

# Sham Surgery: The Problem of Inadequate Medical Evidence



Alan S. Gerber (Yale) and  
Eric M. Patashnik (University of Virginia)

# Motivation of project

- U.S. health care system features rapid pace of medical innovation and highest per capita spending in the world
- Yet physicians' decisions about treatments often based on traditional, anecdotal evidence and personal experience
- Every year, millions of patients receive unproven procedures that increase health care costs without clearly improving outcomes

- Less than half of all medical care rests on rigorous evidence (IOM)
- David Eddy, “The problem is, we don’t know what we’re doing”
- This situation is an open secret among health services researchers, but has received only sporadic public attention



■ What's going on here??

# The New England Journal of Medicine

---

Copyright © 2002 by the Massachusetts Medical Society

---

VOLUME 347

JULY 11, 2002

NUMBER 2



---

## A CONTROLLED TRIAL OF ARTHROSCOPIC SURGERY FOR OSTEOARTHRITIS OF THE KNEE

J. BRUCE MOSELEY, M.D., KIMBERLY O'MALLEY, PH.D., NANCY J. PETERSEN, PH.D., TERRI J. MENKE, PH.D.,  
BARUCH A. BRODY, PH.D., DAVID H. KUYKENDALL, PH.D., JOHN C. HOLLINGSWORTH, DR.P.H.,  
CAROL M. ASHTON, M.D., M.P.H., AND NELDA P. WRAY, M.D., M.P.H.

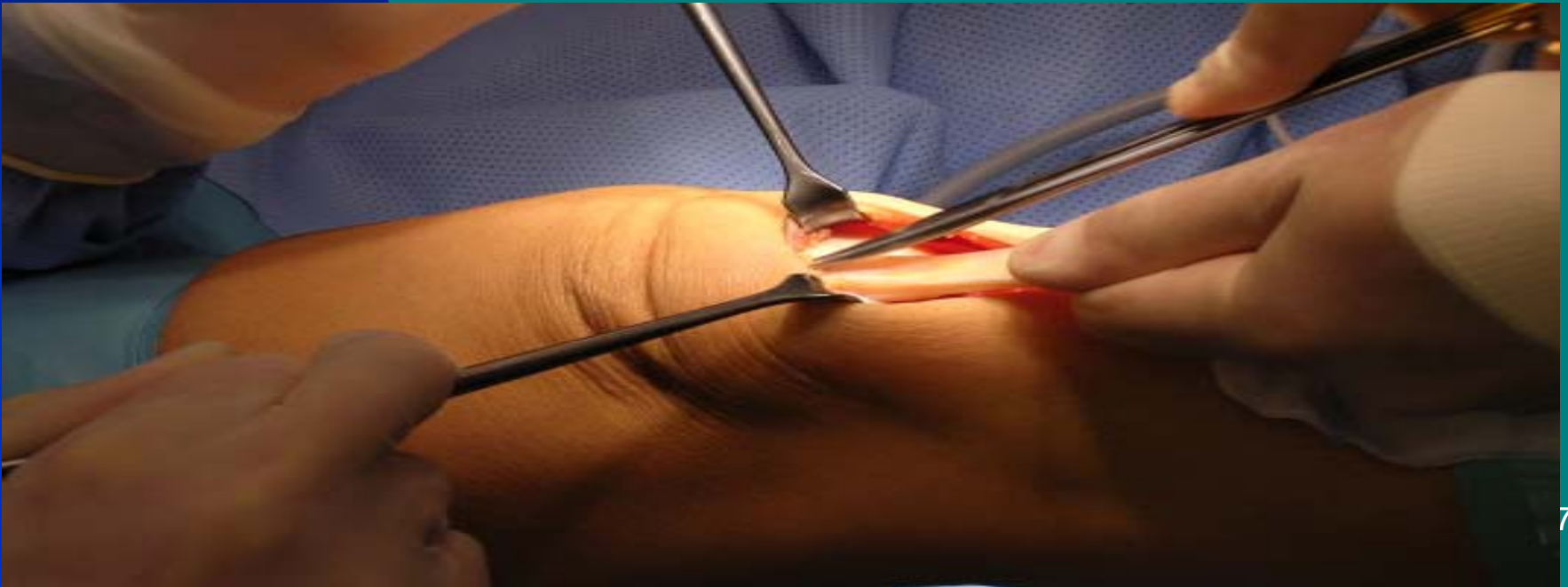
Essay was inspired by a remarkable 2002 randomized, placebo-controlled study in which a common surgical procedure (arthroscopy for osteoarthritis of the knee) found to be no more effective than a fake operation

# Outline of talk

- Design and results of sham surgery experiment
  - ◆ What was the state of evidence *before* the NEJM study?
  - ◆ What have been the responses of the medical and policy communities to the study?
- Is this case typical or unusual?
- Possible explanations
  - ◆ Ethics
  - ◆ Market and political failures
- Questions for future research
- Conclusions

# Landmark sham surgery study

- Arthroscopy for osteoarthritis of the knee
- The experimental design
- Study results

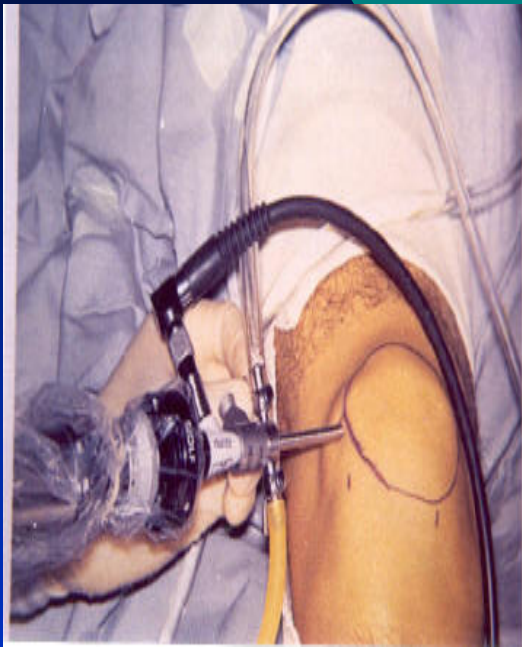


# The condition

- OA is a type of arthritis, affects more than 20 million people
- Among the elderly, knee OA the leading cause of chronic disability
- OA and related arthritic conditions cost more than \$81 billion a year in medical care, lost wages, and other expenses (AHRQ)
- Reducing pain, maintaining mobility main goals of OA treatment
- When drug therapy fails, surgery is often indicated



# Arthroscopy for osteoarthritis of the knee



- Common surgical procedure, perhaps 100,000 per year (at least 25,000 paid by Medicare in 2001)
- \$2.5-\$5K per operation
- Two types: lavage (flush debris from knee) and debridement (shave rough areas of cartilage from joint, then flush)
- Used to reduce pain and delay need for knee replacement

