Framework for Interpreting Clinical Trials

Objectives

- Importance of randomized clinical trials
- Questions to ask when reading a clinical trial report
- Common sources of bias
- Interpreting and applying results
- Limitations/challenges

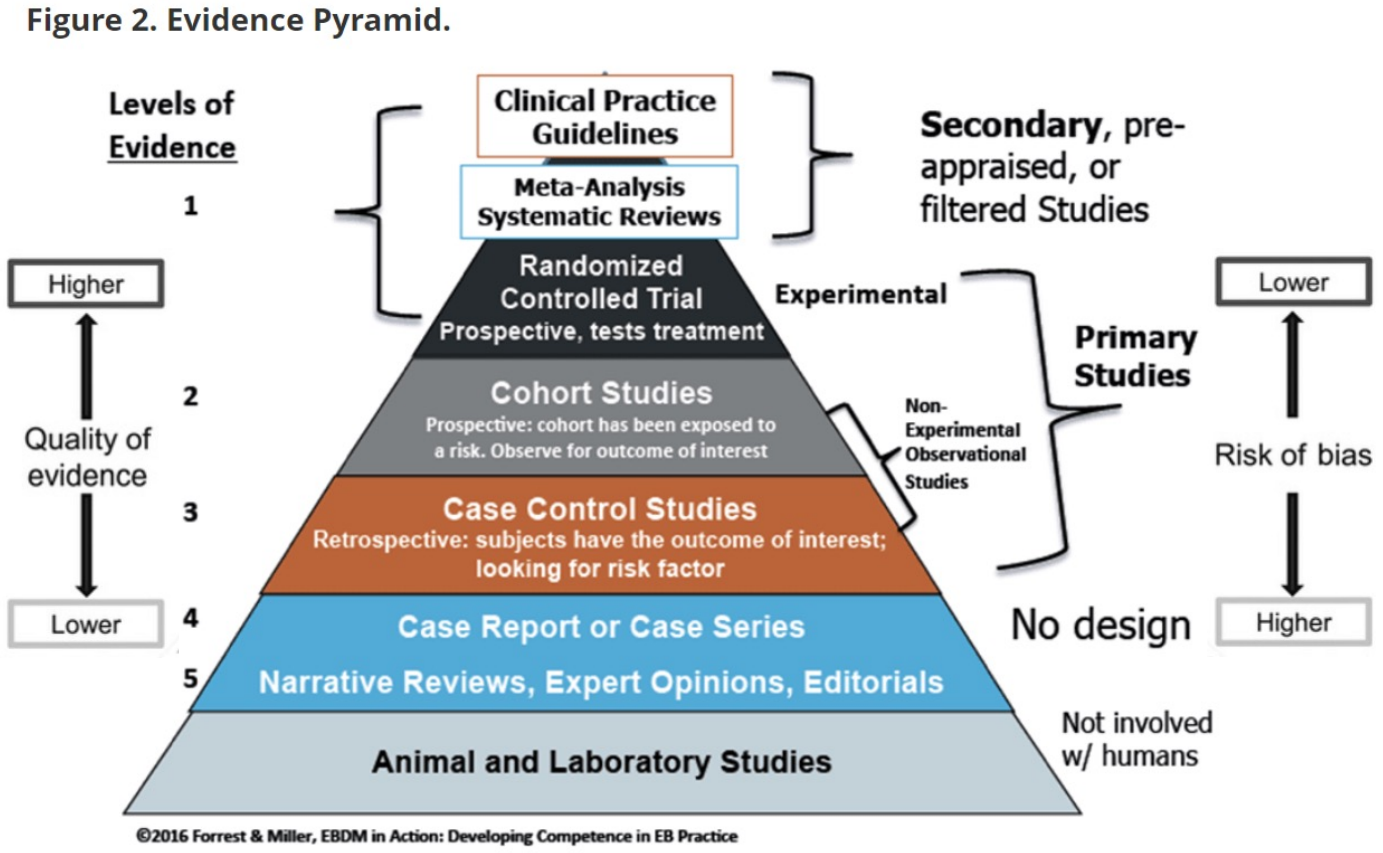
*Superiority RCTs

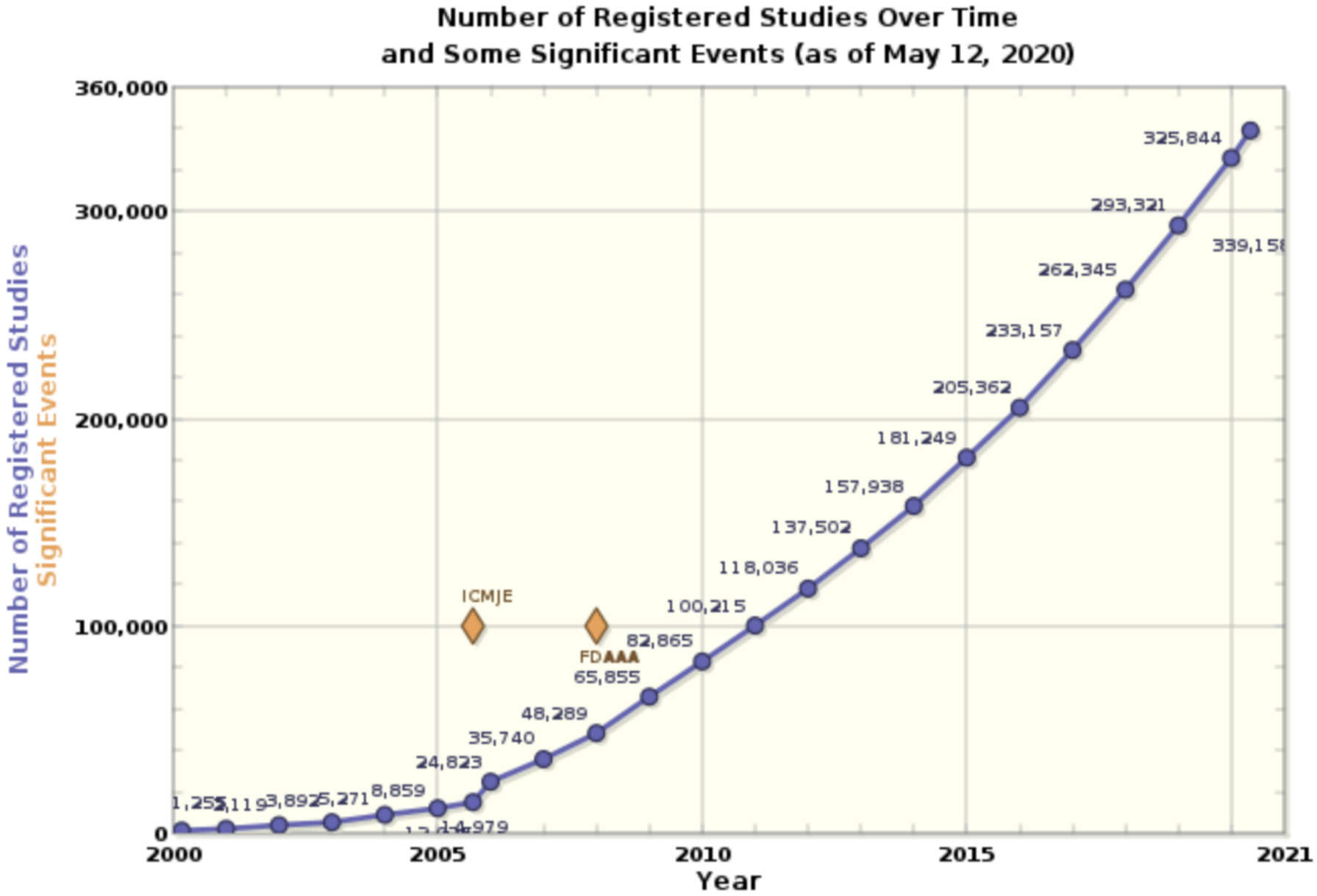
