Table 1 The parameters and costs of each stage

(a) Parameters			
Transition rate		Source	
1-year trastuzumab group			
'Without recurrence' to 'metastatic recurrence'	0.004483	7	
'Without recurrence' to 'local recurrence'	0.001296	7	
Observation group			
'Without recurrence' to 'metastatic recurrence'	0.006916	7	
'Without recurrence' to 'local recurrence'	0.001737	7	
Both groups (after recurrence)			
'MetastasisA1' to 'metastasisA2' (aromatase inhibitor)	0.07109	13	
'MetastasisB1' to 'metastasisB2' (LH-RH agonist plus tamoxifen)	0.07413	14	
'MetastasisA2' to 'metastasisA3' (second-line hormone therapy)	0.1091	Experts' opinion	
'MetastasisB2' to 'metastasisC1' (second-line hormone therapy)	0.1091	Experts' opinion	
'MetastasisA3' to 'metastasisC1' (third-line hormone therapy)	0.1591	Experts' opinion	
'MetastasisC1' to 'metastasisC2' (trastuzumab plus taxane)	0.09558	3	
'MetastasisC2' to 'metastasisC3' (trastuzumab plus vinorelbine)	0.16674	15	
'MetastasisC3' to 'metastasisC4' (trastuzumab plus capecitabine)	0.13191	16	
'MetastasisC4' to 'death' (palliative)	0.1091	Experts' opinion	
'Local recurrence1–3' to 'metastasis C1'	0.008478	17	
Background of patients			
Average age of patients	50	7	
Hormone-receptor-positive	0.5	7	
Premenopausal	0.85	7	
Mastectomy of primary tumor	0.5	7	
Cardiotoxicity			
Severe congestive heart failure	0.006	7	
Symptomatic congestive heart failure	0.02	7	
Asymptomatic congestive heart failure	0.03	7	
Major assumption			
The efficacy period of trastuzumab (base-case, years)	5		
Patients weight (kg)	50–60		
Patient age	50		
Risk ratio of recurrence during the next 5 years compared with first 5 years	0.5		
Discount rate	0.03	11	
(b) Costs			
	Treatment	Cost (JPY)	Unit
Adjuvant trastuzumab			
Weight = 50–60 kg (first cycle)	Trastuzumab (L.D: 8 mg/kg)	300,000	per month
Weight = 50–60 kg (after second month)	Trastuzumab (6 mg/kg/3 weeks)	280,000	per month
Weight < 50 kg (first cycle)	Trastuzumab (L.D: 8 mg/kg)	250,000	per month
Weight < 50 kg (after second month)	Trastuzumab (6 mg/kg/3 weeks)	240,000	per month
Weight = 60–75 kg (first cycle)	Trastuzumab (L.D: 8 mg/kg)	350,000	per month
Weight = 60–75 kg (after second month)	Trastuzumab (6 mg/kg/3 weeks)	350,000	per month
Without recurrence (until 5 years)			
ER positive and premenopausal patients (until 3 years)	Leuprorelin (3.75 mg/4 weeks), tamoxifen (20 mg/day)	70,000	per month
ER positive and premenopausal patients (after 3 years)	Tamoxifen (20 mg/day)	16,000	per month
ER positive and postmenopausal patients	Anastrozole (1 mg/day)	22,000	per month
ER negative patients	Follow-up[b], annual mammography,	700	per month

Table 1 continued

Without recurrence (after 5 years)			
All patients	Annual follow-up and mammography	400	per month
Local recurrence			
Surgery	Mastectomy or resection	800,000	per event
Local recurrence1 (until 5 years)	Exemestane (25 mg/day)	51,000	per month
Local recurrence2 (until 5 years)	Leuprorelin (3.75 mg/4 weeks), tamoxifen (20 mg/day)	70,000	per month
Local recurrence3 (until 5 years)	Only follow-up	700	per month
All patients (after 5 years)	Annual follow-up and mammography	400	per month
Metastatic recurrence			
MetastasisA1	Exemestane (25 mg/day)	51,000	per month
MetastasisA2	Tamoxifen (20 mg/day)	45,000	per month
MetastasisA3	Medroxyprogesterone (800 mg/day)	77,000	per month
MetastasisB1	Leuprorelin (3.75 mg/4 weeks), tamoxifen (20 mg/day)	70,000	per month
MetastasisB2	Medroxyprogesterone (800 mg/day)	77,000	per month
MetastasisC1 (weight = 50–60 kg, first month)	Trastuzumab (L.D: 4 mg/kg, 2 mg/kg/week), paclitaxel (80 mg/ m^2)	600,000	per month
MetastasisC1 (after second month)	Trastuzumab (2 mg/kg/week), paclitaxel (80 mg/ m^2)	480,000	per month
MetastasisC2 (weight = 50–60 kg)	Trastuzumab (2 mg/kg/week), vinorelbine (25 mg/m^2)	370,000	per month
MetastasisC3 (weight = 50–60 kg)	Trastuzumab (2 mg/kg/week), capecitabine (1200 mg/day)	340,000	per month
MetastasisC4	Palliative care	1,100,000	per event
Bone metastasis	Pamidronate (90 mg/4 weeks)	70,000	per month
Adverse event (cardiotoxicity)			
Severe congestive heart failure		810,000	per event
Symptomatic congestive heart failure		170,000	per event
Asymptomatic congestive heart failure		40,000	per event

[a] We assumed coefficient of variation was 0.4 that was arbitrary value but chosen to give a quite large standard deviation based on A. Briggs. [20]

[b] (1) 4-month intervals for 2 years, (2) 6-month intervals for 2–5 years,(3) annually after 5 years based on the follow up guideline by ASCO [21]

metastatic patients is unknown, so we calculated the sum of each treatment cost in constructing this Markov model and estimated only direct medical costs based on the per piece Japanese drug tariff and reimbursement schedule, (Table 1 (b)). This included the cost of supportive care (anti-emetic agents [22], etc.), heart monitoring, routine follow-up [21], diagnostic imaging, blood tests, and so on. In the probabilistic sensitivity analysis, all cost data were modeled as normal distributions with the base-case value as the mean.

Results

Cost effectiveness result

The cost-effectiveness analysis compared the 1-year trastuzumab group with the observation group (Table 2). The ICER of the standard setting (5 years efficacy and 50–60 kg) was JPY 2,740,000 (€18,000) per LYG. The calculation results of other scenario ICER were JPY 1,920,000 (€13,000) and JPY 1,080,000 (€7,200) per LYG for the period of trastuzumab efficacy, respectively, of 10 years (a somewhat optimistic scenario), and throughout the life time (optimistic scenario).

The ICER becomes higher, however, with increased patient weight. At the 60–75 kg weight class, which is heavier than the standard class, the ICER changed from JPY 2,300,000 (€15,000) to JPY 7,400,000 (€49,000) for differing trastuzumab efficacy periods (5 to 50 years) and at 50–60 kg, which is a lighter class, ranged from JPY 1,500,000 (€10,000) to JPY 5,100,000 (€34,000).

Sensitivity analysis

The most influential parameter was the period of trastuzumab efficacy. The results are shown in Fig. 2 for every weight class; and when trastuzumab efficacy continues for more than 2 years, the ICER was less than JPY 7,500,000 (€50,000) for any weight class. Other one-way sensitivity analyses for parameters (discount rate, recurrence rate, cardiotoxicity costs, and terminal costs), showed little change. Among them, however, the discount rate was the most influential parameter, and all the results of sensitivity

Table 2 The result of cost-effectiveness analysis and sensitivity analysis

(a) The result of cost-effectivess analysis (weight = 60 kg)

	C	E	ΔC/ΔE (ICER)	
Observation	¥7,900,000	12.46		
Trastuzumab (conservative)	¥11,500,000	13.06	¥6,000,000	€40,000
Trastuzumab (standard)	¥11,200,000	13.70	¥2,600,000	€17,000
Trastuzumab (optimistic)	¥10,900,000	14.10	¥1,800,000	€12,000

(b) The relationship between weight class and efficacy period of trastuzumab

		Weight class		
Efficacy period of trastuzumab		less than 50 kg	50–60 kg	60–75 kg
	2 years	¥5,100,000	¥6,000,000	¥7,400,000
	(Conservative)	€34,000	€40,000	€49,000
	5 years	¥2,200,000	¥2,600,000	¥3,300,000
	(Standard)	€15,000	€17,000	€22,000
	10 years	¥1,500,000	¥1,800,000	¥2,300,000
	(Optimistic)	€10,000	€12,000	€15,000

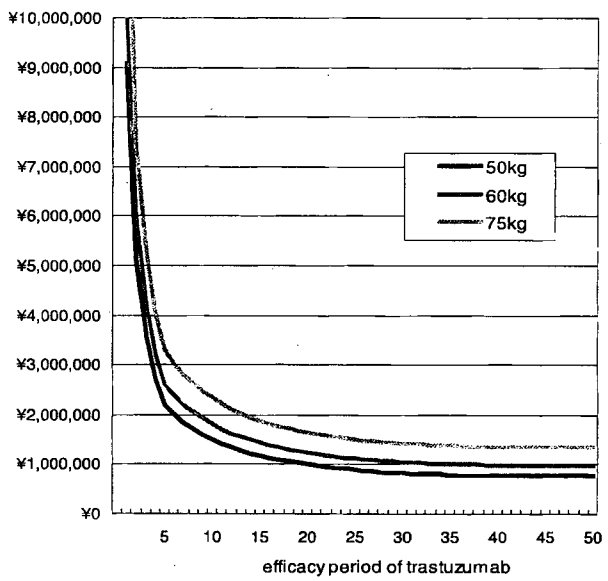

Fig. 2 Uncertainty of cost-effectiveness analysis. The relation between ICER and efficacy period of trastuzumab by weight class

analyses in the standard scenario were less than JPY 5,000,000 (€33,000).

The ICER was changed to ¥4,700,000 (€31,000), ¥1,900,000 (€14,000), and ¥1,300,000 (€9,000) per LYG for the period of trastuzumab efficacy, respectively, of 2, 5, and the10 years, based on the 1-year follow-up data [6].

The acceptability curve of the standard setting is shown in Fig. 3(a), and the probability that the ICER of 1-year trastuzumab was less than JPY 5,400,000 (€36,000) was above 95%. Figure 3(b) shows the incremental cost-effectiveness plane and the 5 percentile and 95 percentile for incremental LYG were estimated to be 0.77 and 1.65. The 5 percentile and 95 percentile for incremental costs were JPY 900,000 (€15,000) and JPY 5,550,000 (€28,000).

Discussion

In Japan, as with other developed countries, the serious social problem of burgeoning medical costs, caused by rapid aging and the evolution of healthcare technology prompts us to consider the efficiency of new expensive healthcare technology.

For metastatic patients, the trastuzumab ICER was estimated as £19,000 (monotherapy) and £37,500 (combination) by NICE, which recommended both trastuzumab monotherapy and combination therapy based on this economic evaluation [23]. Furthermore, NICE issued the guidance recommending trastuzumab as a treatment option based on the 1-year follow-up data of the HERA trial [10]. They estimated that the ICER of 1-year trastuzumab was £18,000 per additional QALY ranging from £16,000 to £33,000.

NICE suggests that the ICER threshold should be £20,000 to £30,000 (=JPY 5,000,000–JPY 7,000,000) per QALY. In the US, $50,000 or $100,000 per QALY is often used as the threshold. In Japan, no cost effectiveness threshold for treatment has been determined; however, it is thought that the values of the NICE's thresholds are acceptable, because of the similar economic and medical environments in UK and Japan.

In US trials, weekly adjuvant trastuzumab therapy had an ICER of US$ 18,970 per QALY [24] and US$ 39,982 per QALY [25] estimating from the joint analysis (NSABP B-31 and NCCTG N9831 trials) [8]. Both papers also conclude that adjuvant trastuzumab is cost-effective. It confirms that

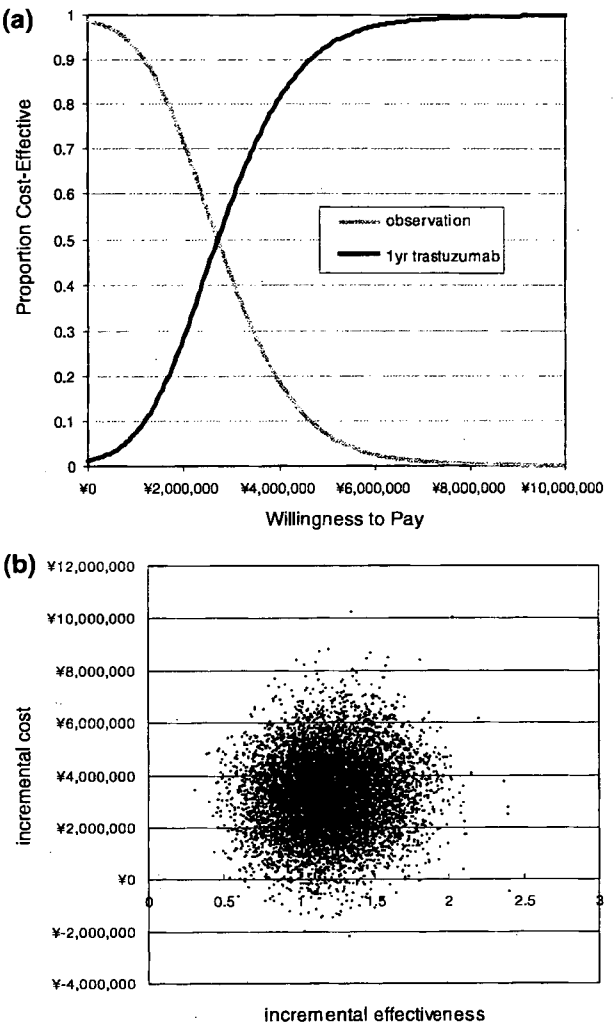

Fig. 3 (**a**) Acceptability curve for 1-year trastuzumab versus observation in a standard setting (5 years of efficacy and 50–60 kg) simulated 10,000 times (**b**) Incremental cost-effectiveness plane for trastuzumab versus observation alone

the conclusion is consistent between the HERA trial and the joint analysis-based economic evaluation.

We calculated the ICER based on the 2-year follow-up data of the HERA trial in this analysis, although other cost-effectiveness analyses of trastuzumab were based on 1-year follow-up data. We had difficulty showing the exact ICER value and predicting the long-term prognosis for the trastuzumab patients' group, which greatly influences the result of this economic evaluation. The results of other large trials, B31/N9831, show that the hazard ratio in the third or fourth year is nearly equal to that in the first year [8]. It is natural to assume that the efficacy of the HERA regimen continues for at least an equivalent period. Thus, we think the base case is 5 years. However, when trastuzumab efficacy continues at least for more than 2 years, which is a conservative setting, the ICER is less than JPY 7,500,000 (€50,000) for any weight class. There is little difference between LYG and QALY in oncology [26], so we could conclude that the 1-year trastuzumab treatment is cost-effective from this analysis, even based on 2-year follow-up data.

The limitation of this result is that our endpoint is LYG not QALY, because in Japan there are no HRQoL data for breast cancer patients applied to our analysis, and we decided it was better to use LYG than the QALY calculated by foreign, not Japanese utility values. In addition, the transition rates derived from published data and costs were calculated by the construction of a standard therapy model, not analyzed by using patient-level data.

In this analysis we adopted the health-care payers' perspective. When the perspective was changed to a societal one, the indirect costs of the trastuzumab group were higher than those of the control group in the first yearly period of trastuzumab therapy. But considering that the expected value of incremental effectiveness is more than 1 year, the indirect costs of the trastuzumab group were lower, as a whole.

It is important to analyze not only the cost-effectiveness but also the budget impact. We considered the incremental cost of 1-year trastuzumab treatment as JPY 2,000,000–4,000,000 (€13,000–26,000) from Table 2. Then, by estimating the number of new breast cancer patients per year to be 40,000, of which 20% are HER2-positive, the total incremental cost was JPY 16–32 billion (€105–210 million), if all the HER2-positive patients were treated by trastuzumab.

In the Finland Herceptin (FinHer) trial [27], 9 weeks of trastuzumab injections yielded a hazard ratio equivalent to that of other studies in which trastuzumab was administered for 1 year. We cannot conclude that 9 weeks of injections is optimal from the FinHer trial, because it included fewer patients ($N = 232$). But this RCT showed a noteworthy result. If the efficacy of 9 weeks of injections is nearly equal to that of 1-year injections, 9-week treatment would be more cost-effective as well as provide greater patient convenience. However, the optimal period of adjuvant trastuzumab treatment has not yet been decided. At present, in the HERA trial, the comparison of 1-year and 2-year treatments has not been demonstrated. When they are available, the cost-effectiveness analyses will play an important role in determining the optimized treatment period.

The price of trastuzumab (¥78,074 (€520) per 150 mg vial) in Japan is lower than that in the UK (£407.40 (€650) per 150 mg vial), which is less than that in other developed countries. At the same time, the Japanese women's average weight is less than that of westerners, and they have the highest life expectancy rates in the world. Though NICE

estimated the average cost per person as £24,600 (JPY 5,900,000) in the UK [10], our calculated cost is JPY 3,390,000 (£14,000) in the standard setting. The Japanese women have an advantage over westerners in cost-effectiveness of adjuvant trastuzumab administration.

Based on the results of some RCTs and our economic analysis, we can conclude that the 1-year trastuzumab adjuvant treatment is superior to observation only in terms of the cost-effectiveness.

Acknowledgement This study was performed as the Comprehensive Support Project for Health Outcomes Research project (CSP-HOR) established by the Public Health Research Foundation (PHRF). We thank the PHRF for the grant supporting our study.

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. Lancet 365:1687–1717
2. Slamon DJ, Clark GM, Wong SG et al (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 235:177–182
3. Slamon DJ, Leyland-Jones B, Shak S et al (2001) Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 344:783–792
4. Marty M, Cognetti F, Maraninchi D et al (2005) Randomized phase II trial of the efficacy and safety of trastuzumab combined with docetaxel in patients with human epidermal growth factor receptor 2-positive metastatic breast cancer administered as first-line treatment: the M77001 study group. J Clin Oncol 23:4265–4274
5. Osoba D, Slamon DJ, Burchmore M et al (2002) Effects on quality of life of combined trastuzumab and chemotherapy in women with metastatic breast cancer. J Clin Oncol 20:3106–3113
6. Piccart-Gebhart MJ, Procter M, Leyland-Jones B et al (2005) Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. N Engl J Med 353:1659–1672
7. Smith I, Procter M, Gelber RD et al (2007) 2-year follow-up of trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer: a randomised controlled trial. Lancet 369:29–36
8. Romond EH, Perez EA, Bryant J et al (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Engl J Med 353:1673–1684
9. Slamon D, Eiermann W, Robert N et al (2006) BCIRG 006: 2nd interim analysis phase III randomized trial comparing doxorubicin and cyclophosphamide followed by docetaxel (AC → T) with doxorubicin and cyclophosphamide followed by docetaxel and trastuzumab (AC → TH) with docetaxel, carboplatin and trastuzumab (TCH) in Her2 neu positive early breast cancer patients. Breast Cancer Res Treat 100(Sup.1):abstr 52
10. National Institute for Health and Clinical Excellence (2006) Trastuzumab for the adjuvant treatment of early-stage HER2-positive breast cancer.
11. Weinstein MC, Siegel JE, Gold MR et al (1996) Recommendations of the panel on cost-effectiveness in health and medicine. JAMA 276:1253–1258
12. Ewer MS, Vooletich MT, Durand JB et al (2005) Reversibility of trastuzumab-related cardiotoxicity: new insights based on clinical course and response to medical treatment. J Clin Oncol 23:7820–7826
13. Suter TM, Cook-Bruns N, Barton C (2004) Cardiotoxicity associated with trastuzumab (Herceptin) therapy in the treatment of metastatic breast cancer. Breast 13:173–183
14. Hortobagyi GN (1998) Treatment of breast cancer. N Engl J Med 339:974–984
15. Mouridsen H, Gershanovich M, Sun Y et al (2001) Superior efficacy of Letrozole versus Tamoxifen as first-line therapy for postmenopausal women with advanced breast cancer: results of a phase III study of the international Letrozole breast cancer group. J Clin Oncol 19:2596–2606
16. Klijn JG, Blamey RW, Boccardo F et al (2001) Combined tamoxifen and luteinizing hormone-releasing hormone (LHRH) agonist versus LHRH agonist alone in premenopausal advanced breast cancer: a meta-analysis of four randomized trials. J Clin Oncol 19:343–353
17. Toi M, Saeki T, Aogi K et al (2005) Late phase II clinical study of vinorelbine monotherapy in advanced or recurrent breast cancer previously treated with anthracyclines and taxanes. Jpn J Clin Oncol 35:310–315
18. Fumoleau P, Largillier R, Clippe C et al (2004) Multicentre, phase II study evaluating capecitabine monotherapy in patients with anthracycline- and taxane-pretreated metastatic breast cancer. Eur J Cancer 40:536–542
19. Le MG, Arriagada R, Spielmann M et al (2002) Prognostic factors for death after an isolated local recurrence in patients with early-stage breast cancer. Cancer 94:2813–2820
20. Briggs A (2001) Handling uncertainty in economic evaluation and presenting the results. In: Drummond M, McGuire A (eds) Economic evaluation in health care: merging theory with practice. Oxford university press, Oxford
21. Smith TJ, Davidson NE, Schapira DV et al (1999) American Society of Clinical Oncology 1998 update of recommended breast cancer surveillance guidelines. J Clin Oncol 17:1080–1082
22. Kris MG, Hesketh PJ, Somerfield MR et al (2006) American Society of Clinical Oncology guideline for antiemetics in oncology: update 2006. J Clin Oncol 24:2932–2947
23. National Institute for Health and Clinical Excellence (2002) Guidance on the use of trastuzumab for the treatment of advanced breast cancer
24. Liberato NL, Marchetti M, Barosi G (2007) Cost effectiveness of adjuvant trastuzumab in human epidermal growth factor receptor 2—positive breast cancer. J Clin Oncol 25:625–633
25. Kurian AW, Thompson RN, Gaw AF et al (2007) A cost-effectiveness analysis of adjuvant trastuzumab regimens in early HER2/neu—positive breast cancer. J Clin Oncol 25:634–641
26. Tammy TO (2004) Cost-effectiveness versus cost-utility analysis of interventions for cancer: does adjusting for health-related quality of life really matter? Value in Health 7:70–78
27. Joensuu H, Kellokumpu-Lehtinen PL, Bono P et al (2006) Adjuvant docetaxel or vinorelbine with or without trastuzumab for breast cancer. N Engl J Med 354:809–820

Breast Cancer
Vol. 14 No. 1 January 2007

Review Article

Recent Topics of Health Outcomes Research in Oncology

Kojiro Shimozuma[*1,2], Hirohisa Imai[*3], Katsumasa Kuroi[*4], Shozo Ohsumi[*5], and Michikazu Ono[*6]

*[1]Department of Healthcare and Social Services, University of Marketing and Distribution Sciences, *[2]Institute for Stress Science, Public Health Research Foundation, *[3]Department of Epidemiology, National Institute of Public Health, *[4]Division of Surgery/Breast Oncology, Nyuwakai Oikawa Hospital, *[5]Department of Breast Oncology, National Hospital Organization Shikoku Cancer Center, *[6]Department of Health Science and Social Welfare, School of Human Sciences, Waseda University, Japan.

This article reviews recent topics in health outcomes research. First, we discuss the concept and importance of 'subjective' assessment of quality of life (QOL), and introduce new guidance, by the respective medical product regulatory authorities in Europe and the United States, for labeling claims of medical products that are assessed for outcomes related to QOL. Second, we address the application of item response theory (IRT) in developing and assessing QOL measures to compensate for several drawbacks of the classical psychometric approach, which has been commonly used to verify the reliability and validity of QOL instruments. Third, the relevance and determination of the minimally clinically important difference (MID) of QOL scores is discussed. Finally, we address the so-called 'response shift' which may affect the reliability of analysis results of QOL scores in longitudinal studies such as randomized clinical trials.

Breast Cancer 14:60-65, 2007.

Key words: Health outcome, Quality-of-life, Patient-reported outcomes, Item response theory, Computer adaptive testing

Introduction

Breast cancer is a disease that most commonly affects women in their 40s - 60s, a time when they are at their peak and most active. This is not only a great loss to society, but also can have negative effects on the social life of the patient herself. Fortunately, breast cancers differ from other types of solid cancer in that there are a number of effective treatments. However, even with successful treatment there may be damage that the breast cancer patient must live with for a long time; not only to physical functions but also psychosocial damage from a perceived loss of femininity. In breast cancer treatment, therefore, assessment of health outcomes such as quality of life (QOL) or health-related quality of life (HRQOL; a concept that excludes domains of QOL such as social environment and spirituality that are difficult to change through medical treatment or care interventions) is an area that has come to be emphasized strongly worldwide.

Thus, there are an increasing number of cases in recent years in which HRQOL is included as a secondary endpoints both in phase III studies of anti-cancer treatments and in some late phase II studies. In clinical trials of supportive or palliative therapy as well, it is not unusual for health outcomes themselves to be the primary endpoints.

As described above, demand for health outcome assessments such as HRQOL is rising considerably in the field of medical technology assessment, but at the same time even the two pillars of the QOL concept on which consensus has been reached among health outcome researchers, namely, multi-dimensionality (multi-domain concepts) and subjectivity (information sources are patients' subjective feelings), are not necessarily well understood by general clinicians or specialists. Moreover, clinicians know very little about the theoretical background for ensuring the reliability and

Reprint requests to Kojiro Shimozuma, Professor, Department of Healthcare and Social Services, University of Marketing and Distribution Sciences, 3-1, Gakuennishi-machi, Nishi-ku, Kobe, Hyogo 651-2188, Japan.
E-mail: Kojiro_Shimozuma@red.umds.ac.jp

Abbreviations:
QOL, Quality of life; HRQOL, Health-related quality of life; PRO, Patient-reported outcomes; IRT, Item response theory; CAT, Computer adaptive testing; MID, Minimally important difference; EBM, Evidence-based medicine; ES, Effect size; SD, Standard deviation; SEM, Standard error of measurement; RS, Response shift

validity of quantitative assessments of health outcomes, or the issues that need to be resolved in order to establish that background more firmly.

This article reviews recent topics in health outcomes research to make them more understandable to clinicians and others.

Organization of Concepts from "Quality of Life" (QOL) to "Patient-Reported Outcomes" (PRO), and Guidance for Using Health Outcomes in Medical Product Labeling Claims

Discrepancies that cannot be Ignored between Clinician-Reported Outcomes and Patient-Reported Outcomes

As mentioned in the introduction, subjectivity is the basis for the concepts of QOL/HRQOL. While important, we shall refrain here from philosophical or phenomenological discussions of subjectivity, such as questions of the range indicated by human subjectivity or whether subjectivity actually exists. What is important in health outcomes, particularly QOL assessments, is not whether or not they are objective but how accurately and precisely patients' subjective feelings (if they exist) are or can be understood.

In the range of a patient subjective experience there are problems that have already risen into consciousness, but there are also those which remain latent in the subconscious. Normally diagnosis proceeds with a search for causes starting with the symptoms the patient complains of, and treatment and care based on those complaints is the basis of medical attention. In assessments of QOL, however, we would like to also understand latent problems in the subconscious. Essentially, one of the aims of psychological measures and other such instruments is to bring to light latent problems that even the patient herself has not noticed.

Most doctors want to believe that they can gain a fairly good understanding of patient complaints from their interviews and observations, but in fact it is known that doctors overestimate or underestimate patient complaints from the nature of their symptoms and problems. For example, in a large-scale study of prostate cancer patients[1] doctors reportedly underestimated bone pain and fatigue/energy, while they overestimated erectile dysfunction. In addition, in studies in which we examined the frequency and level of peripheral neuropathy from various aspects in phase III studies of breast cancer using taxane chemotherapy, doctors' assessments clearly underestimated both sensory and motor disturbances[2,3].

As stated above, the aim of "wanting to assess the patient's subjectivity" is a major aim of QOL/HRQOL assessments, but that aim and the range of included concepts is difficult to clearly communicate with those words, so the term "patient-reported outcomes" (PRO) started being used about 10 years ago. However, until recently this term had not come to be used by many researchers and medical professionals.

A Sign that the Concept and Term PRO is Widely Recognized: Introduction of Guidance in Cases when Health Outcomes are Used in Medical Product Labeling Claims

From last year through this year, two events have occurred that triggered the more widespread use of the term (concept) PRO. They were the issuing of new guidance, by the respective medical product regulatory authorities in Europe and the United States, for labeling claims of medical products that are assessed for outcomes related to QOL. We shall introduce them here briefly while comparing the differences between the two.

1) "Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (HRQOL) Measures in the Evaluation of Medicinal Products"[4]

The European Medicines Agency (EMEA) published this guidance in 2005. It is a short document of about 5 pages, which consists of 4 main sections: I. Introduction (background), II. HRQOL in drug evaluation process, III. Study design for HRQOL assessment, and IV. Statistical analysis. The document defines the concepts of HRQOL and PRO, and describes the points that should be kept in mind when using HRQOL, in particular, in evaluations of medical products.

It should be noted here that in the Introduction section PRO is defined clearly as "Any outcome evaluated directly by the patient himself and based on the patient's perception of a disease and its treatment(s)." Differences with the concept of HRQOL are described as follows: "The term PRO is proposed as an umbrella term to cover both single dimension and multi-dimension measures of symptoms, HRQOL, health status, adherence to

treatment, satisfaction with treatment, etc." Thus, PRO is positioned above HRQOL.

The document also contains a detailed definition of HRQOL, which is that "HRQOL is considered to represent a specific type/subset of PROs, distinguished by its multi-dimensionality. Indeed, HRQOL is a broad concept which can be defined as the patient's subjective perception of the impact of his disease and its treatment(s) on his daily life, physical, psychological and social functioning and well-being. The definition of HRQOL has as a common basis the definition of health given by the WHO in 1948: 'Health is a state of complete physical, mental, and social well-being and not merely the absence of disease' "

A description of II-IV is omitted here.

2) "Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (Draft Guidance)" [5]

This draft guidance was published by the U.S. Department of Health and Human Services in the Food and Drug Administration (FDA) in February 2006. As of August 2006 it remains open to the public for the purpose of gathering public comment, so it may undergo some changes in the final version.

The purpose of this guidance is described as follows: "This guidance describes how the FDA evaluates patient-reported outcome (PRO) instruments used as effectiveness endpoints in clinical trials. It also describes our current thinking on how sponsors can develop and use study results measured by PRO instruments to support claims in approved product labeling. By explicitly addressing the review issues identified in this guidance, sponsors can increase the efficiency of their endpoint discussions with the FDA during the product development process, streamline the FDA's review of PRO endpoint adequacy, and provide optimal information about the patient's perspective of treatment benefit at the time of product approval."

This guidance consists of seven sections: I. Introduction, II. Background, III. PRO - Regulatory perspective, IV. Evaluating PRO instruments, V. Study design, VI. Data analysis, and VII. Glossary. It is a detailed work of 36 pages.

The content differs somewhat from the EMEA document in that the focus is on PRO evaluation guidance rather than HRQOL evaluation guidance.

As in the EMEA document, the Introduction gives a clear definition of PRO: "A PRO is a measurement of any aspect of a patient's health status that comes directly from the patient (i.e., without the interpretation of the patient's responses by a physician or anyone else)." Furthermore, "In clinical trials, a PRO instrument can be used to measure the impact of an intervention on one or more aspects of patients' health status, hereafter referred to as PRO concepts, ranging from the purely symptomatic (response of a headache) to more complex concepts (e.g., ability to carry out activities of daily living), to extremely complex concepts such as quality of life, which is widely understood to be a multidomain concept with physical, psychological, and social components. Data generated by a PRO instrument can provide evidence of a treatment benefit from the patient perspective." In this we see that while QOL is a concept included in PRO, a characteristic of this guidance is that it shows clearly that a broad concept such as QOL can also be used as an end point in clinical trials for medical product labeling claims (however, high levels of reliability and validity are required for the evaluation).

Moreover, in section III. PRO Regulatory perspective, the document says that there are parts of the treatment effect that only the patient can know, indicating the importance of patients' subjective assessment.

As mentioned above, the importance of patients' subjective assessment is one of the most difficult things for clinicians to understand, and so it is worth devoting an entire section to this point.

After the guidance is amended and completed based on public comment, it will serve to bring more widespread awareness of the concept of PRO.

Application of Item Response Theory (IRT) in Developing and Assessing QOL/PRO Measures

The theoretical basis for the reliability and validity of data when developing measures to quantify QOL or PRO, or conducting health outcome evaluations using these measures, comes from psychometry, which has been developed as a field of psychology.

This theory is also called the classical test model, and has been accepted by many resear-

chers. However, it has several drawbacks.

For example, there are inefficiencies in measures developed using this theory and evaluations using these measures. When using these measures in actual clinical trials, it is rare that the study population has exactly the same attributes as the target group when the measure was developed. Therefore, if the same subjects do not have a previously evaluated history, it is necessary to verify whether or not the same psychometric characteristics as at the time of development were maintained in the study, before the results of the analysis of the obtained data can be interpreted.

This problem also occurs when a translated version of a measure is developed in a language other than the original language. Today, in the development of new medical products that are predicted to have a wide market, it is not unusual for multinational clinical trials to be conducted in Japan, the USA, and Europe. If the psychometric characteristics differ with each translation, the QOL/PRO data obtained cannot be used together in analysis.

Moreover, while it is often the case that several measures exist for the same purpose, it is not rare for there to be slight differences in the scale of the measure or the conceptual structure being measured and evaluated. In these cases, there is a problem in that it is difficult to compare data that have been evaluated with different measures.

One method that has started to be considered in order to overcome problems such as those above is to adopt item response theory (IRT; also called new test theory) to health outcome assessments. IRT has been used mainly in the field of education in preparing test problems for TOEFL or shared tests for entrance to overseas universities.

IRT is fundamentally a multiple logistic model. When only the single parameter of "difficulty" is used it is called the "Rasch model," and special software is commercially available in Japan. Additional methods are a two parameter model made by adding the parameter of "discrimination" to that of "difficulty," and a model with the addition of a third parameter of "guess." The most commonly used is a two parameter logistic model.

Although IRT has been applied in the development of short versions of QOL/PRO measures and in cross-cultural validation, it is still not used by many researchers. Therefore, with the wide application of IRT to item analysis and scale scoring of health outcome assessments, we can look forward to increasing its application in new evaluation methods called computer adaptive testing (CAT), similar to the field of education.

CAT is a method of administering QOL/PRO measures by computer using the psychometric framework of IRT. Items are automatically selected by computer from some item banks on the basis of the patient's responses to previously administered items. This process uses an algorithm to estimate a person's score and the score's reliability and then chooses the best next item, enabling scale administration based on specifications such as content coverage, test length, and standard error. The capacity to rank all patients on the same continuum, even if they have not been given any common items, allows for an assessment that is individually tailored to each person. With item banking, each patient need only answer a subset of items to obtain a measure that accurately estimates what would have been obtained by administering the entire set of items[6].

This promising method has the benefit for researchers and clinicians of reducing the labor and cost to develop several measures for each subject group and goal, while for patients who are evaluation subjects it has the benefit of letting them know where they rank in the entire subject group by answering a minimum number of short questions.

However, several premises and constraints can also be predicted when IRT is applied to CAT. For example, IRT relies on strong assumptions, that is, unidimensionality, local independence, and monotonicity. Things that do not fit these assumptions are also included in the data and concepts of QOL/PRO.

In the USA, a large project called Patient-Reported Outcomes Measurement System (PROMIS) was started in September 2004 with the support of the National Institutes of Health (NIH) and the simultaneous contributions of many researchers[6]. This project was begun to solve many of these problems at once, and there are high expectations for the results.

Minimally Importance Difference (MID)

In clinical studies of health outcome assessments using these measures, the many data that are obtained, whether in a cross-sectional study or a clinical trial, are normally analyzed statistically

and applied to the clinical setting based on the results of statistical significance tests. Of course, this is the fundamental method of evidence-based medicine (EBM), which is normally used in analysis of survival periods and other parameters, and it is scientifically correct. However, in regard to outcomes for qualities such as QOL, many clinicians question why feedback on results cannot be given quickly to the subjects assessed. Moreover, some also say that for QOL, perhaps because individual clinicians and patients can to a certain degree imagine the outcomes, findings of a significant difference do not necessarily match the image of clinicians and patients.

The cause of these doubts is that the actual minimum number of points difference needed to indicate a clinical meaning is unknown, as is whether certain differences can be said to be important. In other words, health outcomes researchers did not seem to think seriously about the minimally (clinically) important difference (MID) in the past.

Meanwhile, it has come to be emphasized that, in demonstrating the effectiveness of new medical products with clinical trials that have health outcome as the primary endpoint in the guidance outlined in section I., data on MID are needed in advance in the sample size estimation (in other words, the use of a measure for which MID is already known is strongly recommended) in the stage of developing the study protocol, and debate is increasing.

Various attempts have been made over the past several years to resolve these issues. The methods may be broadly divided into (1) distribution-based methods and (2) anchor-based methods.

In distribution-based methods, the effect size (ES) has been used often. However, the criticism has been made that although standard deviation (SD) is considered, it is sample size dependent. Therefore, indicators such as standard error of measurement (SEM) have come to be considered[7]. These distribution-based methods are simple, but in the end we cannot expect an answer to the fundamental problem of whether they match human perceptions.

Meanwhile, in anchor-based methods detailed investigations using real people have been accumulated for each measure[8,9]. Problems have also been indicated with these methods, such as the possibility of differences occurring with each measure, or when measures have been improved or worsened.

Based on these various studies, Dr. David Osoba, a well-known Canadian health outcome researcher, proposed at an educational workshop of the International Society for Quality of Life Research (ISOQOL) Annual Meeting in 2005 several possible definitions of MID, including "7-8 or 10% change of QOL score', '1/3-1/2 of SD', '0.4 × ES,' and wondered if agreement could not be reached. It is desirable that a consensus among researchers be reached at an early date.

Response Shift (Adaptation to Changing Health)

It has long been known that people's value standards change with experience. This is called a response shift (RS). RS is basically divided into 3 categories[10,11]: "change in internal standards," "change in values," and "reconceptualization."

"Change in internal standards" is the change that occurs in a person's value standards from knowing a higher QOL than before, or conversely a lower QOL, as a result of the occurrence of a major event to that person. "Change in values" is when, for example, a person gives priority to fulfillment in work over family relationships before becoming ill, but after becoming ill comes to place a lower priority on work. "Reconceptualization" is when there is a change in factor structure, such as the inclusion in question items of emotional well-being before an intervention, but inclusion of social well-being after the intervention.

RS is not really a problem in cross-sectional or short-term longitudinal studies, but in randomized clinical trials evaluating the effectiveness of medical products the reliability of statistical significance test results can be affected if a large RS occurs in only one treatment arm of 2 groups.

For all 3 types of RS above, it is important to steadily examine what kinds of problems occur in what situations, and how serious they are. For example, it is necessary to understand the situation from subject attributes, kinds of intervention, study period, baseline QOL, and improving or worsening direction. Clarification of the details of the properties that can occur from these biases may contribute to dramatically raising the reliability of analysis results for QOL/PRO.

Conclusions and Implications

This article outlined the following recent topics and issues in health outcome research, not just in the area of breast cancer, but in medicine in general: (1) introduction of guidance from regulatory authorities related to medical product approval, using the new term of PRO, (2) the possibilities of CAT using IRT, (3) renewed awareness of the importance of MID, and (4) the possibility of improving the reliability of QOL/PRO assessments by promoting RS research.

It is hoped that advances in these research areas will make health outcome research more approachable and beneficial for clinicians and patients alike.

Acknowledgment

This review was partly supported by the grant-in-aid to Kojiro Shimozuma M.D., Ph. D. from the Ministry of Education, Culture, Sports, Science and Technology of Japan (number 08590503).

References

1) Litwin MS, Lubeck DP, Henning JM, Carroll PR: Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: results of the CaPSURE database. *J Urol* 59:1988-1992, 1998.
2) Shimozuma K, Ohashi Y, Takeuchi A, Morita S, Ohsumi S, Sunada Y, Kuroi K, Makino H, Watanabe T, Hausheer FH: Validation of the Patient Neurotoxicity Questionnaire (PNQ) during taxane chemotherapy in a phase III randomized trial in patients with breast cancer: N-SAS BC 02. Proc. of 27th San Antonio Breast Cancer Symposium, 2004.
3) Shimozuma K, Ohashi Y, Takeuchi A, Aranishi T, Morita S, Kuroi K, Ohsumi S, Makino H, Watanabe T, Hausheer FH: Assessment and quantification of taxane-induced neurotoxicity in a phase III randomized trial of patients with breast cancer (AC followed by PAC/DOC vs. PAC/DOC alone): N-SAS BC 02. *J Clin Oncol* 24:473s, 2006.
4) 'Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQOL) measures in the evaluation of medicinal products' European Medicines Agency Pre-authorisation Evaluation of Medicines for Human Use, Committee for Medical Products for Human Use (CHMP), London, 27 July 2005, http://www.emea.eu.int/pdfs/human/ewp/13939104en.pdf
5) 'Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims (Draft guidance).' U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Devices and Radiological Health (CDRH) Feb. 2006 http://www.fda.gov/CDER/GUIDANCE/5460dft.pdf
6) Patient-reported Outcomes measurement Information System (PROMIS). http://www.nihpromis.org/, Aug, 2006.
7) Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD: Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care* 37:469-478, 1999.
8) Jaeschke J, Guyatt GH: Measurement of health status: Ascertaining the minimal clinically important difference. *Control Clin Trials* 10:407-415, 1989.
9) Doyle C, Crump M, Pintilie M, Oza AM: Does palliative chemotherapy palliate? Evaluation of expectations, outcome, and costs in women receiving chemotherapy for advanced ovarian cancer. *J Clin Oncol* 19:1266-1274, 2001.
10) Schwartz CE, Sprangers MAG: Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med* 48:1531-1548, 1999.
11) Sprangers MAG, Schwartz CE: Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med* 48:1507-1515, 1999.

Breast Cancer
Vol. 14 No. 1 January 2007

Review Article

Current Status of Health Outcome Assessment of Medical Treatment in Breast Cancer

Katsumasa Kuroi[*1], Kojiro Shimozuma[*2,3], Shozo Ohsumi[*4], Hirohisa Imai[*5], and Michikazu Ono[*6]

*[1]Division of Surgery, Breast Oncology, Nyuwakai Oikawa Hospital, *[2]Department of Healthcare and Social Services, University of Marketing and Distribution Sciences, *[3]Institute for Stress Science, Public Health Research Foundation, *[4]Department of Breast Oncology, National Hospital Organization Shikoku Cancer Center, *[5]Department of Epidemiology, National Institute of Public Health, *[6]Department of Health Science and Social Welfare, School of Human Sciences, Waseda University, Japan.

Recent research has shown the importance of the patient's point of view on the goals of medical care, and now health-related quality of life (HR-QOL) has become an important endpoint of clinical studies. However, as HR-QOL is essentially a subjective, personal concept determined from the viewpoint of the patient, it is fundamentally important to understand the concept and use the HR-QOL assessment, to express both the subjective and qualitative concept of HR-QOL in an objective and quantitative way that meets the patient's true needs, and also to obtain high-quality information about HR-QOL. In this article, we describe the concept of HR-QOL, the purpose of HR-QOL measurement, the approach to the HR-QOL assessment, instruments used in the measurement of HR-QOL, and general principles of HR-QOL measurements. We also review the current status of HR-QOL assessment of medical treatment in breast cancer.

Breast Cancer 14:74-80, 2007.

Key words: Health outcome, Quality of life, Breast cancer, Medical treatment

Health outcomes are measures or events that define the medical, personal or social consequences of an illness and its treatment[1]. In the field of oncological research, the focus has usually been on assessing the outcomes of cancer treatment and care through the use of objective measures such as length of survival, clinical or pathological response and toxicities. Potential consequences of treatment, such as severe toxicity, patient discomfort, and infrequent mortality, are often viewed as acceptable risks. Since, however, the World Health Organization defined health as being not only the absence of disease and infirmity but also the presence of physical, mental, and social well-being[2], it has become increasingly important to assess the patients' perspective of their symptoms and their impact on the daily life as a tool for determining treatment and a means to assess the outcome of the chosen treatment. This review aims to give a comprehensive insight in the health outcome assessment, especially the health-related quality of life (HR-QOL) of medical treatment in breast cancer.

Reprint requests to Katsumasa Kuroi, Division of Surgery, Breast Oncology, Nyuwakai Oikawa Hospital, 2-21-16 Hirao, Chuo-ku, