High-Grade Spondylolisthesis in Young Patients
Long-Term Results of In Situ Fusion
High-Grade Spondylolisthesis in Young Patients
Long-Term Results of In Situ Fusion
To my family
Contents

Abstract ix

List of papers xi

Abbreviations xiii

1 Introduction 1
   1.1 High-grade isthmic spondylolisthesis ................. 1
   1.2 The controversy .................................. 3
   1.3 Sagittal balance .................................. 4
   1.4 Adjacent segment disk degeneration ................... 9
   1.5 Nonunion ......................................... 10
   1.6 Assessment of health-related quality of life ........... 11
   1.7 Self-image and the SRS-22r questionnaire ............. 13

2 Aims 17

3 Patients and methods 19
   3.1 Patients (I-IV) and controls (IV) ...................... 19
   3.2 Surgical treatment (I-IV) .............................. 20
   3.3 Health-related quality of life (I,IV) ................... 20
   3.4 Physical examination (I) ................................ 21
   3.5 Nonunion (I) ....................................... 21
   3.6 Sagittal balance (II) ................................ 21
   3.7 Slip and slip angle (II) ............................... 23
   3.8 Adjacent segment disk degeneration (III) ............. 23
   3.9 Self-image (IV) ...................................... 24
   3.10 Statistical analysis (I-IV) ............................ 24
   3.11 Ethical review board approval (I-IV) .................... 25
Abstract

The purpose of this long-term follow-up of 38 of 40 consecutive patients was to evaluate the results of uninstrumented in situ fusion for high-grade isthmic spondylolisthesis three decades after surgery. The mean age at surgery was 14 (range 9-24) years. The first of four studies evaluated clinical outcome, function, work status, and health-related quality of life (HRQoL) after in situ fusion in relation to age-matched Swedish population data. The second study evaluated effects on sagittal balance after in situ fusion. The third study evaluated adjacent segment disk degeneration after in situ fusion. The fourth study evaluated self-image and HRQoL after in situ fusion in relation to healthy controls.

The main findings were that (1) young patients fused in situ for high-grade isthmic spondylolisthesis have long-term HRQoL similar to the general Swedish population and controls matched for age and gender, (2) signs of non-compensated sagittal imbalance were observed only in a few individuals whereas compensated sagittal balance was the norm, (3) there was no correlation between any radiographic sagittal balance parameter and HRQoL outcome, (4) there was only a minor reduction in adjacent segment disk height which had no impact on HRQoL outcome, and (5) the only patient reported outcome measure indicating a detrimental effect at long-term follow-up was self-assessed trunk appearance which was slightly negatively affected.

Keywords: Adjacent segment disk degeneration, Fusion in situ, Health-related quality of life, Sagittal balance, Self-image, Spondylolisthesis.
List of papers


3 © 2020 Scoliosis Research Society.
4 © 2018 Scoliosis Research Society.
Abbreviations

ASD Adjacent segment disk degeneration
CI Confidence interval
Dub-LSA Dubouset lumbosacral angle
EQ-5D-3L EuroQol 5 dimensions instrument 3-level version
HRQoL Health-related quality of life
LL Lumbar lordosis
ODI Oswestry disability index
PI Pelvic incidence
PT Pelvic tilt
RCT Randomized controlled trial
r_s Spearman’s rank correlation coefficient
SD Standard deviation
SF-36 Medical outcomes study 36-item short-form health survey
SRS-22r Scoliosis Research Society questionnaire 22r
SS Sacral slope
SVA Sagittal vertical axis
TK Thoracic kyphosis
TPA T1 pelvic angle
T1SPI T1 spinopelvic inclination
VAS Visual analogue scale
ZDS Zung depression scale
1 Introduction

1.1 High-grade isthmic spondylolisthesis

Spondylolisthesis is a spinal deformity characterized by sagittal displacement of one vertebra on the one below (Wiltse and Winter 1983). The deformity is often classified according to Wiltse et al. (1976) into 5 types based on etiology: (1) dysplastic, (2) isthmic, (3) degenerative, (4) traumatic, and (5) pathologic. The dysplastic type involves a congenital deficiency in the facet joints while the isthmic type is due to a lesion (spondylolysis) in the pars interarticularis of the vertebral arch. The degenerative type is secondary to facet joint degeneration. Traumatic spondylolisthesis involves an acute fracture of the posterior elements of the vertebra. The pathologic type is due to a tumor or generalized bone disease (Hu et al. 2008). Isthmic spondylolisthesis is the most common type in young patients (83% in the series of Boxall et al. 1979).

There are several classification systems for grading the severity of spondylolisthesis. The Meyerding (1932) classification grades the severity of the slippage on the basis of translation into 5 grades (1: 0-25%, 2: 26-50%, 3: 51-75%, 4: 76-100%, 5: >100%). The degree of slippage may also be divided into low-grade slips (≤ 50%) and high-grade slips (> 50%) relative to the vertebra below. A different approach is used by Spinal deformity study group (SDSG) classification system where 6 types of spondylolisthesis are defined based on pelvic incidence, pelvic balance and sagittal balance (Labelle et al. 2011).

In this thesis, only high-grade isthmic spondylolisthesis was studied. This is a very uncommon condition. A population study of Virta et al. (1992) found 68 low-grade slips and one high-grade slip (Meyerding grade 3) in a cohort of 1147 adults while Fredrickson et al. (1984) found 13 low-grade slips and no high-grade slip in a cohort of 500 first-grade children. The most commonly affected level is the lumbosacral junction (86% of the cases in the series of Fredrickson and Virta).
The etiology of isthmic spondylolisthesis is unknown. Fredrickson et al. (1984) reviewed radiographs of 500 normal newborns and did not find any case of spondylolisthesis which indicates that spondylolisthesis does not exist at birth. The pars interarticularis, however, is cartilaginous in the newborn which means that spondylolysis could not be diagnosed by the radiographic survey. Fredrickson et al. (1984) also reported, based on radiographs of 500 first-grade children, that slip seems to occur concomitantly with the development of the pars interarticularis defect (spondylolisthesis was seen in 13 of 19 patients with spondylolysis on the initial radiographs) and that it was unusual for progressive slip to develop once the diagnosis of spondylolysis had been made. There is also a hereditary factor. Spondylolysis is more common in first-degree relatives of individuals with spondylolysis (15%) than in general population (6%) (Wynne-Davies and Scott 1979; Fredrickson et al. 1984).

Danielson et al. (1991) found slip progression in 3% of patients (≤ 30 years old) with spondylolisthesis when progression was defined as an increase in slip of ≥ 20% of the sagittal length of the lower end-plate of L5. Seitsalo et al. (1991) and Fredrickson et al. (1984) found slip progression in 15% and 21% respectively in conservatively treated children and adolescent when progression was defined as an increase in slip of > 10%. This illustrates that the definition of progression is critical. Two studies on roentgenologic assessment of spondylolisthesis showed that a true slip progression of < 20% (< 8 mm) was difficult to detect with statistical certainty mainly because of systematic errors introduced by variations in patient positioning (axial rotation and lateral tilt) and inter-/intraobserver errors (Danielson et al. 1988, 1989).

Seitsalo et al. (1991) noted that slip progression increased with increasing slip. Danielson et al. (1991), however, could not find any correlation between slip progression and any radiographic parameter (including slip severity). Seitsalo et al. (1991) also noted that progression was higher in age groups corresponding to growth spurt. Beutler et al. (2003) found no progression (>20%) after the age of 18 years at a 45-year evaluation of the Fredrickson et al. (1984) cohort (diagnosed at age 6-7 years).
1.2 The controversy

There are many conflicting opinions regarding the best treatment of high-grade isthmic spondylolisthesis in young patients (Cheung et al. 2006). Although some authors argue for watchful waiting (Bourassa-Moreau et al. 2013; Lundine et al. 2014), most authors recommend spinal fusion regardless of symptoms to avoid slip progression (Wiltse and Jackson 1976; Lonstein 1999; Hu et al. 2008). There is no consensus on how to surgically address the slip and the lumbosacral kyphosis (surgical reduction and fusion or in situ fusion). Several rationales for reduction have been proposed (Matthias and Heine 1986; Lonstein 1999; Hresko et al. 2009). These rationales can be summarized as (1) to reverse neurological deficiencies, (2) to improve function, (3) to prevent nonunion and slip progression, (4) to restore normal biomechanics, and (5) to improve clinical appearance. The benefits of reduction should be weighed against the risk of neurologic complications due to L5 traction during reduction (Lonstein 1999; Fu et al. 2011).

There are no randomized controlled trials or prospective cohort studies on the treatment of high-grade spondylolisthesis in young patients (Crawford et al. 2017). Such studies are difficult to conduct since high-grade spondylolisthesis is a very uncommon condition (see section 5.2, p. 37 for a note on epidemiology and study design). There are only a few long-term retrospective comparative studies on surgical reduction versus fusion in situ. Boxall et al. (1979), Burkus et al. (1992), Molinari et al. (1999), and Martiniani et al. (2012) are advocates of reduction while Poussa et al. (1993) and Muschik et al. (1997) argue for in situ fusion.

