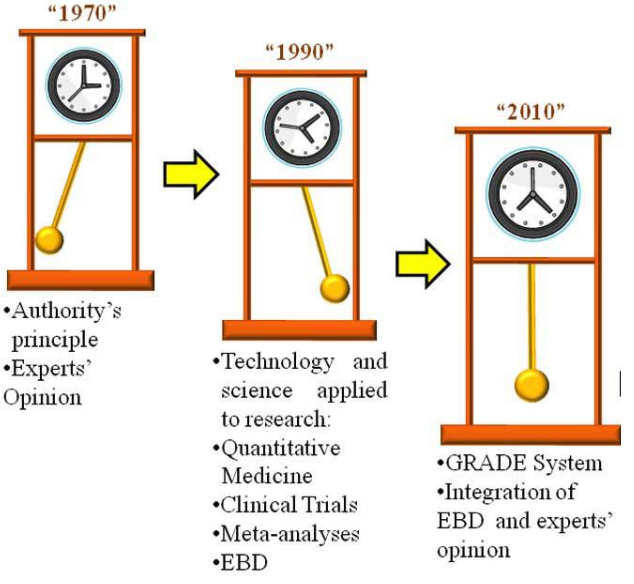
Pro e contro delle sperimentazioni cliniche e dell'evidence-based medicine

Dott. Oscar Bortolami

Prof. Giuseppe Verlato

Sezione di Epidemiologia e Statistica Medica – Università di 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

IL "PARADIGMA STORICO" DELLA FORMAZIONE E DELLA PRATICA MEDICA

Nella Medicina l'apprendimento e l'acquisizione delle competenze professionali è stato basato sulle autorità del settore

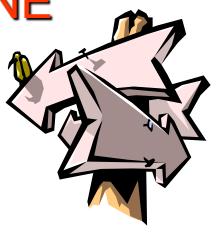
(approccio autoritario ➔ principio d'autorità)

- *ESPERTI*
- *LIBRI DI RIFERIMENTO*
- *CONSENSUS CONFERENCE*

ASSUNZIONI DEL "PARADIGMA STORICO"

- Osservazioni non sistematiche basate sull'esperienza clinica sono valide per basare e formare le nostre conoscenze sulla prognosi, la diagnosi e l'efficacia dei trattamenti;
- La conoscenza dei meccanismi biologici e fisiologici è sufficiente per guidare la pratica clinica;
- Il buon senso, l'esperienza clinica e la conoscenza dei meccanismi biologici sono sufficienti per valutare nuovi strumenti diagnostici e nuovi trattamenti;
- Le autorità del settore (esperti) devono produrre per consenso le "linee guida per il trattamento e la diagnosi delle patologie"

IL NUOVO PARADIGMA "EVIDENCE BASED MEDICINE [EBM]"



Le decisioni e le conoscenze sui test diagnostici, sulla prognosi e sui trattamenti devono essere basate sulle ***evidenze provenienti dalla ricerca clinica***.

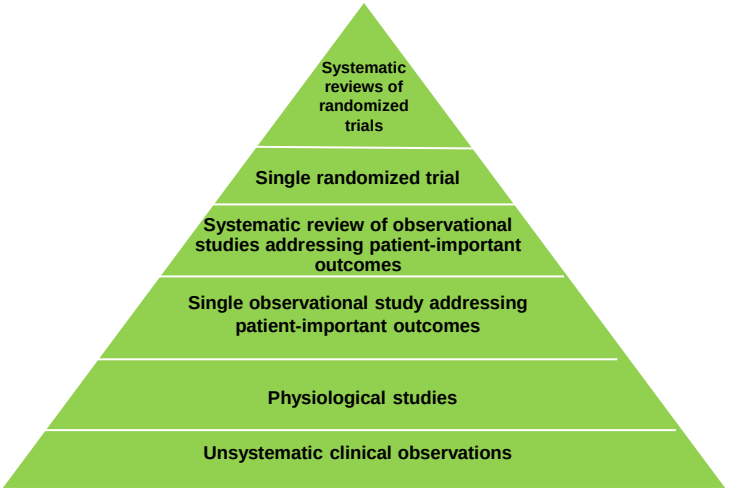
EBM

L'esperienza clinica è fondamentale. Tuttavia l'osservazione clinica deve essere fatta in modo sistematico, riproducibile e quantitativo (adottare la metodologia della ricerca);

La conoscenza dei meccanismi biologici è necessaria, ma non è sufficiente per predire la "miglior" pratica clinica;

La formazione e la pratica medica deve essere basata anche sulla capacità da parte del medico di acquisire l'informazione sulla miglior evidenza possibile (basata sui risultati delle più importanti ricerche cliniche) circa un problema clinico (capacità di consultare la letteratura medica e le banche dati).

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

L'importanza della sperimentazione clinica controllata nel progresso della medicina.

a favore …

1) La terapia ormonale sostitutiva

2) Il fumo

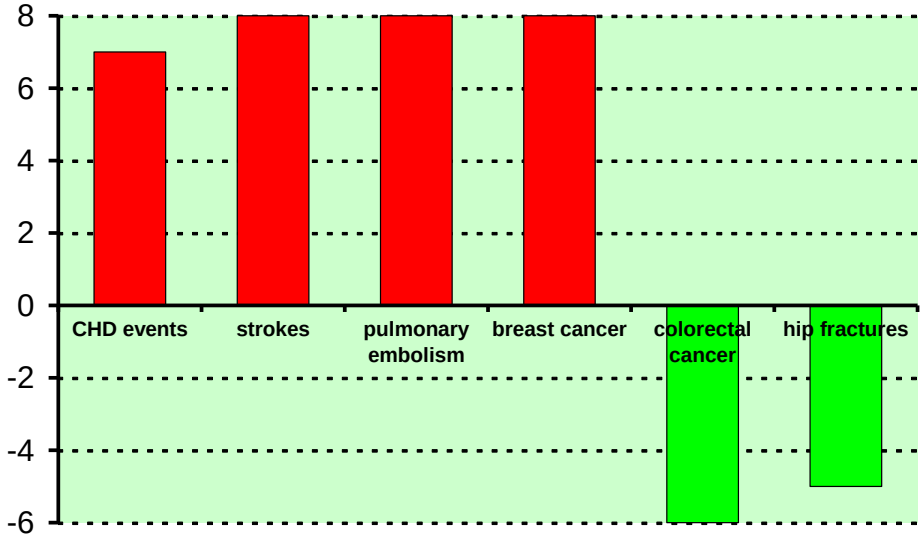
La terapia ormonale sostitutiva

"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)

La terapia ormonale sostitutiva



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

L'importanza della sperimentazione clinica controllata nel progresso della medicina.

contro …

1) La terapia ormonale sostitutiva
2) Il fumo e il cancro del polmone

Il fumo

"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

L'importanza della meta-analisi nel progresso della medicina.

a favore …

1) **<u>Terapia trombolitica e lidocaina nell'infarto miocardico</u>**
2) Linfadenectomia estesa nel cancro gastrico

EFFETTO DELLA TERAPIA TROMBOLITICA SULLA MORTALITA' DOPO INFARTO MIOCARDICO: L'EVIDENZA SCIENTIFICA E L'OPINIONE DEGLI ESPERTI A CONFRONTO

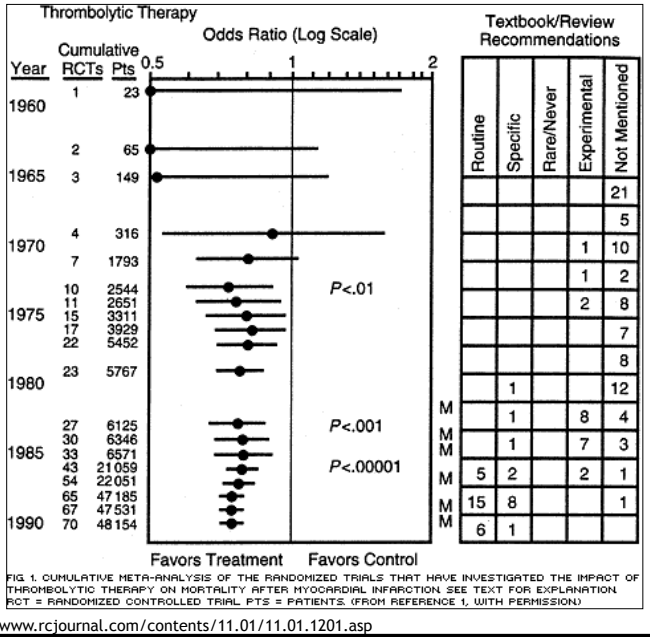


FIG. 1. CUMULATIVE META-ANALYSIS OF THE RANDOMIZED TRIALS THAT HAVE INVESTIGATED THE IMPACT OF THROMBOLYTIC THERAPY ON MORTALITY AFTER MYOCARDIAL INFARCTION. SEE TEXT FOR EXPLANATION. RCT = RANDOMIZED CONTROLLED TRIAL. PTS = PATIENTS. (FROM REFERENCE 1, WITH PERMISSION)

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

EFFETTO DELLA LIDOCAINA SULLA PREVENZIONE DI ARITMIA VENTRICOLARE LETALE DOPO INFARTO MIOCARDICO: L'EVIDENZA SCIENTIFICA E L'OPINIONE DEGLI ESPERTI A CONFRONTO

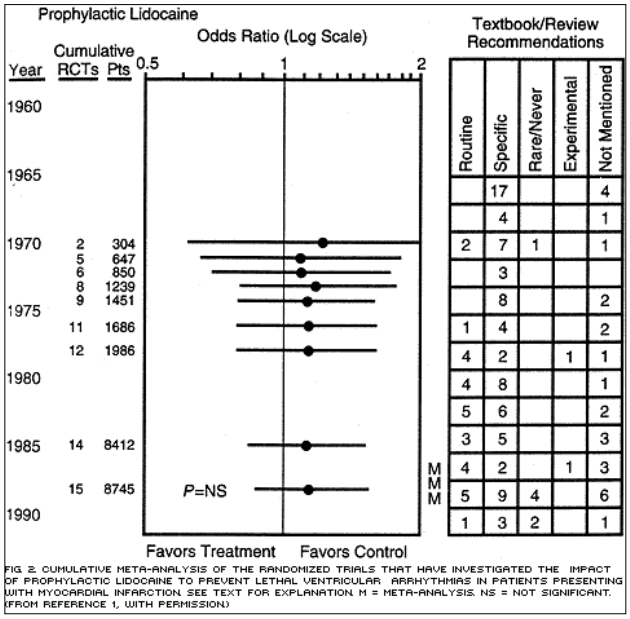


FIG 2. CUMULATIVE META-ANALYSIS OF THE RANDOMIZED TRIALS THAT HAVE INVESTIGATED THE IMPACT OF PROPHYLACTIC LIDOCAINE TO PREVENT LETHAL VENTRICULAR ARRHYTHMIAS IN PATIENTS PRESENTING WITH MYOCARDIAL INFARCTION. SEE TEXT FOR EXPLANATION. M = META-ANALYSIS. NS = NOT SIGNIFICANT. (FROM REFERENCE 1, WITH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

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

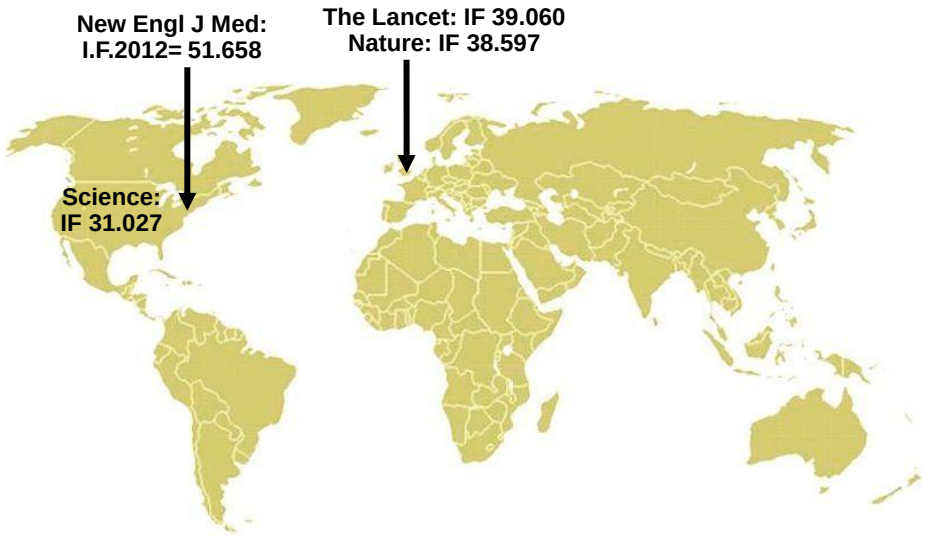
Se la meta-analisi fosse stata adeguatamente riconosciuta, la terapia fibrinolitica sarebbe stata adottata e la lidocaina sarebbe stata abbandonata nell'infarto miocardico 10-15 anni prima.

L'importanza della meta-analisi nel progresso della medicina.

contro …

1) Terapia trombolitica e lidocaina nell'infarto miocardico
2) **Linfadenectomia estesa nel cancro gastrico**

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].

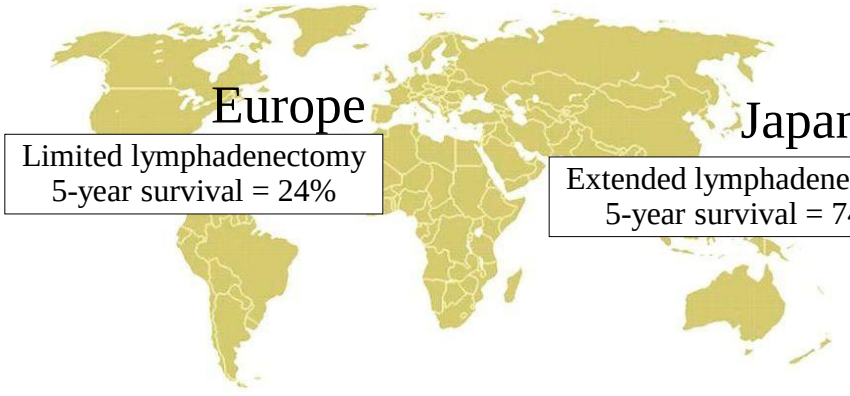
However, at the same time the most widely used intervention in Europe and the States remained a **LIMITED** (D1) lymphadenectomy.

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.

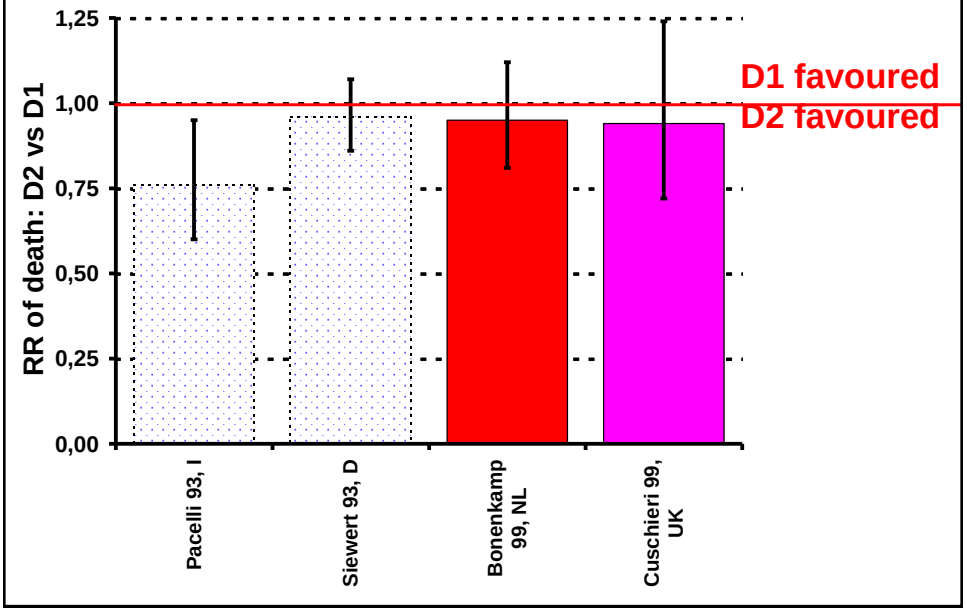
The Dutch and British surgeons organized large trials, where patients were randomly assigned to either limited (D1) or extended (D2) lymphadenectomies.

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].

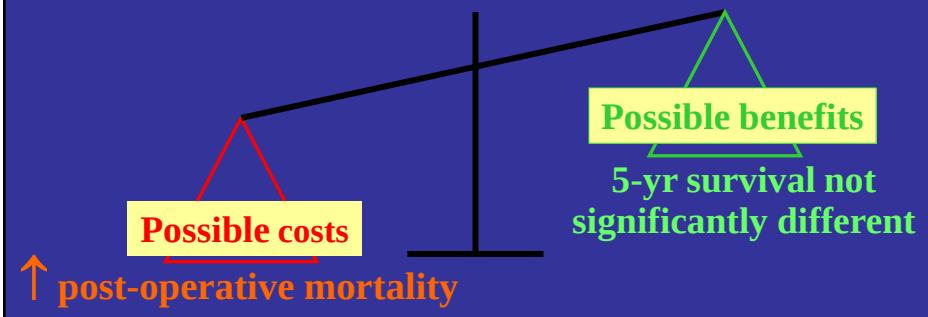
Bonenkamp JJ, Hermans J, Sasako M, van de Velde CJH, for the Dutch Gastric Cancer Group (1999) Extended lymph node dissection for gastric cancer. New Engl J Med 340:908-914

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?



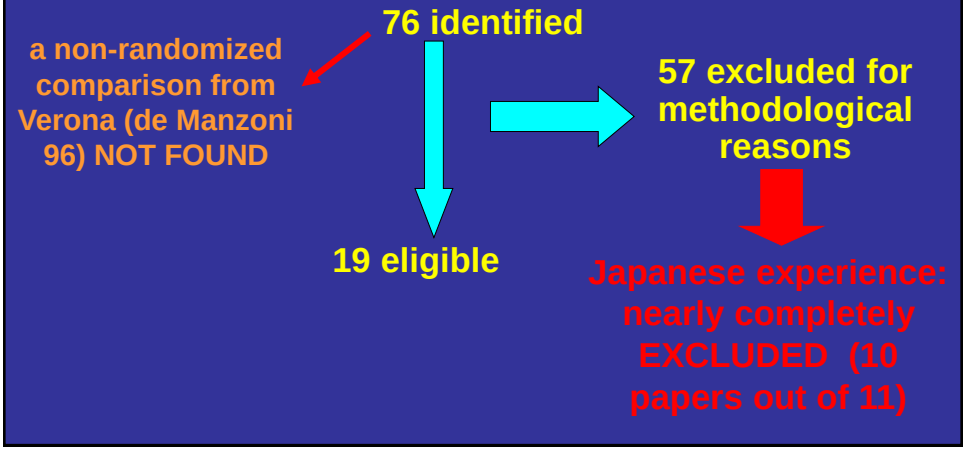
Evidence for D2 dissection is inconclusive:

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.

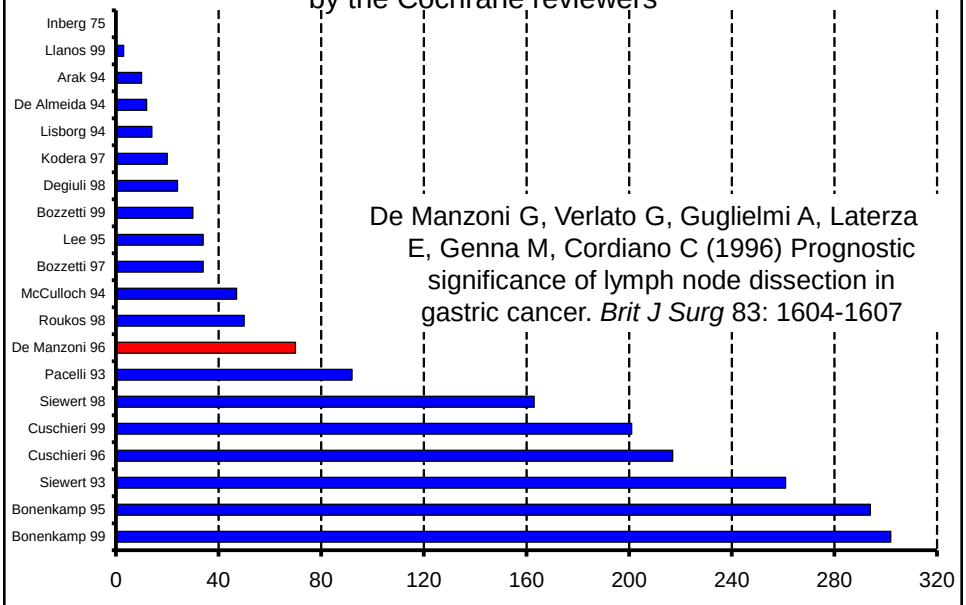
McCulloch, *Brit J Surg*, 2005

McCulloch P, Niita ME, Kazi H, Gama-Rodrigues J (2003) Extended versus limited lymph nodes dissection technique for adenocarcinoma of the stomach (Cochrane Review). In: *The Cochrane Library*, Issue 4. Chichester, UK: John Wiley & Sons.

McCulloch P, Niita ME, Kazi H, Gama-Rodrigues JJ (2005) Gastrectomy with extended lymphadenectomy for primary treatment of gastric cancer. *Brit J Surg*, 92: 5-13



Number of citations achieved by the study from Verona (De Manzoni et al, 1996), compared with citations achieved by the 19 papers selected by the Cochrane reviewers



Source: ISI Web of Science, 21th February 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."

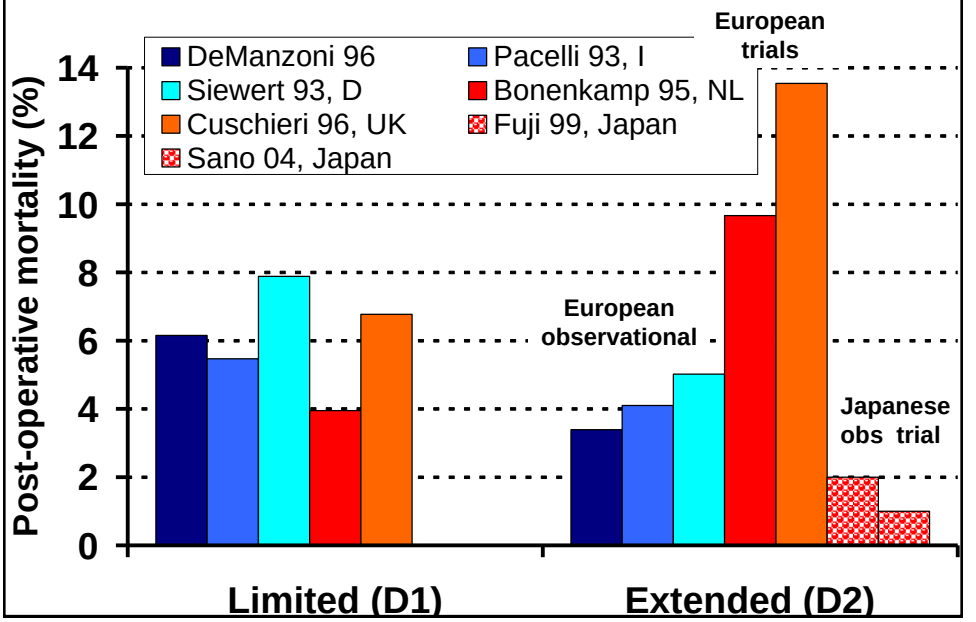
De Manzoni G, Verlato G (2005) Gastrectomy with extended lymphadenectomy for primary treatment of gastric cancer (letter). *Brit J Surg*, 92(6): 784

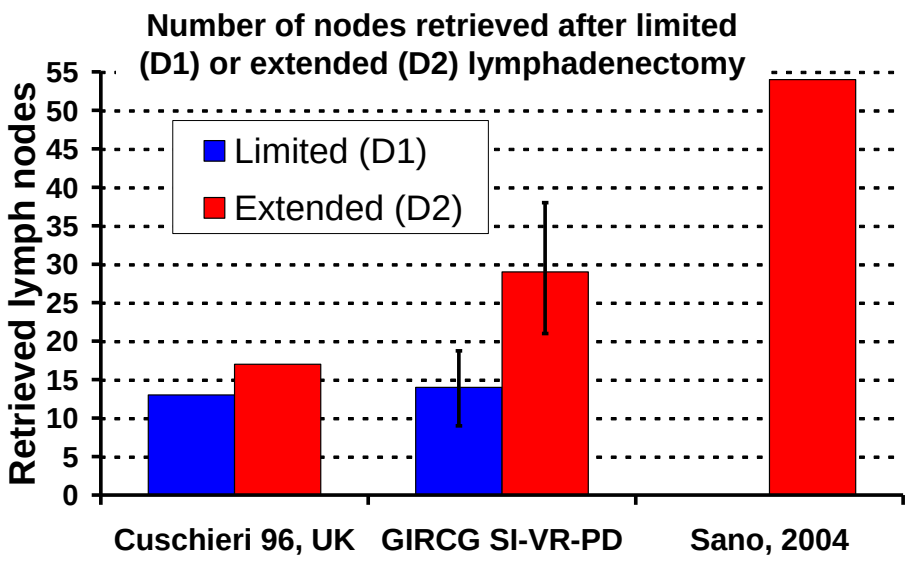
However, the latter trials presented a rather low surgical quality, as they were performed by surgeons without previous training in extended lymphadenectomy, executing less than 5 interventions per year.

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)

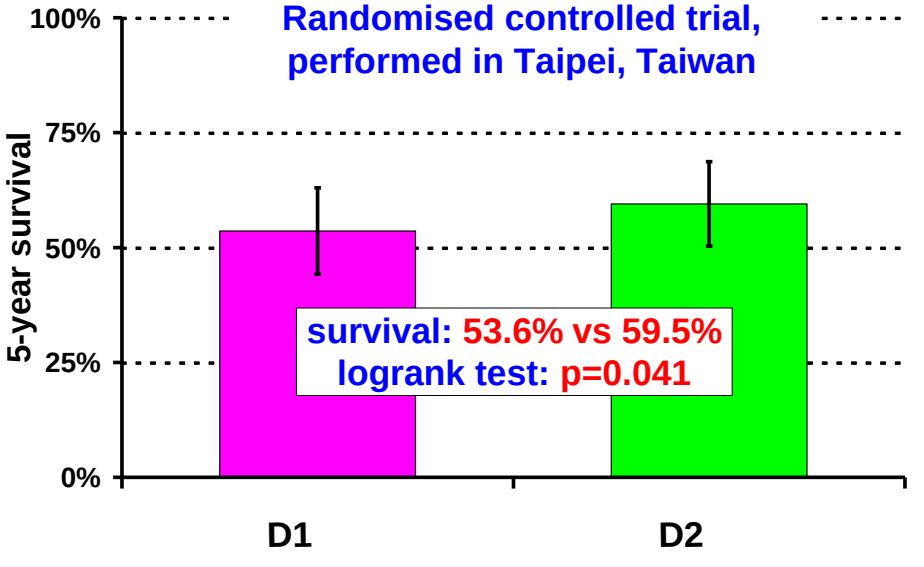
"*In our opinion, it is extremely difficult to ask Japanese surgeons, in whose series post-operative mortality is only 1-2%, to believe in randomized clinical trials where post-operative mortality peaks to 10-14%, irrespectively of methodological quality of those studies*."

