DEPARTMENT OF WOMAN AND CHILD HEALTH
Karolinska Institutet, Stockholm, Sweden

LONG-TERM MAGNETIC RESONANCE IMAGING FOLLOW-UP AFTER PRETERM BIRTH

Zoltán Nagy

Stockholm 2009
If our small minds, for some convenience, divide ... this universe into parts – physics, biology, geology, astronomy, psychology, and so on – remember that Nature does not know it!!

Richard P. Feynman

The sciences do not try to explain. They hardly even try to interpret. They mainly make models. By a model is meant a mathematical construct, which, with the aid of certain verbal interpretations, describes observed phenomena. The justification for such a mathematical construct is solely and precisely that it is expected to work.

John von Neumann
Abstract

Background
Since the introduction of neonatal intensive care units, the survival after preterm birth has improved drastically and development after preterm birth has become an area of intense research. Many studies have found increased rates of neonatal and long-term morbidity. In particular, preterm birth has been identified as a risk factor for cognitive and structural brain development and cardiovascular diseases. However, it is also appreciated that the outcome after preterm birth depends on antenatal, neonatal and postnatal care.

Aims
The aims of these studies were twofold: to use a range of MRI imaging techniques to investigate whether preterm born children and adolescents show differences in structural brain and aortic development as well as to compare these results to those reported from other countries, involving similar cohorts.

Subjects & Methods
The volunteers of interest in these studies were participants in the Stockholm Neonatal Project, which was initiated in 1988 as a prospective follow-up study and included preterm infants if their weight at birth was less than or equal to 1500g. At 5 ½ years of age a group of term-born children were selected as a control group. In a pilot study (Paper V) a subgroup of these individuals were examined. Papers I–IV involve those 74 out of 182 ex-preterm and 69 out of 125 control adolescents that were willing to participate.

The participants underwent cranial MRI examinations at the MR Center at the Karolinska University Hospital to collect 3D anatomical images and diffusion tensor imaging data. In a subgroup of individuals, images were also collected of the descending aorta.

Results
In all studies we found preterm birth to be a risk factor for long-term development. Ex-preterm adolescents possessed more structural brain abnormalities mostly in the form of gliosis and white matter loss (Paper I). Their grey and white matter volumes were also lower with the most preterm showing the largest deviations from that of control adolescents (Paper II). Ex-preterm adolescents also possessed large areas of thinner cortex (Paper III), hindered white matter microstructure (Papers II and V) and smaller diameters within their descending aortas (Paper IV).

Discussion
While these results are not unprecedented, with respect to brain the severity and extent of these findings were milder than that based on similar cohorts. It may be that the care provided to the preterm infants in the prenatal and neonatal periods resulted in these relatively moderate findings. Should that be the case, it is hopeful that further research will identify the appropriate care such that preterm born individuals may one day develop identically to those born at term.
**List of publications**

I. *Cerebral MRI Findings in a Cohort of Ex-preterm and Control Adolescents*
   **Zoltán Nagy** and Baldvin Jónsson
   Accepted in Acta Paediatrica

II. *Structural Correlates of Preterm Birth in the Adolescent Brain*
   In Review at Pediatrics

III. *The Long–Term Effects of Preterm Birth on Cortical Thickness*
   **Zoltán Nagy**, Hugo Lagercrantz and Chloe Hutton
   In Review at Cerebral Cortex

IV. *Preterm birth and maternal smoking in pregnancy are strong risk factors for aortic narrowing in adolescence*
   Anna–Karin Edstedt Bomany, Johan Bengtsson, **Zoltán Nagy**, Hans De Keyzer and Mikael Norman

V. *Preterm Children Have Disturbances of White Matter at 11 Years of Age as Shown by Diffusion Tensor Imaging*
   **Zoltán Nagy**, Helena Westerberg, Stefan Skare, Jesper Andersson, Anders Lilja, Olof Flodmark, Elisabeth Fernell, Kirsten Holmberg, Birgitta Böhm, Hugo Lagercrantz and Torkel Klingberg
## List of Figures

### Introduction
- **Figure 1.1.1** Relationship between GA at birth and IQ  
  p3
- **Figure 1.1.2** Relevant aspects of embryonic and foetal brain development  
  p5

### Methods
- **Figure 1.3.1** A simplified illustration of magnetization  
  p10
- **Figure 1.3.2** A more accurate illustration of the magnetization process  
  p11
- **Figure 1.3.3** Illustration of the T1-relaxation process  
  p12
- **Figure 1.3.4** Examples of T1-weighted images  
  p13
- **Figure 1.3.5** Illustration of the T2-relaxation process  
  p15
- **Figure 1.3.6** Example of a T2-weighted image  
  p16
- **Figure 1.3.7** Illustration of diffusion-weighting using magnetic gradients  
  p19
- **Figure 1.3.8** Illustration of collecting DTI data  
  p20
- **Figure 1.3.9** Examples of FA and colour-coded images  
  p21
- **Figure 1.3.10** Brain tissue segmentation based on MRI images  
  p23
- **Figure 1.3.11** Example templates for anatomical normalization  
  p25
- **Figure 1.3.12** Illustration of susceptibility induced artefacts and their correction  
  p27
- **Figure 1.3.13** Mean FA image of the cohort  
  p28
- **Figure 1.3.14** Testing the accuracy of the aortic segmentation  
  p29
- **Figure 1.3.15** Collection and segmentation of the abdominal aorta images  
  p30
- **Figure 1.3.16** Determination of statistical significance  
  p32
- **Figure 1.3.17** The need for multiple comparison correction  
  p33

### Results
- **Figure 1.4.1** Incidental findings on cranial MRI  
  p39
- **Figure 1.4.2** Results without correcting for multiple comparisons I  
  p40
- **Figure 1.4.3** Results without correcting for multiple comparisons II  
  p41
- **Figure 1.4.4** Checking for confounds in the VBM results  
  p42
- **Figure 1.4.5** VBM results after inclusion of total brain volume as a covariate  
  p43

### Discussion
- **Figure 1.5.1** Reducing variability in the data  
  p52
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.1</td>
<td>Robustness of aortic area measurements</td>
<td>p28</td>
</tr>
<tr>
<td>1.3.2</td>
<td>Accuracy of aortic area measurements</td>
<td>p29</td>
</tr>
<tr>
<td>1.3.3</td>
<td>Demographic data on participants vs. all available</td>
<td>p35</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Neonatal signs of brain injury for participants and non-participants</td>
<td>p36</td>
</tr>
<tr>
<td>1.3.5</td>
<td>Demographic data for the individuals included in Papers I–IV</td>
<td>p37</td>
</tr>
<tr>
<td>1.3.6</td>
<td>MRI acquisition parameters</td>
<td>p38</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>Birth weight</td>
<td></td>
</tr>
<tr>
<td>CNR</td>
<td>Contrast to noise ratio</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebro–spinal fluid</td>
<td></td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion tensor imaging</td>
<td></td>
</tr>
<tr>
<td>EPI</td>
<td>Echo–planar imaging</td>
<td></td>
</tr>
<tr>
<td>FA</td>
<td>Fractional anisotropy</td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td>Gestational age</td>
<td></td>
</tr>
<tr>
<td>GM</td>
<td>Grey matter of the brain</td>
<td></td>
</tr>
<tr>
<td>IVH</td>
<td>Intra–ventricular haemorrhage</td>
<td></td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
<td></td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
<td></td>
</tr>
<tr>
<td>PVL</td>
<td>Periventricular leucomalacia</td>
<td></td>
</tr>
<tr>
<td>ROP</td>
<td>Retinopathy of prematurity</td>
<td></td>
</tr>
<tr>
<td>SGA</td>
<td>Small for gestational age</td>
<td></td>
</tr>
<tr>
<td>TBSS</td>
<td>Tract–based spatial statistics</td>
<td></td>
</tr>
<tr>
<td>VBM</td>
<td>Voxel–based morphometry</td>
<td></td>
</tr>
<tr>
<td>WM</td>
<td>White matter of the brain</td>
<td></td>
</tr>
</tbody>
</table>
# Table of contents

1 **Thesis summary – Main section** 1

1.1 **Introduction** 1
1.1.1 Preterm birth 1
1.1.2 Causes of preterm birth 1
1.1.3 The short- and long-term effects of preterm birth 2
1.1.3.1 Neonatal mortality 2
1.1.3.2 Neurologically relevant neonatal morbidity 2
1.1.3.3 Long-term morbidity 3
1.1.3.4 Neonatal imaging findings 3
1.1.3.5 Long-term cognitive and/or MRI follow-up 3
1.1.4 Relevant aspects of embryonic and foetal brain development 4
1.1.5 Effects of preterm birth on vascular development 6

1.2 **Aims** 8

1.3 **Methods** 9
1.3.1 Magnetic resonance imaging 9
1.3.1.1 T1-weighted contrast 9
1.3.1.2 T2-weighted contrast 14
1.3.1.3 Diffusion-weighted contrast and diffusion tensor imaging 16
1.3.1.4 FIESTA for cardiac and vascular imaging 22
1.3.2 Image processing 22
1.3.2.1 Segmentation 22
1.3.2.2 Normalization 23
1.3.2.3 Estimating cortical thickness from MR images 25
1.3.2.4 Distortion correction in echo-planar imaging data 26
1.3.2.5 Edge detection for estimating aortic cross sectional area 28
1.3.3 Statistical considerations 31
1.3.3.1 Simple random sampling vs. cohort studies 31
1.3.3.2 Statistical significance 31
1.3.3.3 Multiple comparison in MR imaging studies 33
1.3.3.4 Voxel-wise statistics vs. tract-based spatial statistics for DTI data 34
1.3.4 Description of the cohort 34
1.3.5 MRI data acquisition 37

1.4 **Results** 39
1.4.1 Findings on cerebral MRI 39
1.4.2 Mean differences in structural brain images 40
1.4.3 Abdominal aorta imaging results 44

1.5 **Discussion** 45
1.5.1 The pilot study 45
1.5.2 Studies involving the entire cohort 46
1.5.2.1 Radiological evaluation of MRI images 46
1.5.2.2 The VBM and DTI results 46
1.5.2.3 Cortical thickness results 48
1.5.2.4 Abdominal aorta results 49
1.5.3 Discussion of the commonalities among the five studies 50

1.6 **Conclusions** 54

1.7 **Future Directions** 55

2 **Acknowledgements** 56

3 **References** 57
1 THESESUMMARY – MAIN SECTION

1.1 INTRODUCTION

Three themes join together the five papers compiled at the end:

1) In each of the studies the method of data collection was magnetic resonance imaging (MRI). This modality was used because it is non–invasive, it offers relatively high spatial resolution and many different contrasts are achievable.

2) The individuals involved were part of a follow–up study that was initiated in Stockholm in 1988.

3) The examinations took place between the ages of ~10–18y as the studies focused on if and how long–term development was affected by preterm birth.

1.1.1 PRETERM BIRTH

Preterm birth is defined as delivery occurring before the 37th week of gestation. Infants born before the 22nd week are considered foetal loss and are usually not revived, while those born after the beginning of the 37th but before the 43rd week are considered term–born.

It is important to note that the inclusion criterion for the studies compiled here was not based on the number of weeks of gestation. Instead in the Stockholm Neonatal Project, as commonly done at the time, birth weight (BW) was chosen as the variable of distinction with a cut–off of 1500g. Normally, there is a strong monotonic relationship between BW and length of pregnancy; on average foetuses reach 1500g in weight around the 30th week of gestation. Therefore, this inclusion criterion, based on weight, created a confounding effect in that children born closer to the 37th week of gestation tended to be small for their GA (SGA). Consequences of this confound will be elaborated upon in the Discussion.

1.1.2 CAUSES OF PRETERM BIRTH

Preterm birth can occur spontaneously or be induced upon detection of foetal or maternal illness. Spontaneous preterm birth occurs after either preterm labour or the premature rupture of the membrane before labour and can be caused by infection, inflammation, vascular disease, placental ischemia or haemorrhage, maternal and/or foetal stress and some other factors. Preterm birth is medically induced after maternal pre–eclampsia (a hypertensive disorder), intrauterine growth restriction or severe maternal stress.

Therefore, it is important to note that preterm cohorts are usually not homogeneous groups. Such is the case with the Stockholm Neonatal Project. Not only are inclusion criteria difficult to establish due to the many causes of preterm birth, the problem is further complicated by socio–economic status,
race and the occurrence of multiple births. Elucidating risk factors of preterm birth is an area of extensive research but will not be the focus of this thesis. The available cohort will be considered as a whole, representing the population of preterm births of many causes.

1.1.3 The Short-, and Long-term Effects of Preterm Birth

Since neonatal intensive care units were introduced, preterm birth has been identified as a risk factor in numerous studies. Either the causes or process of entering the extra-uterine environment prematurely or the inability or stress of sustaining life there can result in injury to the developing human body.

1.1.3.1 Neonatal mortality

Neonatal mortality rates increase the shorter the gestational period after which the infant is born. For example, in the late 80's in the United Kingdom the mortality rate was about 10% after 31 weeks of gestation but about 70% for 23–24 weeks. In 1999 Hack and Fanaroff reported a review of the literature and found that survival rates of only 2–35% for infants born after 23 weeks of gestation and 35–85% for those born after 25 weeks. In the mid 1990's the overall infant mortality rate in Sweden ranged from about 5% if the infant was born at 31 weeks of gestation to about 56% among those born at 24 weeks. The authors identified an excess risk of mortality if the infant was born at a general hospital as opposed to a university hospital.

In the Stockholm Neonatal Project, which included infants between 1988 and 1993, overall mortality rate was 17.5%, while among the group of infants born between the 23rd–24th weeks of gestation it rose to 25.0%. Half of the deaths occurred within the first two days of postnatal life.

1.1.3.2 Neurologically relevant neonatal morbidity

Common neonatal morbidity includes respiratory distress syndrome, periventricular leucomalacia (PVL), intra-ventricular haemorrhage (IVH) and ventriculomegaly. Because the newborn preterm infant is unable to maintain homeostasis, some form of respiratory support is often given. Because the lung at this age is noncompliant, continuous positive airway pressure is more beneficial and less harmful than mechanical ventilation. The risk of respiratory distress syndrome can be further reduced by the administration of lung surfactant to the infant, or if the preterm birth is anticipated by antenatal corticosteroid treatment to the mother. Note however, that while corticosteroid treatment has proven beneficial for lung development, there are indications that it negatively affects brain growth.