The purpose of this thesis is to evaluate the long-term outcome of in situ fusion for high-grade spondylolisthesis with focus on health-related quality of life (HRQoL), sagittal balance and clinical appearance. The thesis aims to address the concern that fusion in lumbosacral kyphosis might cause long-term problems such as pain or deterioration in HRQoL when degenerative loss of sagittal alignment adds to the original sagittal deformity.
1.3 Sagittal balance

Although good outcomes have been reported 10 to 25 years after in situ fusion for high-grade spondylolisthesis (Johnson and Kirwan 1983; Harris and Weinstein 1987; Freeman and Donati 1989; Seitsalo et al. 1990, 1992; Grzegorzewski and Kumar 2000; Helenius et al. 2006; Poussa et al. 2006; Remes et al. 2006; Lamberg et al. 2007; Helenius et al. 2008) data are limited on patients entering the age interval when the effects of degenerative changes may be clinically more pronounced. Since in situ fusion does not address the lumbosacral kyphosis, there might be some long-term sagittal balance issues when degenerative disk changes, in addition to the original kyphotic deformity, alter the sagittal alignment of the spine. Postural compensatory mechanisms that reduce anterior sagittal malalignment, for example, thoracic hypokyphosis, pelvic retroversion, and knee flexion (Barrey et al. 2013), are commonly observed in high-grade spondylolisthesis (Turner and Bianco 1971; Johnson and Kirwan 1983; Vialle et al. 2007). Consequently, individuals who undergo in situ fusion for high-grade spondylolisthesis may have a limited ability to further compensate for the added degenerative loss of sagittal alignment, for example, by further reducing the kyphosis of a hypokyphotic thoracic spine.

The sagittal balance is evaluated on standing lateral radiographs of the spine and pelvis by using the following steps (Roussouly and Pinheiro-Franco 2011; Barrey et al. 2013): First, the global sagittal balance is analyzed by measuring, for example, the sagittal vertical axis (SVA) (Jackson and McManus 1994) or the T1 spinopelvic inclination (T1SPI) (Fig. 1) (Lafage et al. 2009). Second, compensatory mechanisms such as reduction of thoracic kyphosis or pelvic retroversion are analyzed by measuring pelvic parameters and sagittal curves. The SVA is defined as the perpendicular distance from the C7 plumb line to the posterosuperior corner of S1 (Fig. 1A). This parameter has frequently been used for evaluation of global sagittal balance since several authors have observed that health status scores become poorer with increasing SVA (i.e., forward
shift of the C7 plumb line) (Glassman et al. 2005; Lafage et al. 2009). Some authors, however, have questioned the quantitative relationship between radiographic sagittal alignment and patient-reported outcome (Angevine et al. 2019). An alternative parameter for evaluation of global sagittal balance is the T1SPI (the angle between the T1 plumb line and a line drawn from the center of T1 to the center of the bicoxofemoral axis). Lafage et al. (2009) observed that the T1SPI was the parameter to have the highest correlation with self-reported HRQoL when analyzing over 100 spinopelvic parameters. For sagittal balance, a rule of thumb is that SVA should be < 50 mm. T1SPI should be < 0° (the T1 plumb line falls behind the center of the bicoxofemoral axis, Fig. 1B) (Schwab et al. 2010). T1SPI has the advantage of being an angular measure that does not require calibration of the radiographs (Schwab et al. 2009). Moreover, the T1SPI is independent of the shape of the sacrum. Anatomical landmarks of the sacrum may be difficult to identify because of the convex form of the upper end-plate of S1 often found in high-grade spondylolisthesis (cf. point 2, p. 40).

**Figure 1:** (A) The sagittal vertical axis (SVA). (B) Sagittal balance: T1SPI<0°. (C) Sagittal imbalance: T1SPI≥0°. (D) The thoracokyphosis (TK) and lumbar lordosis (LL).
It is also important to know that SVA might be affected by small changes in the position of the upper (Marks et al. 2003) or lower extremities (van Royen et al. 1998) which means that standardized positioning of the patient is crucial to minimize the potential measurement error.

There are conflicting findings concerning the global sagittal balance in high-grade spondylolisthesis. Mac-Thiong et al. (2008) found no major difference in sagittal balance between young patients with high-grade spondylolisthesis and individuals with no spinal deformity. In contrast, Vialle et al. (2007) found a significant difference in T1SPI between young patients with high-grade spondylolisthesis and individuals without spinal deformity. The mean T1SPI difference, however, was < 3° and the mean T1SPI was < 0° for both groups. Jackson et al. (2003) described a substantial difference in SVA between adults with spondylolisthesis and individuals with no spinal deformity. Harroud et al. (2013) found a major difference in SVA between high-grade and low-grade spondylolisthesis.

The pelvic parameters used in the analysis of compensatory mechanisms are the pelvic incidence (PI), the pelvic tilt (PT), and the sacral slope (SS) (Fig. 2A) (During et al. 1985; Duval-Beaupère et al. 1992; Legaye et al. 1998). PI is often considered to be a constant anatomic parameter (Li and Hresko 2012), unique in each individual, independent of pelvic rotation, while PT and SS depends on pelvic rotation (Legaye et al. 1998). Place et al. (2017) and Schroeder et al. (2018), however, observed a minor variation in PI with pelvic rotation suggesting a possible motion in the sacroiliac joints. The pelvic parameters are related as PI = PT + SS (Fig. 2B) (Legaye et al. 1998). PT is used to quantify pelvic retroversion while PI determines the maximum capacity of pelvic retroversion (Diebo et al. 2015). PT increases with age (Schwab et al. 2006; Diebo et al. 2015). The thoracic kyphosis (TK) (Fig. 1D) is important since younger individuals, with a flexible spine, tend to use reduction of TK as the primary compensatory mechanism while older individuals use pelvic retroversion (Barrey et al. 2013; Diebo et al. 2015). Also, the lumbar lordosis (LL) (Fig. 1D) is important since loss of LL with increasing age causes the C7 plumb line to shift forward with age (Gelb et al. 1995).
Figure 2: (A) Definitions of pelvic incidence (PI), pelvic tilt (PT) and sacral slope (SS). The line PL is the plumb line. (B) Since line X1 is perpendicular to line X3 and line X2 is perpendicular to line X4 it follows that angle A1 equals SS. Line X4 is parallel to line PL and thus angle A2 equals PT. Since A1 + A2 = PI it follows that PI = PT + SS.

Presence of compensatory mechanisms is used to classify the sagittal balance into one of three stages (Barrey et al. 2013): (1) Balanced, which means that there is no forward shift of the C7 plumb line and no compensatory mechanisms are needed, (2) compensated (balanced with compensatory mechanisms), that is, no forward shift of the C7 plumb line but with compensatory mechanisms (e.g., pelvic retroversion or reduced TK), and (3) unbalanced which means that the compensatory mechanisms are not enough efficient to avoid forward shift in C7 plumb line.

For spondylolisthesis, Hresko et al. (2007) introduced the concept of pelvic balance. Based on the values of pelvic tilt and sacral slope, the pelvis is classified as balanced or retroverted according to the formula: balanced if SS > 0.84PT + 25, retroverted otherwise. Hresko et al. (2007) hypothesized, based on a cross-sectional comparison of 133 individuals with high-grade spondylolisthesis, that patients with high-grade spondylolisthesis and an unbalanced pelvis might benefit from a reduction procedure. The concept of pelvic balance has made a large impact on the theories of the biomechanics of high-grade spondylolisthesis. The widely used Spinal deformity study group (SDSG) classification and
treatment recommendation of spondylolisthesis (Labelle et al. 2011), is more or less based on the hypothesis of Hresko et al. (2007). There are reports that restoration of pelvic balance is important when performing reduction procedures (Mac-Thiong et al. 2019; Alzakri et al. 2019).

There are several papers on normative values for the pelvic parameters and the sagittal curves (e.g., Fon et al. 1980; Stagnara et al. 1982; Hansson et al. 1985; Bernhardt and Bridwell 1989; Gelb et al. 1995; Legaye et al. 1998; Vedantam et al. 1998; Jackson et al. 2000; Vialle et al. 2005; Boulay et al. 2006; Roussouly et al. 2006; Schwab et al. 2006; Mac-Thiong et al. 2010; Been et al. 2010) but there is a certain diversity in patient positioning (supine or upright, with or without arm rests, etc.) and angle definitions. Significant differences may occur when different number of vertebral segments are measured. For example, Been and Kalichman (2014) observed that the lumbar lordosis increased > 10° when the angle was measured from the upper end-plate of L1 to the upper end-plate of S1 (as in Fig. 1D) instead of the upper end-plate of L1 to the lower end-plate of L5. Thus, in comparisons of different studies this fundamental fact must be kept in mind. For subjects aged 41 to 60 years, Schwab et al. (2006) reports the following normative values when the subjects are positioned as described by Horton et al. (2005): PT 5-25°, SS 25-55°, PI 35-70°, L1-S1 LL 45-75°, and T4-T12 TK 20-55°. Mac-Thiong et al. (2011) reports the following ratios: PT/PI 0.24 (SD 0.11) and SS/PI 0.76 (SD 0.11).

The anatomic pelvic parameter PL is increased in spondylolisthesis compared with controls (Hanson et al. 2002; Marty et al. 2002; Labelle et al. 2004; Vialle et al. 2007). There are also several reports that PI is correlated to an increased slip (Hanson et al. 2002; Labelle et al. 2004; Harroud et al. 2013). Since PI is an anatomic parameter, PI is unaffected by reduction procedures (Labelle et al. 2008; Martiniani et al. 2012). Hresko et al. (2009), however, found an increase in PI after reduction of the slip and lumbosacral kyphosis. The authors hypothesized that the increase in PI was due to sacral remodeling with growth or that there was motion at the SI joint that lead to flexion of sacrum in the ilium.
Furthermore, several authors have reported that reduction procedures reduce PT (Labelle et al. 2008; Hresko et al. 2009; Martiniani et al. 2012). In contrast, Mac-Thiong et al. (2019) and Alzakri et al. (2019) found only minor changes in PT after reduction. Hresko et al. (2009) noted that there was poor correlation between the amount of surgical correction of the spondylolisthesis and changes in pelvic retroversion.

