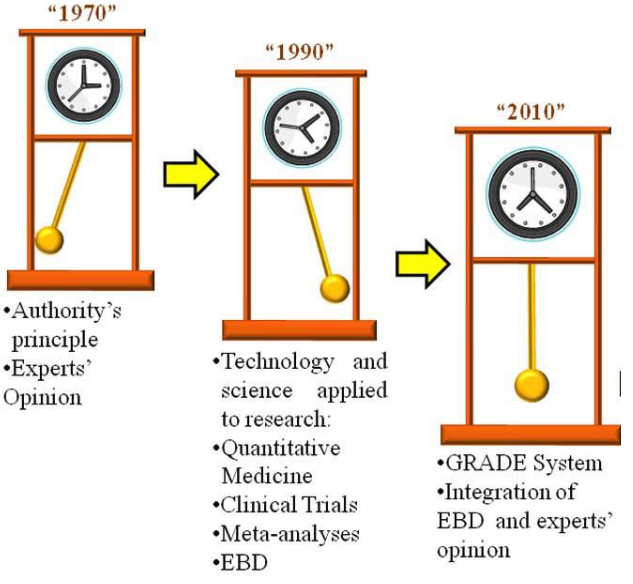
Pros and cons of randomized controlled clinical trials and evidence-based medicine

Prof. Giuseppe Verlato

Unit of Epidemiology & Medical Statistics
University of Verona

Evolution of EBD: from 1970 to Third Millennium



Nocini PF, Verlato G, Frustaci A, de Gemmis A, Rigoni G, De Santis D. **Evidence-Based Dentistry** in Oral Surgery: Could We Do Better? The Open Dentistry Journal 2010;4:77-83

"HISTORICAL PARADIGM" in MEDICAL TEACHING and PRACTICE

Medical teaching and acquisition of clinical skills were based on the knowledge delivered by medical leaders

(authoritarian approach ➔ Authority's principle)

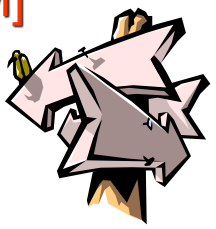
- *EXPERTS' OPINION*
- *REFERENCE TEXTBOOKS*
- *CONSENSUS CONFERENCE*

ASSUMPTIONS of the "HISTORICAL PARADIGM"

- «Unsystematic observations from clinical experience are a valid way of building and maintaining one's knowledge about patient prognosis, the value of diagnostic tests and the efficacy of treatments.
- The study and understanding of basic mechanisms of disease and pathophysiologic principles are a sufficient guide for clinical practice.
- A combination of thorough traditional medical training and common sense is sufficient to allow one to evaluate new tests and treatments
- Content expertise and clinical experience are a sufficient base from which to generate valid guidelines for clinical practice»

Evidence-Based Medicine Working Group et al., 1992, p.2421

The NEW PARADIGM "EVIDENCE BASED MEDICINE [EBM]"



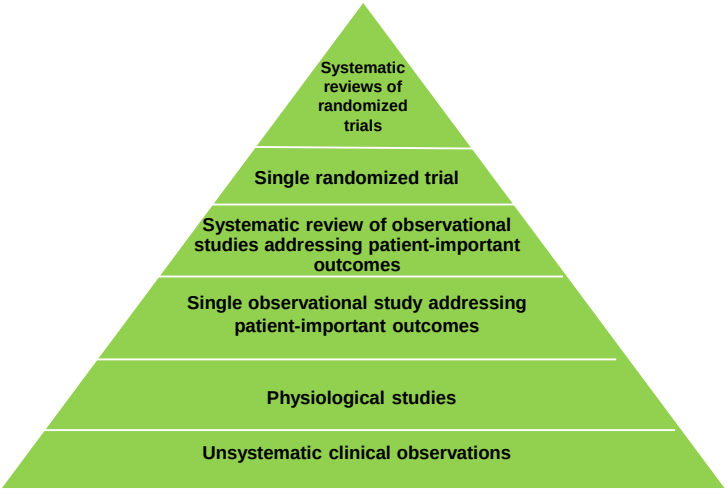
Clinical decision making and knowledge on diagnostic tests, patient's prognosis and efficacy of treatments should be based on the ***evidence derived from clinical research***.

Evidence-Based Medicine

- "While clinical experience and skill are important, **systematic** attempts to record observations in a **reproducible and unbiased** fashion markedly increase the confidence one can have in knowledge about patient prognosis, the value of diagnostic tests, and the efficacy of treatment.
- In the absence of systematic observation, one must be **cautious** in the interpretation of information derived from clinical experience and intuition, for it **may at times be misleading**.
- The study and understanding of basic mechanisms of disease are necessary but **insufficient guides for clinical practice**.
- Understanding **certain rules of evidence** is necessary to correctly interpret literature on causation, prognosis, diagnostic tests, and treatment strategy."

Evidence-Based Medicine Working Group et al., 1992, p.2421

A Hierarchy of Strength of Evidence in Interventional Clinical Trials



Adapted from: Guyatt et al (2000) for the Evidence-Based Medicine Working Group. *JAMA* 284:1290-6

The importance of randomized controlled clinical trials to progress in medicine

In favor…

1) Hormone replacement therapy

2) Smoke and lung cancer

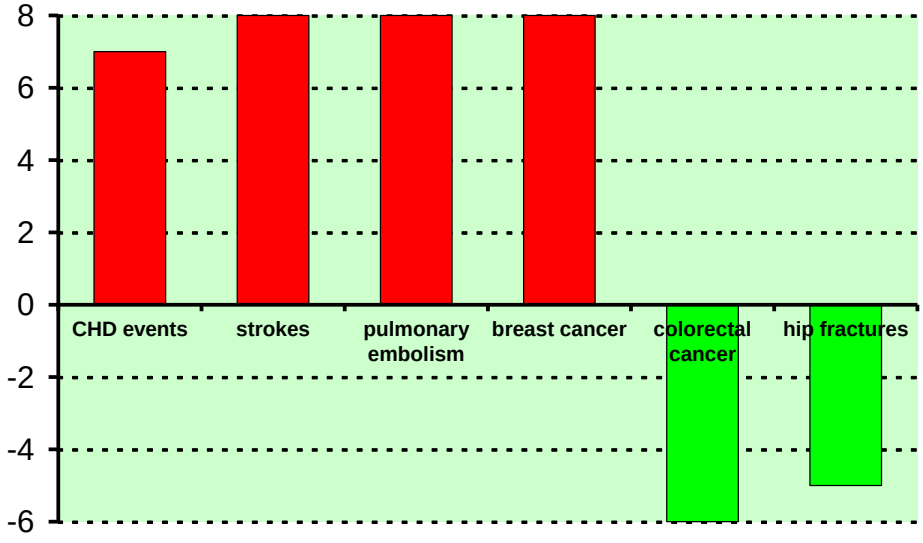
Hormone replacement therapy

"Context: **Despite decades of accumulated observational evidence**, the balance of risks and benefits for hormone use in healthy postmenopausal women **remains uncertain**.

Objective: To assess the major health benefits and risks of the **most commonly used** combined hormone preparation in the United States."

Rossouw JE, Anderson GL, Prentice RI, et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women – Principal results from the Women's Health Initiative randomized controlled trial. JAMA 288:321-333 (citato 4192 volte al 31 marzo 2008)

Hormone replacement therapy



"Absolute excess risks per 10,000 person-years attributable to estrogen plus progestin"

The importance of randomized controlled clinical trials to progress in medicine

against …

1) Hormone replacement therapy
2) Smoke and lung cancer

Smoke and lung cancer

"For example, the studies linking smoking with lung cancer were bitterly criticized by 'conventional' researchers who were not willing to accept evidence from studies where the exposure had not been randomized".

Stolley PD (1991) **When genius errs: Fisher**, R.A. and the lung cancer controversy. *Am J Epidemiol* 133:416-25.

Pearce N (2008) Point-counterpoint. Corporate influences on epidemiology. *Int J Epidemiol* 37:46-53

Importance of meta-analysis to progress in medicine

In favor …

1) **<u>Thrombolytic therapy and prophylactic lidocaine for myocardial infarction</u>**

2) Extended lymphadenectomy in gastric cancer surgery

IMPACT OF THROMBOLYTIC THERAPY ON MORTALITY AFTER MYOCARDIAL INFARCTION: COMPARISON OF SCIENTIC EVIDENCE AND EXPERTS' OPINION

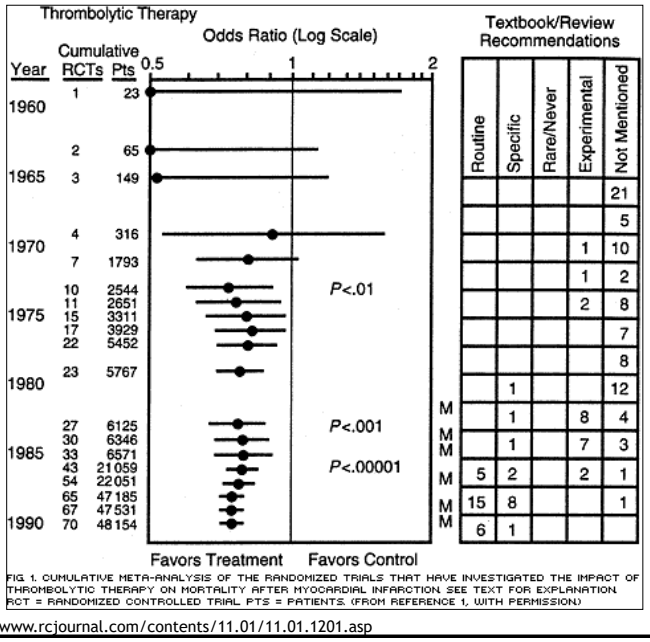


FIG 1. CUMULATIVE META-ANALYSIS OF THE RANDOMIZED TRIALS THAT HAVE INVESTIGATED THE IMPACT OF THROMBOLYTIC THERAPY ON MORTALITY AFTER MYOCARDIAL INFARCTION. SEE TEXT FOR EXPLANATION. RCT = RANDOMIZED CONTROLLED TRIAL. PTS = PATIENTS. (FROM REFERENCE 1, WITH PERMISSION)

FONTE: http://www.rcjournal.com/contents/11.01/11.01.1201.asp

IMPACT OF PROPHYLACTIC LIDOCAINE ON MORTALITY AFTER MYOCARDIAL INFARCTION: COMPARISON OF SCIENTIC EVIDENCE AND EXPERTS' OPINION

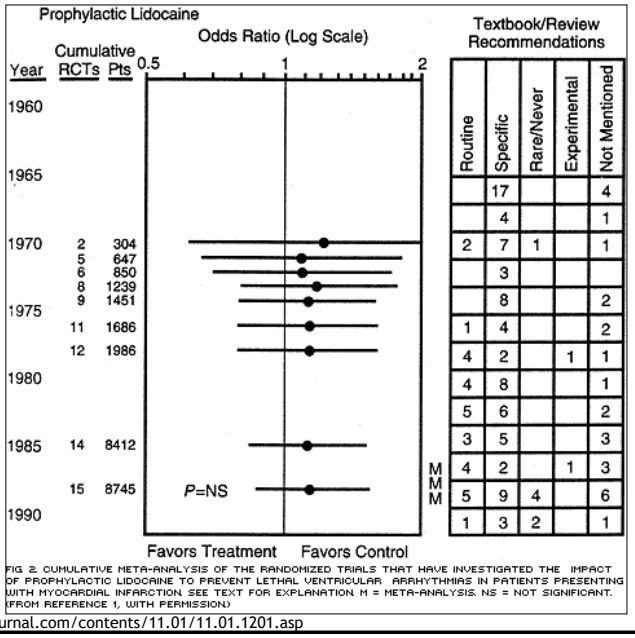


FIG 2. CUMULATIVE META-ANALYSIS OF THE RANDOMIZED TRIALS THAT HAVE INVESTIGATED THE IMPACT OF PROPHYLACTIC LIDOCAINE TO PREVENT LETHAL VENTRICULAR ARRHYTHMIAS IN PATIENTS PRESENTING WITH MYOCARDIAL INFARCTION. SEE TEXT FOR EXPLANATION. M = META-ANALYSIS. NS = NOT SIGNIFICANT. (FROM REFERENCE 1, WITH PERMISSION)

FONTE: http://www.rcjournal.com/contents/11.01/11.01.1201.asp

If the results of meta-analyses had been appropriately acknowledged, thrombolytic therapy would have been adopted and prophylactic lidocaine would have been dismissed 10-15 years earlier.

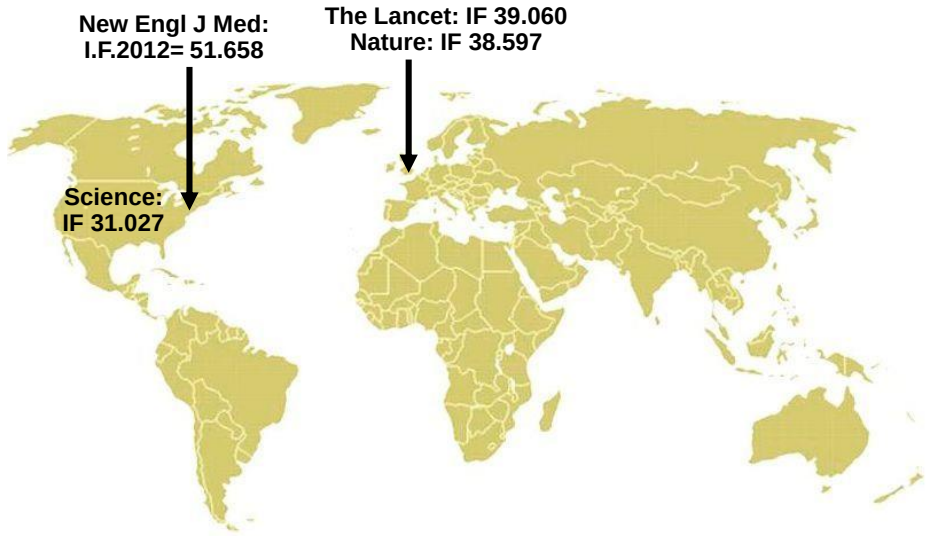
Importance of meta-analysis to progress in medicine

against …

1)Thrombolytic therapy and prophylactic lidocaine for myocardial infarction

2) **Extended lymphadenectomy in gastric cancer surgery**

BACKGROUND - 1



However, in gastric cancer …

New cases of gastric cancer in 2012, according to GLOBOCAN 2012 [Ferlay et al, 2010]

Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer; 2013. Available from: http://globocan.iarc.fr, accessed on 3/1/2014

During the Seventies and Eighties Japanese surgeons developed an aggressive approach to prevent lymphatic spread of the tumor, based on **EXTENDED** (D2) and superextended (D3) lymphadenectomy [Maeta et al, 1999; Kunikasi et al, 2000; Gunji et al, 2003].

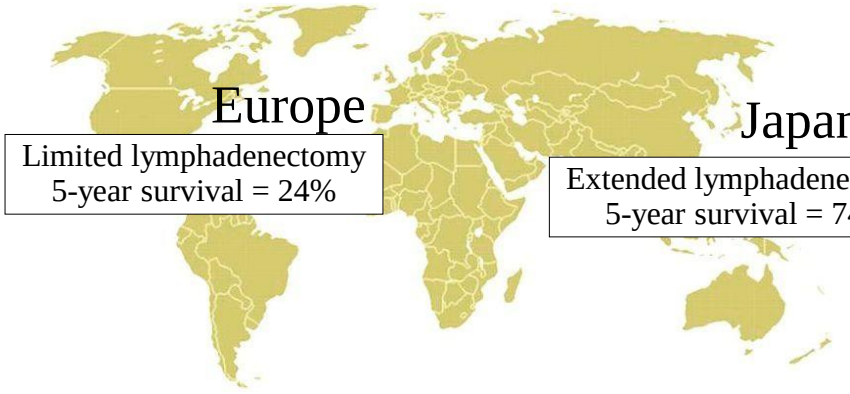
However, at the same time the most widely used intervention in Europe and the States remained a **LIMITED** (D1) lymphadenectomy.

Maeta M, Yamashiro H, Saito H, et al (1999) A prospective pilot study of extended (D3) and superxtended para-aortic lymphadenectomy (D4) in patients with T3 or T4 gastric cancer managed by total gastrectomy. *Surgery* 125:325-331

Kunisaki C, Shimada H, Yamaoka H, et al (2000) Indications for paraortic lymph node dissection in gastric cancer patients with paraortic lymph node involvement. *Hepatogastroenterology* 47:586-589

Gunji Y, Suzuki T, Kobayashi S, et al (2003) Evaluation of D3/D4 lymph node dissection for patients with grossly N2 positive advanced gastric cancer. *Hepatogastroenterology* 50:1178-1182

In the Nineties



Europe

Limited lymphadenectomy
5-year survival = 24%

Japan:

Extended lymphadenectomy
5-year survival = 74%

Berrino F, De Angelis R, Sant M, Rosso S, Bielska-Lasota M, Coebergh JW, et al (2007) Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EUROCARE-4 study. *Lancet Oncol* 8:773-783

Nakajima T (2002) Gastric cancer treatment guidelines. *Gastric Cancer* 5:1-5

However Japanese surgery, in spite of these outstanding achievements, was not considered the benchmark in the States and in Northern Europe, i.e. in those countries considered as the scientific leaders in medicine.

Western surgeons and scientists argued that Japanese results came from retrospective observational studies and attributed the good prognosis, recorded in Japanese series, to a benign tumor-biological behavior of gastric cancer in Japan [Jatzko et al, 1999].

Jatzko G, Pertl A, Jagoditsch M (1999) Chirurgische Therapie und Ergebnisse beim Magen-frühkarzinom. *Chir Gastroenterol* 15:223-226

Is extended lymphadenectomy (D2) advisable in gastric cancer surgery?

During the Nineties a huge effort was made to base the surgical approach to gastric cancer on sounded evidence.

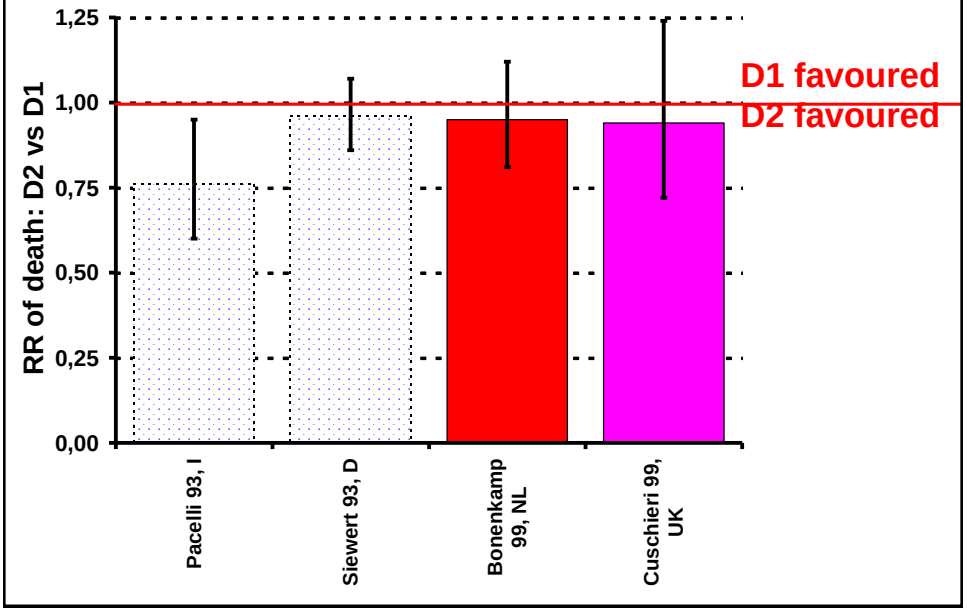
The Dutch and British surgeons organized large trials, where patients were randomly assigned to either limited (D1) or extended (D2) lymphadenectomies.

After 5 years of follow-up these studies showed no evidence of overall survival benefit after extended lymphadenectomy [Bonenkamp et al 1999; Cuschieri et al 1999].

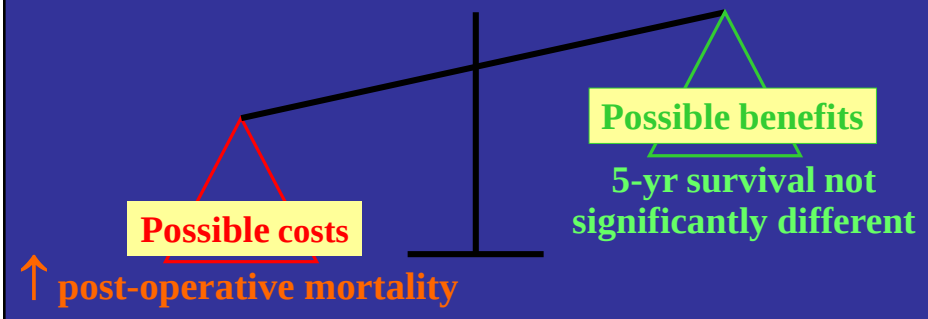
Bonenkamp JJ, Hermans J, Sasako M, van de Velde CJH, for the Dutch Gastric Cancer Group (1999) Extended lymph node dissection for gastric cancer. New Engl J Med 340:908-914

Cuschieri A, Weeden S, Fielding J, Bancewicz J, Craven J, Joypaul V, Sydes M, Fayers P, for the Surgical Co-operative Group (1999) Patients survival after D1 and D2 resections for gastric cancer: long term results of the MRC surgical trial. Brit J Cancer 79:1522-1530

Randomized trials DEMONSTRATE that 5-yr survival is not significantly different after D1 or after D2



Is extended lymphadenectomy (D2) advisable in gastric cancer surgery?



Evidence for D2 dissection is inconclusive:

No overall survival advantage has emerged, but some patients with intermediate stage disease may benefit. Excess operative mortality appears to be associated with pancreatico-splenectomy, low case volume and lack of specialist training.

McCulloch, *Brit J Surg*, 2005

The exclusion of Japanese papers, although justified from a methodological point of view, hinders a lot the development of knowledge.

"At present the Japanese experience in gastric cancer is a kind of benchmark for surgeons throughout the world."

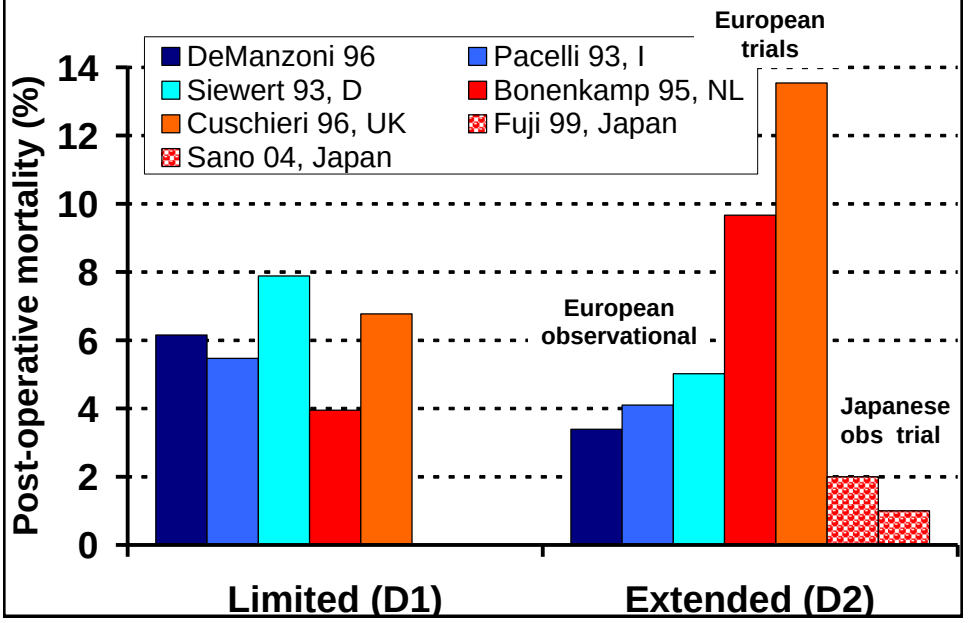
De Manzoni G, Verlato G (2005) Gastrectomy with extended lymphadenectomy for primary treatment of gastric cancer (letter). *Brit J Surg*, 92(6): 784

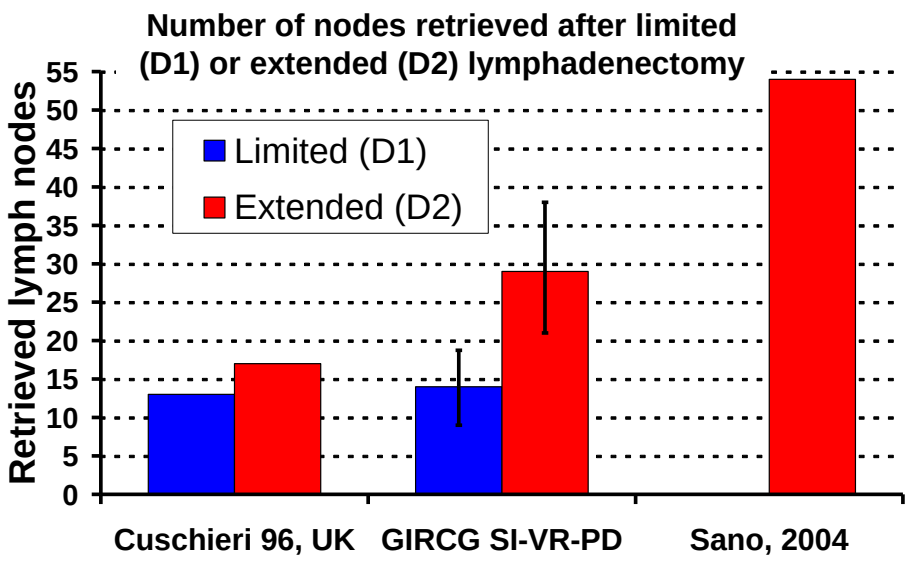
However, the latter trials presented a rather low surgical quality, as they were performed by surgeons without previous training in extended lymphadenectomy, executing less than 5 interventions per year.

The limited surgical experience yielded:

1) a very high post-operative mortality after extended lymphadenectomy (9.7% in the Dutch trial and 13.5% in the British trial),
2) a high percentage of removal of adjacent organs: splenectomies (37% and 65%, respectively) and pancreatectomies (30% and 56%)
3) a low number of nodes retrieved (median of 17 nodes in the British trial).

Post-operative mortality in patients undergoing limited (D1) or extended (D2) lymphadenectomy for gastric cancer





Columns are medians, bars are interquartile ranges (25-75 percentiles)

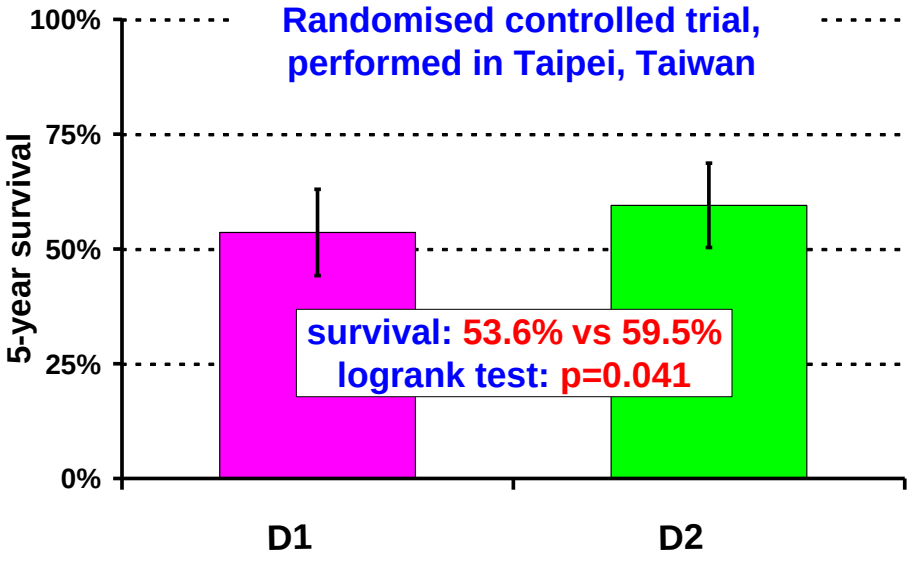
"*In our opinion, it is extremely difficult to ask Japanese surgeons, in whose series post-operative mortality is only 1-2%, to believe in randomized clinical trials where post-operative mortality peaks to 10-14%, irrespectively of methodological quality of those studies*."

Verlato G, Roviello F, Marchet A, Giacopuzzi S, Marrelli D, Nitti D, de Manzoni G (2009) Indexes of surgical quality in gastric cancer surgery: experience of an Italian network. Ann Surg Oncol, 16:594-602

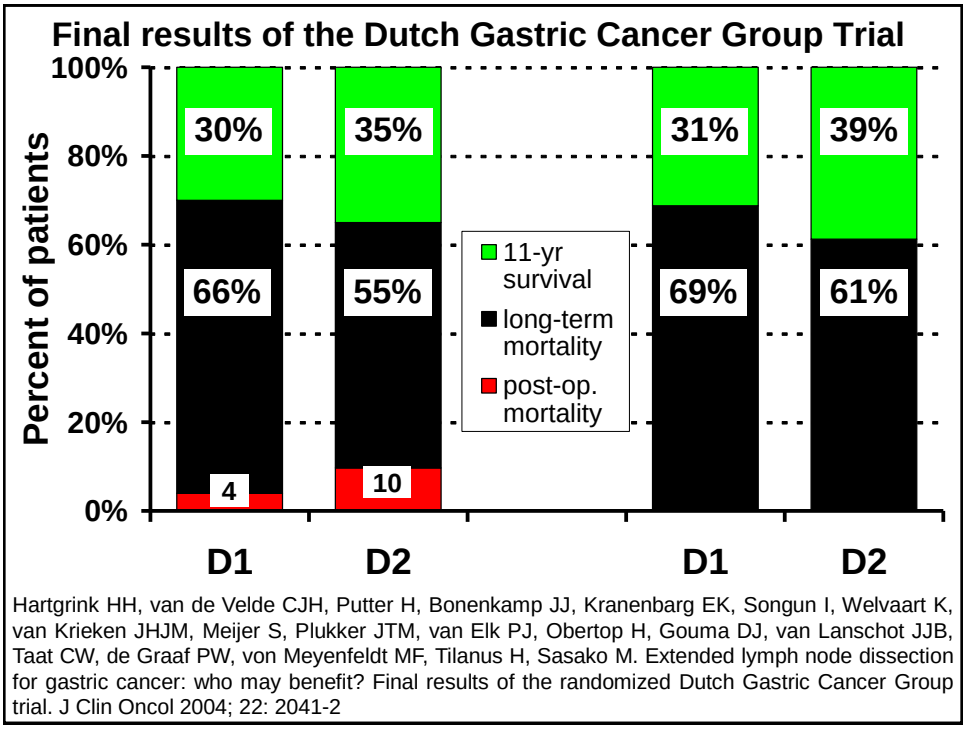
In 2006 another randomized trial [Wu et al, 2006] has been published, showing a mild but significant survival advantage after D2* with respect to D1.

Moreover in the Dutch trial, after 11 years of follow-up, survival was significantly higher after D2 than after D1, when excluding post-operative mortality [Hartgrink et al, 2004].

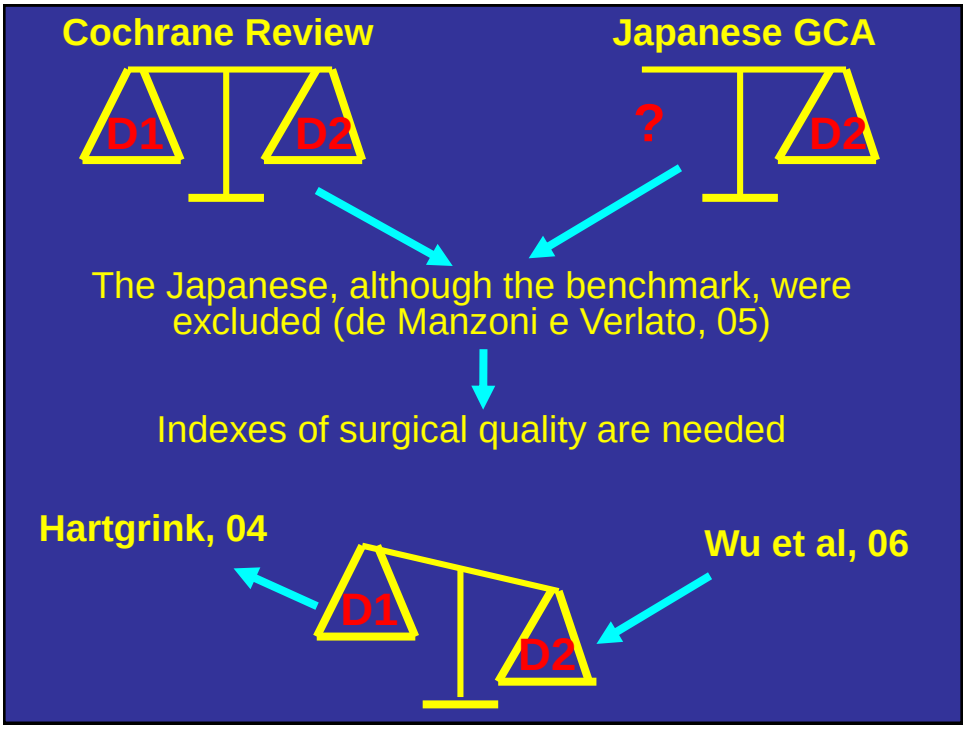
* Wu et al [2006] reported that they compared D1 with D3 lymphadenectomy, but actually their D3 procedure was a slightly extended D2.



Wu CW, Hsiung CA, Lo SS, Hsieh MC, Chen JH, Li AFY, Lui WY, Whang-Peng J. Nodal dissection for patients with gastric cancer: a randomised controlled trial. Lancet Oncology 2006, 7: 309-315



Hartgrink HH, van de Velde CJH, Putter H, Bonenkamp JJ, Kranenbarg EK, Songun I, Welvaart K, van Krieken JHJM, Meijer S, Plukker JTM, van Elk PJ, Obertop H, Gouma DJ, van Lanschot JJB, Taat CW, de Graaf PW, von Meyenfeldt MF, Tilanus H, Sasako M. Extended lymph node dissection for gastric cancer: who may benefit? Final results of the randomized Dutch Gastric Cancer Group trial. J Clin Oncol 2004; 22: 2041-2



The Cochrane review was withdrawn in January 2012.

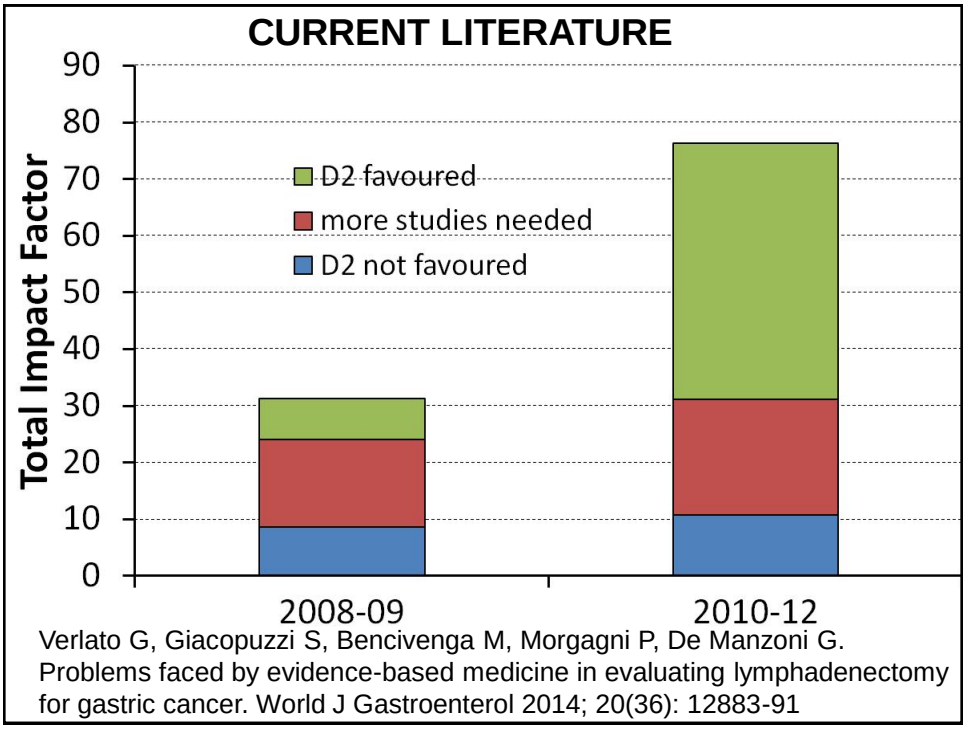
McCulloch P, Nita ME, Kazi H, Gama-Rodrigues JJ. WITHDRAWN: Extended versus limited lymph nodes dissection technique for adenocarcinoma of the stomach. Cochrane Database Syst Rev. 2012 Jan 18;1:CD001964. doi: 10.1002/14651858.CD001964.pub3.

"papers dealing with surgery for gastric cancer cannot be evaluated only according to the quality of the study design, such as the Jadad score, but also the quality of surgical procedures must be taken into account" [Verlato 2009]

"D2 was adopted as the standard of surgical treatment with curative intent by the joint ESMO (European Society for Medical Oncology) - ESSO (European Society of Surgical Oncology – ESTRO (European Society of Radiotherapy and Oncology) guidelines [Waddell 2013]. The ESMO-ESSO-ESTRO guidelines ranked the level of evidence as the highest (I) and the grade of recommendation as B (strong or moderate evidence for efficacy but with a limited clinical benefit).

At variance American NCCN guidelines recommend a D1+ or a modified D2 lymph node dissection, the latter performed by experienced surgeons in high-volume centers [Ajani 2013]."

Verlato G, Giacopuzzi S, Bencivenga M, Morgagni P, De Manzoni G. Problems faced by evidence-based medicine in evaluating lymphadenectomy for gastric cancer. World J Gastroenterol 2014; 20(36): 12883-91



Verlato G, Giacopuzzi S, Bencivenga M, Morgagni P, De Manzoni G. Problems faced by evidence-based medicine in evaluating lymphadenectomy for gastric cancer. World J Gastroenterol 2014; 20(36): 12883-91

Criteria to evaluate the quality of the study design (for instance, the Jadad score based on criteria for randomization and blindness, descriptions of withdrawals and drop-outs) are well-established.

Unfortunately, indexes of surgical quality have not been agreed upon. It would be extremely useful to establish, at international level, quality criteria for this kind of surgery.

For instance, we proposed the following indexes of surgical quality in gastric cancer surgery:

1) number of excised node
2) removal of adjacent organs
3) post-operative morbidity
4) post-operative mortality

Verlato G, Roviello F, Marchet A, Giacopuzzi S, Marrelli D, Nitti D, de Manzoni G (2009) Indexes of surgical quality in gastric cancer surgery: experience of an Italian network. Ann Surg Oncol, 16:594-602

Diabetic foot:
0=absent 1=present

GRADE SYSTEM: integration of EBM and authority principle

Methodology used to organize consensus conference: **Delphi process and Nominal (Expert) Group techniques**

Methodology to derive recommendations from the current literature: **GRADE System**

Delphi technique - 1

The Delphi process is a survey technique for decision making among isolated, anonymous respondents. It aims to guide group opinion towards a final decision, by simultaneously increasing communication within the group while avoiding excess influence by single individuals.

Participants to a Delphi panel are encouraged to produce the most suited ideas to solve a given problem. These ideas are subsequently circulated among all participants to the panel, who can revise their positions without having to defend them in front of the group.

These rounds are repeated until there is a convergence of opinion, according to predetermined criteria on group consensus, stability of individual judgments or the number of rounds. Hence it is possible to achieve consensus, avoiding that solutions proposed by the most influent leaders immediately prevail.

Delphi technique versus nominal group technique

The aim of these techniques is to obtain and synthetize several experts' opinion on a given topic.

Group members communicate by mail or e-mail in the Delphi technique, while in the Nominal technique group members meet together and release their opinion through anonymous leaflets (foglietti).

Hence participants can anonymously release their opinion in both cases.

In the final step a coordinator summarizes the opinion of the whole group, and the grade of agreement is numerically expressed.

Main steps of Delphi and nominal techniques

a) A question (or a group of questions) is formulated and presented.
b) Members give their judgement without communicating with other participants. The judgement can be: "agree, disagree, indifferent"; an order of priority; a score.
c) The coordinator gathers and synthetizes the judgements expressed by individuals and communicate them to group members. Both the global score and individual scores are anonymously released.
d) Thereafter a discussion takes place, whether direct or indirect, for instance through mails. In the discussion group members give their opinion on overall judgement.
e) Group members release a subsequent judgement.
f) The group re-discuss judgements and achieve an agreement. Possible points of agreement are made explicit.

GRADE system

GRADE is the acronym of Grades of Recommendation, Assessment, Development and Evaluation.

