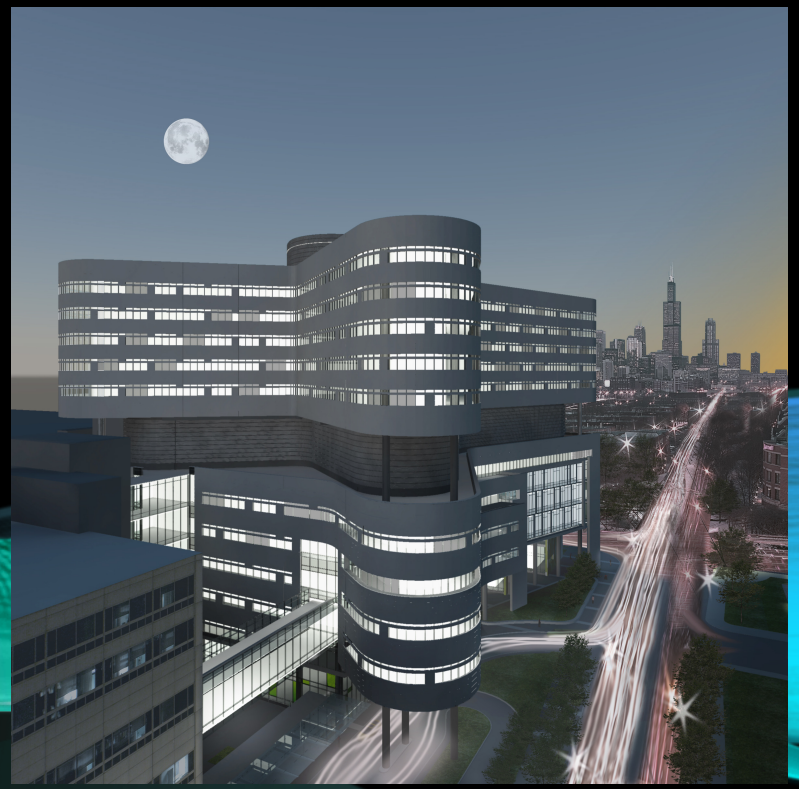
COCHRANE ANALYSIS:
DEFINITIVE STATISTICS OR BIASED OPINION?

Richard G. Fessler, MD, PhD
Professor
Rush University Medical Center
Chicago, IL



RUSH

SURGERY VS NON-SURGERY FOR LUMBAR SPINAL STENOSIS:

AN IN-DEPTH ANALYSIS OF THE 2016 COCHRANE ANALYSIS, THE STUDIES INCLUDED FOR ANALYSIS AND COCHRANE METHODOLOGY

RICHARD G. FESSLER, MD, PHD

JOURNAL OF NEUROSURGERY: SPINE
IN PRESS

THE PROBLEM…

- Over the last several years, "Cochrane Analysis" has come to be regarded as the definitive method of analysis for the quality and reliability of scientific data
  - Results often used by CMS and private insurance for approval of payment
  - Due to the sophistication of the statistical manipulations and relative lack of statistical expertise among clinicians
    - The results have rarely been challenged
- HOWEVER, occasionally the results seem to defy rationality
  - Suggests the possibility that the subjective determination of "quality" could be done in a manner to support a pre-determined conclusion

EXAMPLE

- "Surgical versus non-surgical treatment of lumbar spinal stenosis"
  - Zaina et al. 2016 COCHRANE DATABASE OF SYSTEMATIC REVIEWS
- Analyzed five prospective, randomized, controlled publications comparing surgery to non-surgery for lumbar stenosis
  - Each concluded that surgery was superior to non-surgery
    - This is consistent with surgeons daily observations
- Cochrane conclusion:
  - "On the whole, these studies provide conflicting low-quality evidence on the effectiveness of surgery versus conservative treatment for LSS. Study results preclude conclusions regarding whether surgical or non-surgical treatment provides better outcomes for people with LSS."

HOW IS THIS POSSIBLE?

- This presentation will:
  - Examine the details of each of the five prospective studies
  - Review the mechanics of a Cochrane analysis
    - How it was originally constructed
    - How quality is determined
    - How conclusions are presented
  - Show the conclusions of Zaina, 2016
  - Review recommended changes to Cochrane analysis subsequent to its original construction
  - Re-analyze the five prospective studies in light of the recommended changes
  - Show the conclusions resulting from the re-analysis

FIVE STUDIES

ORIGINAL ARTICLE

A Double-blind, Randomized, Prospective Study of Epidural Steroid Injection vs. The *mild*® Procedure in Patients with Symptomatic Lumbar Spinal Stenosis

Lora L. Brown, MD
Coastal Orthopedics and Sports Medicine, Bradenton, Florida, U.S.A.

Published in final edited form as:
Ann Intern Med. 2008 December 16; 149(12): 845–853.

Surgical Treatment of Spinal Stenosis with and without Degenerative Spondylolisthesis: Cost-Effectiveness after 2 Years

Anna N.A. Tosteson, ScD, Jon D. Lurie, MD, MS, Tor D. Tosteson, ScD, Jonathan S. Skinner, PhD, Harry Herkowitz, MD, Todd Albert, MD, Scott D. Boden, MD, Keith Bridwell, MD, PhD, Michael Longley, MD, Gunnar B. Andersson, MD, PhD, Emily A. Blood, MS, Margaret R. Grove, MS, and **James N. Weinstein, DO, MS [on behalf of for the SPORT Investigators]***
From Dartmouth Medical School, Hanover, New Hampshire; William Beaumont Hospital, Royal Oak, Michigan; Rothman Institute at Thomas Jefferson University, Philadelphia, Pennsylvania; Emory University, Atlanta, Georgia; Washington University School of Medicine, St. Louis, Missouri; The Nebraska Foundation for Spinal Research, Omaha, Nebraska; and Rush University Medical Center, Chicago, Illinois

SPINE Volume 32, Number 1, pp 1–8

Surgical or Nonoperative Treatment for Lumbar Spinal Stenosis?

A Randomized Controlled Trial

Antti Malmivaara, MD, PhD,* Pär Slätis, MD, PhD,‖ Markku Heliövaara, MD, PhD,† Päivi Sainio, PT, MSc,† Heikki Kinnunen, MD,§ Jyrki Kankare, MD, PhD,§ Nina Dalin-Hirvonen, MD,‡ Seppo Seitsalo, MD, PhD,‖ Arto Herno, MD, PhD,¶ Pirkko Kortekangas, MD, PhD,# Timo Niinimäki, MD, PhD,** Hannu Rönty, MD,** Kaj Tallroth, MD, PhD,‖ Veli Turunen, MD,†† Paul Knekt, PhD,‡‡ Tommi Härkänen, PhD,† and Heikki Hurri, MD, PhD,‖ for the Finnish Lumbar Spinal Research Group

SPINE Volume 25, Number 11, pp 1424–1436

Lumbar Spinal Stenosis: Conservative or Surgical Management?

A Prospective 10-Year Study

Tom Amundsen, MD,* Henrik Weber, MD, DrMed,* Helge J. Nordal, MD, DrMed,* Bjørn Magnaes, MD, DrMed,† Michael Abdelnoor, MPH, PhD,‡ and Finn Lilleås, MD§

SPINE Volume 30, Number 12, pp 1351–1358

A Multicenter, Prospective, Randomized Trial Evaluating the X STOP Interspinous Process Decompression System for the Treatment of Neurogenic Intermittent Claudication

Two-Year Follow-Up Results

James F. Zucherman, MD,* Ken Y. Hsu, MD,* Charles A. Hartjen, MD,† Thomas F. Mehalic, MD,‡ Dante A. Implicito, MD,§§ Michael J. Martin, MD,¶ Donald R. Johnson II, MD,‖ Grant A. Skidmore, MD,** Paul P. Vessa, MD,†† James W. Dwyer, MD,†† Stephen T. Puccio, MD,§§§§ Joseph C. Cauthen MD,¶¶ and Richard M. Ozuna, MD‖‖

SPINE Volume 25, Number 11, pp 1424–1436

Lumbar Spinal Stenosis: Conservative or Surgical Management?

A Prospective 10-Year Study

Tom Amundsen, MD,* Henrik Weber, MD, DrMed,* Helge J. Nordal, MD, DrMed,* Bjørn Magnaes, MD, DrMed,† Michael Abdelnoor, MPH, PhD,‡ and Finn Lilleås, MD§

Table 1. Summary of 5 RCT's Evaluated in Zaina, 2016

	Amundsen, 2000			
# Patients	100			
Randomized	31yes, 69 assigned due to severe or minmal pain			
Blinded	No			
Multi-centre	No			
Tyee of surgery	Decompression only			
Follow-up	6 mo, 1, 4, 10 yrs			
VAS-back	with or without			
VAS-leg	yes			
ODI	no			
SF-12/ZCQ	no			
Walking distance	no			
Intent to treat analysis	No			
	Non-surgery-non randomized=50	Non-surgery-randomized=18	Surgery-non-randomized=19	Surgery-randomized=13
	VAS % Good result	VAS % Good Result	VAS % Good Result	VAS % Good Result
6 wk				
12 wk				
6 wk				
12 wk				
6 wk				
12 wk				
6 mo	70	39	79	92
1 yr	64	33	89	69
2 yr				
4 yr	57	47	84	92
10 yr	f remaining group**	of remaining group^	71*	91
worse %	20%	50%	0%	0%
Conclusion	Surgery superior to non-surgery			
	** 20% cross-over to	50% cross-over to su	* Two patients with good results died.	