PI has a relationship to LL of importance when discussing surgical realignment of adult sagittal imbalance. Surgical realignment should attempt to obtain PI-LL < 9° to assure appropriate lordotic alignment (i.e., avoid iatrogenic flatback) (Schwab et al. 2009, 2010). Moreover, Schwab et al. (2013) observed that adult spinal deformity patients with PI-LL > 11° were more likely to have pelvic retroversion or a global sagittal malalignment. PI-LL, SVA, and PT constitutes the foundation of the SRS-Schwab adult spinal deformity classification system (Schwab et al. 2012).

1.4 Adjacent segment disk degeneration

Adjacent segment disk degeneration (ASD) after spinal fusion surgery has been a topic of concern for many years. It has been hypothesized that increased rigidity across previously mobile segments and transfer of biomechanical forces to disk spaces adjacent to the fusion mass may be responsible for ASD after spinal fusion (Schoenfeld 2011). The mean annual incidence for revision surgery for ASD after spinal fusion has been estimated to 2.5% (Sears et al. 2011). Nevertheless, the clinical relevance of radiographic ASD remains unclear. The 10-year follow-up study of Okuda et al. (2018) evaluated 128 of 205 patients who underwent single segment posterior lumbar interbody fusion for L4 degenerative spondylolisthesis and found radiographic ASD in 75% of the patients, symptomatic ASD (leg pain or neurogenic claudication) was found in 31% of the patients and 15% went through revision surgery because of symptomatic ASD. Conversely, several authors have reported that the prevalence of radiographic ASD after spinal fusion is similar to that of conservatively treated controls or normal population (Seitsalo et al. 1997;
Hambly et al. 1998; Wai et al. 2006). Seitsalo et al. (1997) reported no long-term correlation between disk degeneration on radiographs and pain for young patients who underwent fusion surgery for low-grade spondylolisthesis. The randomized controlled trial (RCT) of Ekman et al. (2009a) showed some limited signs of radiographic ASD but with no correlation to outcome 12 years after posterolateral fusion for low-grade isthmic spondylolisthesis. In addition, Endler et al. (2019) could not find any association between radiographic ASD and patient reported outcome when extending the Ekman et al. (2009a) material with a cohort of patients with low-grade isthmic spondylolisthesis (Ekman et al. 2007) treated with posterior lumbar interbody fusion. Similarly, in adult degenerative disk disease, Mannion et al. (2014) reported the results of a combined follow-up of 4 RCTs and found that spinal fusion (with or without instrumentation) is associated with increased radiographic ASD but without influence on clinical outcome.

1.5 Nonunion

Nonunion after fusion surgery for high-grade spondylolisthesis may cause pain, lack of benefit of the surgery and in some cases reoperation (Muschik et al. 1997; Molinari et al. 1999). In a systematic review comparing in situ fusion with reduction and fusion, Longo et al. (2014) found, based on pooled data, that nonunion occurred in of 17.7% of the patients fused in situ compared with 5.5% after reduction and fusion. A weakness of the Longo et al. (2014) review was that the study groups were heterogeneous with respect to type of fusion (anterior, posterolateral, and circumferential). Previous studies have shown that circumferential fusion is superior to posterolateral fusion with respect to nonunion. The long-term follow-up by Lamberg et al. (2007) reported a fusion rate of 86% for posterolateral fusion, 100% for anterior fusion and 96% for circumferential fusion. It should, however, be noted that not all cases of nonunion causes pain. Poussa et al. (1993) found 3 cases of nonunion in a series of 11 patients where all were symptom-free and reoperation was considered unnecessary. Computed tomography scans are by most
authors considered the gold standard for assessing fusion. Other methods include analysis of segmental movement on flexion and extension radiographs (Seitsalo et al. 1992) or evaluation of fusion masses (Boxall et al. 1979; Frennered et al. 1991; Lenke et al. 1992).

1.6 Assessment of health-related quality of life

There are several instruments for the assessment of HRQoL in patients with lumbar spine disorders (Zanoli et al. 2000). This section provides a very short background to the instruments used in this thesis.

The Medical outcomes study 36-item short-form health survey (SF-36) is an 8-dimension (also called 8-scale), 36-item, self-administered HRQoL instrument for the assessment of general HRQoL (Ware and Sherbourne 1992). The results are presented as a health profile (see Fig. 3, p. 27) where the score for each dimension range from 0 to 100 (0 being the worst and 100 the best). There are 2 identical versions of the instrument, one licensed version (SF-36) (Ware and Sherbourne 1992) and one free version (RAND-36) (Hays et al. 1993). There are validation data for the Swedish translations for both versions (Sullivan et al. 1995; Orwelius et al. 2018). Long-term SF-36 data after surgical treatment of high-grade spondylolisthesis was reported by Helenius et al. (2008).

The EuroQol 5 dimensions instrument 3-level version (EQ-5D-3L) is a 5-dimension, 5-item, self-administered HRQoL instrument for the assessment of general HRQoL (EuroQol Group 1990). The dimensions are: mobility, self-care, usual activities, pain and anxiety. Each item has 3 possible answers, coded on the ordinal scale 1 to 3 (1 being the best and 3 the worst). The answers are assembled to a 5-digit health state indicating the score on each dimension (in total $3^5 = 243$ states, 11111 being the best and 33333 the worst). The health states can be converted to a summary index between 0 (states equal to death) and 1 (full health). A Swedish tariff for calculating the summary index was reported by Burström et al. (2014). Before that, Swedish studies often used the UK tariff (Dolan 1997). Diarbakerli et al. (2017) reported similar index
scores when using the UK and Swedish tariffs while Kiadaliriet al. (2015) reported that the Swedish tariff gave higher scores than the UK tariff. This means that EQ-5D indices calculated with different tariffs might not be fully comparable. EQ-5D index values for Swedish population data are reported by Burström et al. (2001). EQ-5D is commonly used for economic evaluations of health care (cf. Burström et al. 2003).

Previous studies have shown that the EQ-5D-3L survey may suffer from ceiling effects (≥ 15% of the respondents achieved the highest possible score, Terwee et al. 2007). To improve the measurement properties, the EuroQol Group has developed a 5-level version of EQ-5D (EQ-5D-5L) (Herdman et al. 2011) which reduces the ceiling effects (Janssen et al. 2013; Greene et al. 2015). The EQ-5D-5L survey was not available when the work on this thesis started. The EQ-5D instrument also has a second part (EQ VAS), a 20 cm vertical VAS, graded 0-100, (0 being the worst imaginable health state and 100 being the best imaginable health state) for assessment of general health. Health assessment with one single question has gained interest in recent years because there are reports of association between single item, self-rated, general health assessment and mortality (DeSalvo et al. 2006).

The Oswestry disability index (ODI) is a single dimension, 10-item, self-administered instrument for assessment of disability (pain, personal care, ability to walk, etc.) (Fairbank and Pynsent 2000). Each item has 6 possible answers coded on the ordinal scale 0 to 5 (0 being the best and 5 the worst). Results are reported as (total score)×100/(5×number of questions answered) and may be interpreted as follows: 0-20 minimal disability, 20-40 moderate disability, 40-60 severe disability, 60-80 crippled and 80-100 bed-bound or exaggerating symptoms (Fairbank et al. 1980). The mean score in the normal population has been estimated to 10.2 (SD 2.2-12) (Fairbank and Pynsent 2000).

The Zung depression scale (ZDS) is a single dimension, 20-item self-administered instrument for assessment of depression (Zung 1965). Each item has 4 possible answers coded on the ordinal scale 1 to 4 (1 being
the best and 4 the worst). Zung (1965) found that patients diagnosed with depressive disorder had the mean total score 59 (range 50-72).

The Million score is a single dimension, 15-item, self-administered instrument for assessment of back pain (Million et al. 1982). For each item, the respondents grade the answer on a 10 cm vertical VAS. The result is reported as the sum of the VAS scores (0 to 150). Anagnostis et al. (2003) suggested the following interpretation of the Million score: 0 no disability, 1-40 mild disability, 41-70 moderate disability, 71-100 severe disability, 101-130 very severe disability and 131-150 extreme disability.

The Scoliosis Research Society questionnaire 22r (SRS-22r) was used in this thesis for assessment of HRQoL and self-image. SRS-22r is described in section 1.7.

HRQoL outcome correlates with work status (disability pension, sick leave, etc.) (Sullivan and Karlsson 1998). Moreover, work status has been reported as a predictor of surgical outcome of spine surgery (Anderson et al. 2006; Ekman et al. 2009b; Khan et al. 2019). Several authors have reported long-term work status data after treatment of spondylolisthesis. Seitsalo et al. (1990) reported a 3% (3/87) disability pension rate 14 years after in situ fusion for high-grade spondylolisthesis. Harris and Weinstein (1987) reported that 1 of 21 patients had missed time from work because of back pain 24 years after in situ fusion for high-grade spondylolisthesis.

1.7 Self-image and the SRS-22r questionnaire

Certain clinical findings, for example, short trunk, transverse abdominal folds, flat back, and lumbar step-off are often seen in patients with high-grade spondylolisthesis (Turner and Bianco 1971; Harris and Weinstein 1987; Hensinger 1989). Since treatment with in situ fusion does not address clinical appearance, self-image, and also HRQoL might be negatively affected in the short-term as well as the long-term perspective. Interestingly, previous long-term evaluations have reported that few patients fused in situ for high-grade spondylolisthesis complain about
their appearance (Johnson and Kirwan 1983; Harris and Weinstein 1987; Freeman and Donati 1989; Seitsalo et al. 1990; Grzegorzewski and Kumar 2000). Few studies, however, have evaluated self-assessed clinical appearance and self-image after in situ fusion for high-grade spondylolisthesis in relation to healthy controls or normative population data (Gutman et al. 2017).