*Statistics, Noninferiority/Adaptive Trials won’t be discussed

Randomized Clinical Trials (Superiority)

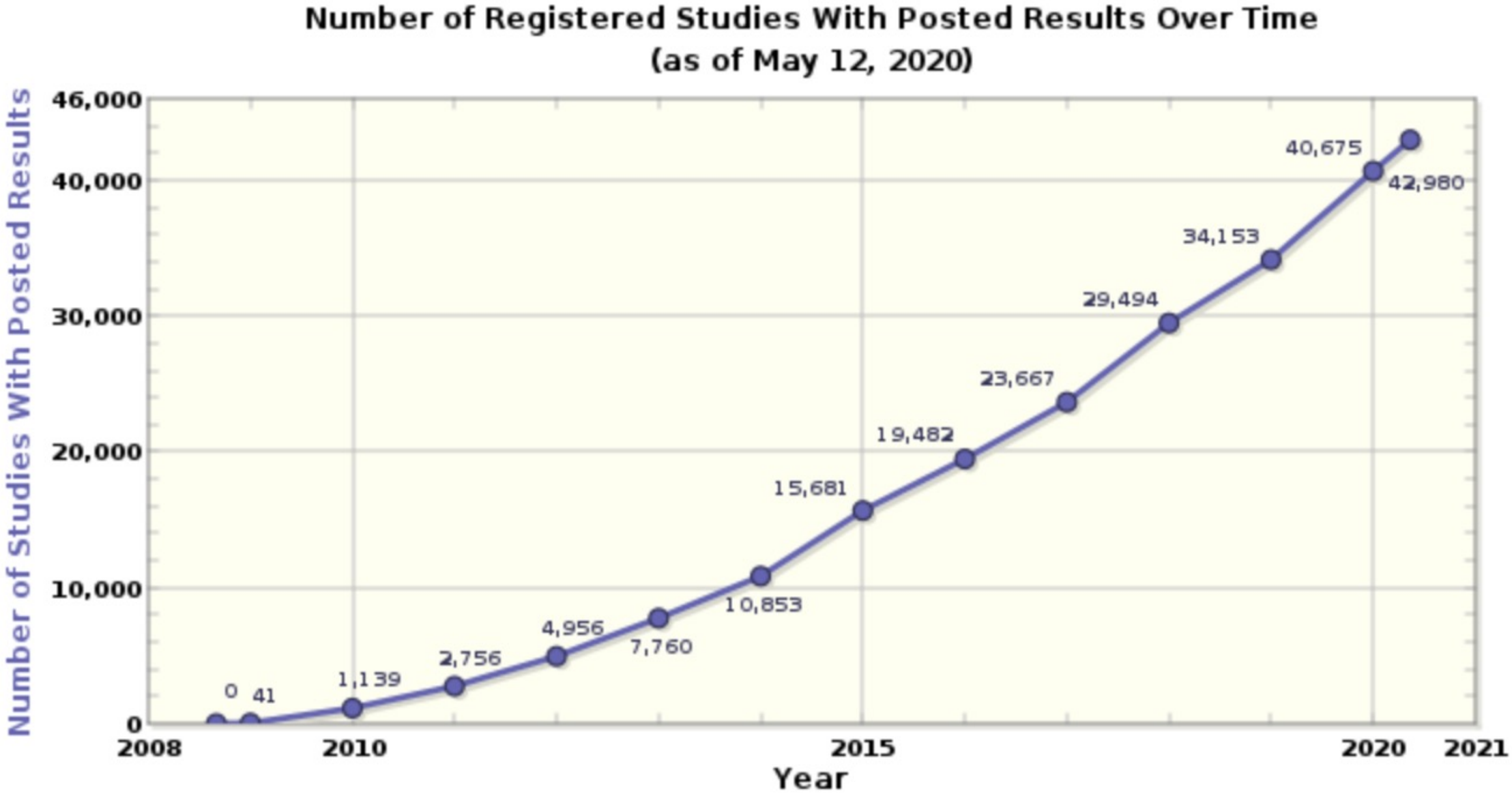
- Prospective human research study assessing the impact of an experimental intervention
  - Drug
  - Device/Procedure
  - Preventive, Screening, or Diagnostic Approach
  - Behavioral Intervention

- Best evidence for causality





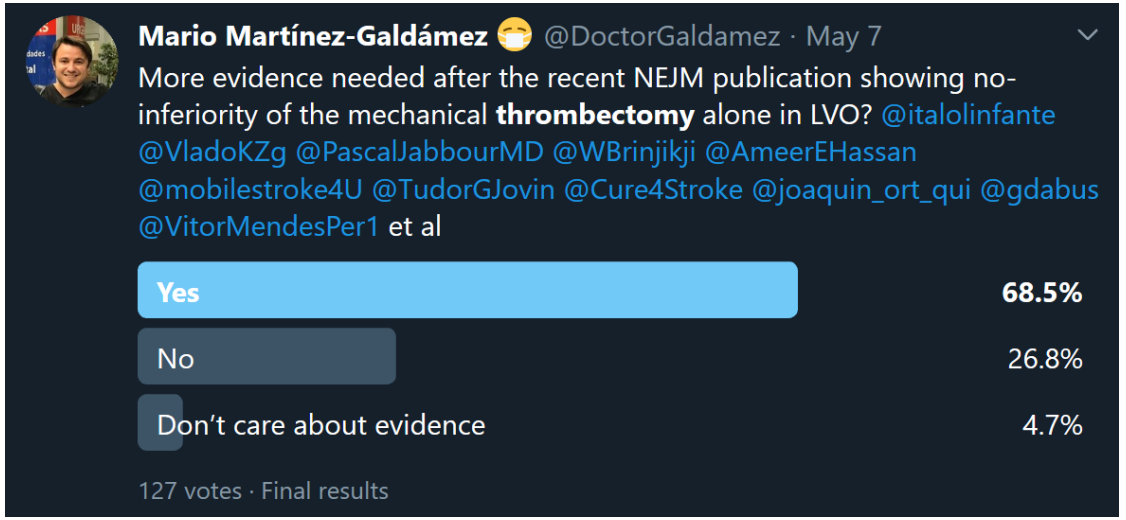
Source: https://ClinicalTrials.gov



Source: https://ClinicalTrials.gov

EBM and Clinical Practice

- Most clinical decisions/practice ≠ high-quality evidence (RCTs)
  - Observational studies or physiological/theoretical basis
  - Historical / institutional bias – “that’s just what we do”
  - Improper incentives
    - Fee-for-service
  - Slow adoption or implementation



A Randomized Trial of Epidural Glucocorticoid Injections for Spinal Stenosis

Janna L. Friedly, M.D., Bryan A. Comstock, M.S., Judith A. Turner, Ph.D., Patrick J. Heagerty, Ph.D., Richard A. Deyo, M.D., M.P.H., Sean D. Sullivan, Ph.D., Zoya Bauer, M.D., Ph.D., Brian W. Bresnahan, Ph.D., Andrew L. Avins, M.D., M.P.H., Srdjan S. Nedeljkovic, M.D., David R. Nerenz, Ph.D., Christopher Standaert, M.D., et al.

BACKGROUND

Epidural glucocorticoid injections are widely used to treat symptoms of lumbar spinal stenosis, a common cause of pain and disability in older adults. However, rigorous data are lacking regarding the effectiveness and safety of these injections.

CONCLUSIONS

In the treatment of lumbar spinal stenosis, epidural injection of glucocorticoids plus lidocaine offered minimal or no short-term benefit as compared with epidural injection of lidocaine alone. (Funded by the Agency for Healthcare Research and Quality; ClinicalTrials.gov number, NCT01238536.)

N Engl J Med 2014; 371:11-21

General Structure of Trial Report

CONSORT
TRANSPARENT REPORTING of TRIALS

- Introduction
  - Background/Rationale
  - Hypothesis
- Methods
  - Design
  - Population studied
  - Procedures/Intervention
  - Outcomes (efficacy, safety)
  - Statistics
- Results
  - Participant enrollment (Figure 1) & baseline characteristics (Table 1)
  - Primary (efficacy) outcome
    - appropriate statistical power
  - Secondary (efficacy) outcomes
  - Safety outcomes
- Discussion
  - Conclusions
  - Results in context of literature
  - Limitations/Strengths
  - Summary

Clinical Trial Interpretation

- Goal of RCT: isolate & understand intervention treatment effect

1. What were potential sources of bias? (prognostic factors comparable between groups?)
    - Baseline: randomization? enrollment?
    - Duration: blinding?
    - Completion: follow-up complete? trial stopped early? analysis in randomized groups (ITT)?
2. What are the results?
3. How do I apply the results?

Users' Guides for an Article About Therapy

How serious was the risk of bias?

- Did intervention and control groups start with the same prognosis?
- Were patients randomized?
- Was randomization concealed?
- Were patients in the study groups similar with respect to known prognostic factors?
- Was prognostic balance maintained as the study progressed?
- To what extent was the study blinded?
- Were the groups prognostically balanced at the study's completion?
- Was follow-up complete?
- Were patients analyzed in the groups to which they were randomized?
- Was the trial stopped early?

What are the results?

- How large was the treatment effect?
- How precise was the estimate of the treatment effect?

How can I apply the results to patient care?

- Were the study patients similar to my patient?
- Were all patient-important outcomes con- sidered?
- Are the likely treatment benefits worth the potential harm and costs?

Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, Third Edition

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 MARCH 19, 2020 VOL. 382 NO. 12

Surgery versus Conservative Care for Persistent Sciatica Lasting 4 to 12 Months

Chris S. Bailey, M.D., Parham Rasoulinejad, M.D., David Taylor, M.D., Keith Sequeira, M.D., Thomas Miller, M.D., Jim Watson, M.D., Richard Rosedale, P.T., Stewart I. Bailey, M.D., Kevin R. Gurr, M.D., Fawaz Siddiqi, M.D., Andrew Glennie, M.D., and Jennifer C. Urquhart, Ph.D.

https://www.nejm.org/do/10.1056/NEJMdo005685/full/?requestType=popUp&relatedArticle=10.1056%2FNEJMoa1912658

N Engl J Med 2020; 382:1093-1102

Groups Start With Same Prognosis?

- Randomized?
  - Prognostic factors (determinants of outcome) balanced
    - Known
    - Unknown
    - Unbalanced → underestimate OR overestimate treatment effect
    - Randomization NOT 100% successful in balancing groups
      - Unbalanced → analyses adjusted for differences?

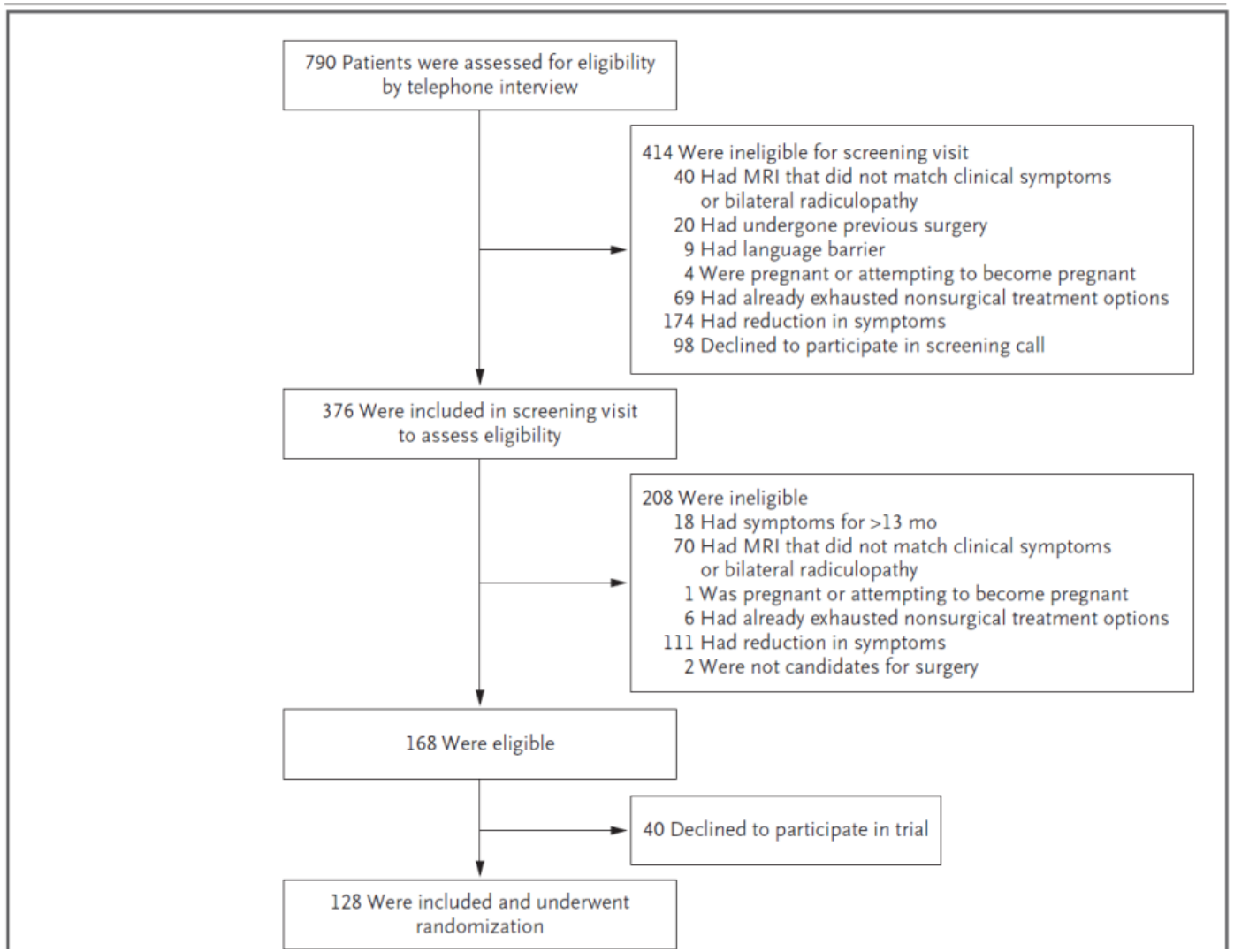
Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

Characteristic	Surgical Group (N=64)	Nonsurgical Group (N=64)
Age — yr	38.0±8.3	37.1±11.9
Body-mass index†	27.1±5.6	27.4±10.5
Female sex — no. (%)	27 (42)	25 (39)
Receipt of workers' compensation — no. (%)	4 (6)	2 (3)
Primary symptom — no. (%)		
Leg pain	49 (77)	55 (86)
Back pain	3 (5)	3 (5)
Both	12 (19)	6 (9)
Neurologic symptom — no. (%)		
Numbness	48 (75)	45 (70)
Tingling	32 (50)	30 (47)
Weakness	15 (23)	13 (20)
Any neurologic deficit — no. (%)		
Asymmetric decrease in reflexes	36 (56)	27 (42)
Asymmetric decrease in sensory response	41 (64)	37 (58)
Asymmetric weakness in motor response	20 (31)	19 (30)
Type of disk herniation — no. (%)		
Protruding	14 (22)	9 (14)
Extruded	46 (72)	47 (73)
Sequestered	6 (9)	10 (16)
Disk herniation level — no. (%)		
L4–L5	17 (27)	20 (31)
L5–S1	47 (73)	44 (69)
Pain-intensity score‡		
Leg	7.7±2.0	8.0±1.8
Back	6.7±2.6	6.5±2.8
Oswestry Disability Index§	49.7±15.8	50.2±15.9
SF-36 score¶		
Physical Component Summary	26.4±7.6	25.3±6.7
Mental Component Summary	36.0±13.8	36.2±12.4

N Engl J Med 2020; 382:1093-1102

Groups Start With Same Prognosis?

- Enrollment?
  - Screened → eligible → randomized
  - Why didn't screened/eligible get randomized?
    - Those randomized unique/different from targeted population? → bias
- Group allocation concealed?
  - Recruiter may enroll those more or less likely to benefit to treatment or control group → bias



N Engl J Med 2020; 382:1093-1102

Prognostic Balance Maintained?

- (Double) Blinding?
  - Enroller ≠ surgeon ≠ outcome adjudicator
  - Cointervention → glucocorticoid injection

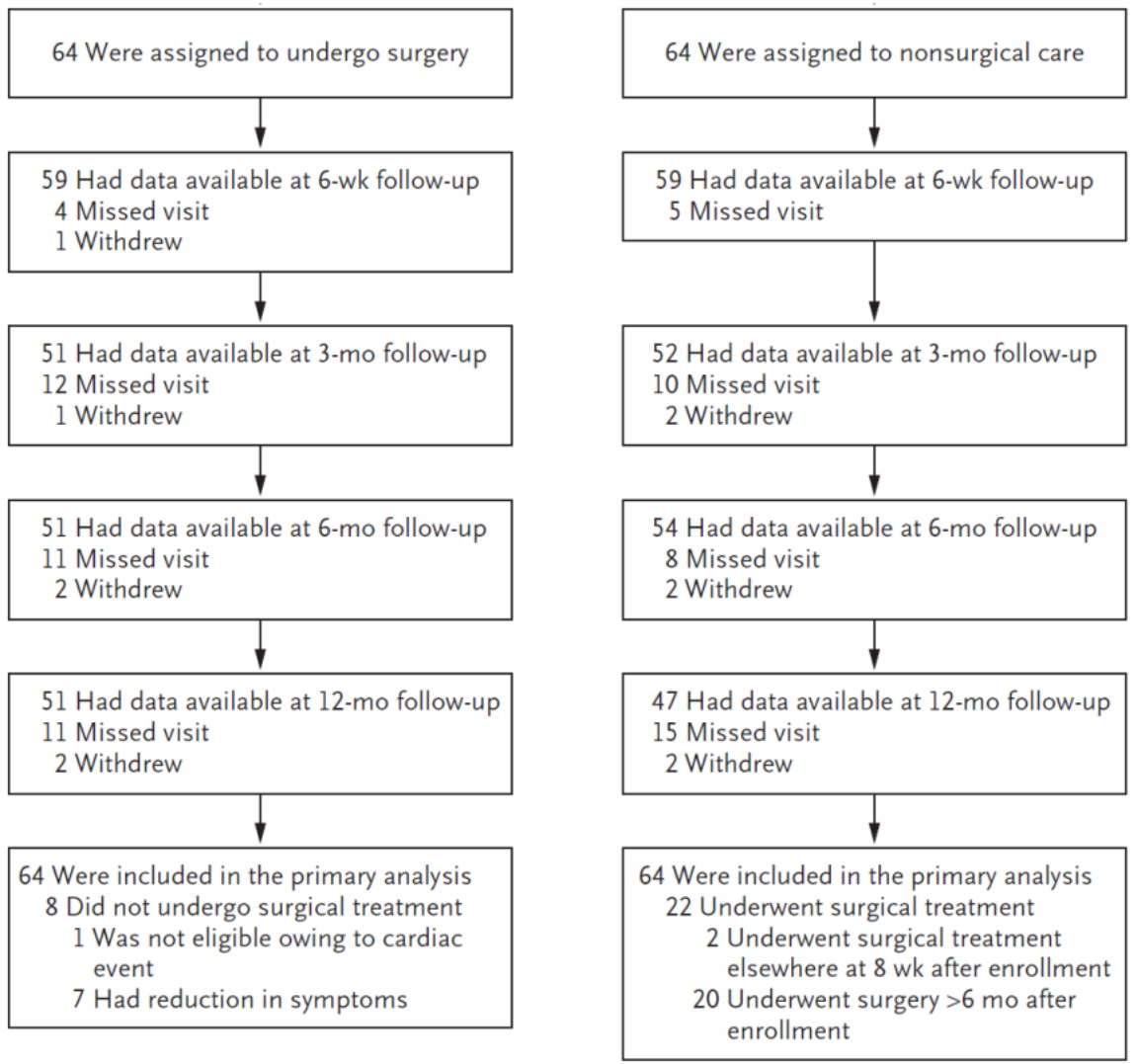
Five Groups That Should, if Possible, Be Blind to Treatment Assignment

Patients	To avoid placebo effects
Clinicians	To prevent differential administration of therapies that affect the outcome of interest (cointervention)
Data collectors	To prevent bias in data collection
Adjudicators of outcome	To prevent bias in decisions about whether or not a patient has had an outcome of interest
Data analysts	To avoid bias in decisions regarding data analysis

Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, Third Edition

Prognostic Balance at Completion?

- Follow-up complete?
  - Outcome differences between retained and lost?
  - Missing data: 20%
- Crossovers
- Trial stopped early?
  - Statistical power compromise
  - Follow-up provide critical info?
- Patients analyzed in groups to which they were randomized? (ITT)
  - Protects baseline comparability
  - Reduces bias
  - Ignores crossovers, nonadherence, dropouts



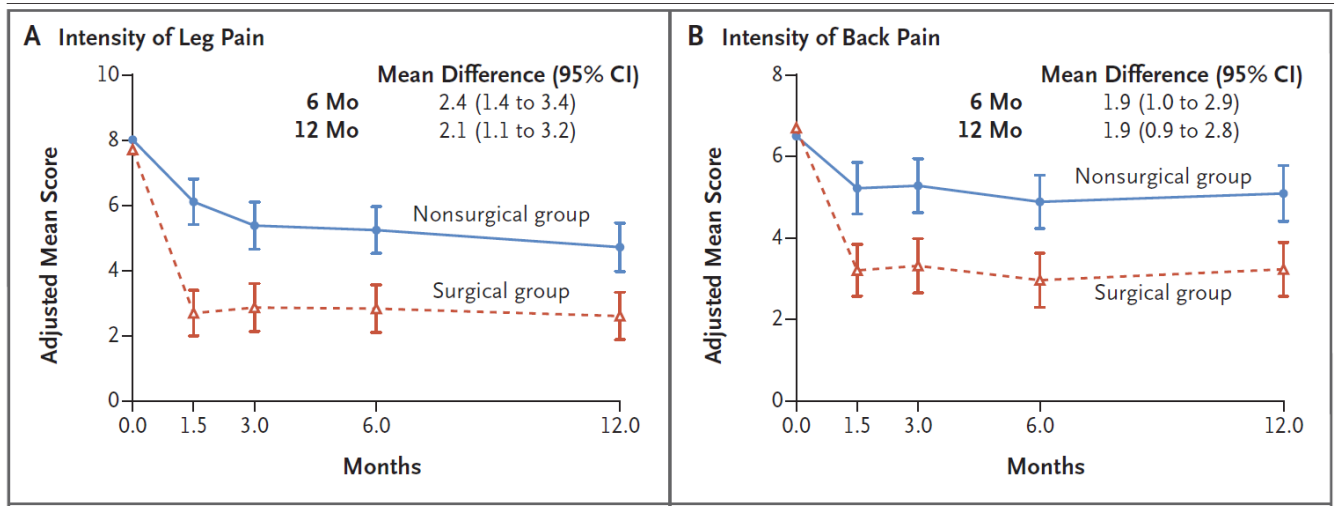
N Engl J Med 2020; 382:1093-1102

What are the primary results?

- How large is the treatment effect?
  - Absolute vs Relative
  - NNT = 1/ARR
- How precise is the treatment effect estimate?
  - Point estimate
  - Confidence intervals
  - Influenced by sample size, events
- 1° outcome → statistically powered
  - Minimal sample size:
    - Estimated effect size
    - Variability
    - False positive risk (α)
    - False negative risk (β); power=1- β
    - Event rate
    - Dropout rate (attrition)

Table 2. Primary and Secondary Outcomes.*

Outcome	Surgical Group		Nonsurgical Group		Difference (95% CI)
	No. of Patients	Value	No. of Patients	Value	
Primary outcome					
Intensity score for leg pain at 6 mo†	51	2.8±0.4	54	5.2±0.4	2.4 (1.4 to 3.4)

