis of Variance ......................................................................................... 55
Abbreviations and Nomelclature

CI  Confidence Interval
CMIV  Center for Medical Image Science and Visualization
CNR  Contrast-to-Noise Ratio
CS  Centrum Semiovale
CSF  CerebroSpinal Fluid
DF  Degrees of Freedom
FC  Frontal Cortex
FLAIR  Fluid Attenuated Inversion Recovery
Ge  Genu
GM  Gray Brain Matter
H₀  Null hypothesis
H₁  Alternative hypothesis
MRI  Magnetic Resonance Imaging
NMR  Nuclear Magnetic Resonance
OC  Occipital Cortex
PD  Proton Density
PDF  Probability Density Function
p-value  Probability that the null hypothesis is true
ROI  Region Of Interest
Sp  Splenium
SS  Sum of Squares
STD  Standard Deviation
T₁  Longitudal spin-lattice relaxation time
T₂  Transversal spin-spin relaxation time
TE  Echo time
Th  Thalamus
TI  Inversion time
TR  Repetition time
WM  White Brain Matter
X  Stochastic variable
x  Observation from the stochastic variable X
\bar{x}  Sample mean
\bar{x}_\cdot  Sum over sample
s  Sample standard deviation
s^2  Sample variance
\mu  Mean value
\sigma  Standard deviation
\sigma^2  Variance
n  Sample size/Group size (if several groups)
N  Total sample size (if several groups)
\alpha  Significance level
\theta  Flip angle
1 Introduction

This chapter provides a brief description of the principles behind MRI followed by the background and aim for this thesis. At the end of this chapter the available resources for the thesis are presented.

1.1 The Principle of Magnetic Resonance Imaging

The Nuclear Magnetic Resonance, NMR, technique is used in the field of Magnetic Resonance Imaging, MRI, to retrieve images of soft tissues in patients. NMR is a method for detecting protons or atom nuclei in materials. In MRI the signal comes mainly from hydrogen nuclei in water. (1)

1.1.1 MRI Physics

Every proton, hydrogen nucleus, has a so called spin. The spin is a magnetic dipole and can be seen as a vector (2). The spin precesses around an axis which can be observed as an oscillating local magnetic field and retrieved as a signal (2), see Figure 1-1. All the proton spins in a human body normally points in different direction and therefore the net signal is zero.

![Figure 1-1. The proton precession and the resulting signal.](image)

If a strong magnetic field is applied over a body the proton spin will align with the direction of the field because it is the equilibrium state, i.e. all spins point in the z-direction. An RF-pulse is then transmitted which will flip the spins in the body to the xy-plane, precession will occur around the z-axis. Since all spins now precess in phase in the xy-plane a strong signal can be retrieved. After the pulse the spins start to relax back to the equilibrium state, i.e. parallel with the magnetic field. The relaxation is called longitudinal spin-lattice relaxation and has the time constant T1. There is another relaxation phenomenon called the transverse spin-spin relaxation which has the time constant T2. The T2 relaxation occurs because the spins experience different local magnetic fields and will therefore precess with slightly different frequencies. This will result in a decreasing signal. (2)

The signal retrieved depends on the relaxation properties and proton density, PD, of the measured volume (1). A high PD result in that more spins can align together, therefore the magnitude of the resulting signal becomes high. The magnitude will decrease with the relaxations and therefore a signal dependent on e.g. the T2 relaxation can be retrieved if the spins are allowed to relax for a certain time. Different tissues in the body relax with different rates depending on the interaction between molecules (2). More compact tissues, like fat, have more interaction between the molecules which increases the relaxation rate. Different relaxation curves for brain tissues can be seen in Figure 1-2.
1.1.2 Examination Settings

Different parameter settings during the examination will highlight different properties. In the instant after one RF-pulse the contrast is primarily dependent on PD since no relaxation has occurred, i.e. a PD weighted image can be retrieved (1). Since tissues have different relaxation rates the contrast will be affected by the relaxation. The repetition time, TR, and the echo time, TE, can be altered in order to highlight other properties than the PD such as T1. TE is the time between sending one pulse and retrieving the signal. All the information for one image cannot be retrieved by using only one pulse and therefore the pulse have to be repeated with certain intervals, TR.

The contrast in a T1 weighted image is highly dependent on the T1 relaxation. The graph to the left in Figure 1-2 illustrates the T1 relaxation as a function of TR in different brain tissues. In the beginning the soft brain tissues relax quickly while CSF relaxes at a much slower rate. If a short TR is used, e.g. 500 ms, then the tissues will not have relaxed back totally before the subsequent pulse. The magnitude in the xy-plane will then correspond to the magnitude in the z-direction before the pulse and if image information is retrieved it will mainly depend on the T1 relaxation. To retrieve a T1 weighted image a short TE is used in order to minimize the dependency of the T2 relaxation, see the graph to the right in Figure 1-2. (1)

To retrieve a T2 weighted image the opposite settings are used, long TE and long TR. If a long TR is used most of the spins are able to relax back to the initial state between the pulses. After an RF pulse the decay of the signal intensity will therefore be similar to the graph to the right in Figure 1-2. Directly after the pulse the intensity is primarily dependent on the PD but after approximately 100 ms the difference in intensity is highly dependent on the T2 relaxation. This is because the intensity in the soft tissues has decayed much more than the intensity in CSF. (1)

It is important to realize that it is not the highest signal intensity that is primarily of interest but the largest difference in intensity between the tissues. This is because the ability to distinguish tissues is the most important aspect of MRI. If only the highest signal intensity was of interest a long TR and short TE should be used but then only PD weighted images would be retrieved.
Inversion recovery is another important approach in MRI. Inversion recovery can be used to extinguish the signal from one tissue which increases the contrast. This is done by first sending one 180 degrees inversion pulse and then sending an additional 90 degree pulse after a certain time of recovery. The time between the first and second pulse is called the inversion time, TI. This parameter is used to decide which tissue to extinguish. FLAIR, fluid attenuated inversion recovery, is used to extinguish CSF in T2 weighted images of the brain. This can be useful if there are lesions, diseased areas, with similar relaxation properties as CSF. (4) Today these T2 FLAIR images have replaced PD weighted images in brain examinations.

Specific sequences with predefined values for TR, TE etc. are used during brain scanning to provide sets of images, so called image stacks. The image stacks contain slices to represent the examined part of the patient, e.g. the brain. This means that T2 weighted images for the whole brain can be retrieved with one sequence.

1.2 Project Background

MRI examinations performed today retrieve contrast images depending on the properties of tissues but the measurements are only qualitative. The radiologists have learned to interpret these images and make statements depending on the visual patterns but they cannot get any diagnostic support by measurements of the intensities in the image (5).

Quantification of the parameters T1, T2 and PD is possible to do within minutes (3) but the resulting images does not resemble the conventional contrast images, see Figure 1-3, and therefore the radiologists are uncomfortable to use them for diagnosis. The quantification can still provide useful information since the relaxation times for a specific tissue do not vary much between individuals and therefore it is possible to segment tissues (3). In the future it may even be possible to automatically recognize diseases.

![Figure 1-3 In the left image a quantification map for the T1 parameter can be seen and in the right image a T1 weighted synthetic contrast image.](image)

If T1, T2 and PD for one voxel in an image are known the signal intensity in the contrast image, S, can be calculated since it depends on these parameters and TR, TE and flip angle (θ) of the pulse (3).

\[
S \propto PD \frac{1 - e^{-(TR/T1)}}{1 - e^{-(TR/T1)} \cos \theta} e^{-(TE/T2)}
\]  

(1.1)
Using Eq(1.1) the regular contrast images can be post-synthesized and provided together with absolute values of the relaxation times and PD, this is the concept of synthetic MRI. Theoretically synthetic MRI and conventional MRI should provide identical images but this has not been validated.

Three parameters commonly used to describe the image quality are resolution, noise and contrast. The signal intensity and contrast are related to the noise since the effect of the noise is dependent on how strong the signal/contrast is. The signal intensity and image resolution can be altered by different scanner settings; generally a higher resolution and signal intensity can be provided at the cost of longer scan time. Images can also contain errors, so called artifacts. Artifacts can arise during image acquisition, e.g. if the patient moves, but it can also arise while post processing the images. In this thesis the focus has been on the contrast and noise since these parameters are important for the ability to distinguish different tissues.

One important aspect in synthetic MRI is that only one sequence is required to retrieve the parameters T1, T2 and PD (3) for synthesizing a variety of different image types. In conventional MRI one sequence is used for each image type, T1 weighted, T2 weighted and T2 FLAIR. Even if the sequence used for quantification is long it is still shorter than the three conventional sequences together.

1.3 Objectives
The purpose of this thesis is to design and conduct a study for evaluation of synthetic MR images and specifically to answer the following questions:

- Do the synthetic and conventional MR images provide the same contrast?
- How does the noise affect the contrast in conventional and synthetic MR images?
- Is it possible to save time by using synthetic MRI instead of conventional MRI?

1.4 Resources
A set of 36 patients were provided for this study. Both conventional and synthetic images were available for each individual. Out of these 36 patients 22 were included in the evaluation; the exclusions were made because of errors in the examinations. All images were from scans of the whole brain in the transversal plane. A radiologist was available during two weeks for medical consulting and verification of the measurements.
2 Statistical Theory

This chapter contains concepts used in statistical analysis. It is primarily directed to those with limited knowledge about statistics and for their possibility to understand the following chapters. The chapter is divided into descriptive statistics and analytic statistics. For description of the specific statistical methods used during this thesis see Appendix A.

2.1 Descriptive statistics

Descriptive statistics is used to get a quick overview of the sampled data. The distribution of the sample can be inspected visually with tables and diagrams, and the average and spread is often pinpointed.

2.1.1 Averages and Spread

Averages and spread is used to present the magnitude and variation of the data. The most commonly used average is the mean which is mathematically defined as (6)

\[ \mu_x = E[X] = \begin{cases} \int_{-\infty}^{\infty} x f_x(x) dx, & \text{for continuous stochastic variables} \\ \sum_k k p_x(k), & \text{for discrete stochastic variables} \end{cases} \]  

(2.1)

where \( E \) is the expected value operator, \( f_x(x) \) is the probability density function, PDF, and \( p_x(k) \) is the probability function. If the average is described with mean value then standard deviation, Eq(2.2), is used to describe the spread (6).

\[ \sigma_x = \sqrt{V[X]} = \sqrt{E[(X - \mu_x)^2]} \]  

(2.2)

\( V \) is the variance operator and the variance, i.e. the squared standard deviation, is sometimes used instead of the standard deviation. If the true mean and standard deviation is not known then they are approximated with the sample mean, Eq(2.3) and sample standard deviation Eq(2.4), \( n=sample \) size (6).

\[ \bar{x} = \frac{\sum_i x_i}{n} \]  

(2.3)

\[ s = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{n - 1}} \]  

(2.4)

If several samples are combined in an analysis and the standard deviation can be assumed to be the same for all samples then a total standard deviation can be calculated as (7):

\[ s_{tot} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_a - 1)s_a^2}{N - a}} \]  

(2.5)

Different denotations for the true values and the approximations are appropriate to use since all approximations have some uncertainty.
Another commonly used average is the median, the middle value in a sample. The mean is almost exclusively used in statistics used in school but there are some cases where the median is preferred (8):

- When the distribution is skewed because the extreme values will affect the mean a lot
- When there are one or two outliers that affect the mean unreasonably
- When the data is categorical since category numbers have no numerical value

If the average is described with the median then the spread should be described with quantile distance. If the arranged observations are divided into equally sized groups then the quantiles are the values of the boundaries between the groups. Quartile is most commonly used where the observations are divided into four groups. (8)

**Skewness**

The skewness can be used as a supplementary measure to the average and spread if the distribution is unknown. The skewness can be useful while evaluating if a sample is normally distributed, if the sample is skewed then normal distribution cannot be assumed (9).

The variance was presented in previous section and it is also called the second moment, $m_2=s^2$. For measurement of skewness the third moment is also used (10; 9).

$$m_3 = \frac{\sum (x_i - \bar{x})^3}{n} \tag{2.6}$$

With these the coefficient of skewness can be calculated as (10; 9):

$$skewness = \frac{m_3}{m_2^{3/2}} \tag{2.7}$$

The skewness is asymptotically zero for a normally distributed sample (10). If the distribution is skewed to the right side the skewness becomes positive (9).