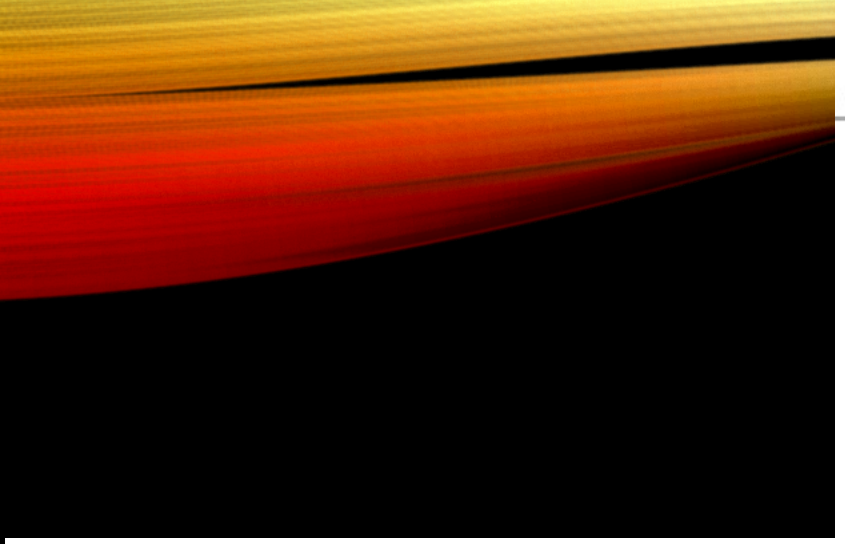
ORIGINAL ARTICLE

A Double-blind, Randomized, Prospective Study of Epidural Steroid Injection vs. The *mild*® Procedure in Patients with Symptomatic Lumbar Spinal Stenosis

Lora L. Brown, MD
Coastal Orthopedics and Sports Medicine, Bradenton, Florida, U.S.A.

Table 1. Summary of 5 RCT's Evaluated in Zaina, 2016

	Brown, 2012	
# Patients	38	
Randomized	yes	
Blinded	yes	
Multi-centre	No	
Tyee of surgery	MILD	
Follow-up	6, 12 wks	
VAS-back	with or without	
VAS-leg	yes	
ODI	yes	
SF-12/ZCQ	yes	
Walking distance	no	
Intent to treat analysis	yes	
	MILD	ESI
	VAS mean improvement	
6 wk	2.5	0.1
12 wk	3.4	Not reported due to 100 % cross-over
	ODI mean improvement	
6 wk	11.4	5.7
12 wk	13.3	Not reported due to 100 % cross-over
	ZCQ mean improvement	
6 wk	2.2	2.8
12 wk	1.8	No reported due to 100 % cross-over
6 mo		
1 yr		
2 yr		
4 yr		
10 yr		
worse %		
Conclusion	Surgery superior to non-surgery	



SPINE Volume 32, Number 1, pp 1–8

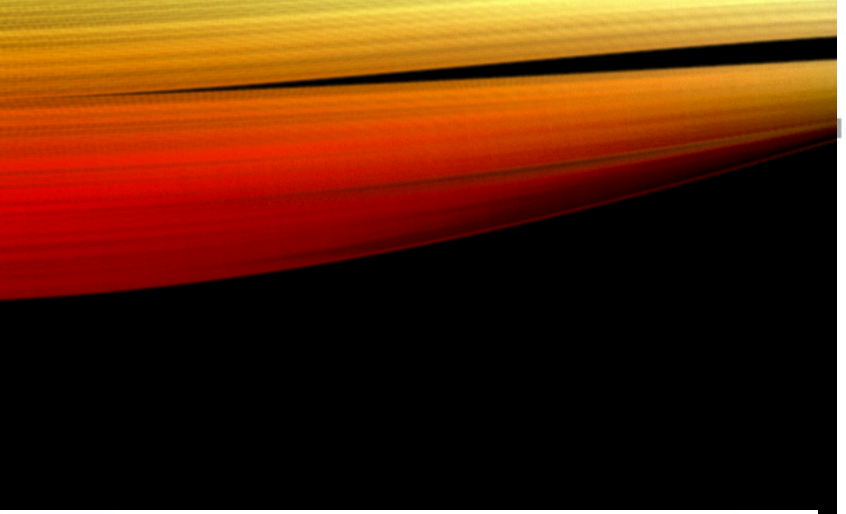
Surgical or Nonoperative Treatment for Lumbar Spinal Stenosis?

A Randomized Controlled Trial

Antti Malmivaara, MD, PhD,* Pär Slätis, MD, PhD,‖ Markku Heliövaara, MD, PhD,†
Päivi Sainio, PT, MSc,† Heikki Kinnunen, MD,§ Jyrki Kankare, MD, PhD,§
Nina Dalin-Hirvonen, MD,‡ Seppo Seitsalo, MD, PhD,‖ Arto Herno, MD, PhD,¶
Pirkko Kortekangas, MD, PhD,# Timo Niinimäki, MD, PhD,** Hannu Rönty, MD,**
Kaj Tallroth, MD, PhD,‖ Veli Turunen, MD,†† Paul Knekt, PhD,‡‡ Tommi Härkänen, PhD,†
and Heikki Hurri, MD, PhD,‖ for the Finnish Lumbar Spinal Research Group

Table 1. Summary of 5 RCT's Evaluated in Zaina, 2016

	Malmivaara, 2007							
# Patients	94							
Randomized	yes							
Blinded	no							
Multi-centre	Yes							
Tyee of surgery	Decompression; Fusion if unstable							
Follow-up	6 mo, 1, 2 yrs							
VAS-back	yes							
VAS-leg	yes							
ODI	yes							
SF-12/ZCQ								
Walking distance	yes							
Intent to treat analysis	yes							
	Surgery	Non-surgery	Surgery	Non-Surgery	Surgery	Non-Surgery	Surgery	Non-Surgery
	ODI % Improvement		VAS Leg % Improvement		VAS Back % Improvement		Walking distance % Improvement	
6 wk								
12 wk								
6 wk								
12 wk								
6 wk								
12 wk								
6 mo	39	19	61	27	59	20	114	94
1 yr	44	13	59	30	61	26	130	107
2 yr	38	16	54	28	60	29	114	104
4 yr								
10 yr								
worse %								
Conclusion	Surgery superior to non-surgery							



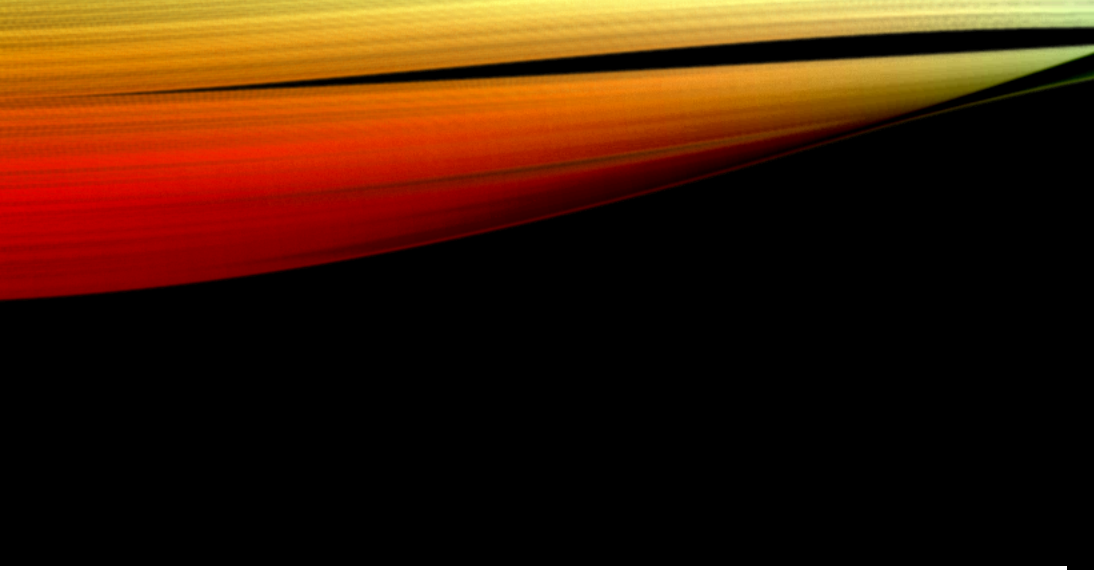
Published in final edited form as:
Ann Intern Med. 2008 December 16; 149(12): 845–853.

Surgical Treatment of Spinal Stenosis with and without Degenerative Spondylolisthesis: Cost-Effectiveness after 2 Years

Anna N.A. Tosteson, ScD, Jon D. Lurie, MD, MS, Tor D. Tosteson, ScD, Jonathan S. Skinner, PhD, Harry Herkowitz, MD, Todd Albert, MD, Scott D. Boden, MD, Keith Bridwell, MD, PhD, Michael Longley, MD, Gunnar B. Andersson, MD, PhD, Emily A. Blood, MS, Margaret R. Grove, MS, and James N. Weinstein, DO, MS [on behalf of for the SPORT Investigators]*
From Dartmouth Medical School, Hanover, New Hampshire; William Beaumont Hospital, Royal Oak, Michigan; Rothman Institute at Thomas Jefferson University, Philadelphia, Pennsylvania; Emory University, Atlanta, Georgia; Washington University School of Medicine, St. Louis, Missouri; The Nebraska Foundation for Spinal Research, Omaha, Nebraska; and Rush University Medical Center, Chicago, Illinois

