VOLUME 26 · NUMBER 7 · MARCH 1 2008

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Developing Clinical Recommendations for Breast, Colorectal, and Lung Cancer Adjuvant Treatments Using the GRADE System: A Study From the Programma Ricerca e Innovazione Emilia Romagna Oncology Research Group

Rossana De Palma, Alessandro Liberati, Giovannino Ciccone, Elena Bandieri, Maurizio Belfiglio, Manuela Ceccarelli, Maurizio Leoni, Giuseppe Longo, Nicola Magrini, Maurizio Marangolo, and Fausto Roila

From the Agenzia Sanitaria Regionale Regione Emilia Romagna, Bologna; Università degli Studi di Modena e Reggio Emilia; Centro Valutazione Efficacia Assistenza Sanitaria, AUSL Modena; Centro Prevenzione Oncologica, Azienda Ospedaliera Molinette, Torino; Consorzio Mario Negri Sud, Santa Maria Inbaro, Chieti, Italy; Azienda Sanitaria Locale Ravenna; and the Azienda Ospedaliero Policlinico, Perugia, Italy.

Submitted April 18, 2007; accepted October 23, 2007.

Supported by Contratto No. 249, Bando Nazionale Ricerca Finalizzata 2005, Ministero della Salute, Italia. Supported in part by the Fondo per l'Innovazione, an unrestricted grant provided by the following companies: AstraZeneca, Chiesi, Cordis Glaxo Smith Kline, Lilly, Pfizer Novartis, Siemens, and Takeda.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Rossana De Palma, MD, PRI-ER Oncology Research Group, Agenzia Sanitaria Regionale, Viale Aldo Moro 21, Bologna, Italy; e-mail: rdepalma@regione.emilia-romagna.it.

0732-183X/08/2607-1033/$20.00

DOI: 10.1200/JCO.2007.12.1608

ABSTRACT

Purpose
In the area of anticancer drugs, the legitimate search for effective interventions can be jeopardized by the strong pressure for accelerated approval, which may hinder the full assessment of their benefit-risk profile. We aimed to produce drug-specific recommendations using an explicit approach that separates the judgments on quality of evidence from the judgment about strength of recommendations.

Materials and Methods
We used the GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) system to develop recommendations for the use of specific anticancer drugs/regimens; 12 clinical questions relevant to adjuvant treatment of breast (three), colorectal (four) and lung (five) cancer have been assessed by multidisciplinary panels supported by a group of methodologists.

Results
For nine of 12 questions, recommendations were produced (one strong and six weak in favor and one weak and one strong against the index treatment); for the remaining three questions no specific course of action could be recommended. The perceived benefits to risk balance of the treatment was the most important and statistically significant ($P < .01$) predictor of panels' recommendations and of their strength, whereas panelists' personal (age, sex) and professional (specialty) characteristics were not statistically associated.

Conclusion
Because the GRADE system sets out an explicit process going from evaluation of the quality of evidence and benefit-risk profile to the judgment of the strength of recommendations, in this experience, it proved very useful to combine methodologic rigor with the interdisciplinary participation that is important in the definition of evidence based clinical policies.

J Clin Oncol 26:1033-1039. © 2008 by American Society of Clinical Oncology

INTRODUCTION

Health care systems are not well equipped to deal systematically with innovation. Traditional health technology assessment (HTA) is based on a posteriori evaluation of interventions already entered clinical practice, and it is, therefore, difficult to carry out the kind of evaluation that would be necessary.[1-3]

The Emilia Romagna Health Care Agency launched a special program—PRI-ER, or Programma Ricerca e Innovazione dell'Emilia Romagna—aimed at systematically introducing evaluative methods within its health care system, targeting promising innovations or interventions whose benefit-risk profile appears still uncertain.[4]

The area of anticancer treatments is an obvious candidate for HTA activities. Anticancer drugs well represent the changes occurring in the field of drug development and registration, where new compounds are often registered with a still largely immature benefit-risk profile.[5-8] New and often expensive molecules, in fact, enter clinical practice with limited evidence of effectiveness and safety and ill defined indication(s), leaving a potential for inappropriate use.

In this project, we focused on tumors whose frequency, health care burden, and use of medical treatments is common and for which evidence-based practice guidelines already exist. Within these tumors (breast, colorectal, and lung) we set out to

produce drug-specific recommendations, similar to Cancer Care Ontario's Program in Evidence-Based Care,[9-11] targeting specific open clinical questions for adjuvant treatment. We also aimed at identifying open research questions where the program could promote confirmatory trials.

We used the GRADE (Grades of Recommendation, Assessment, Development, and Evaluation) system, recently proposed to overcome shortcomings of previous approaches,[12-14] which is based on a sequential assessment of the quality of evidence followed by an analysis of the benefit-risk balance and subsequent judgment about the strength of recommendations.