The Scoliosis Research Society 22r questionnaire (SRS-22r), is a self-administered HRQoL instrument, originally developed and validated for patients with idiopathic scoliosis (Haen et al. 1999; Asher et al. 2003, 2006). SRS-22r has 5 dimensions (called domains): function (5 items), pain (5 items), self-image (5 items), mental health (5 items), and satisfaction with management (2 items). Each item has 5 possible answers coded on the ordinal scale 1 to 5 (1 being the worst and 5 the best). Results are presented as the mean value of each domain respectively, the mean value of all domains (SRS-22r total score) and the mean value of the total score less the 2 satisfaction items (SRS-22r subscore). Specific for SRS-22r compared with other general HRQoL instruments such as EQ-5D-3L (EuroQol Group 1990) or SF-36 (Ware and Sherbourne 1992), and of particular interest for this thesis, is the 5-item domain for assessment of clinical appearance and self-image (cf. Table 3, Paper IV). Although originally intended for assessment of scoliosis, the instrument has no scoliosis specific items. There are, however, spine deformity specific items which limits the usability of the instrument to assessment of spine deformity patients (cf. point 10, p. 42).

Several versions of the instrument have been used to evaluate outcome after surgical treatment of high-grade spondylolisthesis (Boachie-Adjei et al. 2002; Helenius et al. 2006; Poussa et al. 2006; Jalanko et al. 2011; Bourassa-Moreau et al. 2013; Harroud et al. 2013; Lundine et al. 2014; Gutman et al. 2017; Mac-Thiong et al. 2018, 2019; Alzakri et al. 2019; Bourassa-Moreau et al. 2019). The most recent version of the instrument is SRS-22r (Asher et al. 2006). There are two Swedish validation studies on SRS-22r. The study of Danielsson and Romberg (2013) reported validation data for a cohort of scoliosis patients aged 12 to 57 years.
Diarbakerli et al. (2017) reported validation data for a cohort selected from the general Swedish population aged 10 to 69 years. The Swedish version of SRS-22r is not validated for spondylolisthesis. The French-Canadian version (SRS-22fv) is validated for adolescent patients with spondylolisthesis (Gutman et al. 2017). Validation data on adults with spondylolisthesis are lacking.
2 Aims

The overall aim of the thesis was to evaluate the long-term results of in situ fusion for high-grade isthmic spondylolisthesis in young patients. The specific aims were:

1. To study the long-term outcome in terms of clinical findings, pain, function, work status, and HRQoL after in situ fusion for high-grade isthmic spondylolisthesis (I).

2. To study long-term effects on sagittal balance (II) after in situ fusion for high-grade isthmic spondylolisthesis.

3. To study local morphological changes in terms of ASD (III) after in situ fusion for high-grade isthmic spondylolisthesis.

4. To study self-image aspects due to the deformity after in situ fusion for high-grade isthmic spondylolisthesis (IV).
3 Patients and methods

3.1 Patients (I-IV) and controls (IV)

From 1973 to 1985, 40 consecutive patients (30 females and 10 males), mean age 14 (range 9-24, SD 4) years, were operated on with in situ fusion for high-grade isthmic spondylolisthesis at the L5-S1 level. The patients were recruited from an area in western Sweden populated by approximately 1.9 million inhabitants (December 31, 1979) (National central bureau of statistics 1980). The indications for fusion were severe displacement, low back pain, radicular pain, or slip progression. Only minor neurologic deficiencies were found preoperatively.

The long-term follow-up evaluation was made on 2 occasions. Table 1 summarizes the differences in mean duration of follow-up, mean age at follow-up, etc. At the time of the evaluations, one patient had died and the remaining 39 patients were at both occasions invited to participate in the investigations. An 8-year follow-up of the same cohort was reported by Frennered et al. (1991).

Table 1: Characteristics of the study population.

<table>
<thead>
<tr>
<th></th>
<th>Paper I</th>
<th>Paper II</th>
<th>Paper III</th>
<th>Paper IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Follow-up duration, years*</td>
<td>28.5 (3.8)</td>
<td>33.9 (4.0)</td>
<td>33.9 (3.7)</td>
<td>34.0 (3.7)</td>
</tr>
<tr>
<td>Age at follow-up, years*</td>
<td>43.5 (4.5)</td>
<td>48.8 (4.7)</td>
<td>48.7 (4.6)</td>
<td>48.8 (4.9)</td>
</tr>
<tr>
<td>Follow-up rate</td>
<td>35/39</td>
<td>28/39</td>
<td>34/39</td>
<td>38/39</td>
</tr>
<tr>
<td>Female/male</td>
<td>27/8</td>
<td>21/7</td>
<td>26/8</td>
<td>30/8</td>
</tr>
<tr>
<td>Meyerding grade 3/4/5</td>
<td>17/14/4</td>
<td>12/12/4</td>
<td>16/14/4</td>
<td>17/17/4</td>
</tr>
</tbody>
</table>

* The values are given as the mean, with the standard deviation in parentheses.

Seventy-six age and gender matched healthy controls (2 controls for each case) were selected from a Swedish reference sample of healthy individuals (IV) (Diarbakerli et al. 2017). The controls completed the Swedish version of SRS-22r (Danielsson and Romberg 2013) but were not clinically examined.
3.2 Surgical treatment (I-IV)

All 40 patients underwent an uninstrumented posterolateral intertransverse L4 to S1 in situ fusion except for 6 patients in whom Harrington (1962) distraction rods were also used (the rods were removed in all but 1 patient after 8 to 48 months). Thirty-eight patients also had an anterior fusion with a transvertebral L5 to S1 iliac autograft dowel similar to the technique described by Whitecloud and Butler (1988). Thirty-three patients had a Gill et al. (1955) laminectomy. After 1 week of bed rest, the patients were mobilized in a double pantaloon spica cast for 3 months and then in a Boston brace for 3 months. No patient required a second procedure to achieve fusion. In Paper I the surgical treatment was described as similar to the Cloward (1958) technique for cervical fusion but this was changed in Papers II-IV since the Cloward disk approach might give the impression that the kyphosis was reduced which was not the case.

3.3 Health-related quality of life (I,IV)

The patients completed SF-36 (Ware and Sherbourne 1992), EQ-5D-3L (UK tariff) and EQ VAS (EuroQol Group 1990; Dolan 1997), ZDS (Zung 1965), ODI (Fairbank and Pynsent 2000), a back and leg pain VAS (0 mm representing no pain and 100 mm representing maximum pain) (Huskisson 1974), and the Million score (Million et al. 1982). Work status (heavy work, unemployment, sick leave, disability pension) was documented. All male patients were asked if they had retrograde ejaculation.

The SF-36 results were compared with the Swedish normative data of Sullivan et al. (1995). The EQ-5D-3L results were compared with age-matched general Swedish population data (Burström et al. 2001). The proportion of patients at work was compared with age-matched population data from official statistics of Sweden (Statistics Sweden 2010). The Million score results were compared with data from the 8-year follow-up of the same cohort (Frennered et al. 1991).
The patients also answered SRS-22r (Danielsson and Romberg 2013). The SRS-22r results were compared with the results of an age and gender matched control group of healthy volunteers (p. 19).

### 3.4 Physical examination (I)

The patients went through a physical examination using the same protocol as the 8-year follow-up (Frennered et al. 1991). The protocol measured clinical parameters such as length, weight, finger-to-floor distance, neurologic examination of the lower extremities by conventional clinical methods, and pain behavior by the signs described by Waddell et al. (1980). The clinical parameters were compared with the data of Frennered et al. (1991).

### 3.5 Nonunion (I)

All fusion sites were evaluated concerning solidity on standard supine radiographs. A posterolateral fusion was defined as solid when bridging trabecular bone was present. An anterior fusion was considered solid when the graft was bridging the intervertebral space and homogeneously included both vertebral bodies (Frennered et al. 1991).

### 3.6 Sagittal balance (II)

Standing lateral radiographs of the lumbar spine (not including the femoral heads) taken at mean 8 years after surgery and standing lateral radiographs of the spine and pelvis (fists on clavicles position, Faro et al. 2004; Horton et al. 2005) taken at mean 33 years after surgery were used in the evaluation of sagittal balance. Eight-year radiographs were missing in 5 of 28 patients.

Global sagittal balance was evaluated with the sagittal vertical axis (SVA) (Jackson and McManus 1994) and the T1 spinopelvic inclination (T1SPI) (Lafage et al. 2009). The standard definitions of Legaye et al. (1998) for the pelvic parameters pelvic tilt (PT), pelvic incidence (PI), and sacral
slope (SS) were used. The pelvic parameters were all measured, that is, not derived by the formula PI = SS + PT (Legaye et al. 1998). Based on the values of PT and SS, the pelvis was classified as balanced or retroverted according to the formula of Hresko et al. (2007): balanced if SS > 0.84PT + 25, retroverted otherwise. Thoracic kyphosis (TK) and lumbar lordosis (LL) was measured as defined in Fig. 1D, p. 5.

Local compensatory mechanisms (retrolisthesis and hyperextension of adjacent segments) (Barrey et al. 2013; Zhu et al. 2017) were evaluated from T12 to L4 (above the fusion L4 to S1). Hyperextension of adjacent segments (i.e., disk wedging) was measured as the sagittal Cobb angle between 2 adjacent end-plates. Degenerative anterolisthesis and retrolisthesis from T12 to L4 was measured using the slip definition of Danielson et al. (1988). The T1 pelvic angle (TPA), was calculated using the formula TPA = T1SPI + PT (Protopsaltis et al. 2014).

On the standing lateral radiographs of the lumbar spine taken 8 years after surgery, only SS was measured. The posterior border of the roentgenogram was used as vertical reference that was widely accepted in the seventies and eighties (Boxall et al. 1979). PI and PT could not be measured since the femoral heads were not visible on the radiographs. Nor could the LL be measured since the upper end-plate of L1 was missing on several radiographs. Based on the values of the 8-year SS (denoted SS₈) and the 33-year PI (denoted PI₃₃), the pelvis was classified as balanced or retroverted (Hresko et al. 2007) (balanced if SS₈ > 0.84(PI₃₃ - SS₈) + 25).

Dichotomization was used to evaluate association between radiographic parameters and SRS-22r outcome: slip angle > 20° or ≤ 20°, SVA > 5 cm or ≤ 5 cm, TPA > 30° or ≤ 30°, PT/PI >0.4 or ≤ 0.4, and balanced or retroverted pelvis. After the publication of Paper II, the Dubousset (1997) lumbosacral angle (Dub-LSA) ≥ 80° has been suggested as a criterion for adequate reduction (Mac-Thiong et al. 2019). Therefore, the results of this thesis were extended with data on Dub-LSA ≥ 80° or < 80°. A dichotomization on age at surgery > or ≤ mean age at surgery (14 years) was also added.
3.7 Slip and slip angle (II)

The L5-S1 slip and the L5-S1 slip angle was measured on standard supine radiographs of the lumbar spine. For the L5-S1 slip measurement, the definition of Danielson et al. (1988) was used. The Boxall et al. (1979) L5-S1 slip angle was used to measure the lumbosacral kyphosis. Since the lower end-plate of L5 might be deformed in high-grade spondylolisthesis, also the Dub-LSA (Dubousset 1997) was measured. Data on preoperative L5-S1 slip were collected from the research archive at Sahlgrenska University Hospital.

3.8 Adjacent segment disk degeneration (III)

A commonly used method for measuring ASD on plain lateral radiographs is distortion compensated roentgen analysis (DCRA) (Frobin et al. 1997). DCRA has the advantage that the method compensates for variation in image magnification since the disk heights are measured as relative measures (percentages), that is, normalized to the vertebral body height. This thesis used a slightly modified version of DCRA independent of the shape of the inferior vertebra which is often deformed in high-grade spondylolisthesis (Ekman et al. 2009a). For this method, the mean intraobserver error is 2.1% for anterior disk height and 1.95% for the posterior disk height (Ekman et al. 2009a). The mean interobserver error is 10.6% for the anterior disk height and 3.4% for the posterior disk height (Ekman et al. 2009a).

Anterior and posterior disk heights were measured one and two segments above the fusion on standard supine radiographs of the lumbar spine taken 8 years and 29 years after surgery. The same measurements were made on standing lateral radiographs of the lumbar spine (not including the femoral heads) taken 8 years after surgery and on standing lateral radiographs of the spine and pelvis taken 33 years after surgery. Eight-year data were missing in 4 of 34 patients and 33-year data were missing in 7 of 34 patients.
Segmental lordosis (Jackson and McManus 1994) was measured on the standing radiographs one and two segments above the fusion. The measurement error for segmental lordosis has been estimated to 3° (Gelb et al. 1995).

Dichotomization (> or ≤ the mean) was used to evaluate association between radiographic parameters and SRS-22r outcome.

### 3.9 Self-image (IV)

The SRS-22r self-image domain was used to evaluate self-image (Danielson and Romberg 2013).

### 3.10 Statistical analysis (I-IV)

The statistical methods were chosen according to the guidelines of Altman (1991, ch. 9). Normality was checked by graphical exploration with frequency histograms and normal plots (Altman 1991, ch. 7.5.2). A p-value < 0.05 was considered to be significant and two-tailed tests were used. Student's t-tests for paired (Altman 1991, ch. 9.5.2) and unpaired (Altman 1991, ch. 9.6.2) data were used to compare normally distributed data with equal variances. The F-test was used to test variance equality (Altman 1991, ch. 9.6.5). The Mann-Whitney U test (Altman 1991, ch. 9.6.4) was used to compare unpaired non-normally distributed data and the Wilcoxon signed rank test (Altman 1991, ch. 9.5.3) was used to compare paired non-normally distributed data. Spearman's rank correlation coefficient ($r_s$) (Altman 1991, ch. 11.7.2) was used for evaluation of correlations between non-normally distributed variables. The strength of correlation was interpreted as follows: 0.10 to 0.29 small, 0.30 to 0.49 medium, and 0.50 to 1.00 large (Cohen 1988, ch. 3.2.1). Bland-Altman plots (Altman 1991, ch. 14.2.1) with ±2 standard deviations (SD) as limits of agreement were used to evaluate measurement variation. The SRS-22r function, pain, and mental health scores were non-normally distributed (skewed towards higher values) with unequal variances for cases and controls while the SRS-22r self-image data was considered to
be normally distributed with unequal variances. For case-control data in general the recommendation is to use a method for paired data (Altman 1991, ch. 9.5). For the SRS-22r data, however, there were two controls for each case, which means that a paired data method could not be used. The non-parametric Mann-Whitney U test (Altman 1991, ch. 9.6.4) was used for the SRS-22r data.

3.11 Ethical review board approval (I-IV)

The studies were approved by the regional ethical review board in Gothenburg, registration number 640-08 and T975-15 (I,II,III,IV), and the regional ethical review board in Stockholm, registration number 2012/172-31/4 (IV). Informed consent was obtained from all participants.
4 Summary of results

4.1 Health-related quality of life (I,IV)

The scores of the SF-36 (Fig. 3) and EQ-5D-3L were similar to the scores of the general Swedish population. The mean EQ VAS score was 83 (range 16-100), the mean ZDS score was 30 (range 20-52), the mean ODI score was 10 (range 0-34). The mean Million score was 28 (range 0-109) and was slightly worsened compared with 8-year data. The mean back pain VAS score was 13 (range 0-72), and the mean leg pain VAS score was 9 (range 0-60). The proportion of patients at work was the same as that for the age-matched general Swedish population. None of the male patients reported any retrograde ejaculation and all had children.

Figure 3: SF-36 scores three decades after in situ fusion for high-grade spondylolisthesis (squares) and scores of the general Swedish population reported by Sullivan et al. (1995) (n=8930, age range 15-93 years, mean age 42.7 years) (triangles). PF physical functioning, RP role limitation due to physical problems, BP bodily pain, GH general health, VT vitality, SF social functioning, RE role limitations due to emotional problems, MH mental health.
The SRS-22r pain and mental health scores of the patients were similar to the scores of controls. The SRS-22r function and self-image scores were slightly, but statistically significantly, worse in patients compared with controls (Table 2).

**Table 2**: SRS-22r results for patients (three decades after in situ fusion for high-grade spondylolisthesis) and controls matched for age and gender.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Patients</th>
<th>Controls</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function</td>
<td>4.3 (0.75)</td>
<td>4.6 (0.55)</td>
<td>0.04</td>
</tr>
<tr>
<td>Pain</td>
<td>4.1 (1.00)</td>
<td>4.5 (0.71)</td>
<td>0.08</td>
</tr>
<tr>
<td>Self-image</td>
<td>3.5 (0.82)</td>
<td>4.3 (0.51)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Mental health</td>
<td>4.1 (0.74)</td>
<td>4.1 (0.68)</td>
<td>0.7</td>
</tr>
<tr>
<td>Satisfaction with management</td>
<td>4.0 (0.89)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The values are given as the mean, with the standard deviation in parentheses.

There were no statistically significant differences in any SRS-22r domain between patients with grade 3 slip (n=17) and patients with higher than grade 3 slip (n=21). In addition, there were no statistically significant differences in any SRS-22r domain between men (n=8) and women (n=30). There were no statistically significant differences in any SRS-22r domain when SRS-22r data was dichotomized on age at surgery > or ≤ mean age at surgery (14 years).

For the 6 patients in whom Harrington distraction rods were used, the mean SRS-22r scores were as follows: function 4.4 (SD 1.28), pain 4.7 (SD 0.65), self-image 3.5 (SD 1.22), mental health 4.3 (SD 1.06), and satisfaction with management 4.2 (1.25).

The proportion of patients reporting the maximum score for the different domains of SRS-22r were: function 26%, pain 32%, self-image 0%, and mental health 11%. The corresponding proportions for the controls were: function 42%, pain 41%, self-image 13%, and mental health 9%. No patient or control reported the minimum score for any domain of SRS-22r.
4.2 Physical examination (I)

The patients had a mean body mass index of 25 kg/m$^2$ (range 20-39 kg/m$^2$) and a mean finger-to-floor distance of 5.3 cm (range 0-28 cm). Only minor neurologic deficiencies such as minor L4, L5, or S1 sensory disturbances on either lower limb were found. No patient had a positive result on the straight-leg-raising test (> 60°) or a positive Waddell test (three or more of the five signs).

4.3 Nonunion (I)

The radiographic evaluation showed solid fusion in all patients.

4.4 Postoperative complications (I)

Five patients had a superficial wound infection. One patient developed a cauda equina syndrome but recovered and was rated as having an excellent result at the 8-year follow-up (Frennered et al. 1991). One patient had aspiration pneumonia, one patient had postoperative bleeding from the anterior approach leading to urgent surgical intervention and one patient had transient numbness in the left lower limb.