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

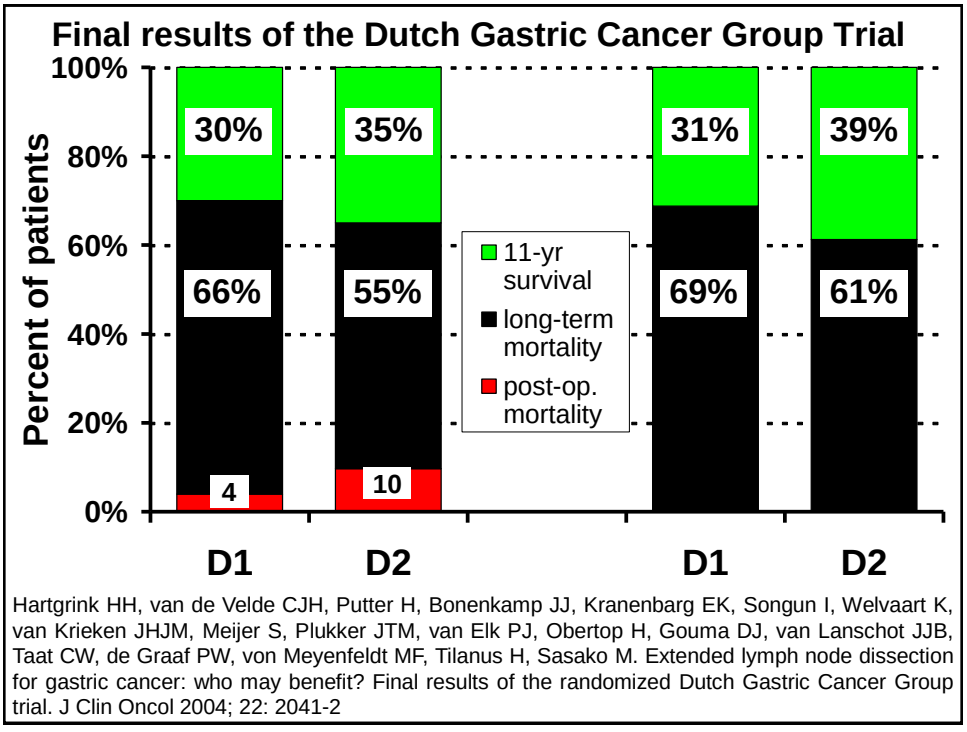
In 2006 another randomized trial [Wu et al, 2006] has been published, showing a mild but significant survival advantage after D2* with respect to D1.

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].

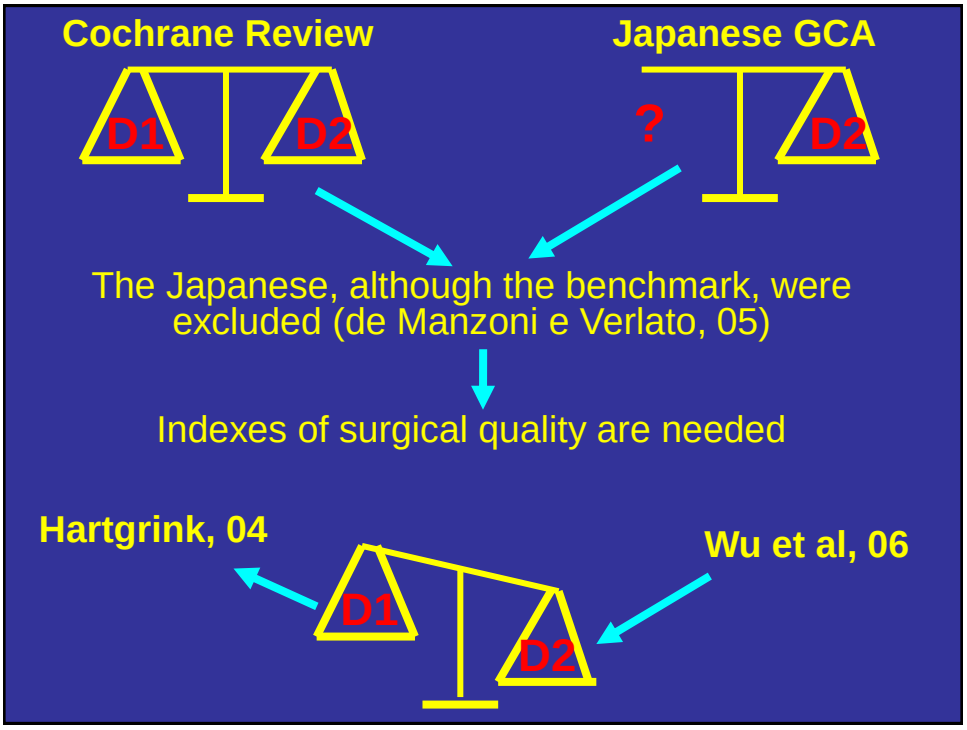
* Wu et al [2006] reported that they compared D1 with D3 lymphadenectomy, but actually their D3 procedure was a slightly extended D2.



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.