The GRADE system is based on the sequential assessment of :

1) Quality of evidence
2) Balance between benefits versus risks, burden, and cost
3) Development and grading of a management recommendations

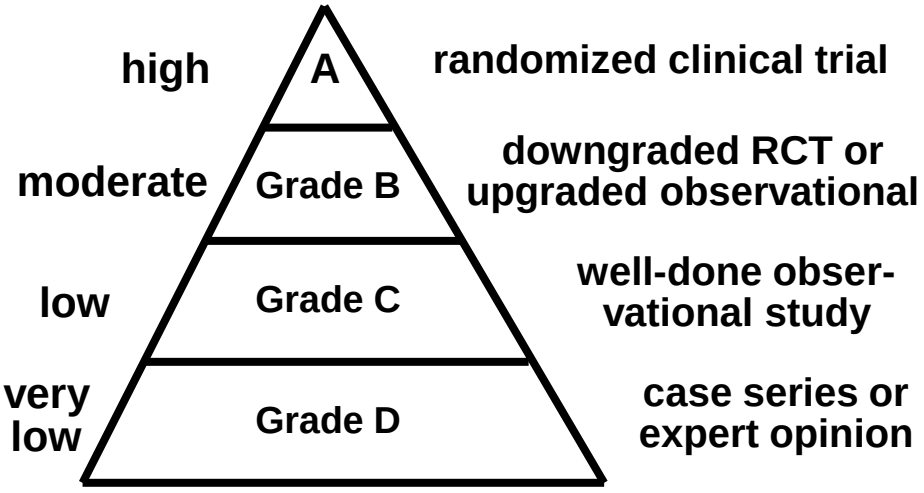
Hence the GRADE system combines EBM (Evidence-Based Medicine) with experts' opinion, expressed in a democratic way through Delphi technique.

GRADE system

The pyramid of evidence slightly differ between traditional Ebm and the GRADE system.

Indeed the GRADE system, although acknowledging the general superiority of experimental studies over observational studies, allows observational studies to be upgraded to B level and experimental studies to be downgraded to the same B level.

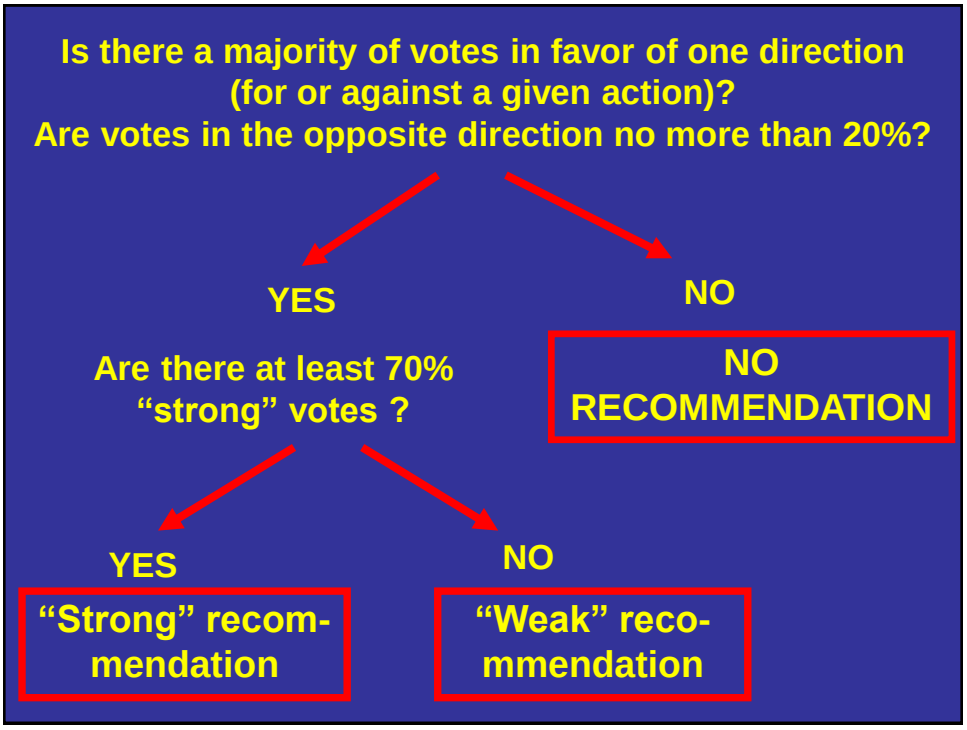
Pyramid of evidence according to the GRADE



Recommendation

"Strong"	"Weak"
- We recommend	- We suggest
Most well-informed patients would accept the intervention	A majority of well-informed patients would accept it (but a substantial proportion would not)
Most clinicians should use it in most situations	Clinicians should consider its use according to particular circumstance

A "strong" recommendation cannot or should not be followed for an individual patient because of that patient's preferences or clinical characteristics which make the recommendation less applicable.



Quality of evidence, graded from A to D, and strength of recommendation, graded as 1 to 2, are combined together

	High-quality evidence	Moderate-quality evidence	Low-quality evidence	Very-low-quality evidence
Strong recommendation	**1A**	**1B**	**1C**	**1D**
Weak recommendation	**2A**	**2B**	**2C**	**2D**

Schünemann HJ, Jaeschke R, Cook DJ, et al (2006) An official ATS Statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. *Am J Respir Crit Care Med* 174: 605-14

Clarity of Risk / Benefit

	High-quality evidence	Moderate-quality evid.	Low-quality evidence	Very-low-quality evidence
Strong recom.	**Benefits clearly outweigh harms and burdens, or vice versa**			
Weak recom.	**Benefits closely balanced with harms and burdens**		Uncertainty in the estimates of benefits, harms, and burdens; benefits may be closely balanced with harms and burdens	Major uncertainty in the estimates of benefits, harms, and burdens; benefits may or may be not balanced with harms and burdens.

Quality of supporting evidence

	High-quality evidence	Moderate-quality evidence	Low-quality evidence	Very-low-quality evidence
Strong recom.	Consistent evidence from well-performed RTC or exceptionally strong evidence from unbiased observational studies	Evidence from RTC with important limitations (inconsistent results, methodologic flaws, indirect or imprecise), or unusually strong evidence from unbiased observational studies	Evidence for at least one critical outcome from observational studies, from RTC with serious flaws, or indirect evidence	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence

Implications

	High-quality evidence	Moderate-quality evidence	Low-quality evidence	Very-low-quality evidence
Strong recom.	Recommendation can apply to most patients in most circumstances. Further research is very unlikely to change our confidence in the estimate of effect.	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.	Recommendation may change when higher quality evidence becomes available; any estimate of effect, for at least one critical outcome, is very uncertain.
Weak recom.	The best action may differ depending on circumstances or patients or social values. Further research is very unlikely to change our confidence in the estimate of effect	Alternative approaches likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.	Other alternatives may be equally reasonable. Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.	Other alternatives may be equally reasonable. Any estimate of effect, for at least one critical outcome, is very uncertain.