4.5 Sagittal balance (II)

The mean values of the pelvic parameters and the spinal curves were PT 35° (SD 9°), SS 46° (SD 13°), PI 81° (SD 14°), PT/PI 0.44 (SD 0.10), LL 46° (SD 17°), and TK 19° (SD 10°). The mean SS 8 years after surgery was 49° (SD 12°). The difference in SS between 8-year and 33-year data was statistically significant (p = 0.04).

Three patients had T1SPI > 0° (i.e., sagittal imbalance, see Fig. 1C, p. 5). These three patients all had SVA > 10 cm. One additional patient had SVA > 10 cm but the T1SPI for this patient was < 0°. Six patients had SVA between 5 and 10 cm, all these patients had T1SPI < 0°. For one patient, the TPA was < 20° and for 16 patients the TPA was > 30°.
Only 6 of 28 patients had a balanced pelvis (SS > 0.84PT + 25). The 3 patients in whom T1SPI > 0° all had a retroverted pelvis.

The 3 patients with T1SPI > 0° had SRS-22r subscore 2.2, 4.5, and 4.8 respectively. There was no statistically significant difference in SRS-22r subscore in the following cases of dichotomization: slip angle > 20° or ≤ 20°, SVA > 5 cm or ≤ 5 cm, TPA > 30° or ≤ 30°, PT/PI > 0.4 or ≤ 0.4, balanced or retroverted pelvis, Dub-LSA ≥ 80° or < 80°, and age at surgery > or ≤ mean age at surgery. There were no statistically significant differences in T1SPI, SVA, TPA, or PT/PI when data was dichotomized on SRS-22r subscore > 4.0 and SRS-22r subscore ≤ 4.0. For the 5 patients (1 dropout) in whom Harrington distraction rods were used, 2 had T1SPI > 0°, all had a retroverted pelvis.

There was no statistically significant difference in SRS-22r subscore at long-term when the data was dichotomized on balanced or retroverted pelvis 8 years after surgery.

Nine adjacent anterolistheses (T12 to L4) were found in 8 patients 33 years after surgery, none of these anterolistheses were present 8 years after surgery. One patient had an L1-L2 retrolisthesis that was present at both 8 and 33 years after surgery. All listheses were < 25%. Hyperextension (> 10°) of adjacent segments was not found more frequently at long-term as compared with the 8-year follow-up.

SRS-22r data was available for 10 of the 11 patients that did not participate in the evaluation of sagittal balance. There were no statistically significant differences in any SRS-22r domain for these 10 patients compared with the 28 participants in the evaluation of sagittal balance.

### 4.6 Slip and slip angle (II)

The mean L5-S1 slip was 74% (SD 19%), the mean slip angle was 25° (SD 18°), and mean Dub-LSA was 79° (SD 14°). Preoperative L5-S1 slip was 66% (SD 16%). The difference in L5-S1 slip between preoperative and long-term data was not statistically significant (p = 0.11).
4.7 Additional operations (II)

The last follow-up revealed that one patient had an additional operation 40 years after the primary operation. The patient now had an additional instrumented posterolateral fusion from L3 to L4, that is, one level above the original fusion from L4 to S1. Review of the patient’s radiographs showed that the original fusion from L4 to S1 was solid. A preoperative MRI from the last surgery revealed a bulging disk at the L3-L4 level and a relative central spinal stenosis. The indication for the additional operation was low back pain.

4.8 Adjacent segment disk degeneration (III)

One segment above fusion the anterior disk height significantly decreased on standing radiographs. Two segments above the fusion the anterior disk height significantly decreased on supine as well as on standing radiographs. The largest reduction was found two segments above the fusion where the disk height was reduced from 33% to 28% of anterior vertebral height between the measurements 8 years and 33 years after surgery. There were no statistically significant decreases in posterior disk heights in any measurement. \( |r_s| < 0.2 \) (\( p > 0.05 \)) for the correlations between anterior disk height reduction (one and two segments above the fusion, supine and standing radiographs) and all domains of SRS-22r. There was no statistical significant difference in any SRS-22r domain when the SRS-22r data was dichotomized, based on anterior disk height reduction (\( > \) or \( \leq \) the mean), one and two segments above the fusion on supine and standing radiographs (i.e., in total 4 cases of dichotomization). There were no statistical differences in disk heights when disk height data was dichotomized on pelvic balance as defined by Hresko et al. (2007). The anterior disk height showed a moderate negative correlation to PT. The decrease in segmental lordosis one segment above the fusion was \( < 3^\circ \). There was no decrease in segmental lordosis two segments above the fusion. Measurements of disk height on supine and standing radiographs showed good agreement and no bias.
4.9 Self-image (IV)

The SRS-22r self-image scores were statistically significantly lower than the scores of the controls (Table 2, p. 28). For all 5 questions of the self-image domain the scores of the patients were statistically significantly lower than the scores of the controls (Table 3, Paper IV). There were no statistically significant differences in SRS-22r self-image scores between patients with grade 3 slip (n=17) and patients with higher than grade 3 slip (n=21). There were no statistically significant differences in SRS-22r self-image scores between men (n=8) and women (n=30).
5 General discussion

5.1 Long-term issues with in situ fusion?

Surgical treatment with fusion in situ does not address the sagittal deformity in high-grade spondylolisthesis. The work with this thesis was driven primarily by a concern that degenerative loss of sagittal alignment, in addition to the original sagittal deformity, could cause long-term problems such as pain or deterioration in HRQoL if the original deformity was left unaddressed.

Clinical findings were reported close to three decades after surgery (I). Only minor neurologic deficiencies were found at the clinical examination and back/leg pain VAS scores were low. These findings confirmed the results of previous evaluations of in situ fusion, e.g. the Finnish studies (Seitsalo et al. 1990; Poussa et al. 2006; Lamberg et al. 2007). The proportion of patients at work was the same as that for the age-matched general Swedish population and was also similar to that reported by Seitsalo et al. (1990). Six validated instruments were used for evaluation of HRQoL, function, and mental health (EQ-5D-3L, SF-36, SRS-22r, Million score, ODI, ZDS) (I,IV). All instruments confirmed good long-term outcome. The widely used gold standard HRQoL instruments SF-36 and EQ-5D-3L gave results similar to Swedish normative data (Sullivan et al. 1995; Burström et al. 2001). The findings were consistent with previous long-term studies of in situ fusion (Johnson and Kirwan 1983; Harris and Weinstein 1987; Freeman and Donati 1989; Seitsalo et al. 1990, 1992; Grzegorzekowski and Kumar 2000; Helenius et al. 2006; Poussa et al. 2006; Remes et al. 2006; Lamberg et al. 2007; Helenius et al. 2008). The thesis extends the results of these previous studies to cover middle age where degenerative changes are superimposed on deformity driven changes.

Solid fusions were reported in all patients based on plain radiographs (I). Since fusion assessment was not a primary goal of the thesis, computed tomography scans were not performed. No statistically significant slip
progression was found, i.e. indirect signs of nonunion were lacking (II). These findings are consistent with those of Lamberg et al. (2007) and confirms the low nonunion rate and low slip progression after circumferential in situ fusion at a young age.

Previous studies have reported that cauda equina syndrome is a potential complication associated with in situ fusion for high-grade spondylolisthesis (Maurice and Morley 1989; Schoenecker et al. 1990). This thesis reports a transient postoperative cauda equina syndrome in one patient (I). There are also reports of cauda equina syndromes after reduction procedures (DeWald et al. 1981; Bradford and Boachie-Adjei 1990; DeWald et al. 2005). The pathogenesis of the cauda equina syndrome in case of in situ fusion is unclear. Several mechanisms have been proposed e.g. slip progression because of peroperative destabilization of the lumbosacral junction (stripping of ligaments, muscle splitting, etc.), additional flexion of the lumbosacral junction due to patient positioning on the operation table (positional impingement as the cauda equina is tethered over the dome of the sacrum or under the lamina of the slipped vertebra) or mechanical trauma during decortication (Maurice and Morley 1989; Schoenecker et al. 1990).

The main findings of the evaluation of sagittal balance were that non-compensated sagittal imbalance were observed in only a few individuals and that no association between any radiographic parameter and HRQoL was found (II). Compensated sagittal balance was the norm. Moreover, except for a slight decrease in SS and reduced TK, the evaluation could not demonstrate any signs of an increased need for compensatory changes such as retrolistheses or hyperextension of adjacent segments when comparing the 8-year and the 33-year radiographs (II). This might be an explanation for the good outcome in terms of HRQoL. The patients adapted to the abnormal lumbosacral kyphosis at a young age, and the pelvic retroversion does not have the same detrimental consequences in the long-term as it does in adult degenerative spinal deformity. This raises the question if timing of surgery, in relation to growth spurt, is of importance for HRQoL outcome. It may be hypothesized that an
immature spine is more adaptive to changes and that young age at surgery could be a positive prognostic factor for long-term HRQoL outcome. This hypothesis, however, could not be verified by this thesis since the dichotomization on age at surgery (section 4.1, p. 28) gave no difference in HRQoL outcome.

No association between any radiographic sagittal balance parameter and long-term SRS-22r outcome was found (II). This finding is partly consistent with the results of Mac-Thiong et al. (2019), who found no association between any radiographic parameter and SRS-22r pain, function, and total score. A systematic review of the Scoliosis Research Society evidence based medicine committee found that there was insufficient evidence in the literature to answer the question what radiographic parameters predict response to medical or interventional treatment of pediatric spondylolisthesis (Crawford et al. 2017).