Only 41% of the participants in the Stockholm Neonatal Project required any mechanical ventilation (only 24% received it from the beginning). The rest were supported only by supplemental oxygen or nasal continuous positive airway pressure.
1.1.3.3 Long-term morbidity

Several conditions affect individuals more often if they were born preterm. These include chronic lung disease, retinopathy of prematurity (ROP) as well as neurologic sequelae and cardiovascular complications, the latter two of which will be dealt with in more detail below.

1.1.3.4 Neonatal imaging findings

Preterm newborns have been investigated both to follow normal development of the brain and to identify and characterize injury or abnormalities. As a result of these studies, preterm birth has been associated with a higher incidence of white matter (WM) injury, IVH, ventricular dilation, disturbances of myelination and WM microstructure, as well as inadequate growth of the cortex, WM or subcortical grey matter (GM).

1.1.3.5 Long-term cognitive and/or MRI follow-up

While it is not evident that the abnormalities identified on neonatal neuroimaging should persist and be identified later in development, follow-up studies have identified preterm birth to be a risk factor even for long-term brain development. On cerebral MRI examination 15–year-old adolescents had a high prevalence of structural brain lesions, larger ventricles and thinning of the corpus callosum. When compared to term-born controls, Peterson et al. have reported wide-spread reduction in brain volumes and an increase in ventricular size at eight years of age. Similarly, Nosarti and co-workers found that ex-preterm 15–year-olds had 6% reduction in total brain volume and a 42% increase in the size of the lateral ventricles. When using diffusion tensor imaging (DTI) it was found that many fibre tracts possess lower fractional anisotropy (FA) in ex-preterm adolescents.

Figure 1.1.1

Relationship between GA at birth and IQ
These structural differences might not be cause for concern unless they also caused (or at least co–occurred with) deficits in cognitive ability. Such deficits have been reported. For example, preterm children have demonstrated disability on psycho–motor tests at 30 months of age\textsuperscript{190}, achieved lower scores on cognitive and neurological assessment at 5–6 years\textsuperscript{26,27,109,118} and 11 years of age\textsuperscript{42,43}; reached lower educational levels and had higher neurosensory impairment in young adulthood\textsuperscript{11}. In a meta–analysis, Bhutta et al. concluded that ex–preterm school–aged children scored lower on cognitive and behavioural tests regardless of when the tests were performed and that the deficits compared to controls were more severe the shorter the gestation or smaller the BW\textsuperscript{24}. The general finding is that the lower the GA at birth the more severe the deficits. An example of this is given in Figure 1.1.1 where IQ scores at 4–6 year old children is plotted against their GA at birth (reproduced with kind permission from and Drs. D. Wolke and N. Marlow but see also\textsuperscript{105}).

In neuroimaging studies a tendency has been to combine neuroimaging with some sort of cognitive/psychological testing\textsuperscript{39,45,85-87,101,119,139,158,166,174}.

1.1.4 RELEVANT ASPECTS OF EMBRYONIC AND FOETAL BRAIN DEVELOPMENT

The development of the central nervous system is a drawn out process that starts during the embryonic stage and continues actively into at least young adulthood. Important stages of this developmental process occur at the time when preterm children prematurely enter the extra–uterine environment (Figure 1.1.2) and this will be the main focus of this short review.

Nerve and glial cells develop from the ectoderm, the outermost layer of the human embryo\textsuperscript{94}. The first visible sign of central nervous system development is the appearance of the neural groove at around 3 ½ weeks into development\textsuperscript{107}. The groove gradually deepens and then closes to become the neural tube. The inner, hollow portion becomes the ventricular system, while the epithelial cells lining the wall of the tube proliferate and differentiate to provide all the cells in the central nervous system. The caudal part becomes the spinal chord whereas the rostral aspect differentiates into the brain, cerebellum and brain stem\textsuperscript{94}.

The cortical neurons are generated around the ventricular zone through cell division between 6–18 weeks of gestation\textsuperscript{151,182}. Therefore, this process is complete before the first half of gestation ends. The new cells that migrate radially away from the ventricular zone form the pre–plate and later the inside out cortex\textsuperscript{150}. The pre–plate becomes visible after about 8–10 weeks of gestation and will eventually split into the subplate and the cortical plate. Neurons that are generated in the ganglionic eminence form the basal ganglia, some of the thalamic nuclei or migrate tangentially before moving up to the cortex\textsuperscript{107}. IVH is harmful because they disrupt the development and function of the ganglionic eminence. Migration can last up to the 26\textsuperscript{th} week of gestation, during which the pre–plate separates into the cortical plate and the subplate\textsuperscript{107,110}.
Between the 15th and 24th weeks of gestation the subplate serves as a waiting zone where neurons differentiate and make dendrites before moving up to the cortex. For example, the thalamo-cortical fibres start to grow around 10 ½ weeks of gestation. They only reach the cortex around the 24th week; therefore, when they start to grow, the neurons that they will eventually innervate may not have even been generated yet. The subplate is most prominent around the 27th-30th week of gestation. It slowly recedes afterwards, lingering longest below the association areas of the cortex.

Figure 1.1.2

Relevant aspects of embryonic and foetal brain development

The 40-week gestational period is depicted from bottom to top. The GAs of the adolescents in this study is illustrated by the shaded rectangle close to the top. Neural development starts with the appearance of the neural groove and proceeds as an interleaved though ordered set of steps as illustrated by the labelled arrows.
Similarly, the thalamic reticular nucleus, lying between the thalamus and internal capsule, is a transient region during foetal development. All thalamocortical and corticothalamic axons must traverse the reticular nucleus. Furthermore, the perireticular nucleus is a group of neurons residing within the internal capsule. Early fibres from the reticular nucleus and perireticular nucleus reach the thalamus before any other afferents. Also, their efferent fibres reach the cortex before or at the same time as thalamic ones. Additionally, the perireticular nucleus has been shown to guide fibres from the cortex to the thalamus [instead of the brainstem], or direct thalamic fibres up to the cortex\(^{181}\). Both the subplate and the perireticular nucleus have been suggested to play essential roles in the proper formation of the adult connections.

After about 10 weeks of gestation some fibres of the commissural mass start to form the anterior portion of the corpus callosum and successively populating the rest of the structure. The last posterior fibres grow in around the 28\(^{th}\) week.

Myelination starts around the 25\(^{th}\) week of gestation in the globus pallidus, the posterior limb of the internal capsule and the ventrolateral nucleus of the thalamus. However, the progress of myelinating the entire brain WM is slow and most of it takes place postnatally\(^{31,97,156}\).

Finally, synapses are formed in a dynamic manner, both before and after birth\(^{92,107}\).

1.1.5 Effects of Preterm Birth on Vascular Development

Similar to the central nervous system, the vasculature also actively develops during the period when preterm children begin to live ex-utero. This thesis focuses only on the descending aorta hence this review will be limited in scope.

Blood vessel walls have three layers, the middle layer providing the elastic properties. Throughout life, the descending aorta acts as a pressure reservoir in order to maintain a driving force even when the heart is relaxing by expanding and storing blood during systole and then using its elastic properties to propel blood in diastole\(^{165}\).

The foetus cannot use its lungs for breathing and hence blood does not need to flow there. Instead, blood is oxygenated and carbon dioxide is eliminated with the help of maternal circulation via the placenta. Two specialized shunts help ensure that blood bypasses the lungs. Following foetal circulation illustrates how this puts an extra load on the descending aorta.

Oxygenated blood that is carried from the placenta by the umbilical vein is emptied into the inferior vena cava and mixes with the deoxygenated venous blood from the foetal periphery. The heart receives this mix in the right atrium. Because the lungs are collapsed they offer high resistance against blood flow. Thus, instead of entering the right ventricle to be pumped into the pulmonary artery, most of the blood is shunted to the left atrium through the foramen ovale. In turn the left atrium empties into the left ventricle to be pumped to the
rest of the body via the aorta. The small amount of blood that enters the right ventricle will be pumped through the pulmonary artery towards the lungs. But a short vessel between the pulmonary artery and the aorta, known as the ductus arteriosus, shunts this blood also to the aorta. The descending aorta not only supplies the foetal body with this mixed blood but a fraction of it needs to visit the placenta again for re-oxygenation.

In infants that are born preterm this extra load is lifted from the descending aorta after the umbilical cord is cut and the lungs open. In other words, the descending aortas of infants born at term will have performed this double duty for several weeks or even months longer and to the demands of a progressively growing tissue mass.

Therefore, it is possible that the size of the descending aorta will differ between preterm and full term individuals. It has been found for example, that in infants the aortic dimension varied directly with GA\(^2\). Similar effects have been shown in the long-term as well. For instance, children born SGA have an increased risk of high diastolic blood pressure\(^62\); in adults the rate of ischemic heart disease was inversely related to BW\(^14\); and that aortic root diameter was smaller in children who had been born with a lower BW\(^89\). Furthermore and in particular, using ultrasound measurements it was found that adolescent girls who had been born preterm had smaller aortas than age matched controls\(^28\).
1.2 AIMS

The overall aim of the works compiled in this thesis was to use MRI and assess the long-term effects of preterm birth on aspects of the structural development of the brain and descending aorta in children and adolescents.

In particular, the studies included at the end compare children and adolescents who had been either born preterm or at term in order to investigate the effects of preterm birth on:

a) voxel-wise WM microstructure
b) total volume of GM and WM
c) voxel-wise GM volume
d) voxel-wise cortical thickness
e) descending aortic diameter

In addition, an important motivational aspect was the chance to compare these results to other reports focused on similar cohorts that were cared for at other centres.
1.3 METHODS

1.3.1 MAGNETIC RESONANCE IMAGING

The principles of nuclear magnetic resonance (NMR) were first reported in the 1940’s by Felix Bloch and Edward Purcell. They won the Nobel Prize in Physics for their discovery (1955). Purcell was also involved with describing how diffusion affects the NMR signal. In the early 1970’s Paul Lauterbur realized that the source of the NMR signal could be spatially encoded and in essence founded the field now known as magnetic resonance imaging (MRI). The Nobel Committee rewarded his contribution to the medical sciences in 2003.

Nowadays most hospitals site an MRI scanner because it is an indispensible diagnostic tool. It is also widely used in the research community. The reasons for the popularity of MRI and why it was used for the studies in this thesis include the following:

a) MRI is completely non-invasive. Physiological effects are observable when human subjects are placed in a strong (and possibly varying) external magnetic field. However, so far no long-term side effects have been observed or reported.

b) Many contrast mechanisms are available in MRI and they are easily changeable by the operator. In medical imaging this wide array of image contrasts means more accurate diagnosis, while in the research community it offers different windows into understanding the anatomical and functional organization of the human body.

c) MRI offers relatively high spatial resolution. With common clinical scanners it is possible to collect images with 1–3mm isotropic resolution. In the research settings the resolution can be improved, although this may be time consuming.

d) Another advantage is that MRI can be utilized in all areas of the body. Although MRI usually requires the immobility of the sample, the methods have also been developed to image regions of the body that are not under voluntary control, such as the heart, intestines or pulsating blood vessels.

In the following, the different image contrasts that were utilized in the studies and relevant aspects of MRI will be reviewed. The review will be referenced sparsely but the following textbooks were used for the knowledge therein.

1.3.1.1 T1-weighted contrast

Nuclear magnetic resonance is a very descriptive term. Simply put, if magnetic nuclei are placed in a strong external magnetic field and subjected to electro–magnetic radiation with a sweeping frequency, a measurable signal is returned from the sample when the frequency matches [i.e. is in resonance with] the inherent frequency of the nuclear magnetic field variation. The
prerequisite for this effect is that the sample be magnetized. By magnetization, it is meant that the nucleic magnetic vectors, usually denoted by $\mu$, align with the direction of the external magnetic field, $B$, as depicted in Figure 1.3.1.

**Figure 1.3.1**

A simplified illustration of magnetization

Atomic nuclei are depicted as spheres with arrows indicating the direction of their nuclear magnetic dipole vector. In the absence of an external magnetic field (i.e. the left and right white regions) the nuclear vectors point in random directions. The central, darker, rectangle is meant to represent a strong external magnetic field pointing in the upward direction (indicated by the large arrow). When subjected to this external magnetic field the nuclear magnetic vectors align with it.

For clarity and simplicity, Figure 1.3.1 hints that the nuclear magnetic vectors align exactly with the direction of the external magnetic field. A more accurate picture has the vectors subtend an angle with respect to the direction of the main magnetic field and precess about it $^{51}$ (Figure 1.3.2) with a frequency given by the Larmor equation

$$\omega = \gamma B$$  \hspace{1cm} \text{Eq. 1.3.1}

where $\omega$ is the precession frequency, $\gamma$ is the gyromagnetic ratio and $B$ is the external magnetic field. The gyromagnetic ratio is a physical constant and is unique to the nucleus. The gyromagnetic ratio for hydrogen results in a resonant frequency in the radio frequency range for a few Tesla of external magnetic field. Thus the electromagnetic radiation required to energize this nucleus does not heat up the body significantly. For this reason, and due to the abundance of hydrogen in biological tissues, MR images of human beings actually usually only observe signal hydrogen nuclei.

Eq. 1.3.1 states that the precession frequency of a given substance will be linearly varying with the external magnetic field. The value of $B$ is a combination of the main magnetic field of the scanner and several other factors originating from the scanner and the subject to be imaged. To
appreciate the following sections it is important to remember that the value of $B$ is not constant, it varies both in space and time.

**Figure 1.3.2**

A more accurate illustration of the magnetization process

Similar to the middle of Figure 1.3.1, the nuclear magnetic vectors align with the external magnetic field (depicted by the large arrow). However, nuclear vectors do not align exactly. Instead they recess about the direction of the external magnetic field (depicted by the circles at the tip of the nuclear vectors) at a constant subtended angle.

In the volume of a usual MRI image, there are a myriad of nuclei and the signal measured by the MRI machine is their combined effect. Because the local magnetic field varies in space and time, the different nuclear magnetic vectors precess at slightly different frequencies (according to Eq. 1.3.1). In fact, they are randomly distributed along the $360^\circ$. In this arrangement, the $z$-component of each nuclear magnetic vector assumes the same value and points along the direction of the external magnetic field. The $x$-, and $y$-components can take on randomly distributed values. Thus the $z$-components of different vectors add while the $x$-, and $y$-components cancel, resulting in an aggregate magnetization vector, $M$, which points exactly along the external magnetic field (i.e. the sample becomes magnetized). It is this aggregate magnetization vector which, after manipulation by on-resonance electromagnetic radiation, provides the measurable MRI signal.