**2.1.2 Visualization of Distributions**

There are two common ways to visualize a distribution, histogram and Box-and-Whiskers plot, see Figure 2-1. These plots are primarily used to visualize spread and skewness. It can be difficult to determine if the sample is symmetrical by looking at the histogram, especially for small sample sizes. The Box-and-Whiskers plot is more sensitive to skewness and therefore more suitable for such investigation (10).

There are also plots specially designed for investigating if a sample belongs to a specific distribution, e.g. quantile-quantile plot (Q-Q plot) or proportion-proportion plot (P-P plot). Q-Q plot compares the distribution of the observed sample with the quantiles for the assumed distribution (10), see Figure 2-2. If the sample belongs to the assumed distribution then the observations will follow a straight line, this line is often marked in statistical software. The P-P plot compares the PDF of the sample with the assumed PDF (10), and the plot is analyzed in similar way as the Q-Q plot. Together these plots are often called normal probability plots because they are frequently used for evaluating normality but they can also be used for other distributions.
2.1.3 Plots for Evaluating Agreement between Two Variables

There are two methods primarily used for visualizing the agreement between paired samples. One method is a so called correlation plot, see Figure 2-3. The samples agree if the observations fit the line with slope one and intercept zero. However observations will never give a perfect match. It is tempting to use correlation analysis but the problem is that a high correlation does not indicate exact agreement and it should therefore be used with caution.

The so called Bland-Altman plot is thought to be more informative than the correlation plot, see Figure 2-3 (11). The difference in each pair is plotted against the mean for each pair. The mean is used as an estimate of the true value and then the spread of the differences can be evaluated (11). Two samples agree if the dots are randomly distributed around zero and no trends can be seen (11). The mean difference together with an approximated confidence interval, \( \mu_d \pm 2 \cdot \sigma_d \), is often marked in the plot to visualize the distribution. Transformation of the data can also be useful if the spread is increasing with the mean, e.g. a logarithmic transformation.

![Figure 2-1. To the left a histogram can be seen and to the right a Box-and-Whiskers plot. The histogram is generated from 100 simulated observations from a standard normal distribution. The histogram is similar to the normal PDF but it is still not obvious that the variable is normal distributed. This illustrates that even with as many as 100 observations the histogram is still not perfect for assuring a certain distribution type.](image1)

![Figure 2-2. Q-Q plot for testing of normal distribution. This plot is generated in SPSS17 and since the observations follow the line well the sample is probably normally distributed.](image2)
Figure 2-3. To the left a Bland-Altman plot can be seen and to the right a correlation plot. The same simulated observations from a standard normal distribution are used in both plots. It can be seen in the Bland-Altman plot that the mean difference is near zero and the differences are distributed evenly on both sides of the mean difference, independent of the size of the observations which indicates agreement between the methods. When a correlation plot, right plot, is used instead it is clearly more difficult to state agreement which is indicated if the observations follow the straight line.

2.2 Analytic Statistics
Analytical methods are often applied to calculate the level of agreement or disagreement. Conclusions from the analysis are stated to be significant if they have a very low probability to be false. However it is important to beforehand state what a low probability is. This section deals with different concepts of analytic statistics, information about different statistical tests can be seen in Appendix A.

2.2.1 Hypotheses and Test Statistic
Before statistical analysis two hypotheses are presented, a null hypothesis and an alternative hypothesis denoted $H_0$ and $H_1$ respectively. The goal is to reject the null hypothesis in favor of the alternative hypothesis hence the goal is to verify the alternative hypothesis. This approach is used since statistics can be used only to reject hypothesis. One common mistake is to state that the null hypothesis is true if it cannot be rejected. (8) The result may indicate that the null hypothesis is true but the test reveals no significant evidence for such a statement.

For each analysis a test statistic is proposed. The test statistic is calculated from the sample and follows a certain distribution if the null hypothesis is true. Comparison with the cut off values from the assumed distribution determines if the null hypothesis can be rejected. (6) The test statistic can be a single value but also a confidence interval. If a confidence interval is used for the analysis then the whole interval has to indicate rejection for the null hypothesis to be rejected.

2.2.2 Significance and Strength
A result is stated to be significant if it has low probability to be false. The significance level is pointed out by the statistician and traditionally 0.05, 0.01 or 0.001 is used (8). In medicine p-value is often used which is the probability that the null hypothesis is true and it is calculated from the observations (9). It is important to differentiate between significance level and p-value. Significance level is a beforehand stated value of the highest acceptable probability to reject a true null hypothesis, also called type I error (6). Rejection of a hypothesis can occur if the p-value is smaller than the significance level (6).
Confidence level, $1-\alpha$, is another commonly used term especially together with confidence intervals. This is a measure of the certainty that the interval contains the examined parameter. (6)

The strength or power of a test is the probability to reject a false null hypothesis (6), often denoted with $1-\beta$. This is connected to the probability of type II error, $\beta$, i.e. not rejecting a false null hypothesis (6). $\beta$ is primarily dependent on the size of the investigated effect, e.g. the difference between two samples, and the sample size (12). Large effects are easier to detect and larger samples can detect smaller effects. The type II and type I errors are connected; if one error is reduced the other will increase.

A simultaneous confidence level is appropriate to use if several confidence intervals are calculated based on the same data, otherwise the intervals cannot be used for a combined conclusion (7). For independent tests the simultaneous confidence level is the product of each independent confidence level, basic probability theory. If independency cannot be assumed then Bonferroni’s inequality (13) can be used to calculate the simultaneous confidence level with Eq(2.8).

$$1 - \alpha_{sim} = P\left(\bigcap_{i} A_i\right) = 1 - P\left(\bigcup_{i} A_i\right) \geq 1 - \sum_{i} P(A_i) = 1 - r\alpha \quad (2.8)$$

$P(A_i) = \alpha$ which denotes the significance level of the individual intervals, $r$ denotes the number of tests and $\alpha_{sim}$ is the simultaneous significance level.

Since Eq(2.8) is based on probability theory it is applicable on all existing tests and confidence intervals. Other estimations of simultaneous confidence level like Schaffé’s method or Tukey-Kramer’s method can be used but then the variable has to be normally distributed (7).
3 Materials and Methods
This chapter contains information about the material used in the study and the different methods applied during the analysis. In the first section the actual parameters for statistical analysis (contrast and CNR) are presented and the following sections contain information about how and which data that were retrieved. The statistical methods used in the analysis and how they are used for evaluating the data are presented in the last sections.

3.1 Subjects
The selection of patients for this work was based on neurological criteria. This thesis is a part of a larger study on multiple sclerosis, MS, and ischemia and therefore the subjects have either the diagnosis MS, the diagnosis ischemia or it is unknown if the patient has MS or ischemia. The diagnoses are based on several different tests and examinations which the patients’ physician has evaluated.

A total number of 36 patients were provided for this study and 22 of these were selected as appropriate to use. Exclusion of patients was due to several different errors like missing data from examination or errors in the image acquisition which generated extreme artifacts. Both conventional MR images and synthetic MR images of the brain in the transversal plane were provided from all patients.

3.2 Evaluated Parameters
It is important in MRI to be able to distinguish different tissues and therefore the contrast is important. The contrast can be obtained through measurements in images since contrast depends on the differences in intensity between two tissues. The standardized contrast can be calculated in as (14):

$$\frac{I_1 - I_2}{I_1 + I_2}$$

(3.1)

$I_i$ represent the signal intensity in tissue type $i$, in this case the mean value from the chosen volume. For all comparisons $I_1$ was the tissue with the highest theoretical intensity, see the different weightings in Figure 3-1. In T1 weighted images the white matter is displayed as brightest followed by the gray matter and the CSF as darkest. T2 weighted images has the opposite relation. The T2 FLAIR images are similar to the T2 weighted images but the CSF is displayed as dark. The choice to always use the tissue with highest intensity as $I_1$ was made to get positive contrasts only. However negative contrast occurred during the evaluation and the data series containing negative values are treated separately in the discussion, see section 6.6.1.

High noise levels in images can make the difference in intensity between tissues less visible which makes the contrast-to-noise ratio (CNR) a relevant supplementary measure to evaluate how the noise affect the contrast. CNR can be obtained through measurement of tissue intensities and the standard deviation of the noise (14):

$$\frac{I_1 - I_2}{\sigma_{noise}}$$

(3.2)
Materials and Methods

Figure 3-1. The three weightings that are primarily used for retrieving conventional images of the brain can be seen above. In T1 weighted images the white matter is brightest, next is the gray matter and darkest is the CSF. In T2 weighted images the intensity relations are reversed and T2 FLAIR images are like T2 images but CSF is dark.

It is difficult to retrieve a reliable measure of the noise in one image stack. One commonly used method is to measure the noise in a signal free region, outside the head. However this is not possible in the synthetic images since these values are zero. The noise is assumed to be homogenous in one image stack which means that the noise should be the same for homogenous tissues independent on the tissue. The standard deviations from each region of interest (ROI) were used to retrieve the median variance as an approximation of the standard deviation of the noise, for information about placement of ROIs see section 3.3. Since all ROIs were placed in as homogenous areas as possible their standard deviation should then represent the image noise. There were though some outliers, probably due to different ROI sizes in different areas and small misplacement of the ROIs. The median was therefore thought to be a more accurate estimate than using some mean calculation, e.g. Eq(2.5). The mean is also difficult to calculate accurately since it demands the number of observations, voxels, in the ROIs which are not given by the software.

Both contrast and CNR are calculations from measured values on a continuous scale making the variables quantitative.

3.3 Sampling

By placing ROIs at interesting positions the intensity and deviation in different tissues can be measured. These measurements are used to calculate the contrast and CNR between different tissues. In the brain it is interesting to distinguish between gray matter, white matter, CSF and possible lesions. For representing gray matter the following structures were used:

- Thalamus, one ROI in each hemisphere
- Occipital cortex, one ROI in each hemisphere
- Frontal cortex, one ROI in each hemisphere
For representing white matter the following structures were used:

- Centrum semiovale, one ROI in each hemisphere
- Splenium
- Genu

For representing CSF the following structure was used:

- Back of the anterior horn, one ROI in each hemisphere

The position of the ROIs can be seen in Figure 3-2. The sizes of the ROIs were approximately 4.5 mm in diameter and the mean and standard deviation were collected from all ROIs. In some cases it was not possible to use a diameter of 4.5 mm due to thin structures. In those cases the aim was to achieve equal ROI size in both hemispheres. The mean value in one ROI was assumed to be the intensity of the marked structure. If two ROIs were placed in the same structure then the mean of the ROIs was assumed to be the intensity of the structure. This is because the mean value is a better estimate of the true intensity of the structure if there are differences between the hemispheres (15). This intensity values were used to calculate the contrast and CNR between different tissues as described in section 3.2.

The contrast, Eq(3.1), and CNR, Eq (3.2), were calculated using different combinations of the presented structures. The contrast and CNR between different structures of the same tissue type, e.g. frontal cortex and occipital cortex, was not used. This is because the difference in intensity between these structures is supposed to be low and therefore the effect of the different methods should be extremely small. Small effects are hard to detect if the sample size is not very large and therefore these comparisons were not thought to give any additional information.

The neurological background was the base for the measurement of intensity in lesions. There are lesion patterns that are common in both MS and ischemia which means if no other parameter is studied it would be impossible to separate between these diagnoses. From each patient with the diagnosis MS or ischemia one lesion of this type was chosen, the patients with unknown disease were excluded since most of them had no lesions. Since the lesions are only present in white brain matter no other tissues were used for calculations of contrast and CNR.

All the relevant information was taken from T1 weighted image stacks, T2 weighted image stacks and T2 FLAIR image stacks. The ROIs in lesions were only placed in T2 weighted images since it is the T2 weighted images that are used to evaluate this type of lesions (5). All image stacks were from brain scans in the transversal plane.

The data was sampled by a master thesis student, Teresa Helmersson, in consult with a radiologist with fellowship in neuroradiology, Ida Blystad.
3.4 Blinding

In this thesis blinding could primarily be done on two levels, image evaluation and statistical analysis. Blinding was not made for the ROI placement because it was thought to be more important to ensure that the regions were placed in the same area for all images from one patient. Therefore the synthetic and conventional images were studied at the same time. It would have been possible to blind the method but the problem is that a trained radiologist can immediately see which image type they are looking at since synthetic images are perceived as noisier. This makes the blinding of method somewhat redundant. The decision not to blind was also a way to do the analysis more time efficient since technical problems occurred.

It is also possible to affect the result during the statistical analysis. To prevent such bias the statistician should also be blinded. Since it was the statistician in this case that planned the analyses and data collection the blinding had to occur afterwards. The blinding was done by coding the different imaging methods and to randomize the order of the rows to minimize the chance of recognition.
Even if the intentions with blinding of the statistician were great, it was in practice redundant in this study. The reason for this is that if outliers are present then the data is studied to ensure that it is reasonable and not due to an error. The measurements of the intensities are very different between the methods since they are obtained from different scales and therefore it is obvious which method that is studied.

### 3.5 Statistical methods

To evaluate the agreement graphically a combination of Bland-Altman plots and correlation plots were used. In the Bland-Altman plots the mean and difference was obtained from each pairwise measurement using the contrast in the synthetic images and the corresponding values from the conventional images.

The pairwise difference of contrast and CNR were statistically analyzed to evaluate if the methods are significantly differing. The difference was calculated by subtracting the contrast in the conventional images from the contrast in the synthetic images. The difference in contrast and CNR were assumed to be normally distributed. This assumption was evaluated by investigating:

- if the mean value was similar to the median value
- if the skewness was very different from zero
- if the shape of the distribution produced with a histogram and Box-and-Whiskers plot was similar to the normal distribution
- if a normal probability plot (Q-Q plot) indicated normal distribution

Also a significance test of normal distribution, Shapiro-Wilks test, was used to evaluate if normal distribution could be assumed.

When normal distribution could be assumed a 95% confidence interval was calculated on the pairwise difference between the methods based on the t-test. The p-value from a paired t-test was used as a supplementary measure. If normal distribution could not be assumed then corresponding evaluation was used based on Wilcoxon signed rank test. This could occur if extreme outliers were present. Outliers makes the distribution somewhat skewed and Wilcoxon signed rank test is more appropriate for symmetric distributions. However rank methods are less affected by outliers and therefore more preferable than a t-test (12). The difference in contrast was calculated in such a way that a totally positive confidence interval indicate that the contrast is higher in the synthetic images.

The confidence interval with corresponding test cannot assure agreement between the methods. Conventional MRI is the method used today and can be seen as the gold standard. The variation between measurements in conventional images can therefore be used to evaluate if the difference between the methods is reasonable. The variation in the conventional images is established with 95% confidence intervals of the mean CNR and contrast for all different tissue comparisons. The range of the confidence intervals gives an approximation of how much two different conventional examinations will differ which can be used for an equivalence test. The methods are agreeable if the difference between them is smaller than this range. This data was also assumed to be normally distributed and the same procedure as for the comparisons of the differences was used. If the assumption of normal distribution holds then the confidence interval is based on the t-test otherwise on Wilcoxon signed rank test. By using the confidence interval of the difference together with the
Materials and Methods

The difference between two conventional examinations should preferably be estimated in the same way as the difference between the methods. This could not be done since two sets of conventional images for each patient is required which were not available.

For the lesions a slightly different approach was used. In principal the same procedure was used but the contrast and CNR can possibly depend on the patients’ disease. This dependency was therefore first evaluated with a one-way ANOVA analysis. The pairwise difference in contrast or CNR between the methods was the dependent variable and the predictor was the disease, denoted with 1 for ischemia and 2 for MS. If the main effect of the disease was not significant the same procedures as for the other contrast and CNR comparisons were used. If the main effect appears significant, the same procedure as before was used but separately on each disease group and with a simultaneous confidence level derived with Bonferroni’s inequality.

All the analyses were done individually for the T1 weighted images, T2 weighted images and the T2 FLAIR images. The significant results were those with a p-value smaller than 0.05 or a confidence interval separated from zero.

3.5.1 Alternative Statistical Methods
The evaluation of similarity between synthetic and conventional MRI was based on confidence interval of the mean difference together with an approximated value of the acceptable difference. Two other methods were considered for the evaluation; Limits of agreement and ANOVA.

Limits of agreement is a method where the confidence interval of the difference is calculated. Afterwards the limits of the confidence interval are used to state how well the methods are agreeing, e.g. measurements using a new method deviates ±0.5 from measurements using the gold standard. (11) The problem with this method is that the difference can be hard to analyze. When evaluating contrast it is difficult to determine if the difference is acceptable without any other measurement.

ANOVA could also be used to evaluate the effect of methods, see Appendix A. Contrast is the observation and it depends on the factors imaging method, studied structures and individual. Since the contrast is highly dependent on which tissues that are compared the ANOVA is preferably conducted individually for each tissue combination; GM-WM, GW-CSF and WM-CSF. The interesting effect is the imaging method. One reason why it is inappropriate to use ANOVA is that a significant result reveals differences between the imaging methods and in this study it is interesting to evaluate if the methods are similar. Another problem is that there is only one observation for each combination of factors and there are many different individuals which can make it difficult to retrieve a significant result.
3.6 Hypothesis
The analyses made in this study aims to evaluate the contrast in the synthetic images versus the conventional images. The question proposed in the objectivities was:

- Do the synthetic and conventional MR images provide the same contrast?

The first test for evaluation of the difference was with confidence intervals. The hypothesis becomes:

\( H_0: \) The difference in CNR/Contrast between the methods is zero

\( H_1: \) The difference in CNR/Contrast between the methods is separated from zero

A significant result, rejection of \( H_0 \), indicates that the imaging methods are different. The desired outcome for the study was to evaluate if the methods are equal and this information cannot be retrieved by the confidence intervals. However it is possible to receive this information with an equivalence test and the hypothesis becomes:

\( H_0: \) There are differences in CNR/Contrast between the methods

\( H_1: \) The variations in the CNR/Contrast between the methods are acceptable

3.7 Software
All images were analyzed in Sectra PACS IDS7. The synthetic image stacks were imported to IDS7 from SyMRI Suite which is an add-in to Sectra PACS. The data was managed in Microsoft Excel 2010 and analyzed in SPSS 17.0. Bland-Altman plots and correlation plots were done in Excel. Wilcoxon signed rank test with confidence interval is not available in SPSS 17.0 and therefore it was implemented in MATLAB R2009b, see code in Appendix B , and corresponding tests were also done in MATLAB.
4 Results from the Statistical Analysis

This chapter includes the results from the statistical analyses. First the mathematical hypotheses are presented and then the results from each image type are presented separately. For all the different image types a graphical analysis is presented to get an indication of the association and correlation. If the observations in the correlation plots follow the straight line the methods are agreeable. The analytical statistics from evaluation of the difference between the methods and an estimation of the reasonable difference between the methods are presented in the end. In all calculations of the difference the contrast in the conventional images was subtracted from the contrast in the synthetic images.

4.1 Hypothesis

The mathematical hypothesis for the confidence intervals is:

\[ H_0 : \mu_d = 0 \quad H_1 : \mu_d \neq 0 \]

The difference between the methods compared to the variation in the conventional images was studied using:

\[ \Delta = \max (CI_{conv}) - \min (CI_{conv}) \]

The range of the confidence interval for the mean contrast/CNR in the conventional images gives an indication of how much the contrasts are varying between these images, which is the upper limit for the difference.

4.2 T1 Weighted Images

Most of the results from the Bland-Altman plots for the T1 weighted images showed no trends. The contrast comparison between occipital cortex and genu deviated from the rest, see Figure 4-1. There were high variations in the correlation plots, see example in Figure 4-1.

![Figure 4-1](image_url)

Figure 4-1. To the left the Bland-Altman plot for the contrast comparison between occipital cortex and genu can be seen and to the right the correlation plot for CNR between frontal cortex and centrum semiovale.
A correlation plot with all contrast measurements can be seen in Figure 4-2 and for the measurements between gray and white brain matter only see Figure 4-3. Corresponding plots for CNR can be seen in Figure 4-4 and Figure 4-5.

![Figure 4-2](image1.png)

**Figure 4-2.** Plot with all contrast measurements in the T1 weighted images. Th=Thalamus, SP=Splenium, Ge=Genu, CS=Centrum Semiovale, OC=Occipital Cortex, FC=Frontal Cortex. Same abbreviations are used in all result plots.

![Figure 4-3](image2.png)

**Figure 4-3.** Plot with all contrast measurements between gray matter and white matter in the T1 weighted images
Figure 4-4. Plot with all measurements of CNR in T1 weighted images.

Figure 4-5. Plot with all measurements of CNR between white and gray brain matter in T1 weighted images.

The result from the analyses of the difference between the methods and the variation in the conventional images can be seen in Table 4-1. In the table the mean and standard deviation for both methods is also presented. The table also contains notations for those data sets where normal distribution could not be assumed. All the comparisons between thalamus and white matter contained negative values as for the comparison between occipital cortex and centrum semiovale. The results are based on images from 22 patients.
Table 4-1. Results from the analyses of T1 weighted images. The confidence intervals are calculated from the pairwise difference between the methods. The variation in the conventional images is an approximation of reasonable disagreement between the methods. Denotations: ¹The difference in intensity were not as the theory in all cases, ²Normal distribution could not be assumed because of outliers, ³H₀ in the equivalence test could be rejected

<table>
<thead>
<tr>
<th>CNR</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td>Tissues</td>
<td>n=22</td>
<td>n=22</td>
<td></td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>3.55 (1.9)</td>
<td>2.79 (2.0)</td>
<td>-0.53 – 1.75²</td>
</tr>
<tr>
<td>Genu¹</td>
<td>4.37 (1.9)</td>
<td>2.33 (1.5)</td>
<td>1.16 – 2.91</td>
</tr>
<tr>
<td>Centrum Semiovale¹</td>
<td>2.31 (1.8)</td>
<td>1.63 (1.7)</td>
<td>-0.11 – 1.48³</td>
</tr>
<tr>
<td>CSF²</td>
<td>19.7 (5.9)</td>
<td>22.9 (6.9)</td>
<td>-6.89 – 0.50</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>4.43 (2.4)</td>
<td>7.05 (3.0)</td>
<td>-3.81 – -1.44</td>
</tr>
<tr>
<td>Genu</td>
<td>5.24 (2.9)</td>
<td>6.59 (1.9)</td>
<td>-2.49 – -0.21</td>
</tr>
<tr>
<td>Centrum Semiovale¹</td>
<td>3.19 (2.7)</td>
<td>5.89 (2.2)</td>
<td>-3.82 – -1.58</td>
</tr>
<tr>
<td>CSF</td>
<td>18.8 (6.2)</td>
<td>18.6 (5.8)</td>
<td>-3.70 – 4.07³</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>7.99 (3.5)</td>
<td>8.65 (4.1)</td>
<td>-3.24 – 1.92³</td>
</tr>
<tr>
<td>Genu</td>
<td>8.81 (3.3)</td>
<td>8.19 (2.9)</td>
<td>-1.34 – 2.58</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>6.75 (2.8)</td>
<td>7.49 (2.9)</td>
<td>-2.55 – 1.08</td>
</tr>
<tr>
<td>CSF</td>
<td>15.2 (5.2)</td>
<td>17.0 (4.8)</td>
<td>-4.28 – 0.73</td>
</tr>
<tr>
<td>CSF vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>23.2 (7.0)</td>
<td>25.7 (8.1)</td>
<td>-6.95 – 2.07³</td>
</tr>
<tr>
<td>Genu</td>
<td>24.0 (6.8)</td>
<td>25.2 (7.0)</td>
<td>-5.14 – 2.83³</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>22.0 (6.0)</td>
<td>24.5 (7.2)</td>
<td>-6.25 – 1.23³</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissues</td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium¹</td>
<td>0.056 (0.03)</td>
<td>0.035 (0.03)</td>
<td>0.007 – 0.034²</td>
</tr>
<tr>
<td>Genu¹</td>
<td>0.069 (0.03)</td>
<td>0.031 (0.02)</td>
<td>0.027 – 0.049</td>
</tr>
<tr>
<td>Centrum Semiovale¹</td>
<td>0.039 (0.03)</td>
<td>0.022 (0.03)</td>
<td>0.006 – 0.029</td>
</tr>
<tr>
<td>CSF</td>
<td>0.491 (0.04)</td>
<td>0.420 (0.06)</td>
<td>0.032 – 0.107²</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>0.071 (0.04)</td>
<td>0.093 (0.03)</td>
<td>-0.042 – -0.001</td>
</tr>
<tr>
<td>Genu</td>
<td>0.085 (0.05)</td>
<td>0.089 (0.02)</td>
<td>-0.029 – 0.020</td>
</tr>
<tr>
<td>Centrum Semiovale¹</td>
<td>0.055 (0.05)</td>
<td>0.080 (0.03)</td>
<td>-0.049 – -0.000</td>
</tr>
<tr>
<td>CSF</td>
<td>0.479 (0.05)</td>
<td>0.372 (0.06)</td>
<td>0.069 – 0.144</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>0.137 (0.05)</td>
<td>0.115 (0.04)</td>
<td>-0.011 – 0.051²</td>
</tr>
<tr>
<td>Genu</td>
<td>0.150 (0.05)</td>
<td>0.111 (0.03)</td>
<td>0.010 – 0.064²</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>0.120 (0.05)</td>
<td>0.102 (0.03)</td>
<td>-0.009 – 0.037²</td>
</tr>
<tr>
<td>CSF</td>
<td>0.426 (0.07)</td>
<td>0.353 (0.05)</td>
<td>0.055 – 0.103²</td>
</tr>
<tr>
<td>CSF vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>0.533 (0.03)</td>
<td>0.448 (0.07)</td>
<td>0.040 – 0.123²</td>
</tr>
<tr>
<td>Genu</td>
<td>0.542 (0.03)</td>
<td>0.446 (0.06)</td>
<td>0.061 – 0.132²</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>0.521 (0.04)</td>
<td>0.438 (0.07)</td>
<td>0.045 – 0.118²</td>
</tr>
</tbody>
</table>
4.3 T2 Weighted Images
The results from the Bland-Altman plots showed no trends, see example in Figure 4-6, except for the contrast comparison between splenium and CSF. All the correlation plots showed less variations compared to the T1 weighted images, see example in Figure 4-6.

Figure 4-6. To the left the Bland-Altman plot for the contrast comparison between thalamus and splenium in T2 weighted images can be seen and to the right the corresponding correlation plot.

A correlation plot with all contrast measurements can be seen in Figure 4-7 and for the measurements between gray and white brain matter only see Figure 4-8. Corresponding plots for CNR can be seen in Figure 4-9 and Figure 4-10.

Figure 4-7. Plot with all contrast measurements in T2 weighted images.
Figure 4-8. Plot with all contrast measurements between gray brain matter and white brain matter in T2 weighted images

Figure 4-9. Plot with all CNR measurements in T2 weighted images
Figure 4-10. Plot with all CNR measurements between gray brain matter and white brain matter in T2 weighted images

The result from the analysis of the difference between the methods can be seen in Table 4-2. In the table the variation within the conventional images can also be seen and which data sets where normal distribution could not be assumed. In the analysis of the T2 weighted images seven on the 15 comparisons contained observations where the relationships of the intensity were not as expected, see Table 4-2. The results are based on images from 22 patients.

4.3.1 Pathology Measurements

The contrast and CNR between normal white matter and pathology, lesions, were calculated for the T2 weighted images. Those data sets contained one missing value (n=15) because one patient diagnosed with MS had no visual lesions in the MR images. The result from the Bland-Altman plots showed no trends and the correlation plots had visible correlation, see example in Figure 4-11.

Figure 4-11. To the left the Bland-Altman plot for the CNR comparison between centrum semiovale and lesions in T2 weighted images can be seen and to the right is the corresponding correlation plot.

A correlation plot for all contrast and CNR measurements can be seen in Figure 4-12.
Table 4-2. Results from the analyses of T2 weighted images. The confidence intervals are calculated from the pairwise difference between the methods. The variation in the conventional images is an approximation of reasonable disagreement between the methods. Denotations: 1The difference in intensity were not as the theory in all cases, 2Normal distribution could not be assumed because of outliers, 3H0 in the equivalence test could be rejected

<table>
<thead>
<tr>
<th>Tissues</th>
<th>CNR</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=22</td>
<td>n=22</td>
<td>95%</td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>3.29 (2.4)</td>
<td>3.06 (2.6)</td>
<td>-0.64 – 1.10</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>4.02 (2.7)</td>
<td>4.23 (2.7)</td>
<td>-1.14 – 0.73</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>-0.28 (2.3)</td>
<td>1.16 (2.0)</td>
<td>-2.16 – -0.71</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>31.4 (8.1)</td>
<td>26.5 (6.5)</td>
<td>2.16 – 7.24</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>3.60 (1.9)</td>
<td>2.61 (2.7)</td>
<td>-0.04 – 2.03</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>4.34 (3.0)</td>
<td>3.78 (3.6)</td>
<td>-0.16 – 1.66</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>0.039 (3.3)</td>
<td>0.707 (3.4)</td>
<td>-1.36 – -0.31</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>31.1 (8.5)</td>
<td>27.0 (6.8)</td>
<td>0.96 – 6.29</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>8.10 (3.3)</td>
<td>7.06 (3.0)</td>
<td>-0.28 – 2.36</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>8.84 (3.6)</td>
<td>8.23 (3.0)</td>
<td>-0.64 – 1.87</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>4.54 (3.0)</td>
<td>5.15 (2.4)</td>
<td>-1.51 – -0.28</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>26.6 (6.6)</td>
<td>22.5 (4.9)</td>
<td>1.53 – 5.85</td>
</tr>
<tr>
<td>CSF vs.</td>
<td>Splenium</td>
<td>34.7 (9.0)</td>
<td>29.6 (7.5)</td>
<td>1.74 – 7.85²</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>35.4 (9.1)</td>
<td>30.8 (7.5)</td>
<td>1.28 – 7.39²</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>31.1 (7.7)</td>
<td>27.7 (6.6)</td>
<td>0.87 – 5.49²</td>
</tr>
<tr>
<td>Contrast</td>
<td>Mean (std)</td>
<td></td>
<td>CI</td>
<td>Variation</td>
</tr>
<tr>
<td>Tissues</td>
<td></td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>n=22</td>
<td>n=22</td>
<td>95%</td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>0.111 (0.06)</td>
<td>0.091 (0.07)</td>
<td>-0.003 – 0.04³</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>0.145 (0.08)</td>
<td>0.131 (0.06)</td>
<td>-0.011 – 0.04³</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>-0.007 (0.06)</td>
<td>0.033 (0.05)</td>
<td>-0.059 – -0.020</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.503 (0.05)</td>
<td>0.423 (0.04)</td>
<td>0.062 – 0.096</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>0.125 (0.06)</td>
<td>0.074 (0.07)</td>
<td>0.029 – 0.072</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>0.158 (0.11)</td>
<td>0.114 (0.09)</td>
<td>0.024 – 0.069³,³</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>0.007 (0.10)</td>
<td>0.016 (0.09)</td>
<td>-0.030 – 0.011³</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.492 (0.05)</td>
<td>0.436 (0.06)</td>
<td>0.040 – 0.074</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>0.242 (0.04)</td>
<td>0.189 (0.06)</td>
<td>0.026 – 0.080</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>0.274 (0.07)</td>
<td>0.228 (0.05)</td>
<td>0.026 – 0.066</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>0.128 (0.06)</td>
<td>0.132 (0.05)</td>
<td>-0.024 – 0.014³</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.396 (0.03)</td>
<td>0.339 (0.02)</td>
<td>0.043 – 0.071</td>
</tr>
<tr>
<td>CSF vs.</td>
<td>Splenium</td>
<td>0.582 (0.03)</td>
<td>0.495 (0.04)</td>
<td>0.068 – 0.107</td>
</tr>
<tr>
<td></td>
<td>Genu</td>
<td>0.604 (0.05)</td>
<td>0.525 (0.04)</td>
<td>0.064 – 0.093</td>
</tr>
<tr>
<td></td>
<td>Centrum Semiovale</td>
<td>0.498 (0.04)</td>
<td>0.450 (0.04)</td>
<td>0.033 – 0.062</td>
</tr>
</tbody>
</table>
No significant effects of the diseases were found in the ANOVA analysis, see p-values for the effect of disease in Table 4-3. The following analyses were therefore conducted in the same way as for the healthy tissues. The result from the confidence intervals of the difference can be seen in Table 4-3. The table also reveals information about which data sets that could not be assumed to be normally distributed. In all observations the relationship of the intensities was as expected.

### Table 4-3. Results from the analyses of lesions in T2 weighted images. The confidence intervals are calculated from the pairwise difference between the methods. The variation in the conventional images is an approximation of reasonable disagreement between the methods. Denotations: ¹Normal distribution could not be assumed because of outliers, ²one missing value,

<table>
<thead>
<tr>
<th>CNR</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
<th>ANOVA p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
<td></td>
</tr>
<tr>
<td>Tissues</td>
<td>n=15²</td>
<td>n=15²</td>
<td>95%</td>
<td>α=0.05</td>
</tr>
<tr>
<td>Lesion vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>19.3 (4.3)</td>
<td>15.2 (4.1)</td>
<td>1.93 – 6.15</td>
<td>4.72</td>
</tr>
<tr>
<td>Genu</td>
<td>19.7 (4.2)</td>
<td>15.8 (3.6)</td>
<td>1.76 – 5.82</td>
<td>4.17</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>15.2 (3.2)</td>
<td>12.9 (3.3)</td>
<td>0.59 – 3.99</td>
<td>3.86</td>
</tr>
<tr>
<td>Contrast</td>
<td>Mean (std)</td>
<td>CI</td>
<td>Variation</td>
<td>ANOVA p-value</td>
</tr>
<tr>
<td>Tissues</td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=15²</td>
<td>n=15²</td>
<td>95%</td>
<td>α=0.05</td>
</tr>
<tr>
<td>Lesion vs.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>0.437 (0.06)</td>
<td>0.346 (0.07)</td>
<td>0.061 – 0.116</td>
<td>0.084</td>
</tr>
<tr>
<td>Genu</td>
<td>0.450 (0.06)</td>
<td>0.363 (0.06)</td>
<td>0.058 – 0.115</td>
<td>0.073</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>0.324 (0.07)</td>
<td>0.280 (0.06)</td>
<td>0.014 – 0.074</td>
<td>0.074</td>
</tr>
</tbody>
</table>

### 4.4 T2 FLAIR Images

The result from the Bland-Altman plots in the T2 FLAIR images showed trends for all the contrasts and CNR comparisons between CSF and soft tissues, see example in Figure 4-13. Similar trends occurred in the Bland-Altman plots for the CNR measurements between the cortexes and splenium. All the correlation plots showed correlation except for the contrast comparison between occipital cortex and CSF where the conventional contrast was fairly the same even if the contrast increased in the synthetic images.
Figure 4-13. The figure shows the Bland-Altman plots for the CNR and contrast comparisons between occipital cortex and CSF in T2 FLAIR images.

Correlation plot with all contrast measurements can be seen in Figure 4-14 and for the measurements between gray and white brain matter only see Figure 4-15. Corresponding plots for CNR can be seen in Figure 4-16 and Figure 4-17.

Figure 4-14. Plot with all contrast measurements in the T2 FLAIR images.
Figure 4-15. Plot with all contrast measurements between gray and white brain matter in the T2 FLAIR images.

Figure 4-16. Plot with all CNR measurements in the T2 FLAIR images.
The confidence intervals from the analysis of the difference between the methods can be seen in Table 4-4. The table also contains information about the variation in the conventional images and which data sets where normal distribution could not be assumed. In the analysis of the T2 FLAIR images nine of the 15 comparisons contained at least one observation where the relationship of the intensity was not as expected, see Table 4-4. These cases are more complex than for the T1 and T2 weighted images and for further information on the impact see section 6.4. The results are based on images from 22 patients.
Table 4-4. Results from the analyses of T2 FLAIR images. The confidence intervals are calculated from the pairwise difference between the methods. The variation in the conventional images is an approximation of reasonable disagreement between the methods. Denotations: 1The difference in intensity were not as the theory in all cases, 2Normal distribution could not be assumed because of outliers, 3H₀ in the equivalence test could be rejected.

<table>
<thead>
<tr>
<th>CNR</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td>Tissues</td>
<td>n=22</td>
<td>n=22</td>
<td>95%</td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$2.02 (1.5)$</td>
<td>$2.40 (2.4)$</td>
<td>$-1.07 – 0.31$</td>
</tr>
<tr>
<td>Genu</td>
<td>$3.04 (2.2)$</td>
<td>$4.26 (3.0)$</td>
<td>$-2.03 – -0.12$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$-1.12 (1.6)$</td>
<td>$-0.29 (1.7)$</td>
<td>$-1.55 – -0.10$</td>
</tr>
<tr>
<td>CSF</td>
<td>$13.2 (3.3)$</td>
<td>$22.9 (6.9)$</td>
<td>$-12.3 – -7.11$</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$1.82 (1.5)$</td>
<td>$0.73 (2.5)$</td>
<td>$0.32 – 1.86$</td>
</tr>
<tr>
<td>Genu</td>
<td>$2.83 (2.4)$</td>
<td>$2.59 (3.5)$</td>
<td>$-0.88 – -1.38$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$-1.32 (2.3)$</td>
<td>$-1.97 (2.9)$</td>
<td>$-0.03 – -1.32$</td>
</tr>
<tr>
<td>CSF</td>
<td>$13.0 (3.0)$</td>
<td>$21.3 (7.0)$</td>
<td>$-10.8 – -5.67$</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$4.44 (1.6)$</td>
<td>$4.02 (3.2)$</td>
<td>$-0.74 – 1.59$</td>
</tr>
<tr>
<td>Genu</td>
<td>$5.45 (2.4)$</td>
<td>$5.87 (3.7)$</td>
<td>$-1.79 – -0.95$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$1.30 (2.1)$</td>
<td>$1.32 (2.3)$</td>
<td>$-0.95 – -0.90$</td>
</tr>
<tr>
<td>CSF</td>
<td>$15.6 (3.2)$</td>
<td>$24.5 (7.9)$</td>
<td>$-11.9 – -5.95$</td>
</tr>
<tr>
<td>CSF vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$11.2 (2.6)$</td>
<td>$20.5 (5.8)$</td>
<td>$-11.7 – -6.99$</td>
</tr>
<tr>
<td>Genu</td>
<td>$10.2 (2.9)$</td>
<td>$18.7 (5.8)$</td>
<td>$-10.4 – -6.58$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$14.3 (3.8)$</td>
<td>$23.2 (7.0)$</td>
<td>$-11.4 – -6.36$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast</th>
<th>Mean (std)</th>
<th>CI</th>
<th>Variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tissues</td>
<td>Synthetic</td>
<td>Conventional</td>
<td>Difference</td>
</tr>
<tr>
<td>Thalamus vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$0.063 (0.05)$</td>
<td>$0.047 (0.05)$</td>
<td>$0.002 – 0.031$</td>
</tr>
<tr>
<td>Genu</td>
<td>$0.099 (0.06)$</td>
<td>$0.091 (0.05)$</td>
<td>$-0.011 – 0.027$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$-0.030 (0.04)$</td>
<td>$-0.006 (0.03)$</td>
<td>$-0.041 – -0.008$</td>
</tr>
<tr>
<td>CSF</td>
<td>$0.656 (0.13)$</td>
<td>$0.821 (0.05)$</td>
<td>$-0.214 – -0.106$</td>
</tr>
<tr>
<td>Occipital Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$0.056 (0.05)$</td>
<td>$0.009 (0.05)$</td>
<td>$0.030 – 0.063$</td>
</tr>
<tr>
<td>Genu</td>
<td>$0.092 (0.08)$</td>
<td>$0.053 (0.06)$</td>
<td>$0.017 – 0.060$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$-0.037 (0.07)$</td>
<td>$-0.043 (0.06)$</td>
<td>$-0.011 – -0.020$</td>
</tr>
<tr>
<td>CSF</td>
<td>$0.654 (0.12)$</td>
<td>$0.810 (0.05)$</td>
<td>$-0.208 – -0.102$</td>
</tr>
<tr>
<td>Frontal Cortex vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$0.130 (0.04)$</td>
<td>$0.074 (0.05)$</td>
<td>$0.036 – 0.076$</td>
</tr>
<tr>
<td>Genu</td>
<td>$0.165 (0.06)$</td>
<td>$0.118 (0.05)$</td>
<td>$0.030 – 0.064$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$0.038 (0.05)$</td>
<td>$0.022 (0.04)$</td>
<td>$-0.004 – 0.035$</td>
</tr>
<tr>
<td>CSF</td>
<td>$0.695 (0.11)$</td>
<td>$0.831 (0.05)$</td>
<td>$-0.183 – -0.089$</td>
</tr>
<tr>
<td>CSF vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Splenium</td>
<td>$0.621 (0.13)$</td>
<td>$0.805 (0.06)$</td>
<td>$-0.232 – -0.124$</td>
</tr>
<tr>
<td>Genu</td>
<td>$0.598 (0.15)$</td>
<td>$0.788 (0.07)$</td>
<td>$-0.247 – -0.122$</td>
</tr>
<tr>
<td>Centrum Semiovale</td>
<td>$0.672 (0.13)$</td>
<td>$0.823 (0.06)$</td>
<td>$-0.203 – -0.090$</td>
</tr>
</tbody>
</table>
5 Radiologists Experience

This chapter describes the pros and cons with synthetic MRI from a radiologist opinion. The information is from a discussion with Ida Blystad, radiologist with fellowship in neuroradiology at Linköpings University Hospital.

5.1 Image quality

The experience is that the synthetic images generally have a lower image quality than the conventional images. Lower resolution is often used in synthetic images to reduce scan time but this also makes the experience of the image quality worse. The noise is experienced to be higher in the synthetic images. This is especially noticeable in CSF in T2 weighted images and T2 FLAIR images. Even if this worsening of image quality is not proved to be clinically decisive the radiologists still rather work with high quality images to be more certain that pathology is not missed.

One large benefit with the synthetic images is that the lesions were experienced more separable from normal tissues. This is primarily for the lesions that are diffuse in conventional images. The advantage is that it is easier to detect and localize these lesions. The drawback is that high intensity lesions in the conventional images were thought to be more diffuse in the synthetic images. One problem with this reversed experience is that if the intensity is used to estimate the age of the lesion it cannot be done in the same way in synthetic images as in conventional images.

5.2 Problems with Clinical Use of Synthetic MRI

The T2 FLAIR images are the main problem with using synthetic MRI clinically. They contain two types of artifacts that can be misinterpreted as disease. One artifact is that the tissues appear swollen which could be an indication that the intracranial pressure is too high, a common side effect after a trauma. This is visual if the grooves in the cortex, sulci, are studied in both conventional and synthetic images, see Figure 5-1. The sulci are small or sometimes not even visual in the synthetic images while they are clearly visual in the conventional images. The other artifact is bright signals that appear in the periphery of the cortex, see Figure 5-1. This is an indication to accumulation of fluids, edema, which is an inflammatory response. T2 FLAIR images are used to detect such edema signals and then it is inappropriate that many incorrect signals are detected.

The problems with the T2 FLAIR images make them inappropriate to use clinically. However it is important to realize that a diagnosis or statement by a radiologist is made using the whole exam. T1 and T2 weighted images reveal evidence if the suspected inflammation is present or not. One alternative to ensure the stability of the diagnosis is to use a conventional T2 FLAIR image stack together with the synthetic image stacks.

5.3 Benefits with Synthetic MRI in Clinical Use

A large benefit with synthetic MRI is that all these three image types can be retrieved in a short time. One MRI examination is quite long, approximately 30-60 minutes, and uncomfortable for the patient due to high noise level. To reduce the time of an examination could make it less uncomfortable for the patient and more patients could be examined in one day. The possibility to increase number of patients a day could reduce the waiting which today is up to 13 weeks in Östergötland (16).
Figure 5-1. The images illustrate the problems with synthetic T2 FLAIR images. To the left a selection from a synthetic image can be seen and to the right the corresponding conventional image selection. It is obvious that the sulci are not as visual in the synthetic image and the circle and arrow show areas where incorrect edema signals are visual.

Another benefit with synthetic MRI is the possibility to get quantified values. This quantification can be used to identify tissues and maybe even to separate different lesions that are visually equal, e.g. certain MS lesions can be visually similar to lesions due to ischemia. Further studies have to be conducted to support such theory but the benefits are large if this can be established.
6 Discussion

This chapter contains discussions of the results, problems with the images and also some ideas for further studies. The contrast and CNR was measured in both synthetic and conventional images for the three different image types T1 weighted, T2 weighted and T2 FLAIR. In the beginning there is a brief discussion of all results and then the chapter is divided into sublevels for the different image types. Since several methods are combined the discussion of each image type is extensive and therefore a summary is provided in the end of each section. Afterwards a rough approximation of the scan time is presented and the problems concerning all weighting schemes are discussed. In the end there is a section with suggestions of developments for similar studies and ideas for future studies.

6.1 Overview of the Results

All the analyses of CNR and contrast had one thing in common; the result from the measurements with CSF deviated from the rest. This can be seen if all the correlation plots are studied. The measurements follow the line that indicates perfect fit fairly well in the plots consisting of only the white and gray brain matter. In all the plots with all contrasts/CNR measurements for each image type at least one of two different trends could be seen. One trend was that the contrasts/CNR between CSF and soft tissues had a higher variation than other measurements and the other was that the contrasts/CNR between CSF and soft tissues were differing between the methods. These effects could also be seen in the analytic statistics. However the results from the CNR measurement in T1 weighted images deviated somewhat from the rest, further discussion can be found in section 6.2.2. The contrast in the synthetic images was revealed as higher for several cases in the analyses of the differences. However this effect was not seen in the CNR which is an indication that both the contrast and noise is higher in the synthetic images.

The results from the analyses with lesions differed from the results using only normal tissues. Both the CNR and the contrast were higher in the synthetic images, the same indications were revealed in all analyses.

6.2 T1 Weighted Images

This section is divided into the discussions about contrast and CNR. A brief summary of the results from the T1 weighted images are presented at the end.

6.2.1 Contrast

An overview of all contrast measurements in the T1 weighted images can be seen in Figure 4-2. This plot indicates that the synthetic images have a higher contrast than the conventional images in comparisons between CSF and soft tissues. All the contrast measurements between white and gray brain matter can be seen in Figure 4-3. The measurements follow the line that indicates agreement between the methods except for some outliers. However the contrast measurements between white and gray brain matter are in a small range and the difference between the methods is almost as large as the range of the measurements. The outliers could occur due to errors during measurement. It is also possible that the ROI has unexpected tissue content due to disease which can result in outliers.

Most of the correlation plots for the separate contrast data series indicated correlation, but not always perfect agreement. When the Bland-Altman plots were studied no trend occurred except for
the comparison between occipital cortex and genu, see Figure 4-1. A vague positive trend can be seen which suggests that the contrast between the synthetic images have a higher variation than the contrast between the conventional images. However since the indication is vague and only seen in one tissue comparison the most probable is that the trend is due to measurement errors. This could occur since the occipital cortex is very thin and white matter could then be included in the ROI by mistake, for further discussion about ROI placement see section 6.6.2.

The analytic statistics indicated that the contrast was higher in the synthetic images than in the conventional images since many of the confidence intervals were completely positive. However the difference between the methods was highly dependent on the studied structures. The result is somewhat ambiguous but there are strong indications that the synthetic images at least produce the same contrast as the conventional images. This is since only two of 15 intervals are negative, i.e. suggest that the conventional images have a higher contrast. Both these two differing comparisons are including occipital cortex for which the measurements were stated as not completely reliable in the previous paragraph. The fact that the synthetic images have a lower resolution also results in more partial volume effects in these images. A higher partial volume effect will probably result in that more white matter is present in the ROIs in occipital cortex, for further discussion about resolution and partial volume effects see section 6.6.3. This will reduce the contrast between occipital cortex and white matter in the synthetic images even if the technique is correct.

The differences in the contrast comparisons were all greater than the approximated variation between conventional images, i.e. $H_0$ in the equivalence test cannot be rejected. When studying the confidence intervals there were only three of the comparisons that showed an insignificant difference between the methods. This combined with the variation in the conventional images is an indication that the contrast is not equal for both methods.

It is important to notice that the synthetic images are not automatically better than the conventional images even if they generate higher contrast. This is because the conventional images are the gold standard and therefore the data from the conventional images are assumed to be the true values.

6.2.2 CNR
An overview of all the CNR measurements can be seen in the correlation plot in Figure 4-4. This plot clearly indicates that the difference between the methods increases with the CNR, i.e. for higher CNR measurements the difference between the methods is larger. This is especially seen in the CNR measurements between CSF and soft tissues. The measurements of the CNR between gray and white brain matter follow the line that indicates perfect fit fairly well, see Figure 4-5. However the variation is quite high compared to the range of values, this is similar to the result for the contrast.

If all the data sets are studied separately the result is differing from the graph with all measurements. All the correlation plots had indications of correlation between the methods but several tissue comparisons did not indicate exact agreement and the high variation made it difficult to state if correlation existed. However all the Bland-Altman plots indicated association, i.e. no trends were visible.

The results from the difference in CNR between the methods were mostly insignificant. There were no strong indications that the conventional images are superior to the synthetic images and most probably both methods generate comparable CNR. The conventional method did also produce higher
CNR in the comparisons including occipital cortex which can be an effect of thin tissue and different resolutions, same to that for the contrast. It is notable that even if the result of the CNR indicates that the methods are similar, the previous result indicated that synthetic images produce higher contrast. The reason could be that the contrast is higher in the synthetic images but the noise is also higher which impair the CNR.

When studying the variation in the conventional images the interpretation gets somewhat different. There were five comparisons with insignificant difference where the difference was larger than the variation between the conventional images, i.e. $H_0$ in the equivalence test could not be rejected. This results in only six comparisons of 15 where the difference was smaller than the variation between conventional images, which indicates less agreement then if only the confidence intervals of the difference were studied. The consensus of this is an indication that even if the difference is insignificant the variations between the methods are too high for stating that the methods are equal.

Finally the difference between the methods is higher for CNR between soft tissues and CSF, compared to CNR between white and gray brain matter. The variation in the conventional images is correspondingly high which was different from all other CNR and contrast comparisons.

6.2.3 Negative Contrast Measurements
Since unintended tissue intensity relation complicates the analysis is the result for comparison between thalamus and centrum semiovale is not totally reliable. The contrast data set between thalamus and centrum semiovale contained many negative values for all image types. One explanation is that the tissue in thalamus is not as gray as the cortexes and centrum semiovale is not an as homogeneous tissue as splenium and genu. The contrast between these tissues can therefore become small or even get reversed relationship. The risk is that the difference in contrast is overestimated. For further discussion about negative contrast see section 6.6.1.

In the other three data sets with negative values in the T1 weighted images only one negative value was found which did not affect the result so much so that the difference would change from significant to insignificant or vice versa. This means that the interpretation of the results is valid even if some values were questionable. It is notable that all data sets of the contrast between gray matter and centrum semiovale contained negative values for all images except the contrast between frontal cortex and centrum semiovale in the T1 weighted images.

6.2.4 Summary
The differences in contrast and CNR between the methods seem to vary much in the T1 weighted images. The CNR seems similar for the methods if all the analyses are regarded but the difference between the methods is somewhat high compared to the variation in the conventional images. The methods are probably correlating since the association analyses did not reveal any anomalies. However it is clear that both methods do not generate images with equal contrast.

6.3 T2 Weighted Images
This section is divided into the discussions about contrast and CNR. A brief summary of the results from the T2 weighted images are presented afterwards and in the end is a discussion about the result from the analyses with pathology.
6.3.1 Contrast

Figure 4-7 shows a correlation plot with all the contrast measurement from the T2 weighted images. That plot together with Figure 4-8 indicates that the contrast between white and gray brain matter follows the line of perfect fit well while the contrast measurement between CSF and soft tissues is below the line, i.e. the contrasts between CSF and soft tissues are higher in the synthetic images.

All correlation plots for the separate data series indicates correlation. It is difficult to determine if the methods are equal by these plots, due to the variation, but they are at least correlating. In the Bland-Altmann plots no trends were visual except for the contrast comparisons between splenium and CSF which had a negative trend. It is unclear why this specific comparison had a negative trend but the results from all plots taken together are indicating association. The Bland-Altmann plots showed that the differences were mostly positive. This indicates that the synthetic images generate a higher contrast than the conventional images.

The indications from the analytical statistics were consistent with the graphical analysis. The contrast seems higher in the synthetic images since ten of 15 comparisons revealed a difference significantly larger than zero and four of the remaining comparisons had insignificant difference. There was only one comparison that suggests that the conventional contrast is higher, thalamus against centrum semiovale. Unfortunately this comparison is not good since there were many observations where the gray matter had lower intensity than the white matter which is not agreeing with the theory, see section 3.2. The most probable reason for this outcome is that the two tissues have almost the same intensity and therefore small differences in placements of the ROIs can affect the result unreasonable, same as for the T1 weighted images. For the other data sets containing negative values there are no indications that these have affected the outcome and should therefore be seen as valid.

The difference was also compared to the variation in the conventional images contrasts which generated almost the same result. Those data sets with a significant difference between the methods had also a difference larger than the variation in the conventional images, i.e. $H_0$ in the equivalence test could not be rejected. One exception was the contrast between occipital cortex and genu which was significantly higher in the synthetic images but the difference was still not larger than the variation in the conventional images. All contrasts between CSF and soft tissues were both significantly higher in the synthetic images and had a difference that was larger than the variation in the conventional images. The reason for this could be that the intensity of CSF is overestimated in the synthetic images. There were five of the 15 comparisons that had a difference smaller than the contrast variation in the conventional images which suggests that the agreement between the methods is not very good. This suggestion is also supported by the confidence interval of the difference where significant difference was revealed in eleven cases.

6.3.2 CNR

An indication of the CNR in total can be achieved by studying the correlation plots in Figure 4-9 and Figure 4-10. The CNR between gray and white brain matter seems similar in both method but difference between the methods increases in the comparisons between CSF and soft tissues. The CNR in the synthetic images was high in those comparisons. Interpretation of the correlation plots from the individual data series gives almost the same result except for that some series with contrast...
between gray and white brain matter indicates correlation but not agreement. The Bland-Altman plots did not show any trend which suggests that association exists between the methods.

The analytical statistics indicates that the CNR is similar in synthetic and conventional images for comparisons between gray and white brain matter. There is only one comparison that suggests that the CNR is higher in the conventional images, thalamus against centrum semiovale. This is however not a reliable comparison for the same reason as explained in section 6.3.1. For the other data series containing negative values there were no indications that these affected the outcome and should therefore be seen as reliable. The CNR was significantly higher in the synthetic images in the comparisons between CSF and soft tissues. This is an additional indication that the intensity in CSF is overestimated in the synthetic images since no other CNR analysis in the T2 weighted images had similar result.

The difference compared to the variation in the conventional images reveals similar results as the former analyses. Where significant difference between the methods was found the difference was also larger than the variation in the conventional images, i.e. \( H_0 \) in the equivalence test could not be rejected. The CNR between frontal cortex and splenium showed an insignificant difference but the difference was still larger than the variation in the conventional images. It is notable that six of the eight comparisons where \( H_0 \) in the equivalence test could not be rejected were comparisons with CSF. This suggests that CSF is the primary reason why synthetic and conventional images do not generate equal CNR.

The results indicate that the synthetic T2 weighted images have at least as high contrast and CNR as the conventional T2 weighted images. The contrast is more favorable for the synthetic images which suggest that one main reason for potential low CNR is that the noise is high, same as for the T1 weighted images.

6.3.3 Summary
In total the contrast and CNR in T2 weighted images were not varying as much between the methods as for the T1 weighted images. If CSF is not accounted for, the methods could be seen as fairly equal. The association analysis showed correlation but exact agreement could not be assumed. The main problem lies within the CSF and if the intensity is corrected to create a similar contrast as in the conventional images there is high probability that agreement can be assured.

6.3.4 Pathology Measurement
The results from the pathology measurement are clearly stating that the imaging methods cannot be assumed to be equal. Both the graphical and numerical analysis suggests that the contrast and CNR is higher in the synthetic images. It is interesting that the lesions seem to be more separable from white matter in synthetic images which correspond to the radiologist, Ida Blystad, experience of the images. However the result from the graphical analysis showed that there was association between the methods but exact agreement could not be assured.

It is important to realize that only one lesion was selected from each patient. All these lesions were also of the same type. For a more general statement different types of lesions would preferably be selected, especially since Ida Blystad noted that it was the lesions that were more diffuse in the conventional images that were clearer in the synthetic images and vice versa.
6.4 T2 FLAIR Images
This section is divided into the discussions about contrast and CNR. A brief summary of the results from the T2 FLAIR images are presented in the end.

6.4.1 Contrast
A correlation plot with all contrast measurements from the T2 FLAIR images can be seen in Figure 4-14 where it is obvious that the contrast between white and gray brain matter are differing from the contrast between CSF and soft tissues. The data series containing CSF strongly suggests that the contrast in the conventional images is higher than the contrast in the synthetic images. There are also indications that the variation is higher between the synthetic images. The measurements of the contrast between gray and white brain matter can be seen in Figure 4-15. These values follow the line of perfect fit better but there are small indications that the synthetic contrast is higher.

Most plots of the separate data series indicated correlation. The correlation plot for the contrast comparison between occipital cortex and CSF showed that the contrast in the conventional images were fairly the same even if the contrast in the synthetic images were varying. This could be an indication that the variation is higher between the synthetic images. When also studying the Bland Altman plots a slightly different result was retrieved. All contrast series where CSF was compared to soft tissue had a positive trend. This is probably a result of different variations in the different methods. The contrasts in the synthetic images are varying much more and it is primarily this variation that affects the mean and difference in the Bland-Altman plot. This is also consistent with the standard deviations for the different methods in the comparisons with CSF, see Table 4-4. The standard deviation was more than twice as large in the synthetic images for all contrast evaluations between CSF and soft tissues.

The results from the analyses of the difference between the methods were consistent with the graphical analyses. All confidence intervals with the contrast between CSF and soft tissues were significantly less than zero which indicated that the conventional images had a higher contrast. This could indicate that the intensity in CSF is overestimated in the synthetic images, similar to the results from the T2 weighted images. For the contrasts between gray and white matter the results were different. Four intervals indicated that the contrast in the synthetic images is higher, three did not reveal any significant difference and only the contrast between thalamus and centrum semiovale was significantly higher in the conventional images.

When taking into account the variation in the conventional images not much additional information was retrieved. For the contrasts where the difference was significantly separated from zero the contrast was also larger than the variation in the conventional images, i.e. $H_0$ in the equivalence test could not be rejected. There were though one exception, the contrast between thalamus and splenium. The difference between the methods was significantly larger than zero but still smaller than the variation in the conventional images.

6.4.2 CNR
An overview of the CNR measurements in the T2 FLAIR images can be seen in the correlation plots in Figure 4-16 and Figure 4-17. For the CNR between white and gray matter the synthetic images and conventional images seem relatively equal. However the CNR between CSF and soft tissues are higher in the conventional images. The variation in CNR is also higher in the conventional images than in the synthetic images. This can also be seen if the standard deviations for the different
methods are studied, see Table 4-4. The standard deviations in the synthetic images are less than half the standard deviation from corresponding measurements in the conventional images.

When the different data series were studied separately all of them indicated some correlation, not necessary exact agreement. The Bland-Altman plots agree with the correlation plots for all CNR between white and gray matter except for the CNR between the cortexes and splenium. These two plots and all plots with CNR between CSF and soft tissues had negative trends. The cause of these trends is probably that the variation is much higher in the conventional images. The mean and difference is therefore mainly affected by the CNR in the conventional images.

The CNR was higher in the conventional images for all the comparisons with CSF according to the confidence intervals of the difference, same as for the contrast. These results are not surprising because the experience of the image quality, especially in the ventricles, is much worse in the synthetic T2 FLAIR images than in the conventional images, see Figure 6-1. It seems that the synthetic T2 FLAIR images contain much noise that will affect the intensity and thereby also affect the contrast and CNR. This also indicates that the intensity in CSF is overestimated in the synthetic images, same as for the T2 weighted images.

For the remaining comparisons of CNR the differences were mainly insignificant. This indicates that the CNR in the synthetic images is comparable with the CNR in the conventional images except for the CNR between CSF and soft tissues.

![Image](Image 71x287 to 525x453)

**Figure 6-1.** To the left a selection from a synthetic T2 FLAIR image can be seen and to the right the corresponding conventional image selection. The experience of the images is that the synthetic image has worse image quality, especially visible when studying the noise in the ventricles, black areas.

All the CNR comparisons where the difference was less than the variation of contrast in the conventional images, i.e. H₀ in the equivalence test could be rejected, had an insignificant result in the first analysis except thalamus compared to genu. For the other cases with significant difference the difference was also higher than the variation in the conventional images. Six of the eight comparisons with a difference higher than the variation in the conventional images were CNR between CSF and soft tissues, similar to the result for the T2 weighted images.

Interesting is that all the comparisons with CSF showed significant difference for both T2 weighted images and T2 FLAIR images. Otherwise there were no obvious trends. The noise in CSF in the synthetic T2 FLAIR images are visually obvious and together with the other results this is a clear
indication that CSF is not likewise separable in the synthetic images as in the conventional. It is possible that similar noise exist in the T2 weighted images but this is not as visible, probably because the voxels in CSF have very high intensity in these images. This is probably not of clinical importance since CSF is clearly visible anyway because its intensity is so different compared to other tissues. It is though technically important since one of the intentions with synthetic MRI is to reproduce the conventional images and then the synthetic images fails when it comes to CSF in T2 weighted and T2 FLAIR images.