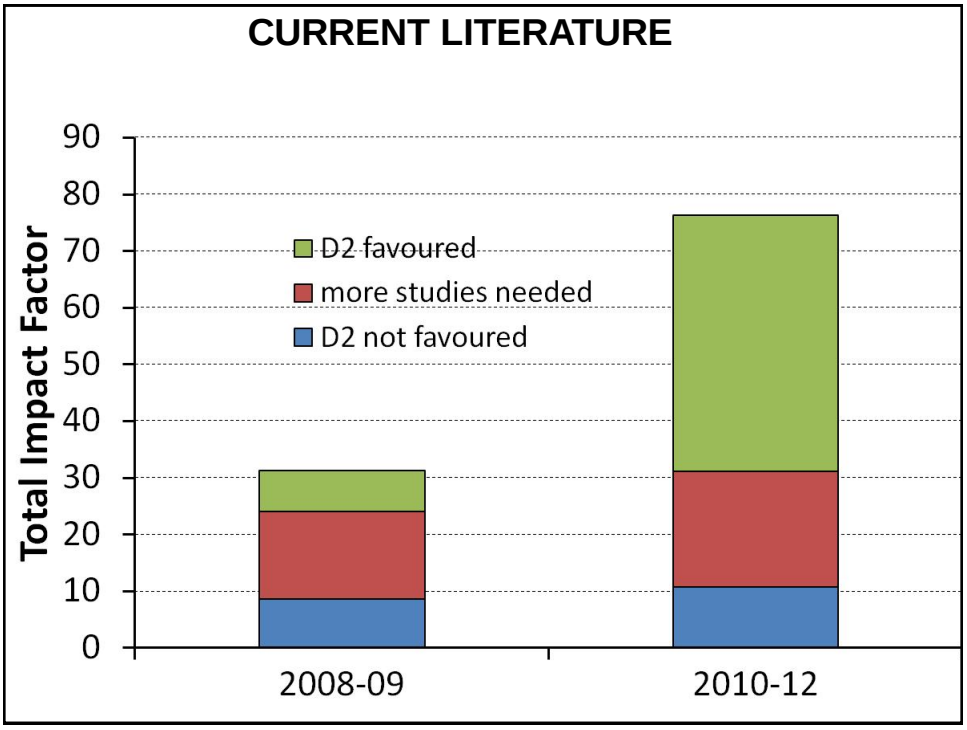
NATIONAL GUIDELINES

D2 was adopted as the standard of surgical treatment with curative intent by the Japanese, German and British national guidelines, by the European Society for Medical Oncology (ESMO) guidelines, by the joint ESMO - ESSO (European Society of Surgical Oncology – ESTRO (European Society of Radiotherapy and Oncology) guidelines.

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).

Of note, D2 is recommended by the Japanese guidelines since 1981 and by German guidelines since at least 2005, and D2 was adopted as the preferred lymphadenectomy within the Italian Research Group for Gastric Cancer (GIRCG) since 1992.

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.



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

Indexes of surgical quality in an Italian series of 1032 patients, according to the extension of lymphadenectomy

	Lymphadenectomy		
	Standard	Extended	Super-extended
Number of excised nodes Median (interquartile range)	14 (9-19)	29 (21-38)	47 (37-57)
Removal of adjacent organs:			
Spleen	6.1%	10.1%	11.4%
Spleen+pancreas	1.8%	2.4%	11.4%
Complications:			
surgical	18.4%	19.2%	21.4%
non-surgical	11.0%	16.3%	11.8%
Post-operative mortality	5.7%	3.6%	2.7%

Piede diabetico:
0=assente 1=presente

METODO GRADE: integrazione tra EBM e principio d'autorità ?

Metodologia utilizzata per organizzare il lavoro di gruppo: **Delphi process and Nominal (Expert) Group techniques**

Metodologia utilizzata per ricavare raccomandazioni dalla letteratura: **GRADE System**

Tecnica Delphi (o Nominale) - 1

La “tecnica” Delphi è un metodo che aiuta a trovare soluzioni a problemi complessi, potenziando la comunicazione all’interno di un gruppo (o panel) e nello stesso tempo limitando il potere di influenza di ciascun individuo.

I partecipanti al panel Delphi vengono stimolati a produrre le idee che ritengono più adatte a risolvere un dato problema e queste idee vengono successivamente diffuse tra i partecipanti al panel, che hanno la possibilità di rivedere individualmente le proprie posizioni, senza però doverle difendere davanti al gruppo.

La riproposizione delle strategie indicate dai pannellisti continua fino al raggiungimento di posizioni condivise. In questo modo diventa possibile raggiungere il consenso su una o più soluzioni al problema proposto, evitando di far prevalere le soluzioni avanzate da coloro che singolarmente hanno maggiore capacità di imporsi.

Tecnica Delphi - 2

Lo scopo di queste tecniche è di ottenere e sintetizzare il parere di più operatori (“esperti”) su un determinato argomento.

Nella tecnica Delphi i membri del gruppo comunicano tra loro per via indiretta, per posta, E-Mail o altro.

Nella tecnica Nominale i membri del gruppo si riuniscono direttamente, ma vengono utilizzati bigliettini anonimi o sistemi simili.

In entrambi i casi viene quindi garantito l’anonimato nella prima fase di giudizio.

Con l’aiuto di un coordinatore, il parere di tutto il gruppo viene sintetizzato esprimendo il grado di accordo in termini numerici.

Fasi principali delle tecniche Delphi e nominale

a) Viene formulata una domanda (o un insieme di domande).
b) I membri esprimono un giudizio senza comunicarlo agli altri. Il giudizio può essere: “accordo, disaccordo, indifferente”; un ordine di priorità da assegnare; un punteggio.
c) Il coordinatore raccoglie e sintetizza i giudizi espressi e li comunica ai membri del gruppo, riferendo sia il punteggio sintetico di tutto il gruppo che i valori individuali in forma anonima.
d) Segue una discussione (diretta o indiretta, ad esempio epistolare) nella quale i membri esprimono un parere sul giudizio complessivo.
e) I membri esprimono un successivo giudizio.
f) Il gruppo ridiscute i giudizi e giunge ad un accordo. Eventuali punti di disaccordo vengono esplicitati.

GRADE system

GRADE è l’acronimo di 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

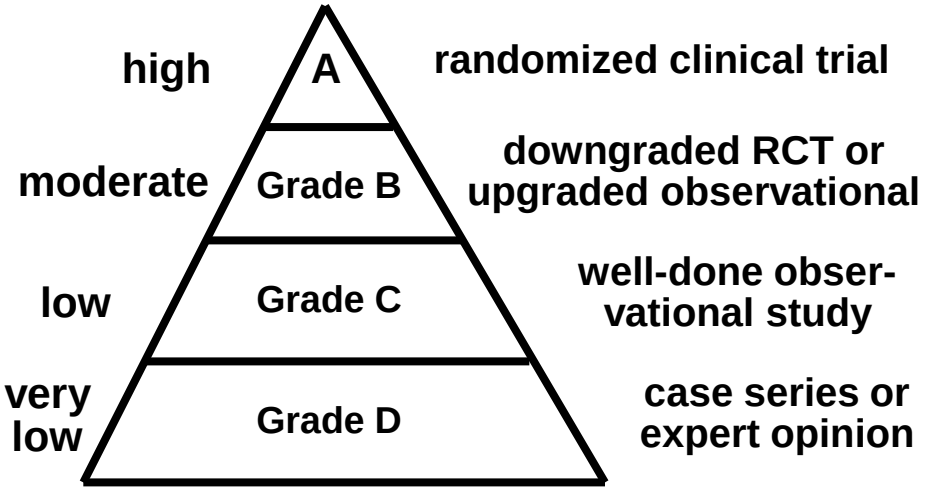
Quindi, il sistema GRADE utilizza sia l’EBM (Medicina basata sulle prove di efficacia) sia il parere di “esperti”, espresso in modo democratico con Tecnica Delphi.

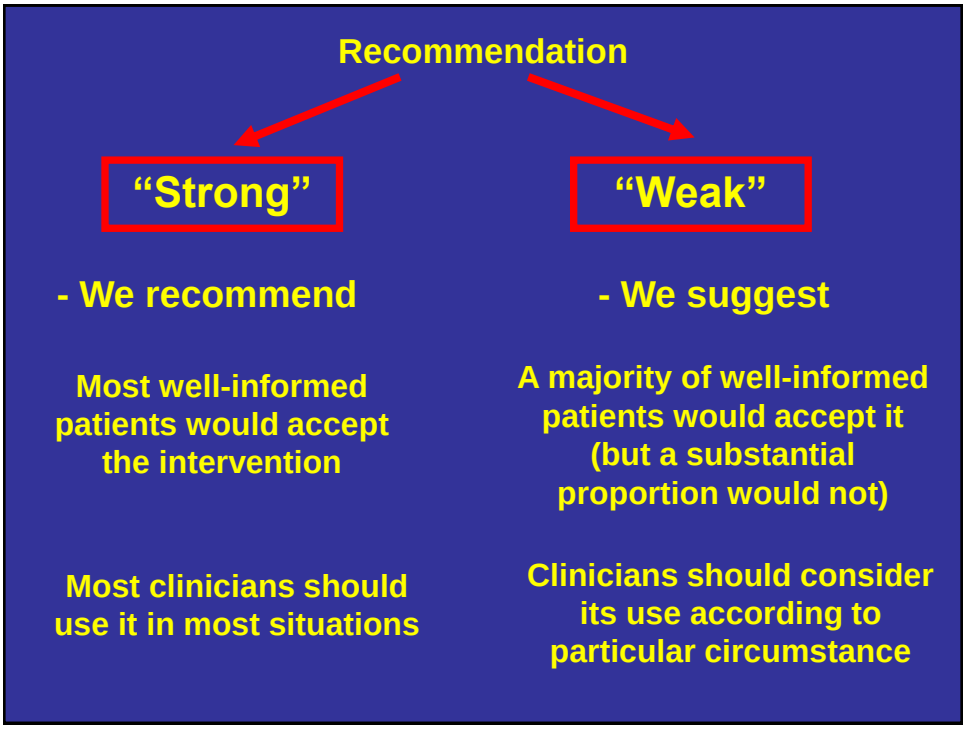
GRADE system

La piramide dell'evidenza nel Sistema Grade è leggermente diversa dalla piramide dell'evidenza dell'EBM classica.

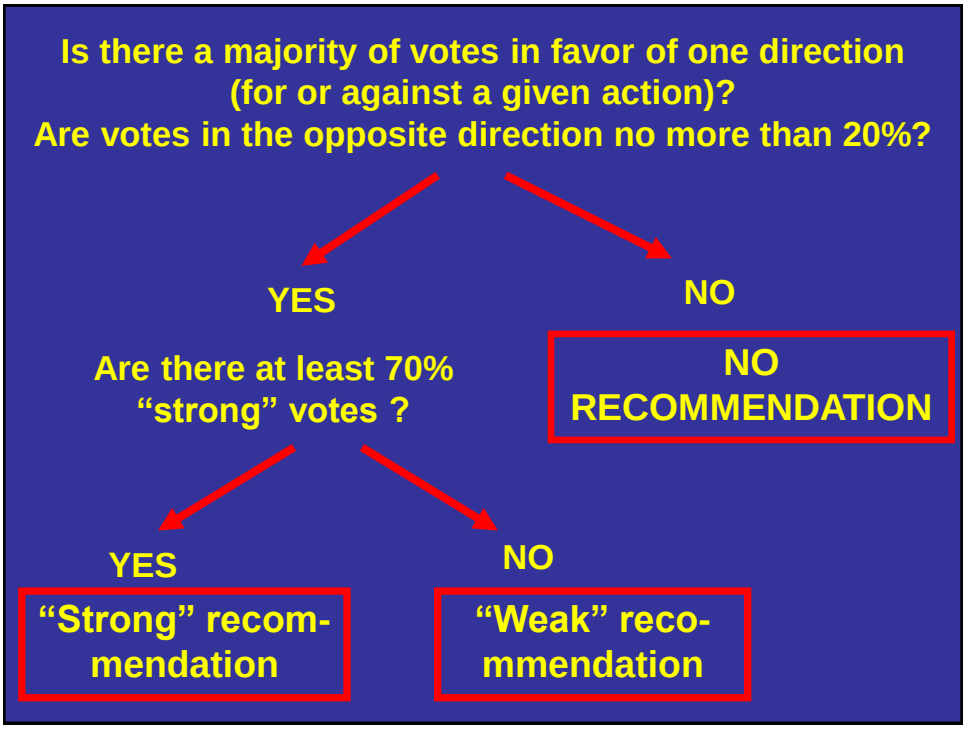
Infatti, pur riconoscendo la superiorità generale degli studi sperimentali rispetto agli studi osservazionali, si ammette che la possibilità che uno studio osservazionale sia promosso ad un livello superiore e uno studio sperimentale retrocesso ad un livello inferiore.

Piramide dell'evidenza





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.