Table 1. Summary of 5 RCT's Evaluated in Zaina, 2016

	Weinstein, 2008											
# Patients	289 randomized; 365											
Randomized	yes											
Blinded	no											
Multi-centre	Yes											
Tyee of surgery	Decompession, fusion if unstable											
Follow-up												
VAS-back	no											
VAS-leg	no											
ODI	yes											
SF-12/ZCQ	yes											
Walking distance												
Intent to treat analysis	yes and as treated											
	Intention to treat analysis-SF-36 mean change bodily pain/PF/ODI						As treated analysis-SF-37 mean change bodily pain/PF/ODI					
	Surgery			Non-surgery			Surgery			Non-Surgery		
6 wk												
12 wk												
6 wk												
12 wk												
	BP	PF	ODI	BP	PF	ODI	BP	PF	ODI	BP	PF	ODI
6 wk	11.2	6.0.	-6.5	7.9.	10.2	-7.9	19.8.	17.1.	-17.0	9,8	8.7	-6.8
12 wk	13.5.	7.4.	-7.6	11.1.	11.6.	-8.1	27.9	24.8.	-21.4	11.8	10.0	-7.6
6 mo	21.	17.6.	-14.6	16.1.	15.1.	-13.7	29.5.	26.9.	-22.9	12.9.	10.6	-8.8
1 yr	23.	18.	-14.9	17.5	16.4.	-12.7	28.0.	26.5.	-21.4	13.5	10.5	-8.9
2 yr	23.4.	17.1.	-16.4	15.6.	17.1.	-12.9	26.9.	23.0.	-20.5	13.3.	11.1	-9.3
4 yr												
10 yr												
worse %												
Conclusion	Surgery is superior to non-surgery											



SPINE Volume 30, Number 12, pp 1351–1358

A Multicenter, Prospective, Randomized Trial Evaluating the X STOP Interspinous Process Decompression System for the Treatment of Neurogenic Intermittent Claudication

Two-Year Follow-Up Results

James F. Zucherman, MD,* Ken Y. Hsu, MD,* Charles A. Hartjen, MD,† Thomas F. Mehalic, MD,‡ Dante A. Implicito, MD,§§ Michael J. Martin, MD,¶ Donald R. Johnson II, MD,‖ Grant A. Skidmore, MD,** Paul P. Vessa, MD,†† James W. Dwyer, MD,†† Stephen T. Puccio, MD,§§§§ Joseph C. Cauthen MD,¶¶ and Richard M. Ozuna, MD‖‖

Table 1. Summary of 5 RCT's Evaluated in Zaina, 2016

	Zucherman, 2005			
# Patients	191			
Randomized	Yes			
Blinded	No			
Multi-centre	Yes			
Tyee of surgery	X-Stop			
Follow-up	6 wk, 6 mo, 1,2 yrs			
VAS-back	no			
VAS-leg	no			
ODI	no			
SF-12/ZCQ	yes			
Walking distance	no			
Intent to treat analysis	no			
	Mean % Improvement Symptom Severity		Mean % Improvement Physical Function	
	Surgery	ESI	Surgery	ESI
6 wk				
12 wk				
6 wk				
12 wk				
6 wk	48	7	53	5
12 wk			54	5
6 mo	46	9	54	9
1 yr	47	7	48	0
2 yr	47	5		
4 yr				
10 yr				
worse %				
Conclusion				

WHAT IS A COCHRANE ANALYSIS?

- 1997
  - 16 statisticians and epidemiologists held a 3 day meeting to develop the tool
    - 7 areas of potential bias were determined
      - Generation of the allocation sequence
      - Concealment of the allocation sequence
      - Blinding
      - Attrition and exclusions
      - Other "generic" sources of bias
      - Biases specific to the trial design
        - Crossover, cluster randomized trials
      - Biases that might be specific to clinical specialty

PROCESS

- A nominated meeting participant prepared a review of the empirical evidence
- Discussion of specific issues occurred
- A set of criteria for assessing protection from bias was proposed
  - Adequate, in-adequate, unclear
- DECISIONS WERE MADE BY INFORMAL CONSENSUS
- Potential biases were then divided into “Domains”
  - Strategies for their assessment were agreed upon by INFORMAL CONSENSUS
- This created the original Cochrane Tool

SUBSEQUENT MODIFICATIONS

2003: van Tulder et al. Updated Method Guideline for Systemic Reviews in the Cochrane Collaboration Back Review Group. Spine 28:1290-1299.

2009: Furlan et al. Updated method guidelines for systematic reviews in the Cochrane Back Review Group. Spine 34:1929-1941.

2011: Higgins et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 33:d5928.

2011 "PRINCIPLES OF RISK ASSESSMENT

1. Do not use quality scales
2. Focus on internal validity
3. Assess the risk of bias in trial results, not the quality of reporting or methodological problems that are not directly related to risk of bias
4. Assessments of risk of bias require judgement
5. Choose domains to be assessed based on a combination of theoretical and empirical considerations
6. Focus on risk of bias in the data as represented in the review rather than as originally reported
7. Report outcome specific evaluations of risk of bias

ZAINA METHODOLOGY

- Used 6 domains
- Divided them into 11 potential sources of bias
- Subjective assessments of risk of bias were determined through consensus of two authors
- Inter-author reliability was not assessed
  - "we reached agreement on each study evaluation"

Cochrane Library

Cochrane Database of Systematic Reviews

Surgical versus non-surgical treatment for lumbar spinal stenosis (Review)

Zaina F, Tomkins-Lane C, Carragee E, Negrini S

Authors' conclusions

We have very little confidence to conclude whether surgical treatment or a conservative approach is better for lumbar spinal stenosis, and we can provide no new recommendations to guide clinical practice. However, it should be noted that the rate of side effects ranged from 10% to 24% in surgical cases, and no side effects were reported for any conservative treatment. No clear benefits were observed with surgery versus non-surgical treatment. These findings suggest that clinicians should be very careful in informing patients about possible treatment options, especially given that conservative treatment options have resulted in no reported side effects. High-quality research is needed to compare surgical versus conservative care for individuals with lumbar spinal stenosis.

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.

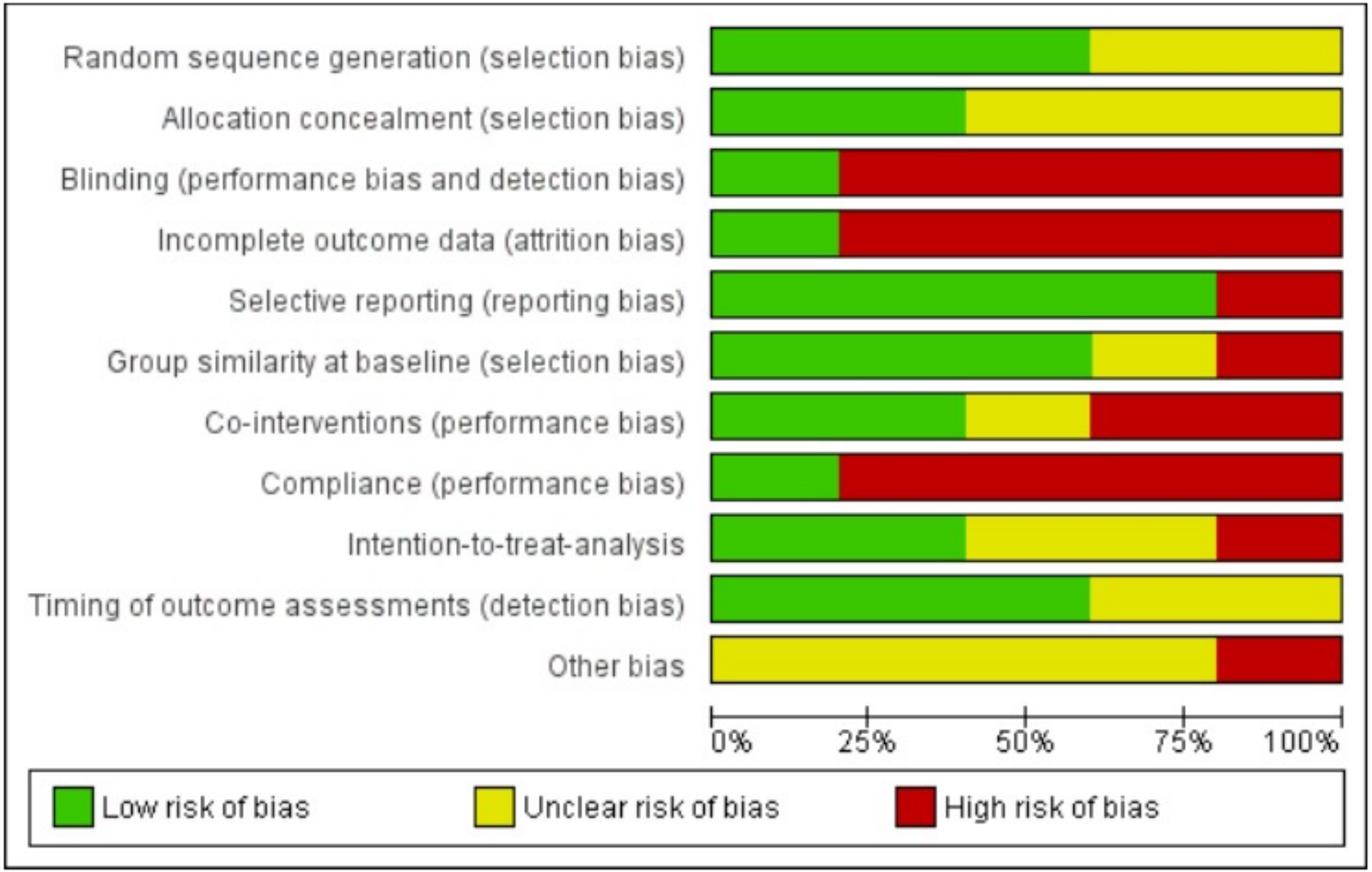
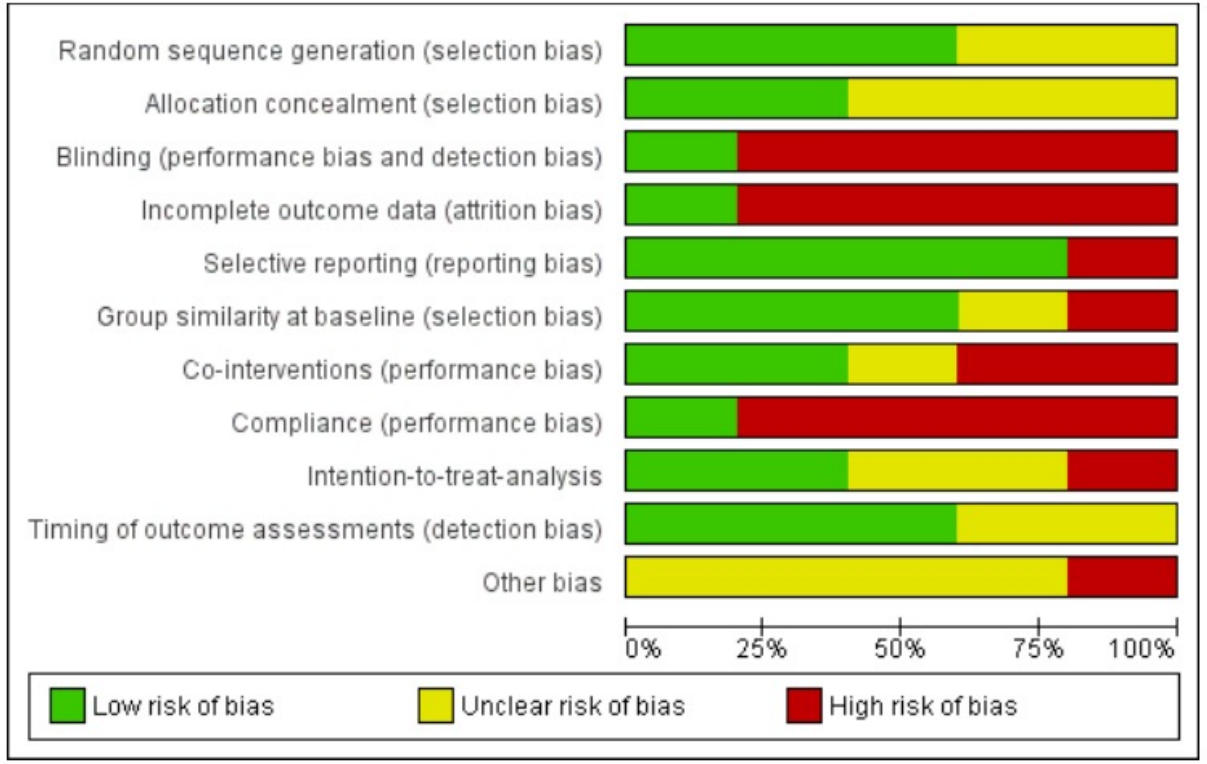


Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



Random sequence generation (selection bias)

Three studies clearly described low risk of bias for the randomisation process (Brown 2012; Malmivaara 2007, Weinstein 2008), and the other two studies (Amundsen 2000; Zucherman 2004) did not provide this information.

Allocation

We considered allocation to be adequate in two studies (Brown 2012; Zucherman 2004) and unclear in the other three studies, given that study authors did not provide the required information (Amundsen 2000; Malmivaara 2007; Weinstein 2008).

Blinding

Blinding is very difficult when surgical and non-surgical treatments are compared because of the nature of the interventions. It is obvious in most cases to participants whether they are undergoing surgical or non-surgical care. Only one study was double-blinded, and it was rated as having low risk bias for this criterion (Brown 2012). The other four studies (Amundsen 2000; Malmivaara 2007; Weinstein 2008; Zucherman 2004) were considered at high risk of bias for this criterionbecause blinding was not possible, given the types of interventions compared.

Incomplete outcome data

Only one study presented complete data and was considered at low risk bias for this criterion (Brown 2012). Three studies (Amundsen 2000; Malmivaara 2007; Zucherman 2004) were considered at high risk of bias because study authors reported only data for completers. One study (Weinstein 2008) was rated at high risk of bias because the number of cross-overs made complete outcome reporting impossible after the first phase.

Selective reporting

Four studies reported all outcomes presented in the protocol and were considered at low risk (Brown 2012; Malmivaara 2007; Weinstein 2008; Zucherman 2004). We considered one study to be at high risk (Amundsen 2000) because not all outcomes were reported.

Group similarity at baseline (selection bias)

Groups were similar at baseline for each comparison.

Co-interventions (performance bias)

We noted no imbalance among co-interventions.

Compliance (performance bias)

Risk of bias was unclear in four studies because compliance was not monitored in the conservative group.

Intention-to-treat analysis

In one study, no intent-to-treat analysis (ITT) was performed because of the high rate of cross-over; therefore, the study was considered to be at high risk (Weinstein 2008).

Timing of outcome assessments (detection bias)

Risk of bias was low because all important outcome assessments for all intervention groups were measured at the same time.

Other potential sources of bias

We found no further risks of bias.

Effects of interventions

See: **Summary of findings for the main comparison**; **Summary of findings 2**

We pooled two of the included studies together for a single outcome - the Oswestry Disability Index (Malmivaara 2007; Weinstein 2008; Analysis 1.1). We presented data for the other three studies (Amundsen 2000; Brown 2012; Zucherman 2004) individually because of heterogeneity of interventions, study populations, outcome measures and duration of follow-up.

Usual conservative treatment versus decompression with or without fusion

Three studies (414 participants) compared usual conservative treatment versus decompression with or without fusion (Amundsen 2000; Malmivaara 2007; Weinstein 2008). The surgical approach consisted of decompression through laminectomy and eventually spinal fusion in cases of risk of instability. Usual conservative treatment consisted of varying approaches including non-steroidal anti-inflammatory drugs, exercise, education, steroid injections and other modalities. For each case, investigators presented no clearly defined standard protocol and no description of the specifics of conservative treatment.

We obtained low-quality evidence from the meta-analysis performed with two trials (320 participants) for the Oswestry Disability Index (pain-related disability), comparing direct decompression with or without fusion versus multi-modal non-operative

Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

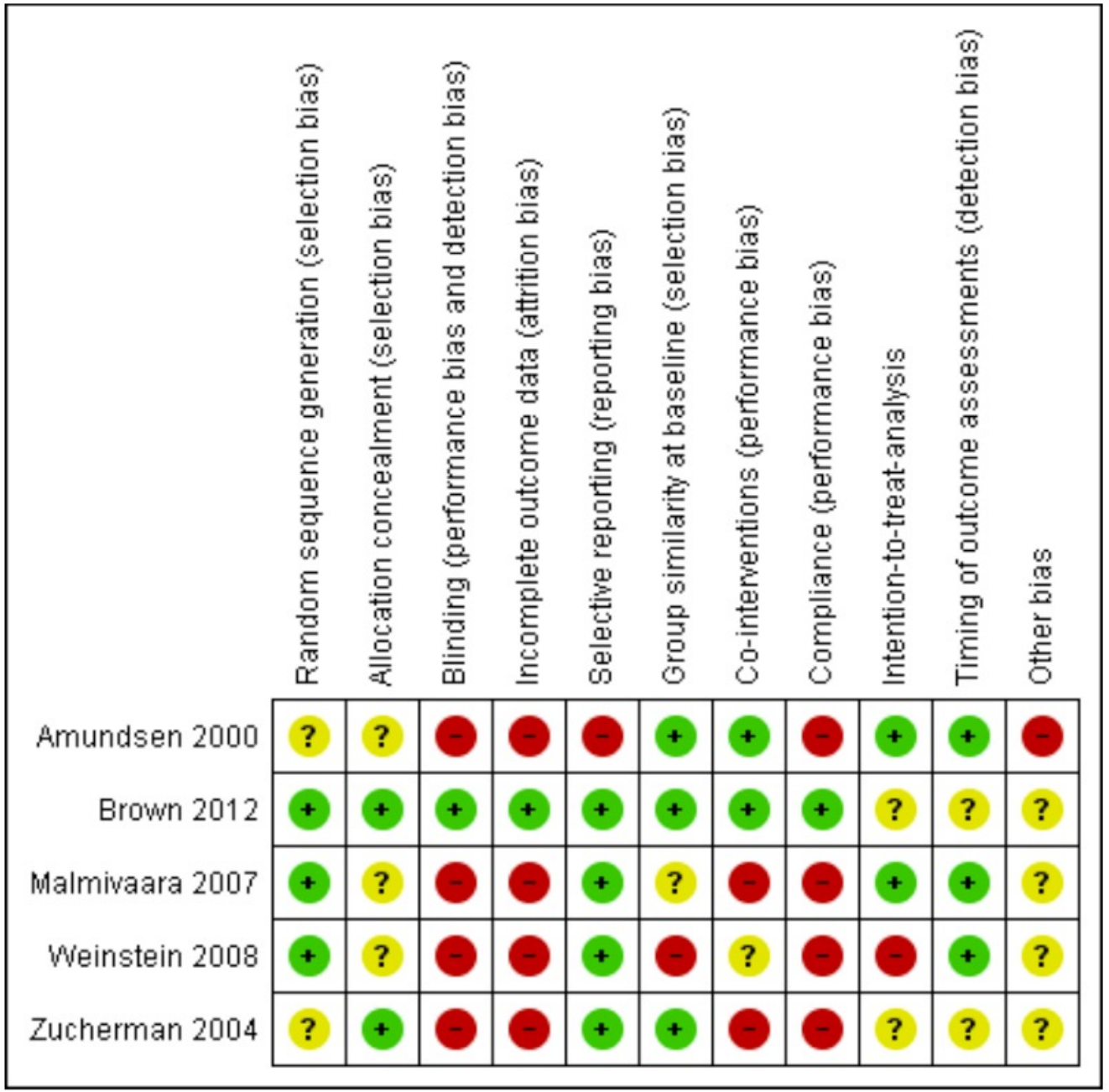


Figure 4. Forest plot of comparison: 1 Decompression ± fusion vs usual non-operative care for Oswestry Disability Index, outcome: 1.1 Oswestry Disability Index [%].

Study or Subgroup	Decompression+/- fixation Mean [%]	SD [%]	Total	Usual nonoperative care Mean [%]	SD [%]	Total	Weight	Mean Difference IV, Random, 95% CI [%]
1.1.1 6 months								
Malmivaara 2007	20.7	15.7	50	28.3	15.39	44	41.2%	-7.60 [-13.89, -1.31]
Weinstein 2008	28.1	12.05	120	29	12.78	135	58.8%	-0.90 [-3.95, 2.15]
Subtotal (95% CI)			**170**			**179**	**100.0%**	**-3.66 [-10.12, 2.80]**
Heterogeneity: Tau² = 16.08; Chi² = 3.53, df = 1 (P = 0.06); I² = 72%								
Test for overall effect: Z = 1.11 (P = 0.27)								
1.1.2 1 year								
Malmivaara 2007	18.9	17.25	50	30.2	17.51	44	43.7%	-11.30 [-18.34, -4.26]
Weinstein 2008	27.8	12.05	120	30	12.35	126	56.3%	-2.20 [-5.25, 0.85]
Subtotal (95% CI)			**170**			**170**	**100.0%**	**-6.17 [-15.02, 2.67]**
Heterogeneity: Tau² = 33.74; Chi² = 5.40, df = 1 (P = 0.02); I² = 81%								
Test for overall effect: Z = 1.37 (P = 0.17)								
1.1.3 2 years								
Malmivaara 2007	21.2	17.25	50	29	17.45	44	21.7%	-7.80 [-14.83, -0.77]
Weinstein 2008	26.3	11.43	108	29.8	11.69	113	78.3%	-3.50 [-6.55, -0.45]
Subtotal (95% CI)			**158**			**157**	**100.0%**	**-4.43 [-7.91, -0.96]**
Heterogeneity: Tau² = 1.60; Chi² = 1.21, df = 1 (P = 0.27); I² = 17%								
Test for overall effect: Z = 2.50 (P = 0.01)								

Test for subgroup differences: Chi² = 0.20, df = 2 (P = 0.90), I² = 0%

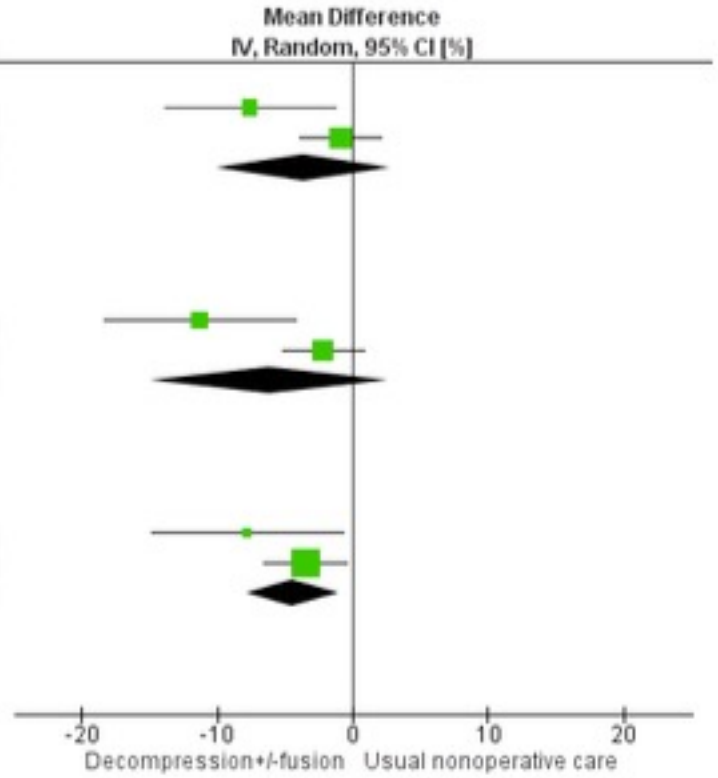
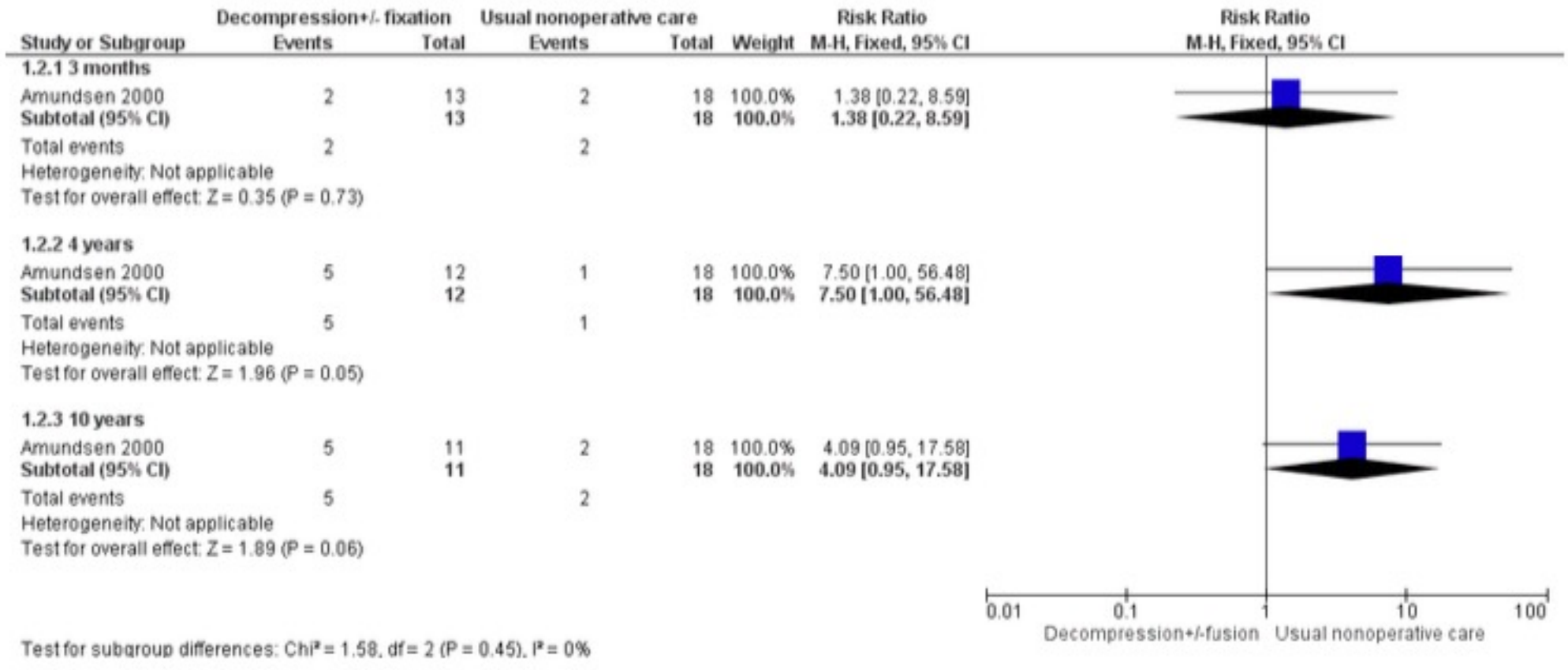


Figure 5. Forest plot of comparison: 1 Decompression ± fusion versus usual non-operative care for adverse events.



MY ANALYSIS

- Utilized same domains and potential sources of bias
- Interpreted the proscribed principles of risk assessment more closely as described by Higgins 2011
- Non-significant differences in baseline values in randomized groups should not be considered as high risk

HIGGINS, 2011 CORRECTIONS OF EARLIER METHODOLOGY

1. ”It would not reasonable to consider a trial “low quality” because of the absence of blinding” in surgical vs non-surgical trials.
2. “Quality (details) of reporting is not directly related to bias.”
3. “Block randomization is an acceptable randomization technique.”
4. “Quality of monitoring of compliance is not directly related to bias.”
5. Attrition is not directly related to bias if statistically accounted for.

Table 2

Risk Assessment in Amundsen as judged by(2007) Zaina et al (2016) vs this manuscript utilizing modifications suggested by Furlan[3] and Higgins[4].

Bias Domain	Source of Bias	Zaina Risk Assessment	Support for Judgement	This manuscript Risk Assessment	Support for Judgement
Selection Bias	Random Sequence generation	Unclear Risk	Block randomization using tables of random numbers	Low Risk	Block randomization using random numbers tables is acceptable
	Allocation Concealment	Unclear Risk	Details not provided	Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011)
Performance Bias	Blinding of participants personnel	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Detection Bias	Blinding of outcome assessment	High Risk	Blinding of assessors not possible	High Risk	Blinding of assessors not possible
Attrition Bias	Incomplete Outcome Data	High Risk	Only completers included	Low Risk	10 year follow up only lost 14 patients (all died)
Reporting Bias	Selective Reporting	High Risk	Data not fully reported	Low Risk	All data reported except for dead patients
Other Bias		High Risk	High Dropout rate	Low Risk	14 % after 10 years (due to death) is a low drop-out rate
Selection Bias	Group Similarity at baseline	Low Risk	Similar Characteristics at Baseline	Low Risk	No longer recommended in Cochrane

					tool (Higgins, 2011)
Performance Bias	Co-interventions	Low Risk	None	Low Risk	Same
Performance Bias	Compliance	High Risk	Not monitored in conservative treatment group	Unclear Risk	Same
Intention to Treat Analysis		Low Risk	ITT performed	Low Risk	Same
Detection Bias	Timing of outcome	Low Risk	Similar for both groups	Low Risk	Same

Table 3

Risk Assessment in Brown (2012) as judged by Zaina et al (2016) vs this manuscript.