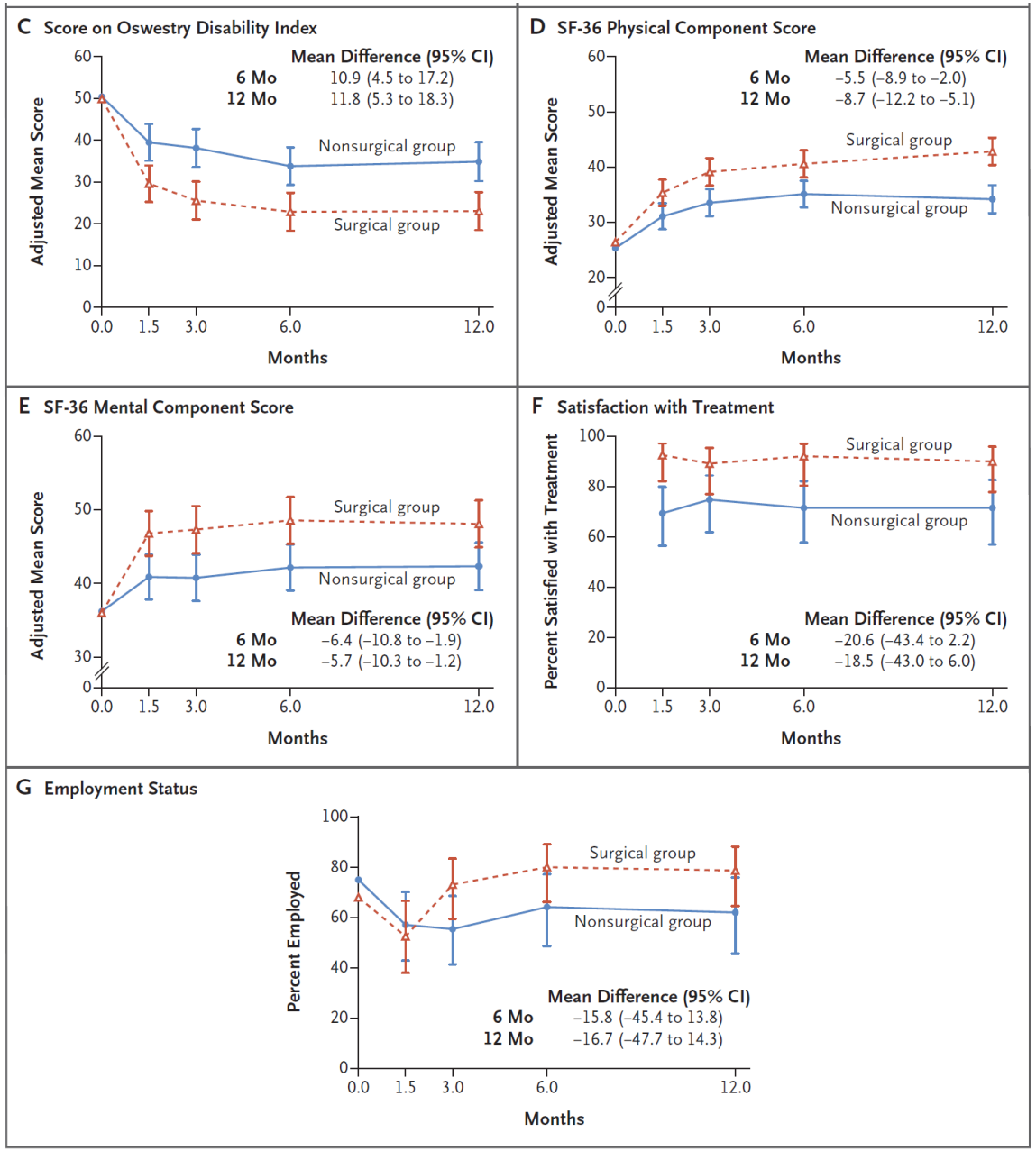


N Engl J Med 2020; 382:1093-1102

Secondary efficacy & safety outcomes

Table 3. Surgery-Related Adverse Events at 1 Year.*

Adverse Events	Surgical Group	Nonsurgical Group
Intention-to-treat analysis		
No. of patients	64	64
No. of patients with at least 1 event (%)	4 (6)	5 (8)
No. of events (event rate)	5 (0.08)	6 (0.09)
As-treated analysis		
No. of patients	56	22
No. of patients with at least 1 event (%)	4 (7)	5 (23)
No. of events (event rate)	5 (0.09)	6 (0.27)
Adverse events — no.		
Dural tear	0	1
Superficial wound infection	2	1
Nerve-root injury	0	1
Postoperative adjacent level condition	0	1
New-onset postoperative neuropathic pain	1	2
Recurrent herniation after surgery†		
No further surgery performed	1	0
Revision surgery performed	1	0



N Engl J Med 2020; 382:1093-1102

What about subgroup analyses?

- Ancillary evidence?
- ↑ risk of false positive w/ ↑ # subgroup analyses
- Statistics

"The essence of tragedy has been described as the destructive collision of two sets of protagonists, both of whom are correct. The statisticians are right in denouncing subgroups that are formed post hoc from exercises in pure data dredging. The clinicians are also right, however, in insisting that a subgroup is respectable and worthwhile when established a priori from pathophysiological principles."

A R Feinstein, 1998

Box 3. Guidelines for Deciding Whether Apparent Differences in Subgroup Response Are Real

Issues for Individual Studies and Systematic Reviews

Can Chance Explain the Subgroup Difference?

Is the subgroup difference consistent across studies?

Was the subgroup difference one of a small number of a priori hypotheses in which the direction was accurately prespecified?

Is there a strong preexisting biological rationale supporting the apparent subgroup effect?

An Issue for Systematic Reviews Only

Is the subgroup difference suggested by comparisons within rather than between studies?

JAMA. 2014;311(4):405-411.

How can I apply the results?

- Who was studied?
  - Inclusion
    - age 18-60 years
    - lumbar radiculopathy x 4-6 months
    - posterolateral disk herniation L4-L5 / L5-S1 disk + root compression (MRI)
  - Exclusion
    - lateral disk or foraminal herniation
    - spinal stenosis
    - herniation level deformity
    - prior LS surgery
    - current treatment – epidural injection or exercise/PT

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

Characteristic	Surgical Group (N=64)	Nonsurgical Group (N=64)
Age — yr	38.0±8.3	37.1±11.9
Body-mass index†	27.1±5.6	27.4±10.5
Female sex — no. (%)	27 (42)	25 (39)
Receipt of workers' compensation — no. (%)	4 (6)	2 (3)
Primary symptom — no. (%)		
Leg pain	49 (77)	55 (86)
Back pain	3 (5)	3 (5)
Both	12 (19)	6 (9)
Neurologic symptom — no. (%)		
Numbness	48 (75)	45 (70)
Tingling	32 (50)	30 (47)
Weakness	15 (23)	13 (20)
Any neurologic deficit — no. (%)		
Asymmetric decrease in reflexes	36 (56)	27 (42)
Asymmetric decrease in sensory response	41 (64)	37 (58)
Asymmetric weakness in motor response	20 (31)	19 (30)
Type of disk herniation — no. (%)		
Protruding	14 (22)	9 (14)
Extruded	46 (72)	47 (73)
Sequestered	6 (9)	10 (16)
Disk herniation level — no. (%)		
L4–L5	17 (27)	20 (31)
L5–S1	47 (73)	44 (69)
Pain-intensity score‡		
Leg	7.7±2.0	8.0±1.8
Back	6.7±2.6	6.5±2.8
Oswestry Disability Index§	49.7±15.8	50.2±15.9
SF-36 score¶		
Physical Component Summary	26.4±7.6	25.3±6.7
Mental Component Summary	36.0±13.8	36.2±12.4

N Engl J Med 2020; 382:1093-1102

Were study patients like my patient?

- Generalizability?
- Does my patient fit these criteria?
  - Subgroup? – can be misleading
- If not, compelling reason why results do not apply to my patient?
  - How is my patient different?
    - Sicker?
    - Age?
    - Absolute benefit and risks?

* Trial patients are selected to maximize chance of benefit and minimize risk of harm