Magnetization, however, does not occur instantaneously. Instead the magnitude of $M$ grows over time in an exponential fashion (Figure 1.3.3) as more and more of the nuclear magnetic vectors align with the external field.
Here, $M_z(t)$ gives the magnitude of the magnetization of the material with respect to time assuming the initial magnetization of zero when the experiment starts (i.e. when $t = 0$) and $M_{eq}$ is the equilibrium value (i.e. the final amount of magnetization that the object will obtain after an infinite amount of time). The time constant, conventionally known as T1 relaxation time, is specific to the sample and represents the point in time when 63% of the maximum magnetization has been reached (i.e. when $t = T1 \Rightarrow 1 - e^{-t/T1} = 0.63$).

**Figure 1.3.3**

Illustration of the T1–relaxation process

Similar to Figure 1.3.1 and Figure 1.3.2, the four rectangles on the top are meant to depict a sample. In this case two different substances are mixed together. Suppose the sample is placed in a strong external magnetic field, $B$. Initially, the nuclear magnetic vectors are randomly oriented (Part A). As time goes on, the nuclear magnetic vectors align with the external magnetic field (depicted by the large upward arrow). Note that in Part C all the red nuclei have aligned while many of the blue ones have not yet.

The bottom of the figure illustrates the evolution of the signal amplitude according to Eq. 1.3.2.
This effect can be used as a contrast mechanism in MRI. For clarity, Figure 1.3.3 depicts the situation when only GM and WM are contained in a voxel. Both fat and water contain hydrogen nuclei and WM contains more fat than does GM due to the compact double cell bi-layer in the myelin sheaths wrapping the axons. Initially (Part A) the nuclear magnetic vectors are oriented randomly since the external magnetic field has just been set up (large grey arrow in the background). As time goes on, the hydrogen nuclei align with the external magnetic field. This process of aligning is faster in fat than in water (Parts A–D). The nuclei that align combine to form a magnetization vector pointing along the main magnetic field. The T1–relaxation time for WM is about 600ms. Therefore, after 2 seconds almost 100% of the nuclear magnetic vectors in WM have aligned with the external magnetic field (Part D). In GM (which contains less fat and more water) there is still some portion of hydrogen nuclei that have not aligned.

If the composite signal is measured at a strategic time (500–900ms) from a sample containing both tissue types, WM will contribute more to the signal amplitude than GM (red and blue arrows respectively at the bottom of Figure 1.3.3). Although it is not explicitly shown in the previous figure, cerebrospinal fluid (CSF) has the highest concentration of water, on average the nuclei align even slower than GM and hence CSF will appear the darkest in a T1–weighted image (Figure 1.3.4).

Note! $T_{1WM} < T_{1GM} < T_{1CSF}$, which are on the order of 0.5–5.0 seconds.

**Figure 1.3.4**

![Examples of T1–weighted images](image)

Two images are displayed from approximately the same axial region in the same subject. Note how the contrast can be varied. For example, the putamen is hardly visible on the left while the contrast between cortical gray matter and WM is reduced on the right.

Three-dimensional T1–weighted images depict anatomy with high resolution and can be used for ascertaining healthy brain structure or
normalization of an individual’s MRI data to a standard space (see section on Normalization on p23).

T1–weighted images were used in all studies except for investigating cross sectional aortic area in Paper IV.

1.3.1.2 T2–weighted contrast

When $\mathbf{M}$ points along $\mathbf{B}$ the individual nuclear magnetic vectors, $\mu$, precess about $\mathbf{B}$ but $\mathbf{M}$ is stationary. For reasons that are beyond the scope of this thesis, the signal from the magnetization vector can only be measured if it is turned away from the direction of the main magnetic field and closer to the x–y plane. In this case, however, $\mathbf{M}$ also precesses about $\mathbf{B}$. Because $\mathbf{B}$ is not uniform in space and time, the individual $\mu$ will precess at different frequencies according to the Larmor equation (Eq. 1.3.1). As they slowly lose coherence and start pointing in different directions in the x–y plane their summed effect will result in a progressive reduction in the length of the magnetization vector, which in this case is conventionally called $M_{xy}$ to distinguish it from $M_z$. The time variation of the magnitude of $M_{xy}$ is an exponential decay

$$M_{xy}(t) = M_0 e^{-t/T_2}$$

Eq.1.3.3

where $M_0$ is the magnetization turned into the x–y plane and the time constant, conventionally known as T2 relaxation time, represents the point in time when $M_{xy}$ has been reduced to 37% of its initial value (i.e. when $t = T_2 \Rightarrow e^{-1} = 0.37$). Similar to T1, T2 is also specific to the sample and can be used to achieve contrast in an MR image. As in Figure 1.3.3, for simplicity Figure 1.3.5 also displays only two spin systems, representing grey (blue) and WM (red). Soon after the nuclear magnetic vectors, and hence their composite effect $M_{xy}$, were turned toward the xy–plane all vectors precess in unison (Figure 1.3.5A) about the main magnetic field (grey upward arrow). Because the external magnetic field that each nucleus experiences is slightly variable, nuclei will precess at a slightly different frequency according to Eq. 1.3.1. As time goes on, this results in the dephasing of the precession – i.e. although the nuclear magnetic vectors will still precess about the external magnetic field, they will start pointing in increasingly different directions. This phenomenon is known as T2–relaxation and occurs faster for WM than for GM (Figure 1.3.5 B). The time constant for this second type of relaxation is shorter than that for T1–relaxation and is on the order of about 80ms for WM. Therefore, after about 250ms almost the entire signal has been lost from this tissue type. After a sufficiently long time the signal is lost even from GM, which has a T2 of about 100ms. Again, if a composite signal is measured from a sample containing both GM and WM at a strategic time of about 70–110ms good contrast can be seen between these two tissue types (depicted by the red and blue arrows in the bottom of Figure 1.3.5). Because CSF has the longest T2–relaxation time it will appear brightest in T2–weighted images (Figure 1.3.6).
Illustration of the T2−relaxation process

The three rectangles on the top are meant to depict a sample where two different substances are mixed together. Suppose the sample is placed in a strong external magnetic field and enough time is allowed for the nuclear magnetic vectors to align with the external magnetic field (depicted by the large arrows). If the so created magnetization \( M_z \) is turned in the x–y plane this magnetization will be composed of the precessing nuclear magnetic vectors and conventionally known as \( M_{xy} \). Initially, the nuclear magnetic vectors precess together (Part A). But due to slight differences in the local magnetic field strength, \( B \), they slowly dephase, according to Eq. 1.3.1, resulting in the reduction in the magnitude of \( M_{xy} \). Note that in Part B the red nuclei point in highly varied directions whereas the blue ones still point more or less in the same direction. After a few hundred milliseconds, both the red and blue nuclear magnetic vectors have completely dephased with the individual vectors pointing in random directions (Part C).

The bottom of the figure illustrates the evolution of the signal amplitude according to Eq.1.3.3.
Note! T2_{WM} < T2_{GM} < T2_{CSF}, which are on the order of 0.05–0.2 seconds.

Figure 1.3.6

Example of a T2–weighted image
A cross sectional image of a volunteer is displayed. Note the inverted contrast where WM appears darkest.

T2–weighted images are usually collected slice–wise and used to investigate histological alterations, such as oedema, glioma or gliosis. In addition, most often T2–weighted images are used for diffusion–weighted imaging (to be discussed next).

T2–weighted images were used for radiological examination in Paper I and for diffusion–weighted images in Papers II and V.

1.3.1.3 Diffusion–weighted contrast and diffusion tensor imaging

Diffusion is the random meandering of particles. Usually, this is thought of as one substance diffusing into another – like sugar molecules in water. But as mentioned above (p10), the MRI signal only observes hydrogen nuclei – so in an MR experiment we speak of self–diffusion of water within itself.

Considering the 2D case, diffusion occurs away from places of high concentration according to Fick’s law

\[ J_x = -D \Delta C_x \]  

Eq. 1.3.4

where \( \Delta C_x \) represents the change in concentration along the x–axis of the substance that is diffusing, D the diffusion constant of the substance in the particular medium in which the diffusion occurs and \( J_x \) the flux through an imaginary surface perpendicular to the x–axis. In the example of sugar diffusing in water, \( \Delta C_x \) could represent the difference in the concentration at the bottom and top of a cup of water. D would denote the diffusion constant of sugar in
water, which is a *scalar* and depends on the temperature and the medium. The flux, $J_x$ would give the amount of sugar that crosses a horizontal surface between the bottom and top of the cup. Clearly, raising the temperature would result in more sugar crossing this surface while changing the medium to gel would lower the diffusion constant and less sugar would cross the surface.

Of course, interesting problems occur in more complex media than pure water and in 3D. For example, diffusion in human tissue would involve diffusion across the hydrophobic lipid bi–layer of cells, which is more difficult than free diffusion in water. Because cells come in many shapes and sizes, and because they are organized in different structural arrangements, the diffusion process will occur along contorted paths and different directions will have their own unique diffusion constants. The mathematical description of such a diffusion process in 3D becomes much more complicated. Full characterization of the flux in a specific direction also includes the concentration differences in the orthogonal directions as well as the respective cross term diffusion constants

$$J_x = -\left[D_{xx}\Delta C_x + D_{xy}\Delta C_y + D_{xz}\Delta C_z\right]$$

$$J_y = -\left[D_{yx}\Delta C_x + D_{yy}\Delta C_y + D_{yz}\Delta C_z\right]$$

$$J_z = -\left[D_{zx}\Delta C_x + D_{zy}\Delta C_y + D_{zz}\Delta C_z\right]$$

Eq. 1.3.5

For example, $J_x$ depends not only on $\Delta C_x$ but also on $\Delta C_y$ and $\Delta C_z$. Also note the subscripts on the diffusion constants, which indicate the relationship between the direction of flux and the concentration difference. Using the rules of matrix multiplication, Eq. 1.3.5 can be written in a simpler and more representative form

$$\begin{bmatrix} J_x \\ J_y \\ J_z \end{bmatrix} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \begin{bmatrix} \Delta C_x \\ \Delta C_y \\ \Delta C_z \end{bmatrix}$$

Eq. 1.3.6

Now, square brackets around the components of the flux (i.e. $J_x$, $J_y$ and $J_z$) on the left represent the 3D flux vector and similarly for the 3D concentration vector on the right. The 3x3 matrix on the right hand side is known as the **diffusion tensor**.

Since there are 9 unknowns, the determination of the diffusion constant requires that the diffusion constant be measured in at least 9 unique directions (i.e. no pair can be co–linear and no three can be co–planar).

When one speaks of DTI, the meaning is the use of MR images for the voxel–wise determination of the 9 unknowns in the diffusion tensor. Eq. 1.3.6 is valid for the description of all particles, even charged ones in an external electric field. But because hydrogen, the element of choice in MRI, is not a charged particle the diffusion constant in any direction is identical to that for the exact opposite direction. In this case $D_{xy} = D_{yx}$, $D_{xz} = D_{zx}$ and $D_{yx} = D_{zy}$, making the diffusion tensor a symmetric matrix with only 6 unique unknowns.
In MRI a reference image is needed for the determination of each of the 6 diffusion constants. The same reference image can be used, making the simplest DTI study consist of 7 images (1 reference image and 6 diffusion-weighted images with a unique diffusion direction for each).

In the absence of noise these 7 images would determine the diffusion tensor in Eq. 1.3.7 exactly, like two points define a unique line. However, MR images are noisy measurements and for this reason, usually many more (i.e. 12 – 60) measurements are made, including several reference images. To fit the diffusion tensor model to this noisy data, usually least square fitting is employed\(^7\).

Because the frequency of the nuclear precession depends on the external magnetic field (see Eq. 1.3.1) and because the local magnetic field is not perfectly uniform in space and time, the diffusive meandering of nuclei in a sample results in the loss of coherence in their precession. This dephasing results in the reduction of the MRI signal amplitude that is returned from the sample. This artefact was noticed soon after the discovery of the NMR phenomenon\(^3\). Interestingly, artefacts can often be exploited for a new type of image contrast. Such is the case with diffusion. As it turns out, the extent of signal loss that results from diffusion within the sample is directly related to the diffusion constant of the sample.

To illustrate an MRI experiment for the measurement of the diffusion constant, consider three hydrogen nuclei initially located at the centre of a voxel (Figure 1.3.7 A). They are all experiencing the same, say 1.5 Tesla, external magnetic field and therefore precess at the same frequency according to Eq. 1.3.1 on p10. The signal amplitude of each hydrogen atom and the sum of their signal is depicted at the bottom of Figure 1.3.7.

Next, a positive gradient, say 40 mTesla/meter, is introduced for a short time such that the magnetic field linearly increases as we move to the right (Figure 1.3.7 B). The three hydrogen nuclei will experience a slightly higher external magnetic field but the same field nonetheless. The higher field induces an increase in their precession frequency (Eq. 1.3.1). But that frequency is still the same for each nucleus, given that they still experience the same, though higher, external magnetic field.

After the gradient is turned off, a certain amount of time is allowed for diffusion to happen. Suppose, during this time one of the hydrogen nuclei moves to the right. Because no magnetic field variation is introduced, the three nuclei still precess at the same frequency, induced by the 1.5 Tesla main magnetic field (Figure 1.3.7 C).

Subsequently another gradient is introduced which is negative in polarity (i.e. \(-40\) mTesla/meter) as we move to the right (Figure 1.3.7 D). In this set-up the single hydrogen nucleus that moved to the right will precess at a different, lower frequency from the other two.

\[
\begin{bmatrix}
D_{xx} & D_{xy} & D_{xz} \\
D_{xy} & D_{yy} & D_{yz} \\
D_{xz} & D_{yz} & D_{zz}
\end{bmatrix}
\]

Eq. 1.3.7
Figure 1.3.7

Illustration of diffusion–weighting using magnetic gradients
The squares at the top are meant to represent a sample placed in an external magnetic field, $B$ (not depicted but assumed to point toward the top). The three nuclei precess about $B$ according to Eq. 1.3.1. This dependence on the external magnetic field is true even when the magnetic field is slightly varied across space (Parts B and D). The bottom illustrates the signal amplitudes and phases of the individual nuclei and their summed effect. For further details please see the text.

When the negative gradient is turned off all three hydrogen nuclei will precess at the same frequency, corresponding to the 1.5 Tesla external field, but only two of the nuclei will precess together. The third one will be out of phase (Figure 1.3.7 E). The diffusion encoding has now finished. If at this point the composite signal is measured from the three hydrogen nuclei the amplitude will be smaller than before the diffusion encoding (compare the amplitude of the composite signal in parts A and E of Figure 1.3.7). Note that the signal from Hydrogen 1 and 3 are of the same frequency and still in phase, as if the two diffusion–encoding gradients had not happened.

The scenario presented in Figure 1.3.7 was highly simplified. If the diffusion process is truly random and involving a large number of particles the loss in the MR signal will be exponential

$$S_b = S_0 e^{-bD}$$

Eq. 1.3.8

where $S_0$ is the signal in a voxel without diffusion weighting [i.e. the reference image], $D$ is the diffusion constant and $b$ is a variable describing the strength of diffusion weighting. The value of $b$ depends on the strength of the diffusion
encoding magnetic gradients, how long they are on and the time allowed for
diffusion between the positive and negative gradients. For example, stronger
gradients will impart a larger phase difference per unit distance, while allowing
more time for diffusion allows for particles to diffuse farther – i.e. both of these
factors will increase the b–value.

**Figure 1.3.B**

![Illustration of collecting DTI data](image)

The top row displays four images. The leftmost one is a normal T2–
weighted imaged without diffusion encoding. In the other three, diffusion is
measured in the three orthogonal directions. The yellow arrows indicate
the voxels from which the signal is displayed at the bottom. The three
rightmost images were collected with a b–value of 1000 s/mm².
Because water diffusion occurs with equal ease in all direction within the
ventricles (i.e. isotropic), the signal drops equally, irrespective of the
direction of diffusion encoding (black curve). On the other hand, in the
posterior corpus callosum the signal amplitude depends on the relative
difference in direction between the axons and the diffusion encoding.

As the bottom Figure 1.3.B shows, the larger the diffusion weighting b–
value, the more extensive the signal loss. The aim is to choose a b–value that
will impart a large enough signal drop for any direction, while still providing
contrast between the different diffusion encoding directions. For a given b–
value the signal is reduced in accordance with the diffusion constant. In the
ventricles, where D is high, the signal loss is very pronounced regardless of
what direction diffusion is measured in. In tissue, D is lower and directionally
variable. For example, in the posterior corpus callosum signal loss will be
pronounced along the left–right [red curve] but not so along the anterior–posterior [blue curve] or superior–inferior [green curve] directions. This variability is attributed to the fact that between the WM fibres water diffuses with ease whereas diffusion across the double bi–layer of the axon wall and the surrounding myelin is restricted.

After compiling the DTI dataset with at least one reference image and at least 6 diffusion–weighted images the diffusion tensor can be fit to the data. Displaying the tensor requires at least 6 images and is therefore not practical. Instead, the information in the tensor is usually reduced to one of several scalar values. The one used in this thesis was FA, which takes the value of zero for spherical tensors and approaches the limiting value of 1 for elongated [rugby ball shaped] tensors. Tensors describing diffusion between myelinated axons become more elongated in shape as the fibres become more tightly packed, more myelinated and more parallel in organization (Figure 1.3.9 A). For this reason FA has been used extensively to scrutinize the WM microstructure.

Figure 1.3.9

Examples of FA and colour–coded images
Both parts display data from the same axial slice of a volunteer. Part A depicts an FA map where the higher the value of FA the brighter the image. Although, this quantity highlights the major WM well it cannot discriminate among fibres oriented in different directions. For this reason the tensor direction is usually colour–coded as shown in Part B. The colours red, green and blue indicate fibre directions in the left–right, superior–inferior and anterior–posterior directions respectively.

Another approach to reduce the information in the tensor to a form that is suitably displayed in a 2D image is colour–coding the direction of the long axis of the tensor. In Figure 1.3.9 B green means that the long axis of the tensor, and hence the direction along which water diffuses with the greatest
ease in a voxel, is superior inferior. Similarly, a voxel is coloured blue if a tensor points along the left–right direction, and red if it points along the anterior–posterior direction.

Diffusion–weighted images and DTI were used in Papers II and V.

1.3.1.4 FIESTA for cardiac and vascular imaging

In Paper IV a balanced steady state free precession sequence\textsuperscript{23,159,160} was used to measure the cross sectional area of the descending aorta. The details of this acquisition method are well beyond the technical scope of this thesis. Different vendors assigned different product names to their individual implementation of this sequence. On GE scanners it is called FIESTA. It offers a contrast, which depends on the flip angle, but for any given setting further contrast can be achieved depending on the ratio of T1/T2–relaxation times. In 2D imaging blood flowing into the slice also contributes, highlighting further the lumen of blood vessels, which have a different T1/T2 ratio than the surrounding muscle and vessel walls (See Figure 1.3.15 on p30).

FIESTA (i.e. balanced steady–state free precession) images were used in Paper IV.

1.3.2 IMAGE PROCESSING

The main interest in medical imaging is contrast. One wants to see clear contrast between GM and WM, for example, when imaging anatomy. But differences between GM and WM may be of no relevance when one would like to identify areas, which are affected by oedema or a malignant tumour and separate them from the surrounding healthy tissue. While the different contrasts reviewed above developed from and serve this need, further processing is possible and often necessary after data collection is complete.

1.3.2.1 Segmentation

Images are often segmented into different tissue types. This is useful, for example, when the aim of the study is to calculate brain volume. The image contains signal from the skull, eyes, neck, etc. which are not of interest. Many segmentation methods have been developed and implemented\textsuperscript{12,37,145,170,188,189}, some specifically used for segmenting MR images of the developing brain\textsuperscript{75,148,195}. They often rely either on a tissue classification method or on registration to a template. They, however, all perform best when the contrast to noise [CNR] ratio is high. That is, one needs to maximize the value of

\[
\text{CNR} = \frac{S_a - S_b}{\sigma}
\]

Eq. 1.3.9
where $S_a$ and $S_b$ are the signal levels of the two areas that need to be separated visually and $\sigma$ is the standard deviation of the noise in the image. Clearly, increasing the difference in the signal level, decreasing the noise or a combination of both these strategies can lead to an increase in CNR. According to the Rose criterion \cite{rose1971} a CNR of at least 3–5 is required for object recognition in a noisy background. In MR images where one can assume additive Gaussian noise a CNR level of 4 is usually taken as a good value. In that case the signal level of the two tissue classes can be taken as two Gaussian curves centred at $S_a$ and $S_b$ and having the common standard deviation of $\sigma$, which represents the noise level in the image. If the means of the two curves are $4\sigma$ apart, the two Gaussians will overlap only about 2.5% of their total area. For a human observer only this CNR level would be unambiguous.

**Figure 1.3.10**

**Brain tissue segmentation based on MRI images**

The top left displays a 3D T1-weighted MRI image of the brain. Using the algorithm of Ashburner and Friston \cite{ashburner2005} this image was segmented into GM (top right), CSF (bottom left) and WM (bottom right). All parts display approximately the same axial, sagittal and coronal cuts of the image of a volunteer.

**1.3.2.2 Normalization**

It is imperative that homologous regions are considered when the aim is to perform statistical analyses, involving many individuals, on the voxel-wise value of the MRI signal or derivatives of this signal, such as, for example, FA values in DTI. One approach to this would be to perform region of interest (ROI)
analyses where anatomical knowledge is employed to identify homologous regions from which the voxel values are extracted and subsequently put through some form of a statistical analysis. This method is labour intensive and can only be used in a limited set of regions and when specific a priori hypotheses are available. None of the papers included in this thesis used this method.

Instead, a process called normalization was employed in which the images of different individuals are morphed to a common template. After normalization it is assumed that a given voxel contains tissue from identical anatomical regions from different subjects. Of course this assumption hinges on two important factors: the quality of the normalization and the individual variability in functional brain anatomy.

With respect to the first factor, important developments have come out during the time this thesis was compiled. For example, in Paper V, which chronologically was the first to be completed, all images were registered to a common template derived from the template provided by the Montreal Neurological Institute (i.e. MNI in the left column of Figure 1.3.11) using the methods of Ashburner et al. In contrast, the results in Papers II and III rely on more advanced registration methods. The GM segments were normalized using the DARTEL algorithm, where the template was specific to the study group (middle column of Figure 1.3.11). An additional improvement in Paper II compared to Paper V was that the images were modulated by the Jacobian determinant. This corrected for local volume change of the tissue that results from moving from one coordinate system to another (i.e. warping one image to the coordinates of another).

As used in Paper II, the mean skeletonised FA image (green) was created using the tract–based spatial statistics (TBSS) pipeline (right column of Figure 1.3.11). Clearly, using DARTEL or TBSS, the spatial registration of numerous images can be considered superior to the older method of using the much smoother MNI template. Please see Section 1.3.3.4 (p34) for a more detailed description of the TBSS pipeline.

Eliminating the errors introduced by the second factor, individual variability in functional brain anatomy, is a more difficult task and it may not ever be accomplished by improving normalization methods whose only input is pixel intensity in anatomical MRI images. To circumvent type II errors (i.e. false negatives), when voxel–wise comparisons are made between two groups of people, the images are usually smoothed. The amount of smoothing to apply is, however, an open question and it has been specifically criticised in relation to the voxel–wise analysis of DTI data. At least it is clear that with better registration less smoothing is needed, because one only needs to smooth the images in order to average functional localization of homologous areas among individuals.

Finally, note also that the error arising from the individual variability in functional brain anatomy is not only a problem for normalization. Attempting to define homologous brain areas based on prior anatomical knowledge (i.e. region of interest (ROI) analyses) suffer from the same problem.
Example templates for anatomical normalization
Three different templates are displayed at approximately the same anatomical location. The MNI template on the left is the average of a large number of individuals’ brain images after affine registration. It is often used as a target in studies involving a specific study group in question. The middle and right columns display the GM and FA templates that were created from images of the participants in this study using the DARTEL\(^9\) and TBSS\(^{171}\) algorithms respectively.

1.3.2.3 Estimating cortical thickness from MR images
Several methods have been developed for measuring cortical thickness. They can be grouped into two variants, which can be combined or used in isolation. As the name implies, the surface–based method relies on the generation of two 3D surface models: one between the pia mater/GM and another between the GM/WM boundaries. Subsequently, some measure of distance between these two surfaces is used to estimate the thickness\(^{56}\). For group studies the surfaces need to be registered and properly averaged\(^{44,57,58}\).

Another way to estimate cortical thickness is to only use the GM/WM surface and then the thickness value is determined volumetrically\(^{80}\). To compare the thickness among subjects, the latter approach does not require a 3D surface model. Instead the GM/WM boundary is estimated from voxel information.
We used the latter method with the following pipeline: Using the unified segmentation procedure the images were partitioned into GM, WM and CSF (Please see Section 1.3.2.1 on p22 for details on segmentation). A voxel-based cortical thickness map was created for each subject based on the 3 segments from above according to Hutton et al. This method automatically extracts the cortical GM boundaries from T1-weighted images and estimates a value for the cortical thickness at each voxel in the GM segment image. Next, using DARTEL, cortical thickness maps were warped into a common reference space representing an average of all the subjects (Please see Section 1.3.2.2 on p23 for details on normalization). The warped cortical thickness maps were re-sampled to an isotropic voxel size of 1.5 mm using trilinear interpolation, rescaled by the Jacobian determinant of the deformations (to account for stretching and compression) and smoothed with a 6mm Gaussian kernel.

1.3.2.4 Distortion correction in echo-planar imaging data

The data collection of MRI signal for DTI is usually accomplished by echo planar imaging (EPI) methods. Because a 2D EPI image can be collected in 100 milliseconds or so it helps to reduce artefacts, which otherwise occur due to subject movement or pulsatile motion of the tissue. This convenience comes at a price: EPI images suffer from susceptibility distortions, which occur because different materials magnetize to different extents when placed in a magnetic field. In other words, the equilibrium magnetization \( M_{eq} \) in Eq. 1.3.2 is different for different materials. For example, air does not magnetize at all (i.e. it is not susceptible to the external magnetic field) while water-based tissue magnetizes well. In between air and tissue, such as around the ear canals and the facial sinuses, the extent of magnetization is spatially variable and unknown. This local magnetization vector adds to that used for spatially encoding images. The details are outside the scope of this thesis, but the result is that EPI images of the brain are distorted by this unknown local magnetic field. However, the local magnetic field vectors are constant in time. Therefore, by collecting at least two datasets where the magnetic gradients are varied, the effect of the invariant tissue magnetization can be calculated and removed. Figure 1.3.12 illustrates this. The top row displays images when the phase encoding direction was left to right. The left column is an axial cut of the brain stem while the right column is a coronal cut of around the temporal lobes. Note how the temporal lobes are distorted and the brain stem is shifted. In the middle row of Figure 1.3.12 the images were collected with the phase encoding direction being right to left resulting in distortions and shift in the opposite direction. These two datasets can be combined to estimate the local magnetic fields and correct the images (Bottom row of Figure 1.3.12).

In addition to susceptibility distortions, the fast switching of the magnetic gradients for image acquisition and diffusion encoding introduces eddy currents in the structure of the MRI machine. These eddy currents, like any electric current, will result in some unknown magnetic fields generated in the field of view. The result of these spurious magnetic fields is a distortion in the acquired images.
Illustration of susceptibility induced artefacts and their correction

The left and right columns display an axial image of the brain stem or a coronal image of the brain respectively. The top row illustrates the scenario when the MRI data is collected while using magnetic gradients in one direction. Due to the local, unknown magnetic gradients that are present in the tissue the images are distorted (Top row). If the image gradient polarity is reversed the distortions also reverse (Middle row). Collection of both datasets allows for the recovery of the true image using to the method of Andersson et al.

By definition, in a DTI dataset the magnetic field gradients are different in each image because they are used to encode diffusion in different directions. Hence, eddy current distortions will vary from image to image and making tensor calculations voxel-wise problematic in the affected regions. To reduce this problem we used the double refocused diffusion encoding sequence for acquisition and image processing methods afterwards. To illustrate the benefits of these precautions, Figure 1.3.13 displays the mean FA image from 132 subjects.
Mean FA image of the cohort
Axial (top), coronal (middle) and sagittal (bottom) cuts are displayed from the mean FA image of the cohort after movement correction and removal of eddy current and susceptibility–induced artefacts. The images were normalized using the DARTEL algorithm.

1.3.2.5 Edge detection for estimating aortic cross sectional area

The detection of the aortic wall from surrounding tissue and blood was done by a semi–automated active contour procedure adopted to detect image edges. This method had been used previously for aortic segmentation. Manual input for the active contour was only needed for the process to be initialized for the first one of the 25 images. Subsequently, the AC from image one is used as a starting estimate for image two, and so forth. The observer who provided the manual initialization was blind to group adherence.

To test the reliability of the active contour segmentation method two separate analyses were performed. The robustness of the method was tested by imaging an adult volunteer five times. The same MRI sequence was used to collect five slices at different levels of the descending aorta. Each set of images was processed identically. The results were highly reproducible (Table 1.3.1).

Table 1.3.1
Robustness of aortic area measurements

<table>
<thead>
<tr>
<th></th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
<th>Level 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \text{mm}^2 )</td>
<td>188.8</td>
<td>153.8</td>
<td>237.6</td>
<td>214.7</td>
<td>246.0</td>
</tr>
<tr>
<td>Standard Deviation ( \text{mm}^2 )</td>
<td>1.9</td>
<td>0.9</td>
<td>5.4</td>
<td>3.8</td>
<td>4.1</td>
</tr>
<tr>
<td>Coefficient of Variation</td>
<td>1.0%</td>
<td>0.6%</td>
<td>2.3%</td>
<td>1.8%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>
The accuracy of the results was tested by comparing the results of the semi-automated method with that of a completely manual segmentation in a subset of images from the entire cohort. Making comparisons in 90 randomly selected images from all 5 levels we found that on the average the manual segmentation resulted in a 1–6% smaller aortic area depending on the level (Table 1.3.2). For acquisition details see Section MRI Data Acquisition on p37.

<table>
<thead>
<tr>
<th>Level</th>
<th>Mean % Difference (mm²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-5.9%</td>
</tr>
<tr>
<td>2</td>
<td>-3.3%</td>
</tr>
<tr>
<td>3</td>
<td>-1.3%</td>
</tr>
<tr>
<td>4</td>
<td>-1.1%</td>
</tr>
<tr>
<td>5</td>
<td>-2.3%</td>
</tr>
</tbody>
</table>

Although the difference between the manual and semi-automated methods was relatively small, we further investigated whether a systematic bias was present in such a way that larger aortas would be overestimated or small aortas underestimated leading to a false positive result when comparing the groups. This was not the case. Figure 1.3.14 plots the difference between the two aortic area estimates against their mean. More often the manual segmentation resulted in a smaller estimate for the aortic cross section but larger aortas did not tend to be overestimated more than smaller ones.

Testing the accuracy of the aortic segmentation
As Table 1.3.2 indicates the manual segmentation resulted in slightly smaller aortic diameter estimates. This figure plots mean value of manual and automatic segmentation methods against the difference between the two. The automatic segmentation method is more likely to produce a larger estimate than the manual tracing for the smaller aortas.
Upon publication, the figure, originally prepared for Paper IV, needed to be simplified somewhat. Here we insert the original figure because it better illustrates the data acquisition and the edge detection.

Figure 1.3.15

Collection and segmentation of the abdominal aorta images
Part A displays the reference images and the positions of the slices from which data was collected using the FIESTA\textsuperscript{a,b,c,d} sequence. The leftmost image illustrates the use of the iliac bifurcation as a reference for slice positioning. Part B displays four illustrative slices at the level of the aorta from Part A. The white arrows indicate the aorta. Part C illustrates the aortic segmentation for the three rightmost images in Part B. The yellow outline is the result of the semi-automated detection of the aortic lumen.
In the 20th century, scientific scrutiny of observable phenomena has become the playground of statistics. This is especially true for medical sciences. However, statistical theory is complex and therefore comparison of results from different studies requires an understanding of some aspects.

1.3.3.1 Simple random sampling vs. cohort studies

It is commonplace that results of scientific experiments are only accepted as valid if the results obtained could be erroneously found less than 5% of the time in a situation when no effect is expected. In order to produce such percentages, or probability values (p values), elementary statistical theory pre-supposes the process of simple random sampling. In this sampling procedure a population of interest is identified and then a sample is drawn in such a way that each and every member of the population has an equal chance of being selected — effectively that any possible composition of the sample is equally possible. If these criteria are met the usual 5% cut off value has a clear meaning: if the experiment was to be repeated a large number of times, using the same population but a different sample, then only 5% of the results would show such a large effect size.

By definition, cohort studies violate the idea of simple random sampling. Members are not selected by chance but included by default. As a result, the interpretation of p values and generalizing the results become difficult.

1.3.3.2 Statistical significance

In scientific articles usually only statistically significant results are reported. That is, only if the probability of finding such a result is less than 5%, as described in the previous section. It is important to note that not only is the threshold for statistical significance decided arbitrarily, it also depends on several factors, including, but perhaps not limited to:

a) The effect size
b) The variability of the data (i.e. standard deviation)
c) The number of data points
d) The number of multiple comparisons
e) The method of correcting for multiple comparisons (if correction is needed)

Suppose we aim to compare two groups, based on the mean FA value in a voxel located in the posterior corpus callosum. The effect size is simply the difference between the mean values. The larger the effect the easier it is to show statistical significance for a given constant amount of variability in the data. For example, if the groups have 10 members each and average FA values of 0.3 and 0.7 respectively, statistical significance will hardly be missed regardless of the variance [left panel of Figure 1.3.16]. If the variance is smaller, it is less likely to have such a large effect size by chance and hence the p-value will be smaller [center panel of Figure 1.3.16]. Because the number of observations [i.e. the number of participants] factors in the estimation of the variability of the data, larger studies, like the ones in this thesis, will end up with smaller estimated standard errors and in turn be able to identify smaller effect...
sizes or arrive at smaller p-values than smaller studies would for the same effect size. This is visually indicated in the right panel of Figure 1.3.16.

Figure 1.3.16

Determination of statistical significance

Statistically significance may not be equivalent to biggest effect. For example, the calculation of the p value that is usually reported includes the variance in the data (left and middle panels) and the number of subjects involved (right panel). Please see text for further explanations.

The more statistical tests we perform the more likely it is that we hit a (possibly false) positive result. This is illustrated in Figure 1.3.17, where 10 random samples were selected from the same population of FA values with mean 0.5 and standard deviation of 0.2. Note that the means of the samples are hovering around the expected value of 0.5 but each sample is different. By chance sample #2 had the highest mean of 0.62 and sample #4 had the lowest mean of 0.41. Even though these samples were drawn from the same population and using the simple random sampling procedure there is a statistically significant difference (p = 0.0236) between the means of sample #2 and #4. If we kept on comparing each of the 10 samples with all the other samples we would have to make 45 different comparisons but sooner or later we would find this statistically significant, albeit false positive, result. Clearly, somehow we must reduce the chance for finding positive results simply because we are relentless in looking for it. The usual procedure is the Bonferoni approach where the threshold accepted as statistically significant result (i.e. usually 5%) is reduced by the number of comparisons we make. In the above example an effect would only be accepted as statistically significant if the chance of finding this result was 5% divided by 45 (i.e. about 0.1%). A detailed discussion of multiple comparison procedures is beyond our interest. Note only that if many comparisons are made, results that would have been significant at the 5% level are now deemed insignificant with the new level. Thus, making a large number of comparisons can reduce sensitivity and the procedure for correcting for the effect of multiple comparisons can decide what results are declared statistically significant. See Section 1.3.3.4 on p 34.

The previous three paragraphs were only included here to illustrate the difficulty in comparing the results among reports if not all factors from the above list are reported. Please see Discussion for further elaboration on this issue.
The need for multiple comparison correction

To simulate an experiment where FA is collected from 10 individuals, ten random samples were generated from a normal distribution with mean 0.5 and standard deviation of 0.2. Even though the samples are drawn from the same population and are assumed to be random, the mean of the samples varies greatly. In fact, blind but systematic comparison of the 10 groups would identify a statistical significant difference between the means of groups #2 and #4.

1.3.3.3 Multiple comparison in MR imaging studies

As already mentioned in section 1.3.2.2 on p23, often in MR neuroimaging studies the images of different subjects are normalized to a common template and then, under the assumption that voxels contain homologous tissue, statistical tests are performed voxel-wise. The number of voxels in MR images commonly range from tens to hundreds of thousands and therefore, as the previous section points out, finding statistically significant results is not a problem.

The Bonferoni approach\textsuperscript{165} to correct for multiple comparisons is very conservative – it often results in false negatives because it simply makes the threshold of statistical significance more stringent as the number of comparisons increases. To illustrate, if an image contains 50,000 voxels with brain tissue in them, a difference would only be accepted as statistically significant if there was a 5% / 50,000 ≈ 0.0001% chance of finding this effect by chance. In practice it means that the difference between the mean values in the voxel or the number of subjects included have to be extremely large. But for example, by definition FA has values only between 0 and 1 and a difference of 0.2 may be biologically/clinically significant. Only large studies would be able to identify this effect if the Bonferoni approach was used for multiple comparison correction.
Instead, in MR neuroimaging studies random field theory is used to reduce the chance of false positives due to the large number of voxels (i.e. large number of comparisons). A good explanation of random field theory can be found at Cambridge Cognitive and Brain Sciences Wiki page [link](http://imaging.mrc-cbu.cam.ac.uk/imaging/PrinciplesRandomFields) by Matthew Brett et al.

In Paper II the GM analysis relied on family-wise error correction, where setting a 5% threshold means that by repeating the experiment many times a false positive voxel would occur only once out of 20 experiments. In the same paper the analysis of FA images used randomization approaches to identify a proper 5% threshold, which is similar to the bootstrap procedure. In this case, the available data (i.e. all images from both groups) are permuted and the statistical tests are performed on each permutation. This way the distribution of all possible results is generated and statistical significance is set as the 5% most extreme possible values.

1.3.3.4 Voxel-wise statistics vs. tract-based spatial statistics for DTI data

There are several statistical analysis pipelines that can be used for FA images. Here we compare two of them. Paper V used smoothed FA images to compare the groups, using voxel-wise two sample t tests and correction was made for multiple comparisons. This procedure is very similar to the one to which the GM segments were subjected in Paper II.

A more recent method, TBSS, reduces the data to a WM skeleton. The statistical analyses are performed only on those voxels that are within this skeleton, thus reducing the number of multiple comparisons made. In this sense the method becomes more sensitive in the major fibre pathways. Overall however, the method may be less sensitive because voxels outside the major fibre pathways will not even be entered into the statistical analysis.

Also importantly, the TBSS method uses a re-sampling procedure to set the threshold for statistical significance. This is more appropriate because FA values are not normally distributed. In fact, their values can only vary between 0 and 1 and therefore the normal assumptions that make t tests valid may not hold. In the re-sampling procedure the members of the two groups are permuted a large number of times. For each resulting permutation, the statistical tests are performed and the results are saved. Based on the distribution of statistical test results a more appropriate threshold can be set for the given data.

1.3.4 Description of the Cohort

Contact and examination of the participants was carried out according to the guidelines, and with the permission of the local ethical committee. Each subject signed an informed written consent.

The Stockholm Neonatal Project is a population-based, prospective study of the long-term effects of preterm birth. Infants were included at birth between September 1988 and March 1993 if they were born at [n = 233] or
transferred to \( n = 58 \) the Karolinska Hospital or the Löwenströmska Hospital and if the BW was \( \leq 1500 \)g and the GA was \( \leq 37 \) weeks.

Of the 291 infants originally included at birth, 182 were available for clinical, cognitive and psychological follow-up at 5 ½ years of age\(^{26,27} \), the remainder having either died \( n = 55 \), declined participation or moved out of the Stockholm area \( n = 54 \). At that time, in order to form a control group, 125 children were selected from a population-based register according to birth date, birth hospital and having GAs \( \geq 37 \) weeks.

The ex-preterm children involved in Study 5 were part of a functional MRI study in which the inclusion criteria were based on motor activity and distractibility. The NEPSY scale\(^8\) was used to include children with motor-activity and distractibility scores, which respectively corresponded to more than 3 and more than 2 SD above the population mean. In addition we excluded all subjects with evidence of PVL or IVH on neonatal ultrasound or having an IQ < 80 at 5 ½ years of age\(^{27} \). With these criteria we had a group of 10 children available.

Seven of these 10 preterm children had matched control subjects from among the 125 included at 5 ½ years of age. The other three control subjects were recruited through advertising. The inclusion criteria for the control subjects were a GA \( \geq 37 \) weeks, BW > 2500g, and they were matched to the preterm group for age and sex.

We had successful DTI scans for nine children in the preterm group (three girls; age = 10.9 ± 0.29 years, BW = 1098.4 ± 89.2 g, GA = 28.6 ± 1.05 wk) and 10 control subjects (three girls; age = 10.8 ± 0.33 years).

For Papers I–IV each member of the entire group of participants was assigned a unique number and the order of these numbers was randomized. They were initially invited in this random order and then availability of the individual and the MRI scanner introduced further randomization so that the examination of the cases and controls proceeded in a mixed order.

### Table 1.3.3

<table>
<thead>
<tr>
<th></th>
<th>Case Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total available</td>
<td>Included</td>
</tr>
<tr>
<td></td>
<td>(n=182)</td>
<td>(n=74)</td>
</tr>
<tr>
<td>BW (g)</td>
<td>1043 [550 - 1500]</td>
<td>1069 [645 - 1486]</td>
</tr>
<tr>
<td>% of Girls</td>
<td>52 [95/182]</td>
<td>51 [38/74]</td>
</tr>
<tr>
<td>Mother’s Age (years)</td>
<td>30.76 [16 - 42]</td>
<td>30.66 [20 - 42]</td>
</tr>
</tbody>
</table>

All data is given as mean [range] except % of Girls given as % (ratio) and Mother’s Education given as median [range].
Of the 182 cases, 36 did not respond to the initial letter of invitation. Further 54 did not want to participate. Sixteen subjects that otherwise would have participated were excluded because the clip used for the surgical closure of the ductus arteriosus was of unknown composition. The remaining 76 of the cases had an MRI examination. One of these subjects could only tolerate the T2 weighted image, while the neonatal records were lost for another. Of the 125 controls, 22 did not respond to the initial letter of invitation while a further 34 responded negatively. The remaining 69 were imaged. The total available groups and the included subgroups were similar with respect to GA, BW and gender distribution as well as the mother’s age and education level at birth (Table 1.3.3).