Bias Domain	Source of Bias	Zaina Risk Assessment	Support for Judgement	This manuscript Risk Assessment	Support for Judgement
Selection Bias	Random Sequence generation	Low Risk	Determined by independent statistician	Low Risk	Determined by independent statistician
	Allocation Concealment	Low Risk	Neither physician or participant aware	Low Risk	Neither physician or participant aware
Performance Bias	Blinding of participants personnel	Low Risk	Both groups received skin anesthesia	Low Risk	Both groups received skin anesthesia
Detection Bias	Blinding of outcome assessment	Low Risk	All outcomes reported	Low Risk	All outcomes reported
Attrition Bias	Incomplete Outcome Data	Low Risk	All outcomes reported	Low Risk	All outcomes reported
Reporting Bias	Selective Reporting	Unclear Risk	Insufficient details given	Low Risk	No justification given for "unclear risk"
Other Bias		High Risk	No further details available	Low Risk	No justification given for "unclear risk"
Selection Bias	Group Similarity at baseline	Low Risk	Similar characteristics at baseline	Low Risk	No longer recommended in Cochrane tool (Higgins, 2011)
Performance Bias	Co-interventions	Low Risk	None	Low Risk	Same
Performance Bias	Compliance	Low Risk	Similar for both groups	Low Risk	Same
Intention to Treat Analysis		Unclear Risk	ITT performed	Low Risk	Same. No Justification given for "Unclear risk"
Detection Bias	Timing of outcome	Unclear Risk	Similar for both groups	Low Risk	Same. No justification given for "Unclear risk"

Table 4

Risk Assessment in Malmivaara (2007) as judged by Zaina et al (2016) vs this manuscript.

Bias Domain	Source of Bias	Zaina Risk Assessment	Support for Judgement	This manuscript Risk Assessment	Support for Judgement
Selection Bias	Random Sequence generation	Low Risk	Computer generated random blocks	Low Risk	Computer generated random blocks
	Allocation Concealment	Unclear Risk	Details not provided	Unclear or Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011)
Performance Bias	Blinding of participants personnel	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Detection Bias	Blinding of outcome assessment	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Attrition Bias	Incomplete Outcome Data	High Risk	Analysis only performed for completers	Low Risk	Method of handling dropouts referenced (Rubin, 1976)
Reporting Bias	Selective Reporting	Low Risk	All outcomes reported	Low Risk	All outcomes reported
Other Bias		Unclear Risk	Insufficient details	Low Risk	No justification given for "unclear risk"
Selection Bias	Group Similarity at baseline	Unclear Risk	Similar characteristics at baseline	Low Risk	Similar characteristics at baseline No longer recommended in Cochrane tool (Higgins, 2011)
Performance Bias	Co-interventions	High Risk	Co-intervention unbalanced in control group (24 % supplementary exercise)	Low Risk	Bias would favor non-surgery
Performance Bias	Compliance	High Risk	Compliance not monitored in control group	Unclear Risk	No longer recommended in Cochrane tool (Higgins, 2011)
Intention to Treat Analysis		Low Risk	ITT performed	Low Risk	No longer recommended in Cochrane tool (Higgins, 2011)
Detection Bias	Timing of outcome	Low Risk	Similar for both groups	Low Risk	No longer recommended in Cochrane tool (Higgins, 2011)

Table 5

Risk Assessment in Weinstein (2008) as judged by Zaina et al (2016) vs this manuscript.

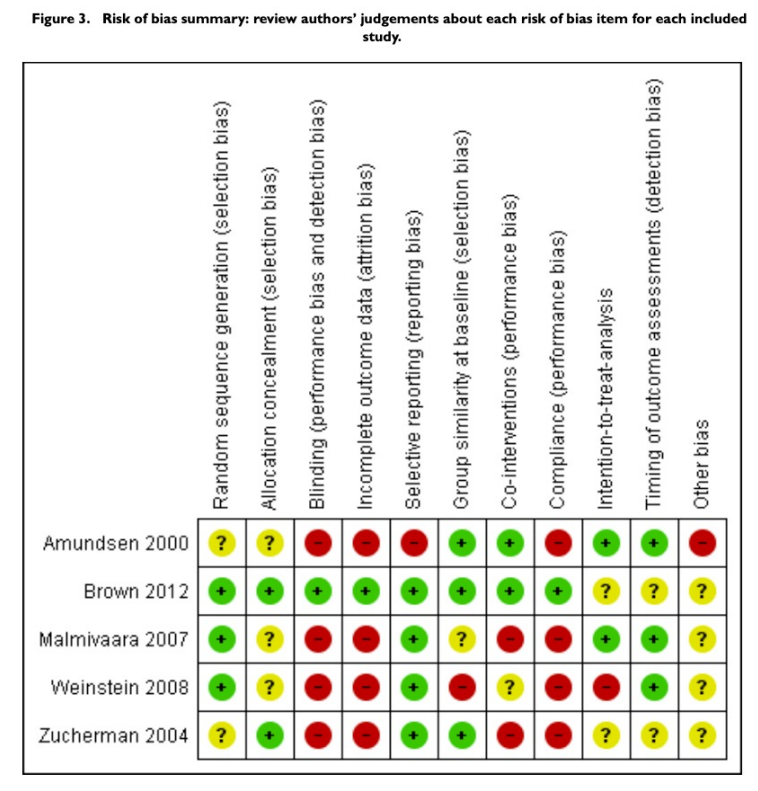
Bias Domain	Source of Bias	Zaina Risk Assessment	Support for Judgement	This manuscript Risk Assessment	Support for Judgement
Selection Bias	Random Sequence generation	Low Risk	Computer generated random blocks	Low Risk	Randomly permuted blocks
	Allocation Concealment	Unclear Risk	Details not provided	Unclear or Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011)
Performance Bias	Blinding of participants personnel	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Detection Bias	Blinding of outcome assessment	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Attrition Bias	Incomplete Outcome Data	High Risk	Large number of crossovers made ITT impossible	Low Risk	Crossovers due to failure of non-surgery treatment
Reporting Bias	Selective Reporting	Low Risk	All outcomes reported	Low Risk	All outcomes reported
Other Bias		Unclear Risk	Insufficient details	Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011
Selection Bias	Group Similarity at baseline	High Risk	Worst pain, function and disability at baseline in surgery group	Low Risk	Similar characteristics at baseline for randomized group No longer recommended in Cochrane
					tool (Higgins, 2011)
Performance Bias	Co-interventions	Unclear Risk	No co-intervention unbalance	Low Risk	No co-interventions. Non-surgery was "usual care"
Performance Bias	Compliance	High Risk	Compliance not monitored in control group	Unclear or low Risk	No longer recommended in Cochrane tool (Higgins, 2011)
Intention to Treat Analysis		High Risk	Large number of crossovers made ITT impossible	Low Risk	ITT performed in randomized group. Crossovers due to failure of non-surgery. No longer recommended in Cochrane tool (Higgins, 2011)
Detection Bias	Timing of outcome	Low Risk	Similar for both groups	Low Risk	No longer recommended in Cochrane tool (Higgins, 2011)

Table 6

Risk Assessment in Zucherman (2004) as judged by Zaina et al (2016) vs this manuscript.

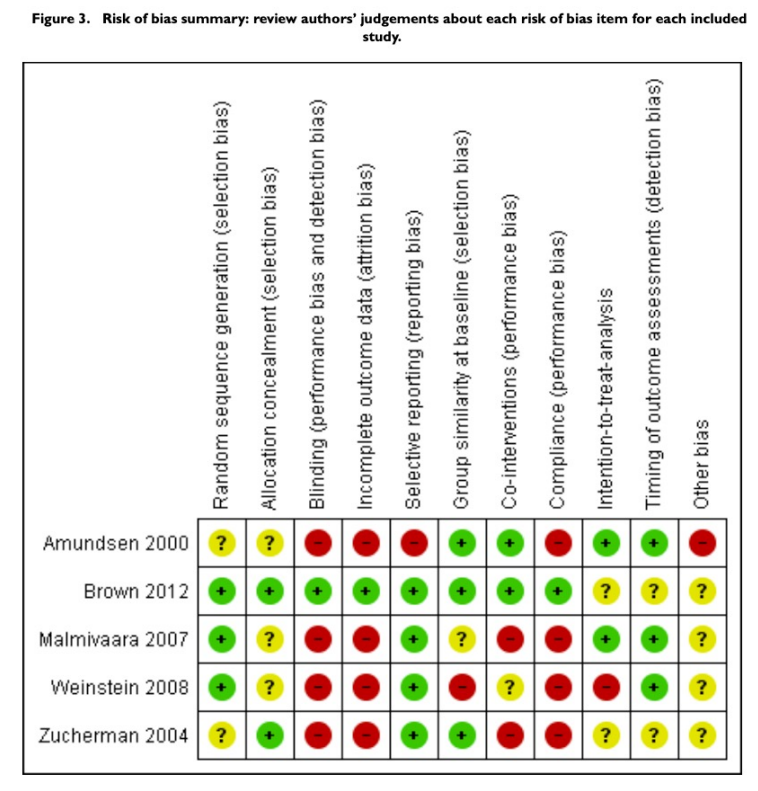
Bias Domain	Source of Bias	Zaina Risk Assessment	Support for Judgement	This manuscript Risk Assessment	Support for Judgement
Selection Bias	Random Sequence generation	Unclear Risk	Block randomization by surgical center	Low Risk	Randomly permuted blocks
	Allocation Concealment	Low Risk	Received randomization from treatment center via phone call	Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011)
Performance Bias	Blinding of participants personnel	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Detection Bias	Blinding of outcome assessment	High Risk	Blinding of participants not possible	High Risk	Blinding of participants not possible
Attrition Bias	Incomplete Outcome Data	High Risk	Only data from completers used	Low Risk	Majority of patients lost to follow up was at 2 years and due to death. Still 93% completion.
Bias	Reporting		reported		reported
Other Bias		Unclear Risk	Not clear	Low Risk	Quality (details) of reporting is not directly related to bias (Higgins, 2011)
Selection Bias	Group Similarity at baseline	Low Risk	Similar groups at baseline	Low Risk	Similar groups as baseline. No longer recommended
					in Cochrane tool (Higgins, 2011)
Performance Bias	Co-interventions	High Risk	Co-interventions not standardized and not properly described	Low Risk	No co-interventions. Control group was ESI only.
Performance Bias	Compliance	High Risk	Compliance not monitored in control group	Low Risk	How do you receive an ESI and not comply? No longer recommended in Cochrane tool (Higgins, 2011)
Treat Analysis				Risk	recommended in Cochrane tool (Higgins, 2011)
Detection Bias	Timing of outcome	Unclear Risk	Similar timing for both groups	Low Risk	If the groups had similar timing, why is this "unclear"? No longer recommended in Cochrane tool (Higgins, 2011)

HOW DOES THIS CHANGE THE ANALYSIS?