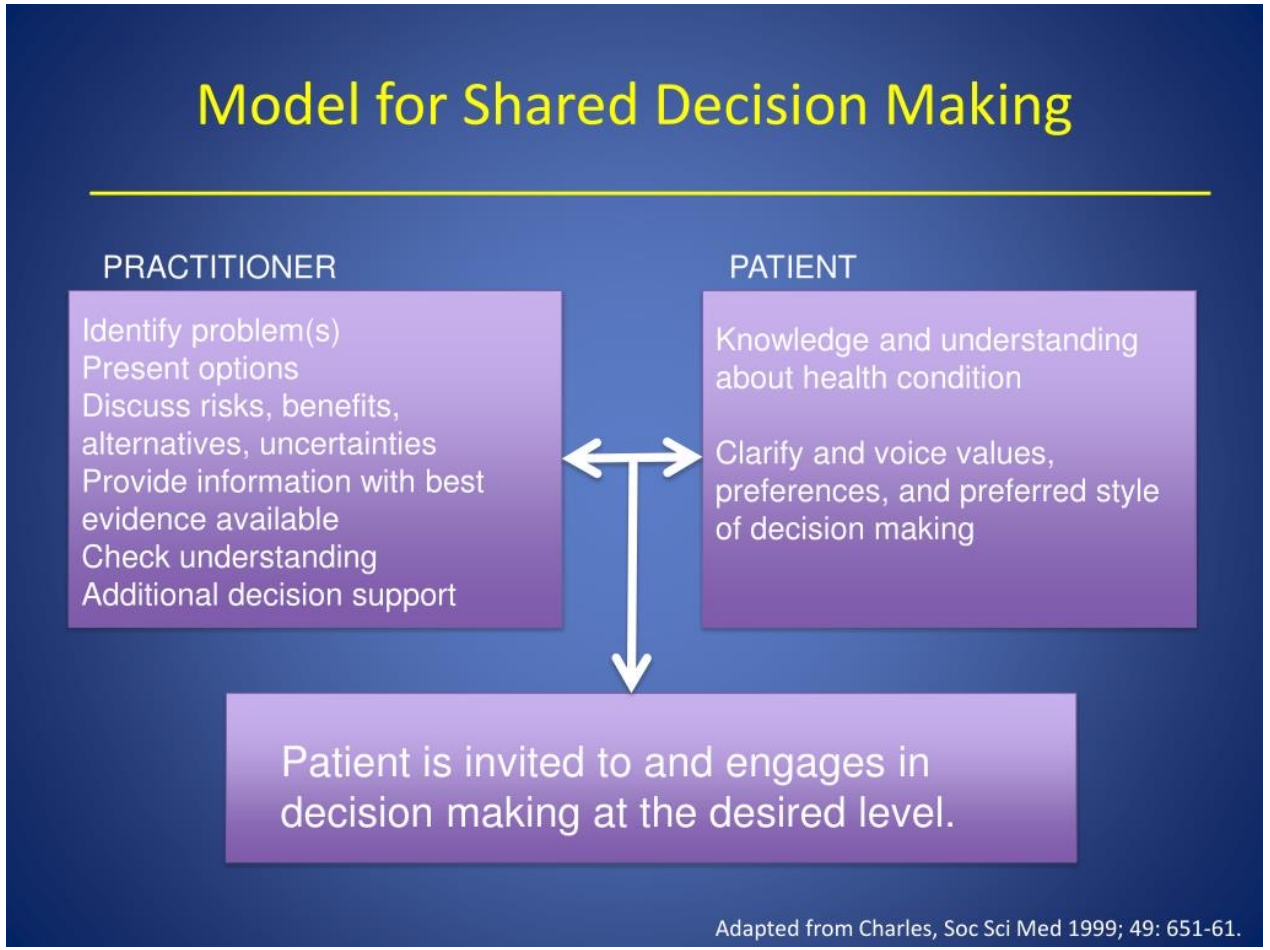
Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

Characteristic	Surgical Group (N=64)	Nonsurgical Group (N=64)
Age — yr	38.0±8.3	37.1±11.9
Body-mass index†	27.1±5.6	27.4±10.5
Female sex — no. (%)	27 (42)	25 (39)
Receipt of workers' compensation — no. (%)	4 (6)	2 (3)
Primary symptom — no. (%)		
Leg pain	49 (77)	55 (86)
Back pain	3 (5)	3 (5)
Both	12 (19)	6 (9)
Neurologic symptom — no. (%)		
Numbness	48 (75)	45 (70)
Tingling	32 (50)	30 (47)
Weakness	15 (23)	13 (20)
Any neurologic deficit — no. (%)		
Asymmetric decrease in reflexes	36 (56)	27 (42)
Asymmetric decrease in sensory response	41 (64)	37 (58)
Asymmetric weakness in motor response	20 (31)	19 (30)
Type of disk herniation — no. (%)		
Protruding	14 (22)	9 (14)
Extruded	46 (72)	47 (73)
Sequestered	6 (9)	10 (16)
Disk herniation level — no. (%)		
L4–L5	17 (27)	20 (31)
L5–S1	47 (73)	44 (69)
Pain-intensity score‡		
Leg	7.7±2.0	8.0±1.8
Back	6.7±2.6	6.5±2.8
Oswestry Disability Index§	49.7±15.8	50.2±15.9
SF-36 score¶		
Physical Component Summary	26.4±7.6	25.3±6.7
Mental Component Summary	36.0±13.8	36.2±12.4

N Engl J Med 2020; 382:1093-1102

How can I apply the results?

- All outcomes important to patients assessed?
  - Intermediate / surrogate
  - Missing relevant outcomes?
- Potential Benefits > Potential Harms/Costs
  - NNT
  - NNH
  - Patient values/preferences → shared-decision making



Potential Sources of Bias

- Selection bias (severe sciatica → < likely conservative care)
- Single (academic) center
  - Sample size too small? (risk failing to demonstrate treatment effect when one exists)
  - Multicenter trials: bigger, variation between sites, more generalizable
- Blinding of enrollers and outcome adjudicators
- Cointervention (epidural steroid injection)
- Crossover
- Missing data

N Engl J Med 2020; 382:1093-1102

Limitations/Challenges of RCTs

- Sources of bias → design/methods, results
- Generalizability (ideal setting, population, procedures, f/u)
- Statistically significant effect ≠ clinically significant
  - Endpoint
  - NNT/NNH
- Generally ≤5% false-positive risk for 1° outcome
- Resource-intensive

Important Considerations

- Not all exposures & interventions can be tested in clinical trials
  - Tobacco
- Single RCT not a panacea → totality of evidence
- Embrace and manage uncertainty
- Think probabilistically and recognize cognitive bias
- Evidence-informed, shared medical decision making