As Table 1.3.4 details, based on neonatal ultrasound examinations, the rate of brain injury was slightly lower in the group of ex-preterm participants compared to the rest of the available group choosing not to attend the MRI in adolescence. Although this difference was not statistically significant, the fact remains that fewer signs of brain injury were noted in those that participated and this could be cause for concern. However, although it is not reported in this thesis, in a separate analysis we found that the predictive value of neonatal ultrasound examinations for long-term MRI examination was low. We reported this at the Pediatric Academic Societies conference in Toronto in 2007.

Table 1.3.4

<table>
<thead>
<tr>
<th>Score</th>
<th>GMH/IVH</th>
<th>PVL</th>
<th>ROP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Included (n=74)</td>
<td>Excluded (n=108)</td>
<td>Included (n=74)</td>
</tr>
<tr>
<td>0</td>
<td>60 [81.1%]</td>
<td>80 [74.1%]</td>
<td>61 [82.4%]</td>
</tr>
<tr>
<td>1</td>
<td>4 [5.4%]</td>
<td>14 [13.0%]</td>
<td>6 [8.1%]</td>
</tr>
<tr>
<td>2</td>
<td>6 [8.1%]</td>
<td>6 [5.6%]</td>
<td>3 [4.1%]</td>
</tr>
<tr>
<td>3</td>
<td>4 [5.4%]</td>
<td>8 [7.4%]</td>
<td>3 [4.1%]</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1 [1.4%]</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1 [0.9%]</td>
<td>0</td>
</tr>
</tbody>
</table>

Numbers are given as count (%). Germinal matrix hemorrhage (GMH) and IVH were classified into 3 categories: 0 = normal, 1 = GMH, 2 = GMH + IVH without ventricular dilation and 3 = GMH + IVH with ventricular dilation.

WM pathology (i.e. PVL) was classified into 4 categories without regards to size, duration or localization: 0 = normal, 1 = subtle WM echodensity, 2 = definite, distinctive WM echodensity, 3 = periventricular cyst formation, 4 = large (≥1 cm) intense echodensity extending into the deep layers of the WM.

ROP was classified into 3 categories: 0 = normal, 1 = grades 1-2 ROP (mild forms), 2 = grade 3+ ROP (severe forms with surgical treatment). One infant was given scores of 5 because of severe findings that did not fit the scoring system above.
After the examination of all willing participants was complete we had useful cranial MRI data from 74 ex–preterm and 69 control adolescents. The two groups did not differ significantly with respect to age, weight or height at the time of MRI scan, or the mother’s age or level of education at birth (Table 1.3.5). Due to excessive movement artifacts or incomplete datasets the DTI data of 11 subjects [7 controls] were excluded. The remaining 70 ex–preterm (51% girls) and 62 control (47%) adolescents still represented the total group well on all demographic variables.

Table 1.3.5

<table>
<thead>
<tr>
<th>Demographic data for the individuals included in Papers I–IV</th>
<th>Case Group</th>
<th>Control Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>74 (51% girls)</td>
<td>69 (49% girls)</td>
</tr>
<tr>
<td>Age (years)*</td>
<td>14.90 (12.38 – 17.7)</td>
<td>14.30 (12.18 – 16.47)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>53.18 (25.60 – 83.80)</td>
<td>55.38 (32.40 – 88.90)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>163.94 (140.00 – 196.40)</td>
<td>165.60 (138.00 – 191.00)</td>
</tr>
<tr>
<td>Gestational age (weeks)</td>
<td>28.54 (24 – 36)</td>
<td>39.72 (37 – 42)</td>
</tr>
<tr>
<td>Birth weight (g)</td>
<td>1069.54 (645 – 1486)</td>
<td>3530 (2750 – 4655)</td>
</tr>
<tr>
<td>Mother’s age at birth (years)</td>
<td>30.66 (20 – 42)</td>
<td>30.86 (22 – 44)</td>
</tr>
<tr>
<td>Mother’s level of education</td>
<td>4 (2 – 6)</td>
<td>4 (2 – 6)</td>
</tr>
</tbody>
</table>

All values are displayed as mean (range) except mother’s level of education given as median (range) and number of participants given as total (% girls).

* As quoted in this table, the age for those born preterm was not corrected for GA at birth. If this correction is made the difference between the ages is not statistically significant.

In a subgroup of participants two examinations were performed. In addition to the cranial MRI examination, abdominal and thoracic images were also collected in 45 ex–preterm and 41 control adolescents. This number was based on power calculations and a stricter set of criteria, which excluded those of multiple births or with a GA of more than 31 weeks. The resulting groups were still well matched. GA, BW and maternal age did not differ between the participants and non–participants in either the case or control groups.

1.3.5 MRI Data Acquisition

All images were collected at the MR Center of the Karolinska University Hospital on two different 1.5 Tesla scanners made by General Electric (Waukesha, WI, USA). For Study 5 a Signa Echospeed, while for Studies 1–4 a
Signa Excite Twinspeed was used. Common imaging parameters are detailed in Table 1.3.6.

The DTI acquisition for Paper V contained 5 reference images and 20 diffusion–weighted images, while for Paper II it contained 4 reference images and 30 diffusion–weighted images. In each case the b value for the reference images was equal to 0 s/mm$^2$ while for the rest it was 1000 s/mm$^2$ and the diffusion directions were isotropically distributed on the surface sphere according to Jones et al.$^{92}$ Note also that for Paper II the DTI data was acquired twice in order to remove susceptibility artefacts$^5$ [See Section 1.3.2.4 on p26].

<table>
<thead>
<tr>
<th>Studies</th>
<th>Type</th>
<th>Mode</th>
<th>Sequence</th>
<th>TE (ms)</th>
<th>TR (ms)</th>
<th>Flip Angle</th>
<th>Voxel Size (mm$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T1</td>
<td>3D</td>
<td>Gradient Echo</td>
<td>6</td>
<td>24</td>
<td>30°</td>
<td>0.98 x 0.98 x 1.5</td>
</tr>
<tr>
<td>2</td>
<td>T2</td>
<td>2D</td>
<td>Turbo Spin Echo</td>
<td>84</td>
<td>6000</td>
<td>90°</td>
<td>0.98 x 0.98 x 4.0</td>
</tr>
<tr>
<td>3</td>
<td>T1</td>
<td>3D</td>
<td>Gradient Echo</td>
<td>6</td>
<td>24</td>
<td>30°</td>
<td>0.98 x 0.98 x 1.5</td>
</tr>
<tr>
<td>4</td>
<td>DTI</td>
<td>2D</td>
<td>Spin Echo EPI</td>
<td>72</td>
<td>Trig</td>
<td>90°</td>
<td>1.96 x 1.96 x 3.0</td>
</tr>
<tr>
<td>5</td>
<td>T1/T2</td>
<td>2D</td>
<td>bSSFP</td>
<td>1.8</td>
<td>4.2</td>
<td>45°</td>
<td>0.94 x 0.75 x 5.0</td>
</tr>
<tr>
<td>6</td>
<td>T1</td>
<td>2D</td>
<td>Spin Echo</td>
<td>8</td>
<td>500</td>
<td>90°</td>
<td>1.72 x 1.72 x 5.0</td>
</tr>
<tr>
<td>7</td>
<td>DTI</td>
<td>2D</td>
<td>Spin Echo EPI</td>
<td>107</td>
<td>Trig</td>
<td>90°</td>
<td>1.72 x 1.72 x 5.0</td>
</tr>
</tbody>
</table>

In Table 1.3.6 the entry “Trig” indicates pulse–triggered acquisition, collecting 1–3 EPI image slices per heartbeat depending on the heart rate. In Study 1 the T2–weighted sequence had turbo factor of 12 (i.e. collecting 12 lines of k–space per excitation).

The entry “bSSFP” means balanced steady state free precession$^{160}$ and the acquisition in Paper IV was retrospectively gated giving 25 images in different phases of the cardiac cycle. This allowed us to identify the end–diastolic (i.e. smallest) diameter of the descending aorta.

For completeness, please note that the T1–weighted sequence in Papers I and II was also one of steady state free precession, but of the spoiled, non–balanced variant. For simplicity it was called gradient echo in this table but on GE scanners it is called SPGR.
1.4 Results

The five papers included in this thesis investigate the same question using several different types of MRI datasets, a large number of image processing tools and advanced statistical methods. Four out of the five papers consider structural brain development and one paper deals with the cross sectional area of the descending aorta. The results of each paper indicate that on the average the long-term structural development of the human body [at least the parts studied] takes a different route after preterm birth.

1.4.1 Findings on Cerebral MRI

Paper I is a prerequisite to Papers II and III in that it examines the members of the cohort with respect to structural brain lesions. We found that out of the 143 participants only 21 had any notable structural abnormalities. These were mostly of gliosis, white matter loss and some incidental findings.

![Figure 1.4.1](image)

**Incidental findings on cranial MRI**

Parts A–D, E–F and H belong to 3 different ex-preterm adolescents while Part G illustrates a control subject. For further details please see text.

Although the results contained in Figure 1.4.1 are compiled in Paper I, the actual figure was left out of the paper due to page constraints. For completeness it is included here. It illustrates a set of unexpected and/or atypical findings. Panels A–D belong to the same ex-preterm subject with both a cerebellar lesion and an occipital/medial temporal infarction that involved moderate substance loss. One subject, also ex-preterm, had a suspected cavernoma in the left insula (Figure 1.4.1 E–F). A control subject had suspected cholesterin granuloma in the right petrous apex (Panel G). One
more ex–preterm adolescent had a suspected craniopharyngeoma at the posterior suprasellar cistern and tuber cinereum (Panel H).

1.4.2 MEAN DIFFERENCES IN STRUCTURAL BRAIN IMAGES

Papers II, III and V report the results of statistical tests performed on the mean differences in a set of variables of interests (e.g. cortical thickness, brain volume, FA) either voxel–wise or at the whole brain level.

Paper V was a pilot study performed about 4–5 years earlier and it involved a small subset (n = 16) of the children that were involved in the other 4 studies [plus 3 additional controls invited through advertisement – for details please see Section 1.3.4 on p34]. The ex–preterm children were selected based on impulsivity scores, a common condition in preterm cohorts. We found that the small group of ex–preterm children had lower FA values in the posterior limb of the internal capsules and the posterior aspect of the corpus callosum. These findings however were slightly confounded by the lower volume of WM in these regions.

Based on the results of the pilot study, all the original participants were invited for more extensive MR imaging. Paper II investigated the mean differences in FA between the 74 ex–preterm and 69 control adolescents that responded, with highly improved image quality and statistical methods. Against our expectations, the initial voxel–wise analyses indicated no differences between the two groups, regardless of whether we used the TBSS\textsuperscript{171} or the more conventional voxel–based morphometry (VBM)–type\textsuperscript{11} pipeline.

Figure 1.4.2

Results without correcting for multiple comparisons I
Results of the voxel–wise statistical analysis testing indicating where FA was lower in the group of ex–preterm adolescents compared to controls.

If correction was not made for multiple comparisons these two pipelines gave similar results. Figure 1.4.2 displays regions where ex–preterm
adolescents had lower FA while Figure 1.4.3 gives the inverse analysis. Reporting results without appropriately correcting for the large number of multiple comparisons is sometimes performed\textsuperscript{172}, but it can lead to false positives. However, it was comforting that the two pipelines provided similar results.

Figure 1.4.3

Results without correcting for multiple comparisons II

Results of the voxel–wise statistical analysis testing indicating where FA was higher in the group of ex–preterm adolescents than in controls.

Combining the TBSS image–processing pipeline\textsuperscript{171} with the randomization tool\textsuperscript{135} for setting the proper threshold for statistical significance yielded lower FA values in the ex–preterm adolescents in the corpus callosum, external capsule, uncinate fasciculus and fornix (Figure 6 of paper II).

In the same paper we reported that the total GM volume and the total WM volume of ex–preterm individuals was 8.8% and 9.4% lower than their term–born counterparts. We also found that the GM and WM volumes of a given individual in adolescence depended directly on his/her GA at birth and BW (Figures 2 & 3 in Paper II).
In addition, a VBM analysis also resulted in large extended regions where the voxel-wise GM volume was lower for the cases than the controls. The areas involved were the temporal lobe, pre–motor cortex, focal regions of the parietal and the orbito–frontal cortices as well as the hippocampi and caudate nuclei, all bilaterally (Figures 4 & 5 and Table 3 in Paper II).

Figure 1.4.4

Checking for confounds in the VBM results

The results of the VBM analysis on the GM from Paper II are displayed as a maximum intensity projection. The left column illustrates the results when only those ex–preterm adolescents were entered into the analysis that had no indication of cerebral lesions upon radiological examination in Paper I. The middle column illustrates the same results as Figures 4 and 5 in Paper II. The right column indicates the results when only those ex–preterm adolescents were entered into the analysis that were born with the appropriate weight for their GA at birth. The high correspondence among the three analyses indicates that our findings were not false positive and can safely be generalized to the entire cohort.
Being born SGA has been identified as a risk factor for structural brain development\textsuperscript{176,184}. Similarly, structural brain lesions can affect the VBM analysis. To avoid the possibility of falsely overestimating the effect of preterm birth we performed two additional analyses. In one we kept only those ex-preterm adolescents in the analysis that had no indication of structural brain lesions in Paper I. In the other analysis we compared only those ex-preterm adolescents that were born with the appropriate weight for their GA to the group of controls. Figure 1.4.4 depicts the results on a maximum intensity projection (i.e. glass brain). When only the healthy ones were kept (left column) the statistically significant regions became more extensive probably due to better normalization of the brains (i.e. smaller variability). When only those were kept that were born with the appropriate weight (right column) the statistically significant regions became slightly smaller in extent, probably due to the smaller number of individuals involved.

Figure 1.4.5

**VBM results after inclusion of total brain volume as a covariate**
The analysis here was similar to that used for Figures 4 and 5 in Paper II except that total brain volume was included as a covariate.
Most of these differences were explained by including the age of the individual at the time of scanning, their gender and their total brain volume as covariates. Only parts of the temporal and parietal lobes remained statistically significant (Figure 1.4.5). Because the two groups were highly matched with respect to age and gender (see Table 1.3.5 on p37) the only variable explaining these differences was the total brain volume. Therefore these findings can be interpreted such that the variability in voxel-wise GM volume was independent of the total brain volume only in the regions depicted in Figure 1.4.5.