6.4.3   Negative Contrast Measurements
There were unfortunately many data series that contained negative values in the T2 FLAIR images. This is probably a result of the low contrast in the T2 FLAIR images and therefore the relation of intensity is not always as expected. For several of these cases there were no indications that the negative values would affect the result significantly but three of the data series had many negative values, the contrast between thalamus and centrum semiovale, occipital cortex and splenium, frontal cortex and centrum semiovale. These comparisons cannot be seen as reliable. The risk is that the difference in contrast is overestimated.

6.4.4   Summary
The main problem with the contrast and CNR measurements in T2 FLAIR images is CSF since all these analyses indicated poor agreement between the methods. The difference in contrast between the methods was quite high in many other tissues and therefore agreement between the methods cannot be stated. It is possible that an improvement of the intensity in CSF would result in higher agreement between the methods but there will probably still be a high variation in the contrast between soft tissues. When also taking account of the experience from a radiologist it is clear that the synthetic and conventional T2 FLAIR images are not similar enough. The artifacts in the images can be interpreted as disease and this is not appropriate for clinical use.

6.5   Resulting Scan Time
When MRI scans are performed at CMIV, Linköpings University Hospital, the approximated time for the different sequences are:

- Conventional T1 weighted image stack: 5 minutes
- Conventional T2 weighted image stack: 3 minutes and 40 seconds
- Conventional T2 FLAIR weighted image stack: 6 minutes
- Quantification stack: 6 minutes

Since the synthetic T2 FLAIR images contain artifacts it is suggested that the conventional method is used to retrieve these images. If the T1 and T2 weighted images are replaced by synthetic images the total resulting scan time for these images would be twelve minutes compared with a total conventional examination which would take 14 minutes and 40 seconds. The time saving would be slightly less than three minutes.

6.6   General Problems
When analyzing the images and data some problems occurred in all weighting schemes. As discussed earlier the contrast was not always positive which would be expected. Different resolutions and slice thicknesses were used in the different image stacks which are difficult problems to estimate the effect of.
6.6.1 Unexpected Relation of the Tissue Intensity

It is difficult to accurately evaluate the result when the intensity difference between two different tissues is not as expected. This could be of natural cause since all studied objects have some disease which can affect the tissue, e.g. in MS where myelin in the white brain matter is degraded which will change the tissue and thereby also the intensity in the MR images. If this occurs the same pattern should be visible in both the conventional and the synthetic images. Another reason for such a result could be out of coincidence. If there are only small differences in intensity between the two tissues that are studied, small errors or ROI misplacement can result in an unexpected behavior. If this is the case the result will not be strongly affected since it should not generate any extreme values. If the unexpected relation only exists in one of the imaging methods the conventional images has to be assumed to be correct since conventional MRI is the gold standard today. This problem occurred more often in T2 FLAIR images than in the other weighting schemes, probably because the contrast is lower in the T2 FLAIR images.

This issue becomes a problem if the negative contrast is in a conventional image and the corresponding synthetic image generates a positive contrast. The difference between the methods will then be highly positive and indicate that the contrast is higher in the synthetic image but the synthetic image have the wrong relation on the contrast compared to the gold standard. The risk is that the contrast difference is overestimated and therefore conclusions about the difference between the contrasts for the image types should be made with caution.

6.6.2 Selected Structures and ROI Placement

During the analyses all structures of different tissues were compared because more information could then be retrieved. Some of the comparisons may have been less informative. One difficulty was to accurately place ROIs in occipital cortex because it is a thin structure. Many of these ROIs contained some white matter and therefore the measurements in occipital cortex were not as stable as for the other structures.

The interest to compare thalamus with centrum semiovale is questionable. This is because centrum semiovale can be found in the upper part of the brain whereas thalamus is in the lower part of the brain. These two structures will never be in the same transversal image slice and therefore it is not an interesting comparison.

Another problem with thalamus is the tissue classification. The center of thalamus consists of gray matter but closer to the outer part of thalamus the tissue is more similar to white matter. Thalamus is still classified as gray matter but as the comparisons between thalamus and white matter indicates, the difference is not large and reversed relation can occur. Small differences in placement of the ROIs in thalamus can therefore give different tissues within the ROI.

6.6.3 Imaging Problems

One issue that complicates the interpretation is that all image stacks do not have the same resolution and slice thickness. Generally the synthetic images have a lower resolution and thicker slices than the conventional images. The ROIs sizes have been equal in both image types in this study which means that the values from the synthetic images are calculated from fewer voxels. Lower resolution during image acquisition also results in more partial volumes. This could both increase and decrease the contrast depending on which tissues the contrast is calculated between. One example is if the contrast between gray matter and CSF is studied in T1 images and the voxels with gray matter partly
contains white matter. This will increase the intensity in the gray matter and also the intensity difference between gray matter and CSF and result in an overestimation of the contrast and CNR.

The noise is also affected by the resolution. Small differences within a voxel disappeared due to averaging which reduces the noise. One voxel can also contain two different tissues, partial volumes. Partial volumes will inaccurately increase the noise estimation. Both the effect of averaging and the effect of partial volumes will increase with lower resolution. The effect of partial volumes is probably the strongest of these effects and therefore the noise is probably overestimated in the synthetic images since they have a lower resolution. The CNR is then underestimated and it is possible that the synthetic images have a higher CNR than this study indicates.

6.6.4 Software Problems
All the images were analyzed in Sectra PACS, IDS7. One problem arose with this software during the evaluation. The images are archived when they have been in the system for approximately one week. If any measurements, like ROI placement, are made before the archiving they will be permanently stuck in that place. The problem was that the measurements can be moved up to half a voxel during archiving which result in slightly different values. This occurred in all synthetic image stacks and one of the conventional image stacks and therefore the values used cannot be exactly checked since they were retrieved before archiving. The reason for using the values before archiving was that some of the archived ROIs had moved and crossed an edge, i.e. two different tissues were in the ROI and these values could therefore not be seen as reliable.

6.6.5 Simultaneous Confidence Level
For each contrast/CNR comparisons two confidence intervals were made which reduced the simultaneous confidence level to 90% (since the intervals had a confidence level of 95% each). It would have been desirable to have a higher simultaneous confidence level. The problem was that the idea of using the range for the conventional images contrast as indication of the variation in the conventional images came after all the difference analysis had been made. To make this variation as comparable as possible with the confidence interval for the difference a 95% confidence level was chosen also for that interval.

6.7 Further Development and Research
Some problems have occurred during this thesis. With the knowledge gained it would be possible to conduct better studies in the future. The indications from this thesis can also be used as a base for future studies. The main focus in the future should be if the difference in image quality is clinically decisive.

6.7.1 Improvements for Similar Studies
The first step in a study is to collect volunteers. Healthy subjects should preferably be used for evaluating the contrast between normal brain tissues. The advantage of using healthy subjects is that if the measurement is unexpected it is reasonable to assume that they occur due to errors in the measurements, which will always be present, and not depend on a disease.

The second step is to collect the data. The same settings should be used for all image acquisitions. This is because the effect of different resolutions and slice thickness is an unnecessary bias which is hard to estimate the size of.
Next issue is management of the data. One problem with archiving in Sectra PACS occurred and therefore the suggestion is to force archiving before any measurement. This is because of the benefit of having the ability to check the values afterwards.

To be able to state equality between the methods a maximal difference that is acceptable is required. The approximation of the acceptable difference was deficient in this study. The differences between two conventional images could be evaluated in the same way as the difference between the methods. This would give a better estimate of the difference in two conventional images of the same subject.

The last suggestion for improvement is to use an independent statistician. The problem in this study was that the statistician had done much preparatory work and was therefore able to see where the data used in the analyses came from. The blinding could therefore not be conducted appropriately.

6.7.2 Suggestions for Future Research
The resolution and slice thickness was discussed earlier and an evaluation of these effects could be useful. If the contrast is much dependent on the resolution, a longer scan time could be reasonable to provide higher image quality. It could also give information about the importance in using the same settings for all scans in scientific studies. This could also be used to ensure stability of the method since the images and contrast should be fairly the same for two directly subsequent scans. Such study could be conducted in a similar way as this study. An ANOVA analysis would be appropriate if the effect of both resolution and slice thickness is studied simultaneously.

The results indicated that the difference in contrast was dependent on the tissues. It could therefore be interesting to evaluate the relative differences in intensity for the different imaging methods. One suggestion is to subtract every voxel in a synthetic image from the corresponding voxel in a conventional image. An optimization algorithm is then required to deal with the differences in scale. The result would be a difference map with only zeros if the images are totally similar. This difference map could be used to evaluate if the difference is varying between different tissues. During this study it was suggested that the intensity in CSF may be overestimated in the synthetic images and such indication could be investigated.

It appeared that the intensity in lesions is not perceived equal in synthetic and conventional images. Diffuse lesions in conventional images appeared brighter in the synthetic images and vice versa. The statistical analysis of the contrast and CNR did not reveal such indications but this could depend on the choice of lesions. A larger study with different types of lesions could be valuable, especially since the intensity of the lesion is used to estimate the age of the lesions. The analysis could be conducted with a two-way-ANOVA where both the lesion type and the disease are factors.

Clinical Effects
Only the measured image quality was evaluated in this thesis. It was not evaluated if the difference in image quality affects the clinical outcome. One suggestion is to use ratings to evaluate the clinical value, e.g. to rate the ability to distinguish different anatomical features like lesions in both conventional and synthetic images. The assessment scale could as a suggestion extend between not sufficient and good, with an arbitrary number of steps. The similarity between the methods can be evaluated with kappa analysis. However the rating could be discriminated into two levels, not
sufficient and sufficient, if the goal is to investigate if the synthetic images provide sufficient quality or not.

The most interesting question is if the imaging method affects the diagnosis or the treatment. MRI examinations are however only one piece of a larger puzzle for diagnosis. It is therefore inappropriate to treat MRI as a diagnostic test. It could be possible to quantify the disease and then the quantification can be used for evaluation of the diagnostic ability. One suggestion is to count the number of lesions in each image stack and evaluate the effect of different imaging methods with Poisson-regression.

To count the number of lesions is applicable only for diseases with well-defined lesions. Many small lesions and larger diffuse areas of pathology are common in patients with MS or ischemia and therefore such counting is not reasonable. An alternative is to count the number of lesions that are only visible in the conventional images; lesions which are only visible in the synthetic images would then result in a negative count. A Kruskal-Wallis H-test can then be used to evaluate if the disease affects the outcome. Since the variable is presumably not normally distributed the suggestions for analysis are the sign test or Wilcoxon signed rank test. With Wilcoxon signed rank test it is possible to create confidence interval which gives indication on how differing the imaging methods are. However it is only useful if there are few ties. It is important to realize that these tests primarily detect if the imaging methods are different. This is because two observations from each patient are required to estimate the agreement.

How the imaging method affects the treatment is difficult to investigate with the same group of patients as in this thesis. One suggestion is to select patients for two groups where one group is evaluated with the synthetic images and the other with the conventional images. The groups have to be fairly similar for the result to be valid. The frequencies of different treatments can be evaluated with Fisher’s exact test or Chi-square test depending on the sample size. The result indicates if the imaging methods have different distributions i.e. if different treatments are used depending on the imaging method.

For information about Kruskal-Wallis H-test see (17) and for the other statistical methods suggested in this section see (12; 18).
7 Conclusions

Synthetic MRI and conventional MRI do not generate equal contrast. The result suggests that the synthetic method provides both higher contrast and higher noise in T1 weighted images and T2 weighted images. The synthetic images are therefore preferable since a higher contrast makes it easier to distinguish tissues. On the other hand investigations whether the difference in contrast is acceptable for clinical use or not are required before assuming that the synthetic images are better. The synthetic T2 FLAIR images have a lower contrast than the conventional images and they also contain artifacts which make them inappropriate for clinical use.

This study indicates that the synthetic images are not equal to the conventional images but it is not known if the differences are clinically decisive. An evaluation of the clinical value could be useful since it is more important than the image quality.
Bibliography


Appendix A

This appendix includes the statistical methods needed for this study. Different approaches for analysis are used depending on the question of the study. Significance analysis is used if the intention is to prove a difference between two groups. In some cases it can be interesting to evaluate how similar two methods are e.g. if a new method is cheaper than the one that is used. Link analysis is used to see if the new method is as good as the more expensive method, for possible replacement. Many different methods exist but only those that are somehow interesting for this study is presented in this chapter.