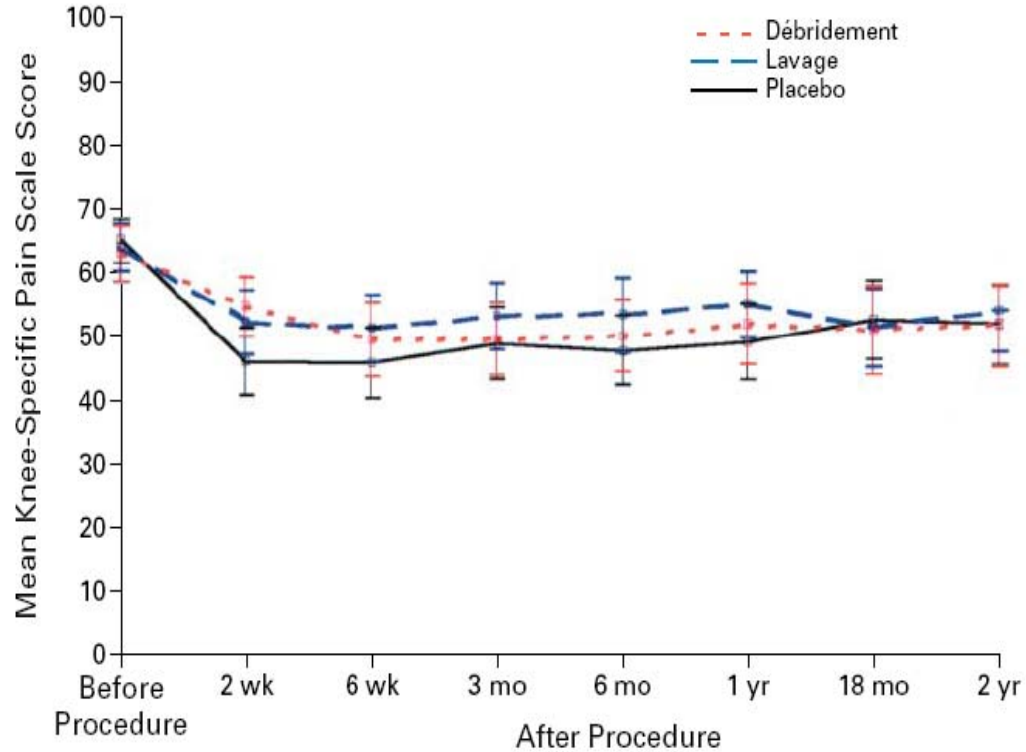
# NEJM study



- 180 patients randomized to debridement, lavage, sham surgery (165 completed trial)
- Patients in placebo group receive skin incisions and underwent simulated debridement
- Patients and assessors of outcomes were blinded
- Informed consent—had to write out statement
- Followed up for 2 years
- Both objective and subjective outcome measures
  - ◆ Self reported pain
  - ◆ Walking and stair climbing

# Study results

- Blinding successful—13 percent of all 3 groups guessed they were in placebo arm
- At no point did either of the treatment groups demonstrate better function or report less pain than the placebo group
- Clinically meaningful differences between the groups excluded at 95 percent confidence level