The results of VBM-type analyses are interpreted as local GM volume effects. But, for example, it is difficult to decide whether a particular change that we observe is a result of changes in the depth of the sulci or the thickness of the cortex. To follow-up the findings in Paper II, we investigated the cortical thickness directly in Paper III. Interestingly, the regions where the cortex was statistically significantly thinner in the group of ex-preterm adolescents were similar to those showing a lower GM volume independent of the total brain volume in the VBM analysis (compare Figure 1 of Paper III and Figure 1.4.5). When the maps were displayed irrespective of statistical significance large areas resulted where the cortex was consistently thinner in the ex-preterm adolescents. Interestingly, these areas differed depending on the BW and GA of the subjects. While for those that had been born with lower BWs or shorter GAs the central, lateral parts were involved, involvement of the anterior and posterior poles became more prominent for those that were born with higher BWs or longer GAs (Figures 2 and 3 in Paper III).

1.4.3 **Abdominal Aorta Imaging Results**

Paper IV is entirely different in that it investigated the effect of preterm birth on the cross sectional area of the descending aorta between the heart and the iliac bifurcation (see Figure 1.3.15 on p30). We found that the average diameter was smaller for the group of ex-preterm adolescents at all levels we measured (See Table 2 in Paper IV). For example the thoracic and abdominal segments were smaller by 16% and 19% respectively. This finding was statistically significant even after adjusting for body surface area and gender. In addition, maternal smoking during pregnancy was associated with a 10%–13% reduction along the entire segment. Finally, although both the systolic and diastolic blood pressures were higher in adolescents born preterm, the extent did not correlate with aortic diameter or maternal smoking.
1.5 Discussion

When comparing ex–preterm children and adolescents to a term–born control group we found that being born preterm affects long–term development. In particular, we found ex–preterm adolescents had

a) higher rates of structural brain injury, mostly gliosis and WM loss
b) lower total GM and WM volumes by 8.8% and 9.4% respectively
c) lower voxel–wise cortical GM volume and thinner cortex in a large set of regions involving all four lobes
d) lower FA in the corpus callosum, the external capsules and the fornix when including the entire group or in the corpus callosum and internal capsules when looking at only a subgroup with attention deficits
e) smaller aortic cross sectional area at all levels examined

Despite this extensive set of results indicating preterm birth to be a risk factor for long–term development, the structural brain imaging findings were relatively minor when compared to reports from other centres involving similar cohorts. Using MR imaging to investigate the dimensions of the abdominal aorta in a cohort of ex–preterm adolescents is unprecedented and therefore the severity of the findings are difficult to assess.

1.5.1 The Pilot Study

Chronologically Paper V was the first study in this thesis. Although originally not planned as such, it can be considered a pilot for the other four papers in several respects. First, we imaged a small but very specific subgroup of the cohort investigated in the other four papers. The inclusion criteria included inattentiveness and hyperactivity\textsuperscript{98}, in a sense comparing controls with those members of the cohort where we were most likely to find structural differences. Although we excluded children with neonatal signs of PVL or IVH, and thus possibly reducing the chance of positive findings, had the results been negative we probably would have been more reluctant to endure the efforts and expenses involved in investigating the entire group. While Huppi et al.\textsuperscript{78} had shown with DTI that preterm infants differed from their term–born counterparts, long–term, follow–up, DTI studies were not available. Therefore, it was not generally clear whether the differences identified in the neonatal period would persist. It was not evident that embarking on a large DTI study, involving the entire cohort, would yield positive results.

Later, when investigating the whole cohort, we found that the long–term predictive power of neonatal ultrasound is rather poor\textsuperscript{131} and therefore the exclusion of children with neonatal signs of PVL and IVH probably had no effect on the results. Even in Paper V we found 3 ex–preterm children with mild reduction of periventricular WM – despite the strict inclusion criterion to exclude them.

Paper V can also be considered a pilot study because the standards of data acquisition and processing are highly improved for Papers I–IV.
The final aspect which makes Paper V a pilot study is that due to the strict criteria to include only those ex–preterm children that had psychological deficits, extending the findings of Paper V to the general population of ex–preterm children was not possible – as it was discussed in that paper.

1.5.2 Studies Involving the Entire Cohort

1.5.2.1 Radiological evaluation of MRI images

Paper I is a prerequisite for Papers II and III because proper description of the involved individuals required a radiological evaluation of the cerebral T1–, and T2–weighted MRI data.

Only 21 of the 143 datasets showed any positive findings and 4 of these belonged to controls. Of these 17 (2 controls) had suspected findings of gliosis (see Figure 1 of Paper I), reduction of WM volume or remnant signs of PVL while the other 4 (2 controls) had incidental findings. These findings were relatively scarce. For example Steward et al. reported that only 17 of 72 preterm born individuals had normal MRI findings and Maalouf et al. only found 13 of 41 participants normal on MRI. Correspondingly, Skranes et al. found MRI pathology in 84% of the cohort they studied.

Note that, although ventricular dilation had been shown to co–occur with preterm birth, in this study it was not scored separately unless it was associated with WM loss (see Figure 2 in Paper I). In addition, among those that were classified as normal above, 19 subjects (4 controls) had enlarged ventricles. Even if these are added to the total tally, the rate of positive findings is still relatively low.

1.5.2.2 The VBM and DTI results

Only the aspects dealing with DTI data in Paper II were the direct follow–up of the pilot study in Paper V. It was interesting, though unexpected, that the results differed to such a large extent. The only commonality was the finding that FA was lower in the posterior corpus callosum among the ex–preterm adolescents. But while inattentive individuals tended to possess lower FA values in the posterior limb of the internal capsule (Figure 1 and 3 of Paper V), the entire cohort did not, and instead possessed lower FA in the external capsules, the uncinate fasciculus and the fornix (Figure 6 of Paper II). This supports our earlier claim that generalizing the results from Paper V to preterm born individuals in general would have been inappropriate. This also provides warning that hypotheses must be clearly identified. The morbidity of preterm birth is quite variable. While one individual may suffer from ROP, another might have periventricular gliosis and a different person still may suffer from neither of these but be abnormally inattentive. Therefore, if we are interested in the effect of preterm birth in general, it is less likely that we find wide–spread results. That is because even if 10 individuals show an effect, say a reduction of FA values in an anatomical region, if another 133 individuals possess no such reduction, the effect will be washed out. In this case, adding more individuals to the study does not increase statistical significance (see the Statistical significance section on p31).
It is also important to realize that both the ex-preterm and the control adolescents will be variable on most outcome measures. If we can identify a statistically significant difference between the groups based on given effect size, we could probably find a statistically significant result within the control group alone with the same effect size. For example, at 5 ½ years of age the ex-preterm members of this cohort were found to have a lower IQ than the controls. The effect size was less than 1 standard deviation. Such an effect size can easily be found by taking the top and bottom halves of the control subjects.

The VBM analysis and the estimation of the total brain volume were added to the DTI analysis in Paper II by making sure that the 3D T1-weighted images that were collected to aid spatial normalization of the FA images were of good enough quality for segmentation.

That ex-preterm adolescents had smaller total GM and WM volumes (by 8.8% and 9.4% respectively in our study) had been reported before and hence was expected. However, studying Figures 2A and 3A in Paper II provides a large overlap between the groups. Many individuals in one group could well belong to the other group. While the results are statistically significant, it may be a mistake to establish clinical significance on that criterion alone. As introduced in the Statistical significance section on p31, given large enough groups effects of any size, regardless how small, can be shown to be statistically significant. If a causal relationship is assumed between preterm birth and brain volume, reduction in brain volume of 8–9% seems significant. However, considering that the natural variability within the control group alone can be over 40%, the clinical and biological significance of 8–9% difference in brain volume between two groups may not be alarming. This reasoning is, of course, only meant for the isolated occurrence of total brain volume differences. For example, if the entire 9% difference between two groups happens to be the missing hippocampus the clinical significance would be obvious.

In a very recent, longitudinal study from earlier in 2009, Ment et al. reported different dynamical brain growth between preterm and term born children. When correlating age at the time of scan and total GM or total WM volumes in Paper II we did not find a difference between the two groups on growth rate. This may be due to differences between the study designs. The mean age of the participants in the study by Ment et al. was about 5 years lower and although in Paper II the ages of participants at the time of scan ranged between 12–17 years the design was still cross sectional in nature. Also, there were fewer participants with a larger difference between the two groups based on IQ scores in the study by Ment et al. and these may have resulted in a more specific and coherent group.

The total GM and WM volumes were related to the GA and BW of the individual (see Figure 2C–D and Figures 3C–D in Paper II), suggesting a causal relationship. Maybe the causal relationship is not between preterm birth and brain volume but between the causes of preterm birth and brain volume. While it is obvious that being born earlier or with a smaller weight is related to having
smaller brain volumes the variance introduced is similar in extent to the variability in brain volume that term–born control children have.

Interestingly, if the total brain volume was not included as a covariate, the VBM analysis indicated that the ex–preterm adolescents had lower voxel–wise GM volume in extended regions, where the spatial specificity was in agreement with previous studies. However, after the variance due to total brain volume was accounted for, independent statistical significant effects were reached in regions of much reduced spatial extent (see Figure 1.4.5 on p43). Not all, but most investigators remove the effect of total brain volume variability. It is debatable whether this correction should be made. The idea is straightforward. For example, one would not like to report a difference in the size of the corpus callosum between two groups if the groups also differ in total brain volume. This reasoning only seems valid if the measurement of the corpus callosum size was made independent of the rest of the brain. If a voxel–wise, exploratory analysis identifies the corpus callosum as the sole difference between the groups, then it is clear that the effect is largest in that region. In addition, if the results of the voxel–wise analysis are as extensive as in our case (see Figures 4 and 5 in Paper II), it is expected that the total brain volume will co–vary. Therefore, removing the effect of the total brain volume will automatically remove the local effects, as it happened in our case (see Figure 1.4.5 on p43).

The correspondence between the VBM and DTI results was promising. For example, the regions of the corpus callosum where we found lower FA in the ex–preterm adolescents contain fibres that innervate the bilateral motor, sensory, parietal and temporal cortices as well as the parahippocampal gyrus. Also, the fibres running in the external capsule run to the striatum and those in the fornix pass the hippocampi and the uncinate fasciculus connects the orbitofrontal and anterior temporal areas. Further studies are required to investigate the more accurate correspondence between the VBM and DTI findings (see Future Directions on p55).

Importantly, most of the main findings remained unaffected when excluding individuals born SGA or having positive findings on radiological examination in Paper I. This ensures that our findings are not false positives that result from the large effects of small subgroups within the cohort, which should not be generalized.

1.5.2.3 Cortical thickness results

Similar to Paper II, the investigations of the effect of preterm birth on cortical thickness were also provided as an additional measure based on the higher quality 3D T1–weighted image that afforded good segmentation (see the Segmentation section on p22).

Statistically significant differences were found in very restricted regions between the two groups (see Figure 1 in Paper III). Few neuroimaging studies have been aimed at investigating the long–term effects of preterm birth on cortical thickness. Our findings indicate a much more restricted set of regions than that reported by Martinussen et al.

Still the effect seems wide–spread, and even though it is not significant voxel–wise, these results deserve further analysis. In neuroimaging, statistical
tests are usually performed independently for each voxel and then a threshold is chosen that assures a correct rate of false positives despite the large number of multiple comparisons. By convention, results are only displayed if they surpass the threshold set. This approach is easy to understand when the effect sizes are random except for the area that surpasses the threshold. However, in our case a large set of regions was uniformly thinner on the average for the ex-preterm group. While any individual voxel may not be statistically significant, it is highly unlikely that the finding of thinner cortex in such a large area arose by chance. Subjecting the data to a different analysis emphasizes this effect. For example, the sign test assigns a score of +1 to a voxel if the control group has thicker cortex there and a −1 if that group has a thinner cortex. The scores are summed and in the absence of any differences the total score should be close to zero. In our case, however, it is far from zero because in a large set of regions the mean cortical thickness for the control group is higher than that for the preterm group.

The spatial arrangement of the regions where the cortex was thinner for the group of ex-preterm adolescents depended on GA and BW (see Figures 2 and 3 in Paper III). This effect is a combination of true dependence of cortical thickness on GA and BW, statistical power and diluting the effects by studying a heterogeneous group. For example, if there is a true effect between individuals with low GAs and controls that does not exist between individuals with higher GAs and controls, then combining the groups can result in the loss of statistical significance. But in general the subgroup with lower GA and BW had thinner cortices in the central regions of the brain. In the subgroup with higher GA and BW this effect shifted towards the posterior and anterior poles of the brain.

Cortical thickness has been found to relate to the general intelligence. Interestingly, the relationship was negative for young children but changed to a positive one in late childhood and stayed as such until late adolescence. In addition, the rate of change in cortical thickness was also related to score on intelligence tests. The more dynamic cortical development represented those with superior scores while more modest changes were observed in those with average scores. Although the range of individuals’ ages was more restricted in our study (i.e. 12–17 years) the effects in the above study by Giedd et al. were highly significant only when the children were subdivided according to age (not at the level of the whole group). Therefore, if in the currently ongoing cognitive follow-up of the cohort a difference is found in the IQ scores between the terms and preterm participants, cortical thickness correlations are hopeful.

1.5.2.4 Abdominal aorta results

Finally, Study IV was an even more extreme addition to the original design for the follow-up in that it did not even involve an investigation of the brain. While admittedly it loses focus, it importantly recognizes that there is more to the developing body than the brain and broadens the methodological requirements and applications of this thesis.

We found that the cross-sectional area of the descending aorta was significantly smaller in individuals born preterm at all levels between the heart and the iliac bifurcation. Similar to the findings on the brain, there was a
correlation between GA and the size of the aorta. While this finding was not unprecedented\textsuperscript{29,108} the data in the current study does not afford to investigate the underlying mechanisms resulting in this difference. One working hypothesis is that in the ex-utero environment aortic development must occur under varied stresses. For example, the systemic pressure of the neonate is higher than when connected to the placenta through the umbilical cord. The presence of a significant patent ductus arteriosus can reduce aortic blood flow because it allows oxygenated blood to flow back to the lungs, which now offer reduced resistance\textsuperscript{155}.