The evaluation of sagittal balance (II) could not demonstrate any long-term benefits of a balanced pelvis. This finding questions the relevance of the Spinal deformity study group (SDSG) classification and treatment recommendation of spondylolisthesis (Labelle et al. 2011), which is more or less based on the hypothesis of Hresko et al. (2007) that pelvic balance should be restored. Contrary to the findings of this thesis, Alzakri et al. (2019) and Mac-Thiong et al. (2019) found a weak association between pelvic balance and SRS-22r self-image (but no association between pelvic balance and SRS-22r function, pain or mental health domains).

The study on ASD (III) found only moderate signs of ASD (reduction of the anterior disk height) using 3 different measuring techniques. Disk height reduction had no impact on HRQoL. The results support the concept that lower lumbar fusion does not accelerate symptomatic adjacent disk degeneration (Seitsalo et al. 1997; Ekman et al. 2009a; Mannion et al. 2014; Endler et al. 2019). There was no statistically significant reduction in posterior disk heights in any measurement. In contrast, Ekman et al. (2009a) found a more pronounced posterior disk height reduction. Ekman et al. (2009a) hypothesized that the loss of poste-
rior disk height may be a reflection of a compensatory hyperextension mechanism triggered by fusion in a kyphotic position and that fusion in kyphosis leads to compensatory cranial hyperlordosis and ASD. This hypothesis could not be verified by this thesis.

The current literature on clinical appearance after surgery for high-grade spondylolisthesis is inconclusive. Postoperative SRS-22r data for in situ fusion are lacking. The few long-term comparative studies on surgical reduction versus fusion (see section 1.2, p. 3) reported contradicting findings. Burkus et al. (1992) found that no patient complained of the cosmetic appearance or of persistence of abnormal posture or gait after the operation. Poussa et al. (1993) reported that the results were similar regarding pain, neurologic findings and patient satisfaction, except for cosmetic considerations where 4 of 11 patients of the reduction group (sic) were dissatisfied compared with none in the in situ group. The diverging results might reflect the pronounced subjectivity of cosmesis.

Based on the results of a prospective 2-year follow-up of 61 patients, age 8-21 years, Mac-Thiong et al. (2019) suggested that residual retroversion of the pelvis is particularly detrimental to self-image after surgery for high-grade spondylolisthesis. This thesis, however, could not find any association between pelvic retroversion or pelvic back tilt and SRS-22r outcome (II).

Interestingly, for the patients of this thesis, the lowest score of the self-image domain was reported for SRS-22r Question 10, a specific trunk appearance question (IV). This might reflect some disappointment with the typical clinical findings of high-grade spondylolisthesis, for example, a short trunk, flat back, or waistline skin folds. The satisfaction with the appearance of the trunk might also be negatively affected by the 10 to 15 cm abdominal and lumbar surgical scars.

The findings of this thesis all point in the same direction. Fusion in situ for high-grade spondylolisthesis has no obvious long-term shortcoming of importance irrespective of outcome measure. The concern that fusion of a young spine in lumbosacral kyphosis could cause problems appears
to be unfounded from a long-term functional and HRQoL perspective. To summarize, this thesis shows that there is no negative effect on HRQoL, sagittal balance, or disk degeneration above the fusion three decades after in situ fusion for high-grade spondylolisthesis. A minor negative effect on self-image was, however, observed.

5.2 Epidemiology and study design

This thesis evaluated 40 consecutive patients, that went through surgical treatment for high-grade spondylolisthesis at young age between 1973 and 1985. The patients were recruited from an area in western Sweden populated by approximately 1.9 million inhabitants (December 31, 1979) (National central bureau of statistics 1980). Given these numbers, the annual incidence for surgery for high-grade spondylolisthesis could be estimated to approximately 0.16 operations per 100 000 people. This illustrates that high-grade spondylolisthesis is a very uncommon condition and it will be difficult to power RCTs for the evaluation of surgical treatment options (in particular surgical reduction and fusion versus in situ fusion). Crawford et al. (2017) found no level 1 or level 2 evidence regarding the treatment of pediatric lumbar spondylolisthesis. As a result, surgical treatment of high-grade spondylolisthesis in young patients has to rely on level 3 evidence (retrospective cohort studies and case-control studies) and level 4 evidence (case series) (Marx et al. 2015) which, however, is not uncommon in the field of musculoskeletal medicine and surgery (Sangeorzan and Swiontkowski 2016). The level of evidence classification of this thesis is level 4.

5.3 Comment on gender

The female/male ratio of the study population was 30/10. This confirms previous observations that high-grade slippage is more common in girls (Boxall et al. 1979; Seitsalo et al. 1991; Lonstein 1999; Lamberg et al. 2007). The present thesis could not find any long-term differences in SRS-22r scores between females and males (IV).
5.4 Patient positioning in disk height measurements

There is no gold standard for the positioning of the patient in disk height measurements. For example, Seitsalo et al. (1997) and Mannion et al. (2014) were using standing lateral radiographs, Ekman et al. (2009a) and Endler et al. (2019) were using supine lateral radiographs and the normative material of Frobin et al. (1997) used a mix of supine and standing radiographs. The analysis of degree of agreement between measurements for supine and standing radiographs in the assessment of disk height reduction after fusion surgery showed good agreement when comparing measurements on supine and standing radiographs (III). The measurement variation was in the same magnitude as the measurement error for anterior disk height measurements reported by Ekman et al. (2009a). One reason for the good agreement in our measurements could be that all radiographs for each patient were reviewed simultaneously in order to identify identical landmarks on a series of radiographs. Moreover, the interobserver error is eliminated if all measurements are made by the same radiologist (Danielson et al. 1988). Consequently, standing radiographs may not be necessary when measuring disk height.

5.5 Assessment of global sagittal balance

There are several methods for assessment of global sagittal balance. The SRS-Schwab classification, primarily intended for adult degenerative spinal deformity classification, evaluates global sagittal alignment by analyzing the SVA (Schwab et al. 2012). The SDSG spondylolisthesis classification defines the spine as sagittally imbalanced when the C7 plumb line falls in front of the femoral heads (Labelle et al. 2011). This thesis analyzed the T1SPI, a parameter commonly used for assessment of global sagittal balance (Vialle et al. 2007; Schwab et al. 2009, 2010; Ames et al. 2012). The spine is imbalanced when T1SPI ≥ 0°, i.e. the T1 plumb line falls in front of the center of the bicoxofemoral axis (Fig. 1C, p. 5). The results of this thesis confirm the findings of previous studies that patients with high-grade spondylolisthesis have an increased SVA com-
pared with individuals with no spinal deformity (Jackson et al. 2003) or low-grade spondylolisthesis (Harroud et al. 2013) (II). The most obvious reason for the increased SVA in high-grade spondylolisthesis is that the lumbosacral deformity in spondylolisthesis shifts the spinal column above the sacrum forward, which results in an increased SVA. In this thesis, 10 of 28 patients had SVA > 5 cm but only 3 of these patients had T1SPI ≥ 0°. No patient with SVA ≤ 5 cm had T1SPI ≥ 0°. This means that the criterion SVA < 5 cm used in adult spinal deformity cases for the assessment of global sagittal balance (Schwab et al. 2010) may not apply to cases involving high-grade spondylolisthesis.

T1SPI offers several potential advantages as the primary measure of global sagittal balance: (1) angular (as opposed to distance) measurements does not require calibration of the radiographs (Schwab et al. 2009), (2) T1SPI is independent of the shape of the sacrum (anatomical landmarks of the sacrum may be difficult to identify because of the convex form of the upper end-plate of S1 often found in high-grade spondylolisthesis, cf. point 2, p. 40), (3) the measurement uses the center of the bicoxofemoral axis as distal reference i.e. it compensates for axial pelvic rotation, (4) high reliability, the SD of the mean difference between repeated measurements of T1SPI has been estimated to 0.8° (Table 2, Lafage et al. 2015), and (5) T1SPI is a continuous variable that may be used to quantify the amount of sagittal imbalance.

There are some minor but important differences between the T1SPI and the SDSG definition of sagittal imbalance. Since the C7 plumb line falls in front of the T1 plumb line there might be a risk that some patients are classified as balanced using T1 as reference but imbalanced using C7 as reference. SDSG, however, uses the femoral heads and not the center of the bicoxofemoral axis as distal reference. Parsons (1914) reported an average femoral head diameter of about 46 mm and Francis (1955) measured the sagittal diameter of the body of C7 to about 16 mm. Consequently, it seems unlikely that the T1SPI definition underestimates the number of sagittally imbalanced spines compared with the SDSG definition.
5.6 Internal validity

There are several factors that may compromise the internal validity of this thesis.

1. Different forms of biases affect the results. For this thesis, the work was based on a series of patients treated at a single center which may result in selection or indication bias.

2. All radiographic evaluations are affected by measurement errors. Measurements in patients with high-grade spondylolisthesis are difficult to perform because of the convex form of the upper end-plate of S1 and the trapezoid shape of the L5 body (Boxall et al. 1979). Legaye (2007) reported an interobserver variability (1 SD) of 5° when repeated measurements of PI and SS were made on a dome-shaped sacrum. The measurement variability was higher for a dome-shaped sacrum than for a sacrum with a normal sacral end-plate. Vialle et al. (2006) found an inter- and intraobserver variability (1 SD) of 2.8° and 2.7°, respectively, when evaluating the reliability of measurements of PI in patients with high-grade spondylolisthesis. The SD of the mean difference between repeated measurements of T1SPI has been estimated to 0.8° (Table 2, Lafage et al. 2015). For L5-S1 slip measurements, Boxall et al. (1979, p. 480) reported an interobserver error of 1-7%. Danielson et al. (1988) reported the inter- and intraobserver error both to be about 16% when using a modified version of the Boxall et al. (1979) method, systematic errors due to variations in patient positioning (axial rotation and lateral tilt) were estimated to 15%. Postoperative measurements may be even more imprecise since fusion masses or instrumentation might obscure landmarks. Timon et al. (2005) reported an interobserver agreement of 73% for the Meyerding (1932) classification. It is important to be aware of the size of the measurement error when assembling study populations based on slip or the Meyerding classification since some patients might be misclassified (i.e., information bias, Grimes and Schulz 2002; Stenson and Kepler 2019).
3. No correction for multiple testing was made which increased the probability for type 1 errors (Altman 1991, ch. 9.8.4).