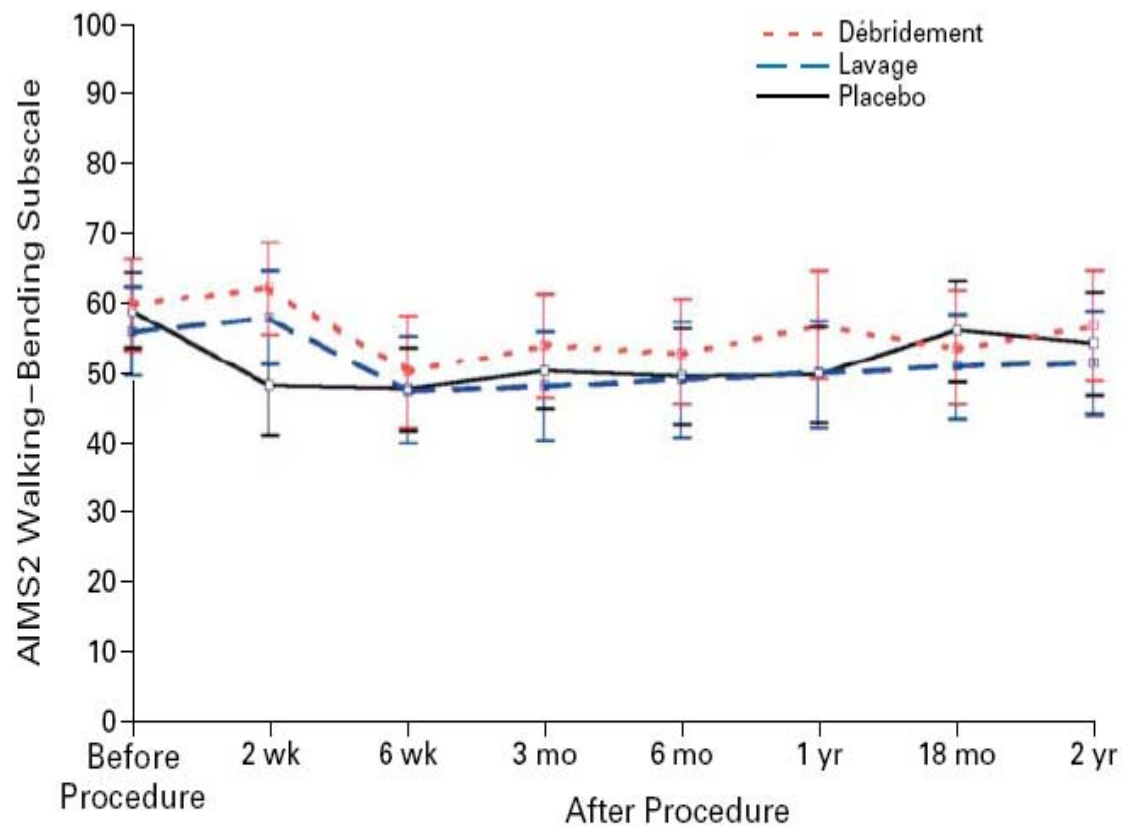


No. AT RISK

Placebo	60	59	57	56	57	53	52	55
Lavage	61	59	57	59	59	57	56	55
Débridement	58	59	59	58	56	50	51	53

**Figure 1.** Mean Values (and 95 Percent Confidence Intervals) on the Knee-Specific Pain Scale.

Assessments were made before the procedure and 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. Higher scores indicate more severe pain.



No. AT Risk

Placebo	60	59	57	56	57	54	52	55
Lavage	61	59	57	59	59	57	57	56
Débridement	58	58	59	58	55	51	51	53

**Figure 2.** Mean Values (and 95 Percent Confidence Intervals) on the Walking-Bending Subscale of the Arthritis Impact Measurement Scales (AIMS2).

Assessments were made before the procedure and 2 weeks, 6 weeks, 3 months, 6 months, 12 months, 18 months, and 24 months after the procedure. Higher scores indicate poorer functioning.

It looks like I've had  
an operation...

That's  
the main  
thing...



# State of evidence *before* 2002 NEJM study

- Little scientific evidence to support operation for this indication
- Pre-2002 studies were primarily case series evidence
- Example: Yang and Nisonson (1995) evaluated 105 cases (called them up and asked how they felt)
  - ◆ Not blinded
  - ◆ No control group

- Goodfellow at orthopedists' meetings:  
These are “pseudo treatments”
- Notes that no one had performed the controlled trial to distinguish between placebo effects and direct benefits of the operation



- After publication of the study, one leading evidence-based medicine group pointed out the results should not be surprising based on the literature
- “There never was any good evidence that lavage and debirdement were useful things to do”

# Responses to study

- Study praised in *NEJM* editorial
- Conclusion: 80-90 percent of these operations should not be done
- Some surgeons insisted the study *must* be wrong:
- “I’ve done thousands of these in people with osteoarthritic knees, and they really are better,” said Robert W. Jackson

# Professional associations go on the attack

- Claim: subjects unrepresentative of population
- Procedure works for patients with “mechanical” symptoms
- Support for this claim very weak
  - ◆ Only four prior case series studies
  - ◆ All lacked control group and used unvalidated assessment scales
- Appropriate way to resolve any lingering doubt would be to replicate the study
- Professional groups *called* for replication studies, but didn't conduct or sponsor any

# CMS issues Medicare coverage determination

- CMS initiates national coverage determination
- Tremendous pressure from provider groups to maintain coverage
- Determination: Lavage noncovered, but debridement remains generally covered
- AAOS: “The coverage decision parallels the position of the musculoskeletal societies. CMS recognized that arthroscopy is appropriate in virtually all circumstances in which the orthopedic community now employs this technology”

# Recent developments

- Recent medical studies indicate that arthroscopy remains a widely used treatment for OA of the knee
- 2007 AHRQ evidence report concludes that evidence of benefit is lacking for many common ways of treating OA of the knee, including arthroscopy, and that additional placebo-controlled RCTs showing clinically significant advantage for arthroscopy would be necessary to refute Moseley
- But no such follow-up studies have been performed

# Is this case atypical?

- Medical procedures, which account for a much larger share of total spending on health care than drugs and devices do, regularly achieve widespread use without extensive clinical evaluation
- There is no FDA for surgery
- Average of 17 years required for findings of RCTs to be incorporated into practice, and application is highly uneven
- RCTs rarely used in certain areas of medical practice (orthopedics)

# Evidence matters: Case of bone marrow transplants with high-dose chemotherapy to treat breast cancer

- Trials performed in the 1990s showed that contrary to belief, conventional therapy superior to ABMT/HDC.
- Some 30,000 women were unnecessarily subjected to ABMT/HDC, and an estimated 600 women died prematurely as a result

# What accounts for “medical guesswork” problem?

- Ethical imperatives?
- Market failures?
- Political failures?



# Ethical imperatives?

- Often said that physicians have an obligation to help patients, which limits kinds of studies can be done
- Hardest test: sham surgery trials
  - ◆ Ligation of internal mammary arteries for angina (1959)
    - ★ Common procedure, yet turned out to work no better than placebo operation involving skin incision under local anesthesia
    - ★ But this trial lacked fully informed consent

- AMA position: If there is no consensus on the use of a procedure, the scientific need for blinding is pressing, participants are fully informed, and the risks to subjects in the control group are sufficiently minimized, then sham surgery trials may be ethically appropriate.

- Spodick: “[T]he omission of adequate standards for surgical therapies should be especially surprising, since even the most essential operation involves inevitable trauma—physical, metabolic, and psychic—not to mention the risks of anesthesia.”

# ■ Market failures?

- ◆ Effectiveness research is costly for insurance firms and other private actors to carry out.
- ◆ Once studies have been conducted, the findings will be widely disseminated.
- ◆ Since private actors usually can obtain the benefits of the research without paying for it, little incentive to produce it
- ◆ “Fouling the nest” problem—A provider who claims she only performs “effective procedures” is implying that some medical procedures are ineffective, which could reduce consumption (Posner)

# Political Failures?

- ◆ In sum, information about the effectiveness of a given procedure is a public good, and the market will not supply the socially efficient amount of it
- ◆ But the market failure explanation of the problem is incomplete, because when markets fail, public policy often provides missing incentives

- In fact, we spend less than 1/10<sup>th</sup> of 1 percent of overall health expenditures on medical effectiveness research
- Public agencies that have issued reports questioning the effectiveness of specific procedures have seen their budgets cut and authority curtailed (e.g, Agency for Health Care Policy Research study of surgery as first-line treatment for low back pain in the 1990s)

## EXHIBIT 1

### Budgets Of The Federal Government's Leading Health Services Research (HSR) And Biomedical Research Agencies, Thousands Of Dollars, Selected Years 1980-2003

Year	HSR budget <sup>a</sup>	NIH budget <sup>b</sup>	Ratio of NIH to HSR dollars	HSR as percent of NIH
1980	\$ 26,916	\$ 3,428,935	127.4:1	0.8%
1990	97,671	7,576,352	77.6:1	1.0
2000	204,000	17,793,587	87.7:1	1.0
2003 <sup>c</sup>	251,700 <sup>d</sup>	27,432,075 <sup>e</sup>	108.0:1	0.9

SOURCES: See below.