1.5.3 DISCUSSION OF THE COMMONALITIES AMONG THE FIVE STUDIES

The introduction detailed some reports indicating wide spread neurological deficits in both the short and the long term among individuals that were born preterm. Despite these reports and against our expectation in Papers I–III we found relatively mild effects of preterm birth. Although surprising at first, maybe these findings should not have been unexpected for several reasons. The neonatal mortality rates were much higher in other countries\textsuperscript{70,178} and even in Sweden\textsuperscript{90} than in the Stockholm Neonatal Project cohort\textsuperscript{96}. Also, children with low BW have been shown to have impaired heights and weights compared to their normal BW peers\textsuperscript{147,157} but in the cohort we studied we could not identify such a difference between the ex-preterm and control adolescents (please see Table 1.3.5 on p37). Furthermore, all the participants underwent a psychological evaluation at 5 ½ years of age. Although, on the average, the IQ scores of the ex-preterm children fell below those born at term, the group mean was still within the normal range\textsuperscript{96}. In a recent report Zacharia et al. found no differences in brain volumes in the neonatal period between preterm born infants and term born controls\textsuperscript{195}. Although, their preterm born participants had longer gestation compared to participants in our cohort, due to the mild neonatal mortality and morbidity as detailed earlier in this paragraph, results of Zacharia et al. may be applicable to our cohort.

The relatively mild neonatal mortality and morbidity\textsuperscript{96} and the outcome on psychological follow-up may be considered a bias. As such, these may be grounds for criticism against the generalizability of our findings and will be discussed further in the next paragraph. A definite selection bias was found in that the adolescents that chose to attend the MRI examination for Papers I–III had a lower rate of neonatal signs of brain injury or occurrence of ROP (please see Table 1.3.4 on p36). Regardless of the fact that the difference between the participants and the non-participants was not statistically significant (two-sample Kolmogorov–Smirnov goodness-of-fit test), the participants still had lower rates of PVL, IVH and ROP. And although we found that the neonatal ultrasound examination used for detecting PVL and IVH had a very low predictive value\textsuperscript{131} for the long term–term MRI findings in Paper I, this selection bias may still be considered a limitation of Studies I–III.

Regardless of their severity, our results probably cannot be generalized to other cohorts that have been cared for at other international sites. But that
is not a weakness of this study. In fact, exactly this lack of chance to generalize results across cohorts and countries was the driving force behind these studies. Both the short–, and the long–term outcome after preterm birth can depend on the care received\textsuperscript{3,32,67,72,73,130,137} and the care given differs from site to site \textsuperscript{183}. Even within Sweden mortality rates can vary depending on whether the infant was born at a general hospital or a clinic that is associated with a university\textsuperscript{90}. In addition cohort studies, by definition, do not allow for generalizing results because the involved subjects are not randomly selected but instead included based on a pre–defined condition they have. The only way to generalize results internationally would be to allow any preterm born infant to be included and select those to be studied by chance. Of course, this would be logistically impossible to undertake. Even if it were possible, the high variability of outcome in preterm children would reduce the sensitivity of any finding. For example suppose at site A for some reason children tend to have smaller corpus callosum cross sectional area on MRI examinations. If this is not a problem at the other sites then on average it will not be found. Still, if one wants to find the commonalities of preterm birth the multi centre randomized trial is the proper way to do so.

We stated in Papers II and III that our findings were milder than those reported elsewhere. However, comparing severity of outcome between cohorts based on statistical analysis of MR is very difficult. Usually, the display indicates only statistically significant results. Often the regions are colour–coded but the range of colours corresponds to the p–value rather than the effect size. As detailed in the Statistical significance section on p\textsuperscript{31}, the declaration of statistical significance depends on several factors. When saying that our results were milder than that of others’ we meant that the regions where statistically significant differences were found were less extensive. But this does not necessarily indicate a difference in the biological significance. For example the effect size is not mentioned explicitly. Considering the cortical thickness measurements in Paper III, one would also like to know how the difference between the preterm–, and term–born adolescents compares to that in other studies. To safeguard ourselves from false statements we can strive for the situation displayed on the right of Figure 1.3.16 on p\textsuperscript{32}. That is, collect a large set of data and try to minimize all the sources of variance. The first aspect is simple: invite everybody and image everybody that is willing, as we have done. Reducing the variability in the data is more complex and requires care in both the data acquisition and the image processing. On the acquisition side the following precautions were taken:

a) we decided to follow a two–stage design where in the first stage only neuroimaging data were collected but no psychological correlates were investigated
b) the same individual collected all but 2 of the datasets for Papers I–IV
c) the same scanner running the same MR sequences was used
d) twice refocused diffusion encoding\textsuperscript{153} was used to obtain data with reduced eddy current distortions
e) the DTI data was collected twice in order to remove susceptibility distortions
On the post processing side we also took special care to:

a) work together with the groups developing image processing algorithms (SPM, FSL distortion correction of DTI datasets)
b) use the latest available tools of the above
c) visually validate the output of all automatic algorithms
d) remove movement related registration errors
e) remove eddy current artefacts
f) remove susceptibility induced artefacts

To illustrate one aspect of part b) in the list above, Figure 1.5.1 displays the difference that using the DARTEL algorithm for spatial normalization can make. The images were not smoothed. The apparent smoothness of the image on the left is a result of imperfect alignment with the standard MNI template.

Figure 1.5.1

Reducing variability in the data
Both images display the mean normalized FA image of 132 subjects. On the left the older spatial normalization algorithm was used while on the right we used DARTEL.
Although Paper V used only a very specific subgroup of the cohort involved with Paper II, the differences between the results are probably due as well to highly improved data quality and processing methods.

We considered our two–stage design a strength and this needs further qualification. The current trend is toward considering structural brain imaging results only in tandem with one or more psychological test scores. The reason we resisted this tendency was with the aim of improving image quality. Collecting only MRI data allowed for a more extensive dataset per individual and faster completion of the study. Also, during this time there were no major hardware changes in the scanner and hopefully neither any deterioration of the machine. We scanned all 143 individuals between April 2005 and February 2006 and even collected abdominal data in a large portion. The psychological follow–up is currently under way, where each subject is seen at approximately 18 years of age, hopefully reducing unwanted variability in that data as well.

It was stated several times in Papers I–III that being born SGA did not affect our results. Because there are contradicting reports about the risks of being born SGA, it needs to be stated that the studies included in this thesis were not designed to investigate whether being born SGA carries further risk to being born preterm but with the appropriate weight. The adolescents that had been born SGA were not matched to those born with a weight that was appropriate for their GA. But because due to the selection criteria being BW but not GA there was a higher occurrence of being born SGA in the adolescents with longer gestation. Compiling a cohort based on BW alone can be problematic and for this reason we took extra care to ascertain that including those individuals who had been born SGA did not artificially produce positive findings. But these extra checks were only that, checks to validate our results, and should NOT be interpreted as an indication that being born SGA carries no extra risk!

Although Figure 1.1.2 illustrated the prominent neurodevelopmental stages that occur during the period of normal pregnancy when the ex–preterm adolescents were born, it was not meant to reduce the importance of other factors that may cause or co–occur with preterm birth. For example Smith et al. found indications early in pregnancy which correlated with preterm delivery, SGA birth or growth retardation (see also work by Goldenberg et al.). It is plausible that early entrance into the extra–uterine environment causes our findings. For example, it is known that thickness of the cortex remains relatively constant throughout evolutionary brain development. The thinner cortex, identified in Paper III, may be a result inadequate apoptotic activity, while the reduced cortical grey matter volume in Paper II could be due to a disturbance in the synaptic proliferation. In turn these could lead to the hindered connectivity between these areas of the brain (Papers II and V). However, the studies in this thesis were not designed to elucidate the causes of preterm birth. Instead, because it is known that normal brain development reaches well into adolescence, these studies, being cross sectional and observational in nature, were meant to investigate whether ex–preterm individuals catch up to their term–born peers by adolescence.
1.6 **CONCLUSIONS**

Based on the cohort available to us for examination and the methods used we conclude that being born preterm can be a risk factor for long-term development. We found that preterm children tend to have smaller brains, thinner cortices, lower FA and smaller aortas even after allowing 15 years for catch-up. However, when considering the results on the brain, our results indicate a milder deviation from the control group than previously reported while investigating similar cohorts from other centres of the world. The milder than expected outcome may be due to selection bias in that the subgroup of ex-preterm adolescents we imaged was relatively healthier compared to the non-participants. However, it is also possible that the entire cohort was better off since, for example, even in the newborn period the mortality rate was relatively low. It has been shown recently that being born at a university hospital can improve the outcome compared to other hospitals of the same nation. Should this possibility be founded, it would provide hope that neonatal care can perhaps improve to the point where preterm born children can grow up without major abnormalities.
1.7 Future Directions

The results presented in this thesis are but a fraction of the available information contained in the MRI data compiled. Several further studies are planned:

- At the time of writing of this thesis, cognitive data are being compiled from the participants. Once completed, the structural MRI data will be used in conjunction with those to investigate whether the structural differences we found have relevant influence or connection on cognitive ability.
- DTI data will be used to investigate whether preterm birth causes differences from the normal connectivity of the brain. In particular we would like to investigate whether the areas where we found lower FA contain WM fibre pathways that connect cortical areas where we identified a reduction of GM or thickness.
- Data from the controls alone can be used for several investigations, both medical and methodological. For example we can study gender variability, developmental trajectories as well as use the data to investigate the robustness of statistical methods on FA or estimate the benefits of distortion correction.

Furthermore, we gained valuable experience in carrying out an MRI study of large cohorts. It would be important to follow this cohort further into young adulthood. The brain develops past 15 years of age and it is unclear whether the differences found in these papers would be exaggerated later in life or reduced below the limit of observability.

Finally, our findings indicate that preterm children, born after the 23\textsuperscript{rd} week of gestation, can grow up to be minimally different from those born at term. Therefore, maybe it is not unreasonable to hope that with the proper neonatal care and, if needed, with rehabilitation in childhood, preterm birth could be eliminated as a risk factor. These results encourage further research into caring for preterm born individuals.
## Acknowledgements

Professors Hugo Lagercrantz, Torkel Klingberg and Hans Forssberg provided guidance during my first years as a PhD student. For the most part Professor Lagercrantz was my main supervisor and most of what I know about neonatology I learnt through his kind help and support.

Ms. Viveca Karlsson is my Swedish mother. It was with her encouragement that I made the move to Sweden from Canada and she took very good care of me whenever I needed it. Kiss!

Drs. Bo Nordell and Stefan Skare guided me through my baby steps as an MRI physicist. It took a lot of steps and I’m very grateful for their patient support.

I am forever indebted to Dr. Jesper Andersson. For one, he got me going with MR image processing. But even more important, he encouraged me to apply for the physicist position at the FIL (Wellcome Trust Centre for Neuroimaging; anno Functional Imaging Laboratory), a post, which I thoroughly enjoy and am not exactly sure how I could live without it before.

For a while, Anders Nordell and I were the only two MR physicists at the Karolinska Hospital in Solna. We were learning as we went along. They were great times, which I’m missing now and I know I always will. Love ya’ Dude!

I would also like to thank Dr. Lu–Yang Wang of the Physiology Department at the University of Toronto. He was not directly involved with this thesis. However, he supervised me through my MSc project and much of what I had learnt from him I relied on regularly even after having left his lab. He is an exceptional scientist, a great educator and dear friend. See you soon Lu.

Drs. Anna–Karin Edstedt Bonamy and Mikael Norman helped me realize that there is more than a brain to the developing human body. Thanks!

When I arrived at the MR Centre at the Karolinska Hospital I was scared to be inadequate for the job. But Yords Österman, Marie Lundberg and Kristina Malmqvist selflessly and tirelessly taught me how to actually examine human beings in an MRI camera. My sincere gratitude goes out to them. Their involvement definitely improved the quality of the data in these studies as well as the comfort and safety of the individuals examined.

Many more people made my time at Karolinska (and in Sweden) fantastic. Even though they are not listed here they probably know who they are. I think of them warmly and will thank them personally for the fun times...

The work herein was supported by grants and scholarships from Sällskapet Barnavård. During my PhD program I completed more projects than are included in the thesis. I would like to acknowledge further support from Stiftelsen Frimurare Barnhuset, Stiftelsen Sven Jerrings Fond and Stiftelsen Solstickan.
3 References


(10) Ashburner J and Friston KJ. Nonlinear spatial normalization using basis functions. Hum Brain Mapp 1999: 7[4]:254-266.


[34] Chen J and Smith LE. Retinopathy of prematurity. Angiogenesis 2007: 10(2):133-140.


Nosarti C, Al Asady MH, Frangou S et al. Adolescents who were born very preterm have decreased brain volumes. Brain 2002: 125(Pt 7):1616-1623.


[142] Perlman JM. White matter injury in the preterm infant: an important

[143] Peterson BS, Vohr B, Staib LH et al. Regional brain volume

[144] Petrides M and Pandya DN. Association fiber pathways to the frontal

[145] Pham DL, Xu C, Prince JL. Current methods in medical image

[146] Poletti CE and Creswell G. Fornix system efferent projections in the

[147] Powls A, Botting N, Cooke RW, Piling D, Marlow N. Growth impairment
in very low birthweight children at 12 years: correlation with perinatal
and outcome variables. Arch Dis Child Fetal Neonatal Ed 1996:
75(3):F152-F157.

MR images of the developing newborn brain. Med Image Anal 2005:

[149] Purcell EM, Torrey HC, Pound RV. Resonance Absorption by Nuclear

[150] Rakic P. Neurons in rhesus monkey visual cortex: systematic relation
between time of origin and eventual disposition. Science 1974:

[151] Rakic P. A small step for the cell, a giant leap for mankind: a hypothesis
of neocortical expansion during evolution. Trends Neurosci 1995:
18(9):383-388.

[152] Ranke MB, Vollmer B, Traunecker R et al. Growth and development are
similar in VLBW children born appropriate and small for gestational

[153] Reese TG, Heid O, Weisskoff RM, Wedeen VJ. Reduction of eddy-
current-induced distortion in diffusion MRI using a twice-refocused spin

[154] Reiss AL, Abrams MT, Singer HS, Ross JL, Denckla MB. Brain
development, gender and IQ in children. A volumetric imaging study.

[155] Rosner B. Fundamentals of biostatistics. 5th ed ed. Australia: Duxbury,
2000.

[156] Rutherford MA. MRI of the neonatal brain. London : W.B. Saunders,
2002.