4. The small sample size increased the probability for type 2 errors (Altman 1991, ch. 8.5.3, ch. 15.3.1).

5. Eleven of thirty-nine patients were lost to follow-up in the evaluation of sagittal balance which compromises internal validity (II). Only one patient was lost to follow-up in the evaluation of self-image and HRQoL which strengthens internal validity (IV).

6. The lower limbs were not visible on the radiographs which means that compensatory mechanisms of the lower limbs (for example knee flexion or ankle extension, Barrey et al. 2013) could not be evaluated.

7. Nonunion was evaluated on supine radiographs and not on standing flexion/extension radiographs or computed tomography scans.

8. ASD was not evaluated on MRI which could have given more details on disk degeneration and spinal stenosis proximal to the fusion. Frobin et al. (1997), however, reported that the reductions in disk height measured on plain radiographs are in general comparable to those assessed by MRI, although MRI also detects earlier signs of disk degeneration. The same investigators also showed that classification of height loss in individual disks based on MRI was imprecise.

9. The investigated cohort of patients included 6 patients with Harrington rods. The diversity weakens internal validity. These 6 patients were not excluded since previous investigations have shown that Harrington rods negatively affects lumbosacral kyphosis (Bradford and Gotfried 1987) and are, from a biomechanical perspective, inappropriate for the treatment of high-grade spondylolisthesis. This means that inclusion of the Harrington rod patients would probably not improve outcome on group level. Notably, 2 of 3 patients with global sagittal imbalance had a Harrington rod instrumentation. The SRS-22r subscore of these three patients was > 4 in 2 of 3 patients.
10. The Swedish version of SRS-22r is not validated for spondylolisthesis. This means that there might be a risk that the questionnaire does not measure appearance properties specific to spondylolisthesis. SRS-22r has 2 trunk/shape specific questions, Question 4 (If you had to spend the rest of your life with your back shape as it is now, how would you feel about it?) and Question 10 (Which of the following best describes the appearance of your trunk defined as the human body except for the head and extremities?). These questions are not scoliosis specific and most probably also apply to a sagittal plane deformity such as spondylolisthesis.

11. The function and pain domains of SRS-22r demonstrated substantial ceiling effects for patients and controls (≥ 15% of the respondents achieved the highest possible score, Terwee et al. 2007). Also Danielsson and Romberg (2013) observed a high percentage of patients with ceiling effects for the SRS-22r function and pain domain. Ceiling or floor effects may compromise sensitivity (the ability to detect differences between groups) or responsiveness (the ability to detect changes when a patient improves or deteriorates) (Fayers and Machin 2016, p. 119).

5.7 External validity

The external validity is limited by the fact that the thesis evaluated uninstrumented circumferential in situ fusion with external bracing postoperatively and, in a few patients, instrumentation with Harrington distraction rods. With the advent of modern spinal instrumentation techniques, in situ fusion with external bracing postoperatively, which was a common choice in the 1970s and 1980s, is now rarely used in the surgical treatment of high-grade spondylolisthesis. However, to evaluate new methods, it is important to understand what was actually accomplished with the treatment methods from previous eras in spine surgery. Also, the study design (i.e., case series, level of evidence 4) limits external validity.
5.8 Statistical notes

1. Hypothesis testing was used for the statistical inference (I-IV). Exact p-values were used as a measure to decide whether to reject the null hypothesis or not. This approach is widespread in medical research (Altman 1991, ch. 8.8). In 1991 Altman argued that inference based on estimation and confidence intervals is superior to hypothesis testing since confidence intervals also gives information about the magnitude of the effect of interest (Altman 1991, ch. 8.5, ch. 8.8). Also Ranstam (2019) argues in favor of estimation and confidence intervals and p-value-free manuscripts.

2. The research questions were discussed in the introductory sections of Papers I-IV but an overall null hypothesis was not explicitly stated. This is recognized as a limitation of this thesis. Formulating an overall primary outcome measure could be elegant from a methodological perspective but could also be misleading and risk an oversimplification of a complex set of data (HRQoL instruments, clinical and radiographic findings, etc.).

3. Hypothesis testing and dichotomizations were used in favor of more complex statistical methods (e.g., regression analysis) for statistical inference (II-IV). The main reasons for this simplistic approach were (1) the limited number of patients of the thesis and (2) most data were discrete, skewed and non-normally distributed, and could not easily be transformed into e.g. a linear regression model.

4. The SRS-22r function, pain, and mental health scores were non-normally distributed (skewed towards higher values) with unequal variances for cases and controls while the SRS-22r self-image data was normally distributed with unequal variances. The non-parametric Mann-Whitney U test was used in the statistical inference for all domains of SRS-22r (IV). Alternatively, the Welch test (t-test modified for the case with unequal variances) could have been used (Altman 1991, ch. 9.6.5). The Welch test is fairly robust against violation of
the normality assumption and more robust against unequal sample sizes (two controls for each case) and unequal variances than the Mann-Whitney U test (Stonehouse and Forrester 1998; Zimmerman 1998).

5. The evaluation of self-image did not report any effect size measure for the difference in self-image between patients and controls (IV). For independent groups Fayers and Machin (2016, p. 534) recommend the standardized mean difference (SMD) as effect size measure. In short, the SDM is the difference in means in terms of SDs (the difference is normalized with the SD) (Cohen 1988, eq. 2.2.1, p. 20). For example, an SMD of 0.5 means that the means differs by 0.5 SDs. Formulas for SMD and the standard error of SMD are given in Fayers and Machin (2016, table 20.1, p. 535). An approximate 95% CI for SMD is given by Fayers and Machin (2016, eq. 20.4, p. 542). If these formulas are applied on the data of Paper IV, the SMD in self-image between patients with spondylolisthesis and controls is -1.27 (95% CI -1.69 to -0.85). Likewise, if the formulas are applied on the data of Gutman et al. (2017), the SMD in self-image between patients with spondylolisthesis and controls is -0.49 (95% CI -0.77 to -0.22). The SMD in self-image between high-grade and low-grade spondylolisthesis is -0.42 (95% CI -0.77 to -0.08). Guidelines for the interpretation of effect sizes are given by Cohen (1988, ch. 2.2.3) and Fayers and Machin (2016, p. 499). Note that the effect size can be limited by the ceiling effects of SRS-22r as discussed in point 11, p. 42 (Fayers and Machin 2016, p. 119).
6 Conclusions

1. Young patients fused in situ for high-grade isthmic spondylolisthesis have long-term HRQoL similar to the general population (I) as well as to age and gender matched controls (IV).

2. Fusion in situ for high-grade isthmic spondylolisthesis in young patients does not result in sagittal imbalance in adulthood (II).

3. Radiographic sagittal balance parameters do not correlate with long-term SRS-22r outcome after fusion in situ for high-grade isthmic spondylolisthesis (II).

4. Fusion in situ for high-grade isthmic spondylolisthesis in young patients does not result in symptomatic ASD in adulthood (III).

5. The self-image of patients fused in situ for high-grade isthmic spondylolisthesis at a young age is negatively affected in adulthood (IV).
7 Acknowledgements

This thesis could not have been written without the support of many people. I owe a particular debt of gratitude to:

Dr. Karin Frennered, supervisor, who introduced me to research and very generously gave me access to unique data on high-grade spondylolisthesis. Thank you, Karin, for your endless enthusiasm and support.

Dr. Per Wretenberg, supervisor, for encouragement, help, and advice during my period of research. Thank you, Per, for persuading me to write this thesis and for providing an excellent research environment.

Dr. Rune Hedlund, co-author, for enthusiastically giving me far more help and advice than I could ever have expected or imagined. Thank you, Rune, for your consistent support throughout the years.

Dr. Barbro Danielson, co-author, for good collaboration and for invaluable help in reviewing radiographs.

Dr. Paul Gerdhem, co-author, for good collaboration and for most kindly providing access to previously unpublished SRS-22r data.

Dr. Elias Diarbakerli, co-author, for good collaboration.

Dr. Ewald Ornstein, head of the Department of Orthopaedics, Örebro University Hospital, for supporting my research.

Dr. Emma Turesson, for reviewing several drafts of this text.

Financial support was provided by the Dr. Felix Neubergh Foundation and the Region Örebro County Research Committee (OLL-639771).
8 References


Fayers PM, Machin D. (2016). Quality of life: The assessment, analysis and reporting of patient-reported outcomes. 3rd ed. Chichester: John Wiley and Sons Ltd. [42, 44]


Grimes DA, Schulz KF. (2002). Bias and causal associations in observational research. Lancet, 359(9302), 248-52. [40]


Martiniani M, Lamartina C, Specchia N. (2012). "In situ" fusion or reduction in high-grade high dysplastic developmental spondylolisthesis (HDSS). Eur Spine J, 21(Suppl 1), S134-40. [3, 8, 9]