NOTES: The 1980 HSR budget is for the National Center for Health Services Research (NCHSR). The 1990 HSR budget is for the Agency for Health Care Policy and Research (AHCPR). The 2000 HSR budget is for the Agency for Healthcare Research and Quality (AHRQ). Percentages in final column are approximate.

<sup>a</sup> Source for 1980, 1990, and 2000 data: Jonathan Lawniczak, Coalition for Health Services Research, December 2002.

<sup>b</sup> NIH, *NIH Almanac 2001*, Pub. no. 01-5 (Bethesda, Md.: NIH, Editorial Operations Branch, Office of Communications and Public Liaison, December 2001).

<sup>c</sup> President Bush's proposed 2003 budget. The 2003 HSR budget is for AHRQ.

<sup>d</sup> AHRQ, "AHRQ Fiscal Year 2003 Budget in Brief," February 2002, [www.ahrq.gov/about/cj2003/budbrf03.htm](http://www.ahrq.gov/about/cj2003/budbrf03.htm) (20 May 2003).

<sup>e</sup> NIH, "Summary of Appropriations," [www4.od.nih.gov/officeofbudget/CJ2003/App%20sum%20w%20AIDS%20break.PDF](http://www4.od.nih.gov/officeofbudget/CJ2003/App%20sum%20w%20AIDS%20break.PDF) (20 May 2003).

Source: Gray, Gusmano and Collins, *Health Affairs* Web Exclusive 2002

## ■ Factors driving weak political response

### ◆ *Power of the medical lobby*

- ★ First face of power: active lobbying
- ★ Agenda manipulation: Information actively discredited, not just passively ignored
- ★ The political strategies used to defend existing practices are subtle, and include
  - obfuscation (using complex language to hide clear conclusions)
  - false promises (e.g., calling for replication studies but not performing them)
  - exploitation of public trust in the medical profession (doctor knows best)



- ◆ *Weak incentives for political entrepreneurship*
  - ★ Elected officials may support evidence based medicine in principle, but are usually loathe to intervene in debates over the effectiveness of specific treatments for fear of attracting powerful enemies
  - ★ This issue hasn't offered partisan or ideological dividends
  - ★ The public has a diffuse interest in learning what treatments work best, but politicians rarely have an incentive to invest their time and energy in it

## ◆ *The silence of the elites*

- ★ Unlike the national health insurance crisis, whose costs are disproportionately borne by marginalized clienteles, the harms from the use of unproven treatments affect all Americans, *including the affluent and well-insured.*
- ★ Evidence that public policy is disproportionately responsive to preferences of elites (Bartels, Gilens)
- ★ A puzzle: Why haven't the rich and powerful used their ties to public leaders to demand better system performance?

- ★ One hypothesis: wealth elites may believe they are personally immune to the systemic problems in the evidentiary basis of American medicine
- ★ About nine in ten Americans report being very or somewhat satisfied with their last visit to their physician
- ★ Would not be surprising if elites are especially satisfied with their health care, because they enjoy access to the nation's "best doctors" and "top hospitals" or for other reasons
- ★ Gerber and I are planning an elite survey to probe this question

# Many questions for future research...

- ◆ Why does the medical industry produce better evidence in some practice areas than others?
- ◆ What can we learn from the experiences of other nations?
- ◆ How can we promote better evaluation of procedures without chilling innovation?
- ◆ How can we maximize the placebo effect ethically?

# Conclusions

- There are clearly inadequate incentives to generate and disseminate systematic evidence about the benefits and risks of many treatments
- The problem results from democratic failures as well as market failures, and requires a political – not just a technical – solution
- Recent calls for creation of a new comparative effectiveness agency are certainly welcome, but challenge remains to shield such an agency from inevitable political attacks and build supportive constituencies
- Political sustainability is key