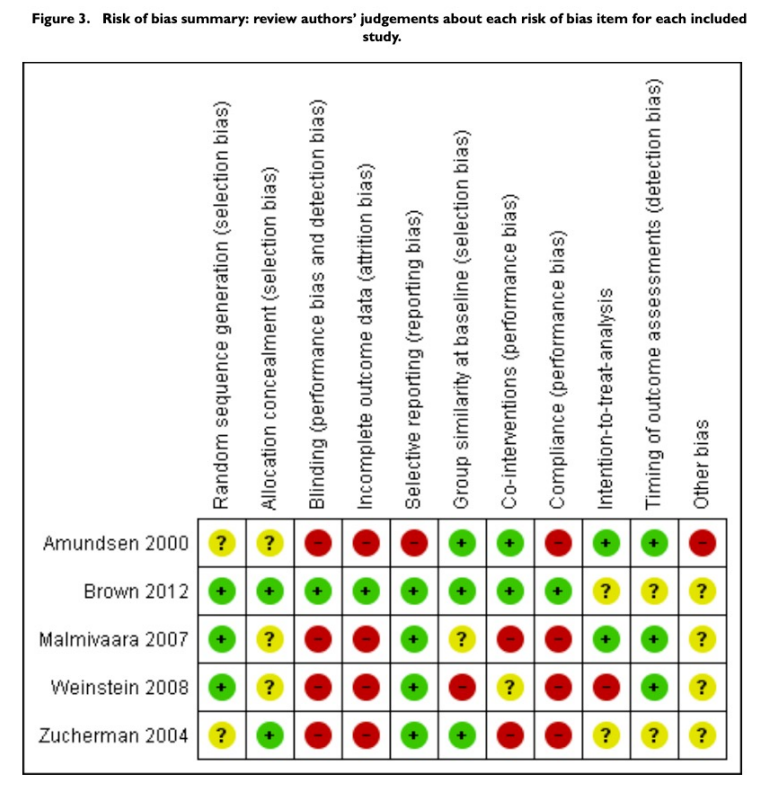
- Under "allocation concealment" Zaina assesses this as "unclear risk"
  - Justifies it with "details not provided"
- Higgins specifically states
  - "assess the risk of bias in trial results, not the quality of reporting"
- Therefore, I assess this as "low risk"

HOW DOES THIS CHANGE THE ANALYSIS?



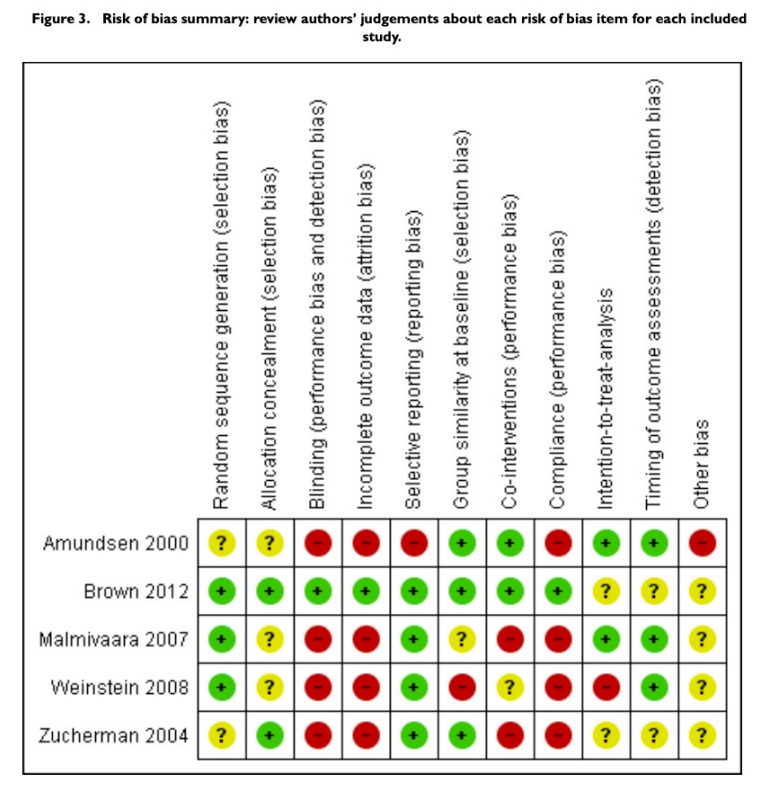
- Both Zaina and I assess "blinding" as "high risk"
- Higgins clearly states:
  - "blinding may not be feasible in many non-drug trials, and it would not be reasonable to consider the trial a low quality because of the absence of blinding"
- This issue was also thoroughly addressed by O'Toole and Traynelis
  - Journal of Neurosurgey: Spine 14:555-560, 2011

HOW DOES THIS CHANGE THE ANALYSIS?



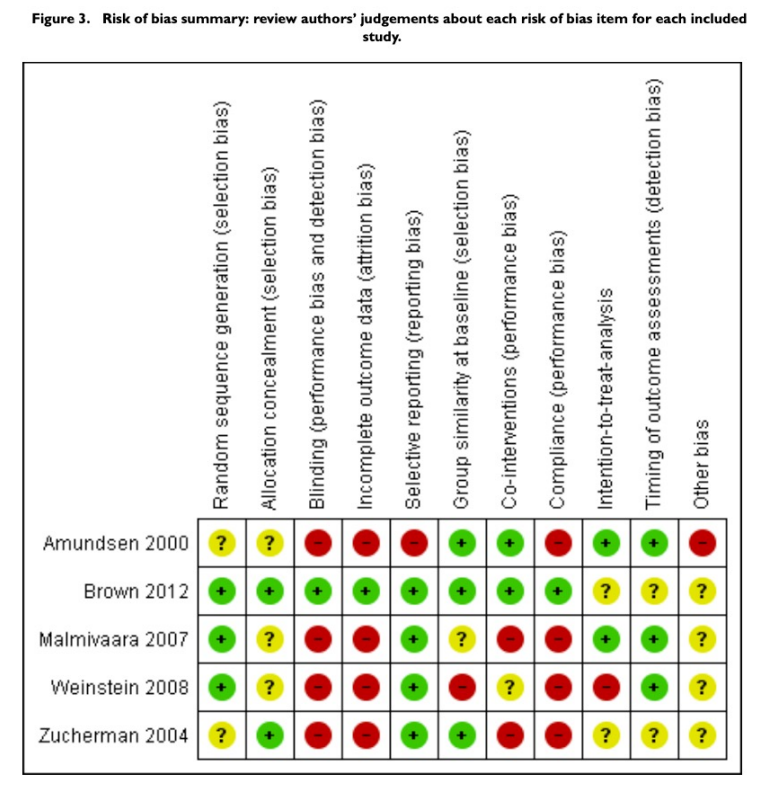
- Zaina assesses high risk of bias in Amundson
  - "only completers were assessed"
- Reality
  - 14 of 100 patients were lost to follow–up at 10 years due to DEATH
  - Not reasonable to judge this as high risk
  - This would make it impossible for ANY studies of the elderly EVER to reach high quality data
- The same argument is true for "other bias"

HOW DOES THIS CHANGE THE ANALYSIS?



- Zaina rates compliance as "high risk"
  - compliance in the non-surgical group was not closely monitored
- BUT…this analysis of surgery to non-surgery,, not to a specific method of non-surgery
  - Compliance therefore is irrelevant in that "no treatment" is an equally valid form on non-surgery
- I assessed this as low risk

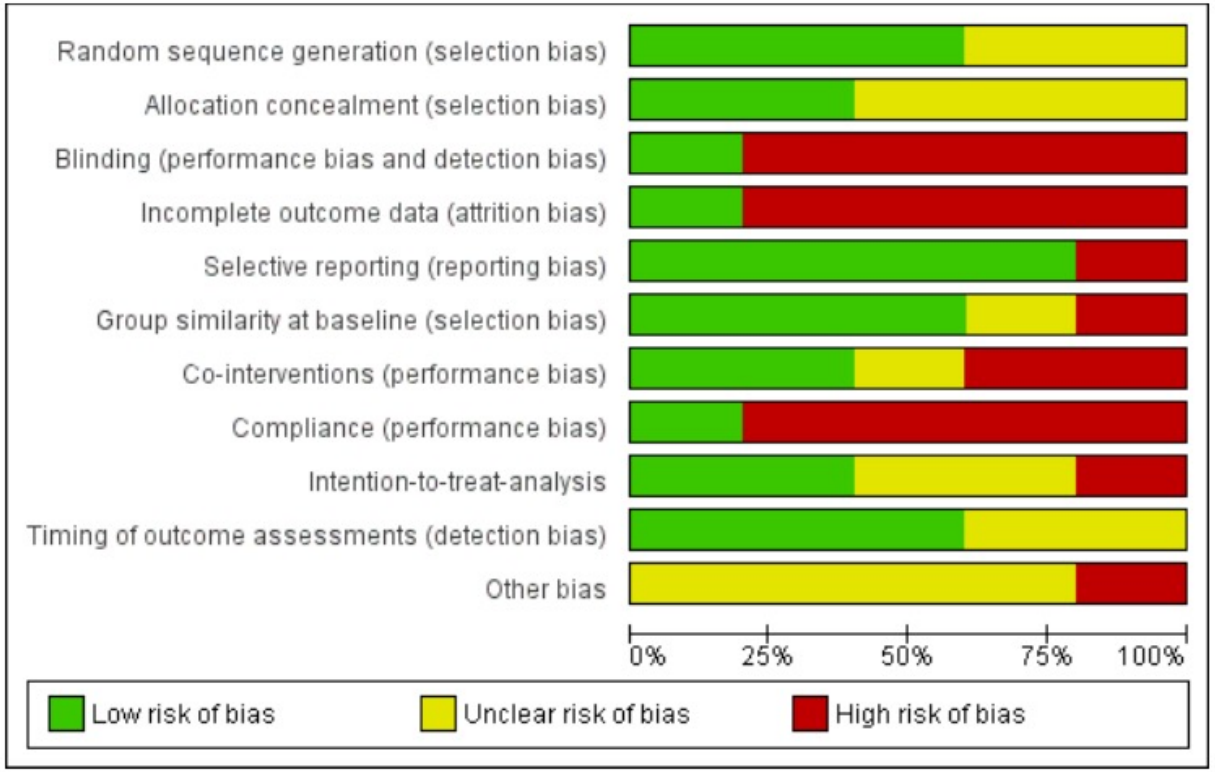
HOW DOES THIS CHANGE THE ANALYSIS?



- Under "other potential sources of bias"
  - Zaina in the text specifically states
    - "We found no further risks of bias"
- However, in Zaina's figures 2 and 3 Zaina reports this as "unclear risk of bias"
  - Justifies it as "insufficient details given"

ZAINA, 2016

Figure 2. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



FESSLER, 2021

Figure 1: Risk of bias graph. Author's judgement about each risk of bias item presented as percentages across all included studies utilizing modifications suggested by Furlan[3].

RSG-Random Sequence Generation, Bl Part-Blinding Participant, Bl Ceso-Blinding Assessments, Inc Data-Incomplete Outcome Data, Sel Rep-Selective Reporting, Other-Other Bias, Sp Sim- Group Similarity at Baseline, Co-Amt-Co-Interventions, ITT-Intention to treat analysis, Time Amt-Timing of Outcome Assessments

ZAINA, 2016

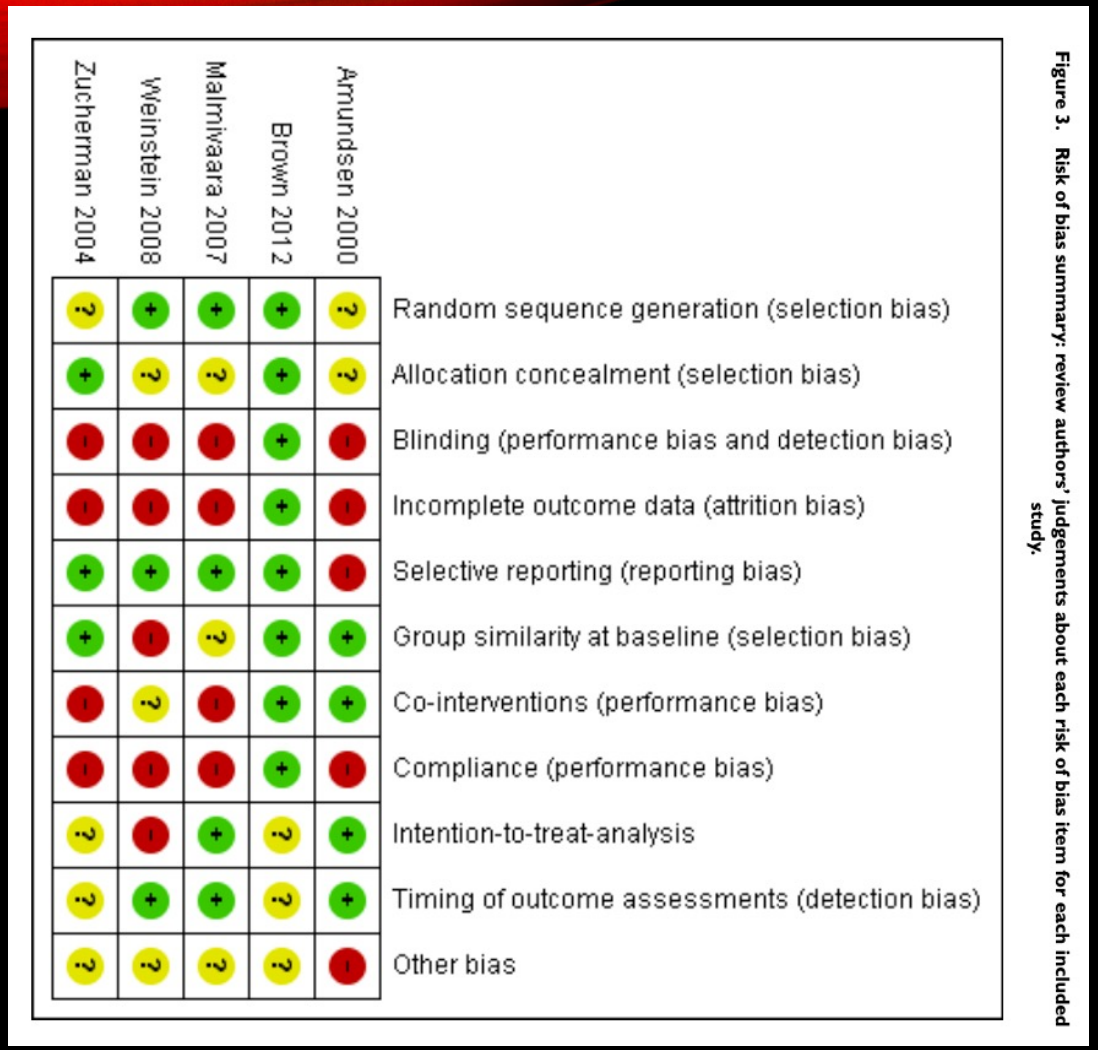
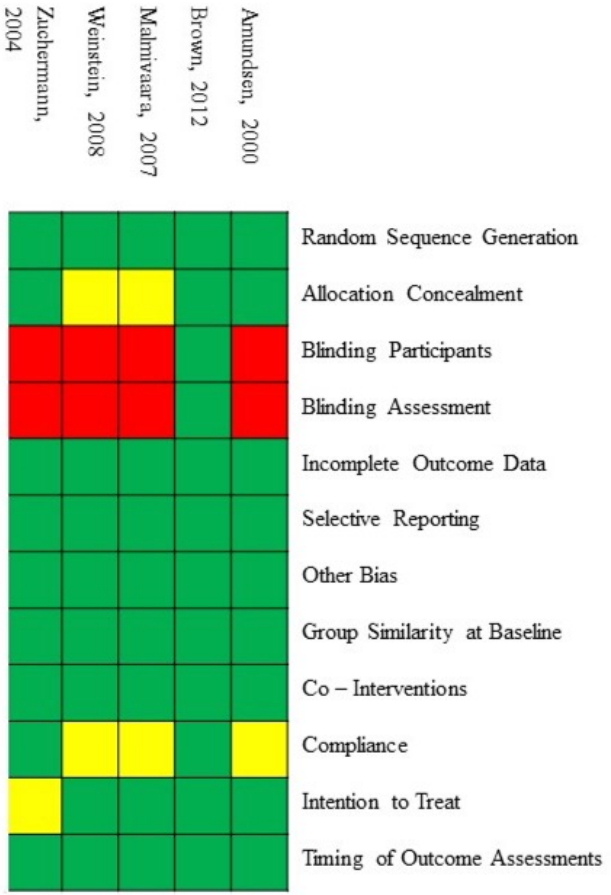


Figure 3. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.

FESSLER, 2021

Figure 2: Risk of bias summary. Author's judgements about each risk of bias item for each included study utilizing the modifications suggested by Furlan[3].



OTHER METHODOLOGICAL ISSUES

- This report includes only 5 manuscripts in total
- Performs outcomes meta-analysis on only 2, complications on only 1
  - Outcome: Meta-analysis on 2 studies is meaningless
  - Complications: Reports on only 1 study
    - This is certainly not a "systematic review"
- Studies in meta-analysis should have relatively comparable methodology
  - These five range from MILD to open fusion/instrumentation
  - Follow-up ranges from 6 weeks to 10 years

WHAT IS EVIDENCE BASED MEDICINE?

- "Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients." (Sackett et al. BMJ 1996)
  - It is intended to create a hierarchy of evidence, but not to rationalize a failure to evaluate lower level evidence if that evidence is the best available
- IN THIS CONTEXT THE COCHRANE METHODOLOGY IS A MIS-APPLICATION OF EVIDENCE-BASED MEDICINE PRINCIPLES IN THE DOMAIN OF MOST SURGICAL TRIALS

IMPACT

- What is the potential impact of misleading and potentially biased reports on patient care and public health decisions?
  - Lumbar stenosis is a painful condition which leads to decreased mobility, depression, compromised quality of life, accelerated morbidity and mortality
- Any practicing physician sees first hand the rapid, positive impact of decompression
  - This is verified by these 5 randomized trials
- Physicians and/or government or private payers who read the report of Zaina and deny their patients surgical decompression would certainly be doing them a disservice.

SUMMARY

- Cochrane analyses have come to be regarded as the state-of-the-art method of unbiased evaluation of published data
- We report
  - a detailed evaluation of the Cochrane analysis of Zaina evaluating surgical vs non-surgical treatment of lumbar stenosis,
  - the prospective, randomized, controlled studies in their analysis
  - the Cochrane analysis tools themselves

CONCLUSIONS

- Unlike what is widely considered the most objective evaluation of clinical reports
  - What is revealed is a remarkably subjective methodology
  - Which is strongly subject to authors biases
- The conclusions of the original published manuscripts cited in the report of Zaina are much more reliable than the Cochrane analysis itself
- Like all published manuscripts, Cochrane analysis must be reviewed with appropriate skepticism
- Before accepting the validity of the conclusions, readers must:
  - Closely analyze and consider the quality of the analyzed data
  - The statistics utilized
  - The review and Cochrane methodology itself

THANK YOU