A.1 Shapiro-Wilk test - Test for Normality

Several ways have been proposed for testing if a sample is normally distributed. The Shapiro-Wilk test is based on correlation and if the sample is normally distributed then the test statistic \( W \) is near 1 (10).

\[
W = \frac{1}{D} \left[ \sum_{i=1}^{k} a_i (x_{n-i+1} - x_i)^2 \right] \quad (A.1)
\]

\[
D = \sum_i (x_i - \bar{x})^2 \quad (A.2)
\]

The test statistic is used to calculate \( z \) which p-value is given by comparison with the standard normal distribution. The assumption of normal distribution is rejected if the p-value is low.

\[
z = b_n + c_n \ln \left( \frac{W - d_n}{1 - W} \right) \quad (A.3)
\]

All the parameters \( k, a, b, c \) and \( d \) are collected from tables and \( n \) is the number of observations.

The test can be made for sample sizes between 3 and 5000 with the present approximations and algorithms (10). Suggestions have been made that the test statistic should be corrected for the sample size since it affects the outcome, generally a larger sample size gives a more correct result (10). However this is not always practical since this feature is not provided in all software.

The Kolmogorov-Smirnov test which also exists with Lilliefors’ correction is a more conservative test of normality. However indications that it is not as robust as the Shapiro-Wilk test have been found (10).

A.2 t-test – Significance Analysis for Normal Distributed Variables

The t-test is used to evaluate parameter estimation from one or two normally distributed variables with unknown variances. If only one variable is studied then the usual approach is to investigate if it is significantly differing from zero. The same approach is used if the sample is paired, e.g. measurements before and after a treatment. The pairwise difference, \( d \), is evaluated and treated as one sample which gives the test statistic (7):

\[
z = \frac{d}{S_d \sqrt{\frac{1}{n}}} \quad (A.4)
\]
The stochastic variable \( Z \sim t(n - 1) \) if the null hypothesis is true. The null hypothesis \( H_0: d = 0 \) is rejected for large and/or small values of \( z \), cut off values are collected from the standard normal PDF.

A confidence interval can also be created based on the test statistic. The probability that the estimated parameter, the mean difference, is in the interval is then \( 1 - \alpha \). Some remodeling of the test statistic gives the confidence interval:

\[
CI(\bar{d}) = \left( \bar{d} \pm \frac{t_{\alpha} s_d}{\sqrt{n}} \frac{1}{\sqrt{n}} \right)
\]

This is a two way confidence interval with \( t_{\alpha} \) as cut off value for the significance level \( \alpha \), same for the negative and positive side since the t distribution is symmetric around the zero. If the confidence interval is used for testing the null hypothesis then rejection occur if the interval is totally separated from zero.

The t-test could also be used for two independent samples. Some remodeling is then demanded and slightly different approaches are used if the variances of the different samples are equal or not.

**A.2.1 Equivalence Testing – Normal Distributed Variables**

When dealing with normally distributed variables it is easy to reject a specific value but to determine equivalence between two groups is more complicated. For equivalence testing it is necessary to first define how different the two groups could be and still be assumed as equal. This acceptance value is denoted as \( \Delta \) and then the null hypothesis becomes \( H_0: \delta \geq \Delta \) where \( \delta \) is the true difference between the two groups (18). The observed mean difference is denoted \( \bar{d} \) and a confidence interval can be calculated in the same way as for the t-interval. The null hypothesis is rejected if (18):

\[
|z| = \left| \bar{d} \pm t_{\alpha} s_d \frac{1}{\sqrt{n}} \right| < \Delta
\]

\( |z| \) is the limit of the confidence interval with highest absolute value.

To define the acceptable difference can sometimes be difficult. It is possible to create a confidence interval and afterwards discuss if the difference is acceptable but it is preferable to use the equivalence test since then the interpretation effect of the result is smaller.

**A.3 Wilcoxon’s Signed Rank Test – Nonparametric Test of Symmetrical Distributions**

Wilcoxon’s signed rank test is a non-parametric option to the t-test and it can therefore be used for non-normal distributions. The test evaluates if the distribution is symmetric around a specific median, \( m_0 \) (19).
First all the differences, \( y_i = x_i - m_0 \), is calculated and arranged in order of magnitude based on \( |y_i| \), observations where \( y_i = 0 \) are ignored. Ranks from 1 to \( n \) are then assigned to the arranged values. In the presence of any ties they both get the mean of the following ranks. The test statistics are (19):

\[
T_+ = \text{sum of ranks for the positive observations } y_i
\]

\[
T_- = \text{sum of ranks for the negative observations } y_i
\]

If the null hypothesis is true and the distribution is symmetric around \( m_0 \) then \( T_+ \) and \( T_- \) have the same distribution. \( W_s \) is the stochastic variable with the signed rank distribution for sample size \( n \) which is used to retrieve the cut off value \( c \) (19):

\[
\frac{\alpha}{2} = P(T_+ \leq c) = P(T_- \leq c) = P(W_s \leq c)
\]  

(A.8)

The null hypothesis is rejected if \( T_+ \) or \( T_- \) is smaller than \( c \). If the sample size is smaller than 15 the cut off value can be obtained from a table but if the sample size is larger than 15 then \( W_s \) is approximately normal distributed with (19):

\[
E[W_s] = \frac{n(n + 1)}{4}
\]

(A.9)

\[
s_{W_s} = \sqrt{V[W_s]} = \sqrt{\frac{n(n + 1)(2n + 1)}{24}}
\]

(A.10)

This test can also be conducted for paired samples where \( y_i = x_{1i} - x_{2i} \) and the null hypothesis is that the difference is symmetrically distributed around zero.

**A.3.1 Confidence Interval Based on Wilcoxon’s Signed Rank Test**

If a confidence interval is of interest than the original observations (or pairwise differences) are ordered on size where \( O_1 < O_2 < \ldots < O_n \). Then mean values are calculated with (19)

\[
\frac{O_i + O_j}{2} \text{ where } j \geq i
\]

(A.11)

The ordered mean values are denoted \( A_1, \ldots, A_N \) and it is possible to show that (19)

\[
P(A_k < m < A_{N-k+1}) = 1 - 2P(W_s \leq k - 1)
\]

(A.12)

\( k \) is derived in the same way as for Wilcoxon signed rank test with the condition:

\[
\frac{\alpha}{2} = P(W_s \leq k - 1)
\]

(A.13)

The confidence interval \( I_m = (A_a, A_{N,k+1}) \) has then the confidence level \( 1-\alpha \).

**A.4 Analysis of Variance – ANOVA**

If the response variable is quantitative and dependent of one or more factors then analysis of variance, ANOVA, is commonly used. The response is then modeled to depend on the mean of the total sample and effects of different group belongings. A factorial design with more than one factor can be modeled in two ways, additive or full factorial. The difference is that the full factorial design
takes the interaction between factors into account. One-way ANOVA, one factor design, is handled in this section but all the equations can be generalized for multi factorial design. However if many factors are involved it will require a large amount of data to be able to analyze all the effects.

The easiest way to explain ANOVA is with an example which is used through this whole section. The strength of different rods is examined which is thought to depend on the material of the rod. The observed strength, $y_{ij}$, can be modeled as:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (A.14)$$

The total mean is denoted $\mu$, the effect of the $i$:th material is denoted $\tau_i$ and $\varepsilon_{ij}$ denotes the standard deviation. $\varepsilon_{ij} \sim N(0, \sigma)$ and the stochastic variable $Y_{ij} \sim N(\mu + \tau_i, \sigma) \quad (7)$.

All the parameters in the model have to be estimated. If the $i$:th material is represented with $n_i$ observations and $N$ is the sum of all $n_i$ the estimations becomes (7):

$$\hat{\mu} = \bar{y}_i \quad (A.15)$$

$$\hat{\tau}_i = \bar{y}_i - \bar{y}_L \quad (A.16)$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_L \quad (A.17)$$

$$\bar{y}_i = \frac{1}{n_i} \sum_{j} y_{ij} \quad (A.18)$$

$$\bar{y}_L = \frac{1}{N} \sum_{i} n_i \bar{y}_L \quad (A.19)$$

$\varepsilon$ has to be normally distributed for the model to be accurate. Evaluation of normality can be conducted in different ways. A normal probability plot, see section 2.1.2, can be used and plotting the residuals, $\hat{\varepsilon}$, against the predicted value $\bar{y}_L$ can assure that the error does not depend on the magnitude of the measurement. The residuals for the different materials can be plotted to reveal if the error is larger for some groups. Severe deviations can impact the result and conclusions (7). Note that $\varepsilon$ cannot be estimated if there is only one observation for each material since $\varepsilon_{ij} = 0$.

It is the different effects that are analyzed during an ANOVA. Variances are analyzed with sum of squares. The total sum of squares for a sample is the numerator in the variance estimation (7):

$$SS_T = \sum_{i} \sum_{j} (y_{ij} - \bar{y})^2 \quad (A.20)$$

The total sum of squares can be divided into several components describing variations between the groups ($SS_{TREAT}$) and variation within the groups ($SS_E$) (7). The effect of the material can be evaluated with the sums of squares, great variation between materials indicate that the material affect the response much.

All these sums of squares are $\chi^2$-distributed with the degrees of freedom shown in Table A-1. An ANOVA also contains F-statistics with its corresponding p-value, the test is described in section A.5. The F-distributed variable is created from $SS_{TREAT}$ and $SS_E$. Rejection of the null hypothesis indicated that material is important for predicting the strength of a rod.
Table A-1. Two-Way Analysis of Variance (7). N denotes the total sample size, \( n_i \) denotes the sample size for the \( i \)-th material and \( a \) denotes the number of materials that are tested.

<table>
<thead>
<tr>
<th>Source</th>
<th>Sums of squares</th>
<th>Df</th>
<th>Mean squares</th>
<th>F statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Material</td>
<td>( SS_{TREAT} = \sum_i n_i (y_{il} - \bar{y}_l)^2 )</td>
<td>( a-1 )</td>
<td>( MS_{TREAT} = SS_{a}/df_{TREAT} )</td>
<td>( MS_{TREAT}/MS_E )</td>
</tr>
<tr>
<td>Error</td>
<td>( SS_E = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 )</td>
<td>( N-a )</td>
<td>( MS_E = SS_E/df_E )</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( SS_T = \sum_i \sum_j (y_{ij} - \bar{y}_\cdot)^2 )</td>
<td>( N-1 )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### A.5 F-test – Test of Equal Variances

It can be of interest to compare two samples that are normally distributed. If the variances of the samples can be assumed to be equal then the calculations gets easier. This type of analysis can be made with an F-test. Another important use of F-test is to investigate if different effects in an ANOVA analysis are of importance.

All the sums of squares in an ANOVA analysis are assumed to be \( \chi^2 \)-distributed. The quotient of two \( \chi^2 \)-distributed stochastic variables divided by their degrees of freedom gives an F-distributed variable. The test statistic for the ANOVA example becomes (7):

\[
\nu = \frac{SS_{TREAT}/(a - 1)}{SS_E/(N - a)} \sim F(a - 1, N - a)
\]  
(A.21)

The null hypothesis is that the effect of treatment (TREAT) is not important (\( H_0: SS_{TREAT} = SS_E \)) which can be rejected if the test statistic is high since large variation between groups generates a large \( SS_{TREAT} \). Cut off values are retrieved from the F-distribution.
Appendix B

function [CI, stat, pr] = signrankinterval(X, alfa)

[n c] = size(X);

if c==2
    D=X(:,1)-X(:,2);
elseif c==1
    D=X;
end

Ds=sort(D);
N=n*(n+1)/2;
[pr h stat]=signrank(Ds); % implements a Wilcoxon signed rank test

g=alfa/2;
if n>15
    t=icdf('normal',g,0,1);
    k=round(t*sqrt(n*(n+1)*(2*n+1)/24)+n*(n+1)/4)+1;
    % k is related to the cut off value with approximated normal distribution
else
    k=20; % this is table value if n=14, which were the case for pathology
end

A=zeros([N 1]);
p=1;
for i=1:n
    for j=i:n
        A(p)=(Ds(j)+Ds(i))/2;
        p=p+1;
    end
end
A=sort(A);
CI=[A(k) A(N-k+1)]; % the confidence interval
end