This article reports on our experience and discusses the relationship(s) among panel members' characteristics, quality of scientific evidence, benefit-risk balance of any given drug regimen, and the direction and strength of the recommendation(s).

MATERIALS AND METHODS

This project was based on the steps briefly described in detail in the following sections and in Table 1.

Definition of the Project's Objectives

A kick-off workshop was convened in February 2005 to discuss experiences in cancer guidelines programs in the United Kingdom, Canada, and France.[15] This helped focusing our aims on the production of drug-specific recommendations involving multidisciplinary panels.

Identification of the Coordinating Group

A 10-person coordinating group (CG)—five members with expertise in oncology and five in critical appraisal and research synthesis—oversaw the process. The CG was appointed by the regional agency with the tasks of (a) undertaking the initial literature review; (b) preparing the training material and the "summary of findings tables" needed for the panel to formulate recommendations; and (c) chairing panel meetings and drafting the initial versions of the recommendations. Details on the literature review are available from the authors.

Table 1. The Process Leading to the Development of the Clinical Recommendations

Project Step
Definition of project's objectives
Appointment of the Coordinating Group (5 physicians and 5 clinical epidemiologists)
Appointment of the three multidisciplinary panels (including different specialists, public health doctors, and patients representatives)
Presentation and illustration to panel members of the GRADE system
Definition of the clinical question(s) and discussion of all the relevant outcomes
Individual rating of the importance (relevance) of each outcome
Systematic literature search and preparation of the evidence and summary of findings Tables for each relevant outcome
Individual rating of the quality of evidence for each relevant outcome and overall
Individual rating of the balance of benefits and harms for each relevant outcome and overall
Individual rating of the direction and strength of the recommendation(s)
Distribution to panel members of the recommendations drafted by the Coordinating Group
Plenary meetings to discuss on the final version of the recommendations
External review of the format and presentation of the text of the recommendations
Publication and dissemination of the recommendations

Abbreviation: GRADE, Grades of Recommendation, Assessment, Development, and Evaluation.

The Multidisciplinary Panels, Disclosure of Conflicts of Interest, and Clinical Questions

We convened three multidisciplinary panels on adjuvant treatment for breast, colorectal, and lung cancers. Panel members were chosen to include representatives of hospitals from around the region and to represent all relevant specialties/expertise (medical oncologists, radiotherapists, surgeons, pathologists, internists, pneumonologists, pharmacists, and patient representatives). All but five of those invited agreed to participate (Table 2). Each panel members was asked to disclose any tie he/she had in the last 5 years with pharmaceutical companies manufacturing the drugs considered in the recommendations. Of 57 panel members (16 medical oncologists and 41 others), none was a regular consultant, two (3%) received a research grant of more than €30,000, and 15 (26%) participated in trials using the index drugs (12 were medical oncologists).

The identification of clinical questions was guided by consideration of (a) the relative importance of the treatment; (b) the lack of conclusive recommendations in existing guidelines; and (c) the interest of the local oncology community. This way, 12 clinical questions (three for breast, four for colorectal, and five for lung cancer) were identified (Table 3).

The GRADE System and Its Application in Our Study

To develop our recommendations, we used GRADE because it represents an explicit assessment of the quality of evidence, the balance between benefits and risks, and the strength of recommendations. Separation of the judgments on quality of evidence and strength of recommendations is a critical and defining feature of GRADE. Five limitations—related to study quality, consistency, directness, precision, and reporting bias—may lead to its downgrading. Large effects and dose-response gradient can lead to upgrading quality of evidence.[12] Given the type of clinical questions we addressed (relatively new drugs with only a few trials available) and the type of studies eligible (only randomized controlled trials) we could downgrade the evidence only if one of the aforementioned drawbacks occurred. The main criteria used were presence of serious limitations in study conduct, duration of follow-up, and type (relevance) of end points used. We downgraded quality from "good" to "fair" whenever one or both of the following occurred: (a) follow-up less than 5 years or (b) disease-free instead of overall survival as the main outcome.

Making recommendations then involves tradeoffs between benefits and harms and therefore four elements should be considered: tradeoffs, quality of

Table 2. Distribution of the Characteristics of the 57 Panel Members

Characteristic	No.			Total	
	Breast	Colorectal	Lung	No.	%
Age, years					
< 55	11	11	7	29	50.9
≥ 55	9	10	9	28	49.1
Sex					
Male	13	15	14	42	73.7
Female	7	6	2	15	26.3
Specialty					
Oncology	5	6	5	16	28.0
Radiation therapy	2	3	3	8	14.0
Surgery	3	2	1	6	10.5
Internal medicine	—	2	2	4	7.0
General practice	1	1	1	3	5.3
Public health/administration	1	1	1	3	5.3
Patients' representatives	4	3	—	7	12.3
Pharmacist	1	1	1	3	5.3
Others	3	2	2	7	12.3
Total	20	21	16	57	100.0

Table 3. Clinical Questions and Strength and Direction of the Recommendations Formulated by the Three Multidisciplinary Panels Using the GRADE System

Clinical Question	Recommendation
Breast cancer	
In women with HR+ breast cancer in postmenopause, are aromatase inhibitors recommended instead of tamoxifen?	Probably use it, weak positive
In women with positive nodes, should a taxane be used as adjuvant therapy?	Probably use it, weak positive
In women with HER-2+ breast cancer (HER-2 3+ in immunohystochemistry or FISH test +) without cardiac impairment, is trastuzumab recommended as adjuvant therapy?	Probably use it, weak positive
Colorectal cancer	
In patients with stage II colon cancer, is adjuvant chemotherapy recommended?	No recommendation
In patients with stage III colon cancer, should oxaliplatin be used in association with FU + folinic acid?	Probably use it, weak positive
In patients with stage III colon cancer, is capecitabine recommended instead of FU + folinic acid?	Probably use it, weak positive
In patients with stage II and III rectal cancer, is chemoradiotherapy recommended before surgery instead of postsurgery?	Use it, strong positive
Non–small-cell lung cancer	
In patients with stage Ib-II NSCLC, should chemotherapy with cisplatin be recommended instead of nontreatment?	Probably use it, weak positive
In patients with stage IIIa NSCLC, should chemotherapy with cisplatin be recommended instead of nontreatment?	No recommendation
In patients with stage Ib-II NSCLC, should postsurgical radiotherapy be recommended?	Don't use it, strong negative
In patients with stage IIIa NSCLC, should postsurgical radiotherapy be recommended?	No recommendation
In patients with stage IIIa NSCLC, should postsurgical chemoradiotherapy be recommended instead radiotherapy or chemotherapy alone?	Probably don't use it, weak negative

Abbreviations: GRADE, Grades of Recommendations, Assessment, Development, and Evaluation; HR, hormone receptor; HER-2, human epidermal growth factor receptor 2; FISH, fluorescent in situ hybridization; FU, fluorouracil; NSCLC, non–small-cell lung cancer.

evidence, translatability of evidence into a specific setting and uncertainty about the baseline risk for the population.

Using these evaluations, recommendations can be classified into four mutually exclusive categories: Do it, probably do it, probably don't do it, don't do it. In this study, we also allowed panels to abstain from making a recommendation—adding specific suggestions for new studies to be undertaken—when evidence was too sparse. All steps in the process are shown in Table 1.

Panel Activities

After identification of panel members, a first meeting was held to introduce the GRADE system and its key features in comparison with other existing approaches. Overall, the panels had seven meetings as they completed the following tasks.

During the first meeting, panel members refined the clinical questions and choose the outcomes of interest relevant to deciding whether a given adjuvant treatment is worth recommending. Then, they individually voted using a scale of 1 to 9 on whether each outcome should or should not be considered in the assessment.

Between the second and third meeting, the CG identified relevant studies and prepared for each relevant outcome "evidence tables," with short comments on all the predefined dimensions of quality ("ie, study design, study quality, consistency, and directness); quantitative summaries of effect for each outcome were also provided (copies of this material are available, in Italian, from the authors). The CG was also in charge of producing "summaries of findings tables" providing data on absolute and relative risk reduction on the outcomes previously identified as critical for the decision. At the third meeting, the material was presented and discussed. Between this and the subsequent meeting, panel members were asked to individually rate the quality of evidence (for each item separately, and then across all items), the balance between benefits and risks, and the draft recommendation. Provisional results were presented in draft form to panel members, highlighting agreement and disagreement. Final adjudication of the recommendation (s) was made after extensive discussion and, if unanimity could not be reached, by majority rule.

External Review of the Recommendations Before Dissemination

The CG prepared a draft of the final document. This was fed back to panel members and subsequently presented to external reviewers (n = 18) asked to comment on the best format but not on the content of the recommendations. The material was prepared in hard-copy and electronic format.[16]

Data Analysis

Data were collected though hard-copy forms during the panels face-to-face meetings, and via e-mail between meetings. Information about panel members' personal characteristics (age, sex, specialty) and individual judgment of the quality of evidence, benefit-risk profile and judgment about the recommendations were collected individually.

To evaluate the potential influence of individual panel members' characteristics and of the main variables considered by the GRADE system on the final strength and direction of the recommendations, we analyzed all 189 valid ratings obtained by three panels (comprising a total of 57 members) who assessed the evidence and voted for 12 recommendations. Thirty-five votes (of the total of 234 that should have been expressed) were not available because individual panel members did not vote or were absent at a meeting.

Logit regression models were used to analyze predictors of three dependent variables: (a) quality of evidence (high or low); (b) balance between benefits and risk (positive, uncertain, negative); (c) strength and direction of recommendations (strong positive, weak positive, uncertain, weak negative, strong negative). Adjusted odds ratios and 95% CIs were estimated with binary or ordered (for dependent variables with more than two ordinal outcomes) logit models, with a robust variance estimator. Individual panelists were used as clusters.

RESULTS

Panels completed their activities through seven face-to-face meetings (including two initial training sessions). Panel members went through the steps illustrated in the Materials and Methods section (and in Table 1), and each member undertook an individual appraisal of the summaries of findings tables provided by the CG.

Analysis of Panels' Performance of the Steps of the GRADE System

Within different clinical questions, the distribution of panel judgments on the quality of evidence, balance of benefits and risk, and strength of recommendation for each clinical question varied (Fig 1). Overall, there was always variation in the assessment of quality and evaluation of the balance between benefits and risks. This, in turn, led to differences in the strength of recommendations, suggesting different criteria used by panel members in assessing the quality of available

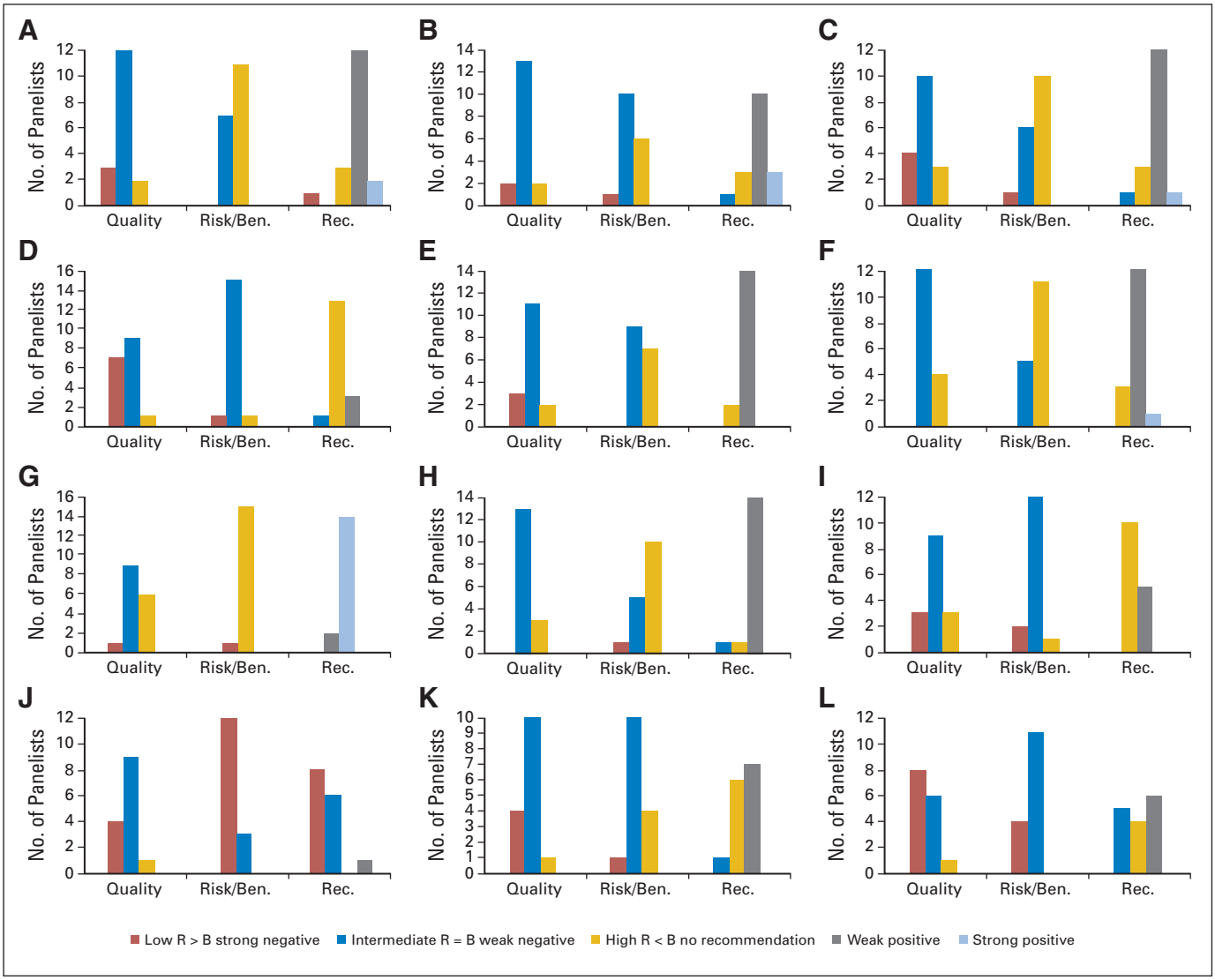

Fig 1. Distribution of the judgments (*x*-axes) on quality of evidence (Q), balance of risks/benefits (R/B) profile and the type of recommendation (R) stated by panel members (*y*-axes) for each of the clinical questions for breast, colorectal and lung cancer. (A) In women with hormone-receptor–positive breast cancer in postmenopause, are aromatase inhibitors recommended instead of tamoxifen? (B) In women with positive nodes, should a taxane be used as adjuvant therapy? (C) In women with human epidermal growth factor 2 (HER-2)-positive breast cancer without cardiac impairment, is trastuzumab recommended as adjuvant therapy? (D) In patients with stage II colon cancer, is adjuvant chemotherapy recommended? (E) In patients with stage III colon cancer, should oxaliplatin be used in association with fluorouracil (FU) + folinic acid? (F) In patients with stage III colon cancer, is capecitabine recommended instead of FU + folinic acid? (G) In patients with stage II and III rectal cancer, is chemoradiotherapy recommended before surgery instead of postsurgery? (H) In patients with stage Ib-II non–small-cell lung cancer (NSCLC), should the chemotherapy with cisplatin instead be recommended of nontreatment? (I) In patients with stage IIIa NSCLC, should chemotherapy with cisplatin be recommended instead of nontreatment? (J) In patients with stage Ib-II NSCLC, should postsurgical radiotherapy be recommended? (K) In patients with stage IIIa NSCLC, should postsurgical radiotherapy be recommended? (L) In patients with stage IIIa NSCLC, should postsurgical chemoradiotherapy be recommended instead of radiotherapy or chemotherapy alone? Ben., benefit; Rec., recommended.

information and the influence of the supporting evidence (or lack thereof).

Table 4 illustrates the distribution of the row data for the 189 ratings expressed by panel members relative to the 12 clinical scenarios. Overall, quality of evidence was rated high/intermediate in 79% of cases. The benefit-risk balance was rated "uncertain" in approximately half of the cases (48%), with little variation across disease sites (range, 43% to 53%). By contrast, wide variation emerged in the proportions indicating a positive benefit-risk balance for breast and colorectal (53% and 52% respectively) versus lung (20%) cancer treatments.

A "weak positive" recommendation was the most frequent option chosen by panelists (51%; range, 65% for breast to 45% for lung). The "uncertain" category accounted for 47 votes (25%).

None of the panelists' characteristics (specialty, age, sex) was associated with any of the judgments analyzed. Oncologists rated quality of evidence higher compared with all other panelists, whereas no differences resulting from specialty were detected in the judgments of the benefit-risk balance and strength of recommendations.

Data reported in Table 4 helps explore the internal consistency of the method. When quality of evidence was "high-intermediate" (n = 150), the benefit-risk profile was more often "favorable" (74% to

Table 4. Frequency (%) of the Total 189 Judgments on Quality of Evidence, Benefit-Risk Balance, and Strength of Recommendation by Panel Characteristics and Evaluation

Criterion	Quality of Evidence				Benefit-Risk Balance						Strength of Recommendation										Total No.
	Low		High		Greater Benefit		Uncertain		Greater Risk		+ +		+		Neutral		−		− −		
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	
Panel characteristics																					
Panel																					
Breast	9	18	42	82	27	53	22	43	2	4	6	12	33	65	9	18	2	4	1	2	51
Colon-rectum	11	17	54	83	34	52	30	46	1	2	15	23	31	48	18	27	1	2	0	0	65
Lung	19	26	54	74	15	20	38	53	20	27	0		33	45	20	27	12	17	8	11	73
Specialty																					
Nononcologist	30	24	96	76	48	38	65	52	13	10	15	12	61	48	34	27	10	8	6	5	126
Oncologist	9	14	54	86	28	44	25	40	10	16	6	9	36	57	13	21	5	8	3	5	63
Age, years																					
< 55	20	21	75	79	38	40	49	52	8	8	10	10	51	54	26	27	2	2	6	7	95
≥ 55	19	20	75	80	38	40	41	44	15	16	11	12	46	49	21	22	13	14	3	3	94
Sex																					
Male	28	20	110	80	49	35	70	51	19	14	14	10	69	50	36	26	12	9	7	5	138
Female	11	22	40	78	27	53	20	39	4	8	7	14	28	55	11	21	3	6	2	4	51
Panel evaluation																					
Quality of evidence																					
Low	—		—		2	5	29	74	8	21	1	3	5	13	22	56	9	24	2	4	39
High	—		—		74	49	61	41	15	10	20	13	92	61	25	17	6	4	7	5	150
Benefit-risk balance																					
Positive	—		—		—		—		—		20	26	56	74	0	0	0	0	0	0	76
Uncertain	—		—		—		—		—		1	1	40	44	45	50	3	4	1	1	90
Negative	—		—		—		—		—		0		1	4	2	9	12	52	8	35	23
Total	39	21	150	79	76	40	90	48	23	12	21	11	97	51	47	25	15	8	9	5	189

49%) compared with when quality was "low" (2% to 5%). Better quality of evidence was associated with positive recommendations: 61% weak and 13% strong as opposed to 13% weak and 3% strong when quality was low.

Table 5 reports on the analysis of the relationships between panel members' characteristics and the judgments on quality of evidence, benefit-risk profile, and direction and strength of recommendations. Overall, no statistically significant association emerged. Further exploration of data indicates that the last three variables were, as expected, significantly associated. Specifically, a high quality of evidence is associated with a positive benefit-risk balance (the probability of a higher rating on the benefit-risk balance is 4.89 times higher if quality of evidence is high) and with a stronger positive recommendation (odds ratio = 3.52; 95% CI, 1.78 to 6.98). The benefit-risk balance was the most important predictor of the direction and strength of the recommendation.

Content and Presentation of the Recommendations

The assessments described herein led to the recommendations reported in Table 3. Overall, there were two strong and seven weak recommendations, and three instances where panels concluded that no recommendation could be formulated.

The final template of the recommendations includes: (a) clinical question and its target population; (b) recommendation including its strength; (c) main reason(s) for grading; (d) distribution of panel members' votes on quality of evidence, benefit-risk profile, and strength of recommendation; and (e) summaries of finding tables. Moreover, the full text of the recommendation included a session labeled "evidence in context" in which panels described the target population for the treatment and the information to be given to patients to facilitate their choices.

DISCUSSION

In setting up this project, we sought to assess whether a participatory mechanism in the production of clinical recommendations would work and whether GRADE was suitable for this purpose, adding scientific rigor to the process.

Our data provide a positive answer to both questions. Producing evidence-based recommendations for the appropriate use of anticancer treatments in everyday practice is a challenge that needs an innovative, scientifically sound, and participatory process. High patient expectations, commercial pressures, clinical and organizational constraints, and availability of resources need to be considered together if one bets on the survival and sustainability of universal health care systems.

Anticancer drugs are a hot topic in the discussion about the adequacy of current standards for approval of new drugs. While the current US Food and Drug Administration and European Medicine Agency (EMEA) legislation requires as a prerequisite that a drug it is found effective in well conducted clinical trials before approval, the reality is that a new drug is often approved only on the basis of its effects on surrogate outcome, with limited follow-up and sometimes using data obtained from phase II rather than phase III studies.[6] Although strong concerns have been raised about the need of more

Table 5. Association Between Panel Member Characteristics and Judgment on Quality of Evidence (low *v* high), Benefit-Risk Balance (positive, uncertain, negative) and Grading of Recommendations (strong positive, weak positive, uncertain, weak negative, strong negative).

	Dependent Variables					
	Quality of Evidence*		Benefit-Risk Balance†		Grading of Recommendation†	
Predictive Variable	Odds Ratio	95% CI	Odds Ratio	95% CI	Odds Ratio	95% CI
Panel						
Breast	1	—	1	—	1	—
Colon-rectum	1.01	0.38 to 2.71	1.03	0.54 to 1.93	1.37	0.72 to 2.61
Lung	0.58	0.21 to 1.66	0.20	0.09 to 0.42	0.58	0.29 to 1.15
Specialty						
Nononcologist	1	—	1	—	1	—
Oncologist	1.87	0.71 to 4.95	1.05	0.52 to 2.10	0.96	0.50 to 1.82
Age, years						
< 55	1	—	1	—	1	—
≥ 55	1.12	0.46 to 2.74	1.17	0.59 to 2.30	1.29	0.74 to 2.25
Sex						
Male	1	—	1	—	1	—
Female	1.01	0.34 to 3.02	1.57	0.85 to 2.90	0.72	0.37 to 1.41
Quality of evidence						
Low	NA		1	—	1	—
High			4.89	2.77 to 8.64	3.52	1.78 to 6.98
Benefit-risk balance						
Positive	NA		NA		85.74	9.95 to 738.81
Uncertain					1	—
Negative					0.01	0.00 to 0.06

NOTE. Odds ratios and 95% CIs estimated with binary or ordered logit models including all the variables listed cluster = individual panelist).
*Binary logit regression model.
†Ordered logit regression model.

coherent standards from regulatory bodies,[5-7] health care systems should be equipped to deal with the introduction of yet experimental interventions. This could be done by creating and supporting a framework so that the information that is missing is produced through pragmatic studies while managing the introduction in clinical practice through guidelines. This is the approach taken by Regione Emilia-Romagna[17], supporting clinical trials,[18] and other HTA activities[19-21] in oncology. Almost simultaneously a highly innovative funding scheme called National Research Program for Independent Research has been implemented by the Italian Drug Agency (AIFA).[22] Two large randomized controlled trials emerged from questions identified by PRI-ER's panels have been approved and funded within the 2006 call.

Our experience suggests that GRADE is feasible and facilitates a multidisciplinary interaction: first, because it fostered a team atmosphere among health professionals of the regional oncology network; and second, because it allowed reconciliation of the traditional separation between clinicians', methodologists', and administrators' points of view as well as allowing for patients to play a more active role.[23]

That said, the variation(s) found in the way panel members appraised the quality of evidence (Fig 1) could be seen as a drawback of the system and a fundamental limitation to its viability. On the contrary, we believe that the in-depth assessment of the evidence that underlies an intervention is one of GRADE's distinctive features. To avoid losing this richness, the recommendations should therefore not be presented as a Yes/No conclusion, but the results of the assessment process should be presented transparently in all its determinants. This led us to choose the template for the presentation of our recommendations described in the Results section.

Given its in-depth assessment, GRADE seems likely to produce "more conservative and justified" recommendations. Noticeably, at the same time as ours, other guidelines organizations (National Institute of Clinical Excellence [NICE] and Cancer Care Ontario) issued recommendations for the use of trastuzumab and aromatase inhibitors for breast and oxaliplatin for colon cancer. Compared with those produced by others,[24-28] our recommendations offer an explicit rationale and justifications as well as a full account of the amount of existing uncertainty and disagreement among panel members.

Less clear is how to make sense of the different interpretations on the benefit-risk profile as a function of the judgment of the quality of evidence and determinant of the strength and direction of the recommendations. Although it would be hard to expect that the judgment of the benefit-risk profile would be homogeneous among panel members, our results suggest that there is coherence between the different steps required by GRADE (Table 1).

In general, differences in ratings within and among panels are not surprising because the type of available literature and the results of the studies were different. In lung cancer, the panel reviewed a topic (adjuvant radiotherapy in completely resected stage I and II NSCLC) with studies showing, with good confidence, a detrimental effect of treatment; this lead to a strong negative recommendation. In breast cancer, we dealt with clinical questions fraught with uncertainty, and this may explain the internal spread of judgments (Fig 1) as well as the strength of the recommendations. Moreover, it must be borne in mind that we purposely chose controversial questions resulting either from conflicting results among primary studies or from lack of relevant evidence, with recent drugs still under confirmatory investigations.

The influence of panel composition (by age, sex, and specialty) did not predict the strength of recommendations, even though we may not have had sufficient statistical power to identify important differences. On the other hand, the quality of evidence and, even more, the balance between benefit and risk, are the only predicting factors of the final strength and direction of recommendations. This is reassuring because it suggests that no major bias occurred as a result of the composition of panels or of potential conflicts of interest of panelists, phenomena that have both been already documented in previous research.[29-34]

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

The author(s) indicated no potential conflicts of interest.

AUTHOR CONTRIBUTIONS

Conception and design: Rossana De Palma, Alessandro Liberati, Giuseppe Longo, Nicola Magrini, Maurizio Marangolo, Fausto Roila
Collection and assembly of data: Rossana De Palma, Alessandro Liberati, Giovannino Ciccone, Elena Bandieri, Maurizio Belfiglio, Manuela Ceccarelli, Maurizio Leoni
Data analysis and interpretation: Rossana De Palma, Alessandro Liberati, Giovannino Ciccone, Maurizio Belfiglio, Maurizio Leoni, Giuseppe Longo
Manuscript writing: Rossana De Palma, Alessandro Liberati, Giovannino Ciccone
Final approval of manuscript: Rossana De Palma, Alessandro Liberati, Giovannino Ciccone, Elena Bandieri, Maurizio Belfiglio, Manuela Ceccarelli, Maurizio Leoni, Giuseppe Longo, Nicola Magrini, Maurizio Marangolo, Fausto Roila

REFERENCES

1. Gabbay J, Walley T: Introducing new health interventions. BMJ 332:64-65, 2006
2. Stevens A, Milne R: Health technology assessment in England and Wales. Int J Technol Assess Health Care 20:11-24, 2004
3. McDonnell A, Wilson R: Evaluating and implementing new services. BMJ 332:109-112, 2006
4. De Palma R, Liberati A: Il Programma Ricerca e Innovazione della Regione Emiliaqqhyphen Romagna (PRI ER): Un approccio nello sviluppo della ricerca nel e per il servizio sanitario. Techne 10:171-174, 2006
5. Roberts TG, Chabner BA: Beyond fast track for drug approvals. N Engl J Med 351:501-505, 2004
6. Wood AJJ: A proposal for radical changes in the drug-approval process. N Engl J Med 355:618-623, 2006
7. Avorn J: FDA standards: Good enough for government work? N Engl J Med 353:969-972, 2005
8. Garattini S, Bertelè V: Efficacy, safety and cost of new anticancer drugs. BMJ 325:269-271, 2002
9. Cancer Care Ontario Program: Chemotherapy Regimens: Breast Adjuvant. http://www.cancercare.on.ca/index_chemoRegimensbyDisease.htm
10. Browman GP, Levine MN, Mohide EA, et al: The practice guidelines development cycle: A conceptual tool for practice guidelines development and implementation. J Clin Oncol 13:502-512, 1995
11. Pater JL, Browman GP, Brouwers MC, et al: Funding new cancer drugs in Ontario: Closing the loop in the practice guidelines development cycle. J Clin Oncol 19:3392-3396, 2001
12. Atkins D, Best D, Briss PA, et al: Grading quality of evidence and strength of recommendations: A new method by the GRADE Working Group. BMJ 328:1490, 2004
13. Schünemann HJ, Jaeschke J, Cook D, et al: An official ATS statement: Grading the quality of evidence and strength of recommendations in ATS guidelines and recommendations. Am J Respir Crit Care Med 174:605-614, 2006
14. Schünemann HJ, Hill S, Kakad M, et al: Transparent development of the WHO rapid advice guidelines. PLoS Medicine 4:e119, 2007
15. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Area di programma Ricerca e Innovazione: I progetti. http://asr.regione.emilia-romagna.it/wcm/asr/eventi/20050223_ws2_prier.htm
16. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Raccomandazioni sulla terapia adiuvante del tumore della mammella, del colon retto e del polmone. http://asr.regione.emilia-romagna.it/wcm/asr/ric_inn/prier/gr_v/pr_oncologia/stpr_farmaci_onco/raccomandazione.htm
17. Grilli R, Taroni F: Managing the introduction of expensive medical procedures: Use of a registry. J Health Serv Res Policy 11:89-93, 2006
18. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Pubblicazioni: L'innovazione in radioterapia: trattamenti più brevi ma ugualmente efficaci per il tumore del seno? http://asr.regione.emilia-romagna.it/wcm/asr/collana_pillole/pill025/link/pill25.pdf
19. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Pubblicazioni: Impatto diagnostico della Tomografia Computerizzata multistrato nella malattia coronaria. http://asr.regione.emilia-romagna.it/wcm/asr/collana_pillole/pill027/link/pill27.pdf
20. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Pubblicazioni: Le indicazioni all'uso appropriato della PET in oncologia:uno strumento per il governo clinico. http://asr.regione.emilia-romagna.it/wcm/asr/collana_pillole/pill029/link/pill29.pdf
21. Programma Ricerca e Innovazione Emilia Romagna (PRI-ER): Pubblicazioni: Cosa devono fare i pazienti dopo le cure iniziali? Un progetto per l'appropriatezza del Follow-up in oncologia. http://asr.regione.emilia-romagna.it/wcm/asr/collana_pillole/pill026/link/pill26.pdf
22. Italian Drug Agency (AIFA): http://www.agenziafarmaco.it/section8983.html
23. Barrett A, Roques T, Small M, et al: How much will Herceptin really cost? BMJ 333:1118-1120, 2006
24. National Institute for Health and Clinical Excellence: Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer. Tech Appraisal Guidance 107:1-16, 2006
25. Trudeau M, Madarnas Y, McCready D, et al: The role of trastuzumab in adjuvant and neoadjuvant therapy in women with HER2/*neu*-overexpressing breast cancer: A clinical practice guideline. Cancer Care Ontario Program 1:1-5, 2007
26. Jonker D, Rumble RB, Maroun J, et al: The role of oxaliplatin combined with 5-fluorouracil and folinic acid in the first and second-line treatment of advanced colorectal cancer: A clinical practice guideline. Cancer Care Ontario Program 1:1-4, 2005
27. National Institute for Health and Clinical Excellence: Hormonal therapies for the adjuvant treatment of early oestrogen-receptor-positive breast cancer. Tech Appraisal Guidance 112:1-37, 2006
28. Herrin J, Etchason JA, Kahan JP, et al: Effect of panel composition on physician ratings appropriateness of abdominal aortic aneurysm surgery: Elucidating differences between multispecialty panel results and specialty society recommendations. Health Policy 42:67-81, 1997
29. Kravitz RL, Park RE, Kahan JP: Measuring the clinical consistency of panelists' appropriateness ratings: The case of coronary artery bypass surgery. Health Policy 42:135-143, 1997
30. Liberati A, Apolone G, Nicolucci A, et al: The role of attitudes, beliefs and personal characteristics of Italian physicians in the surgical treatment of early breast cancer. Am J Public Health 81:38-42, 1991
31. Fitch K, Lazaro P, Aguilar MD, et al: Physicians recommendation for coronary revascularisation: Variations by clinical specialty. Eur J Publ Health 9:181-187, 1999
32. Ayanian JZ, Laundrum MB, Normand SL, et al: Rating the appropriateness of coronary angiography: Do practicing physicians agree with an export panel and with each other? N Engl J Med 338:1896-1904, 1998
33. Choudhry NK, Stelfox HT, Detsky S: Relationship between authors of clinical practice guidelines and industry. JAMA 287:612-617, 2002
34. Frye CB: Disclosing conflicts of interests involving clinicians who prepare therapeutic guidelines. Am J Health Syst Pharm 62:361-362, 2005

Appendix

The Appendix is included in the full-text version of this article, available online at www.jco.org. It is not included in the PDF version (via Adobe® Reader®).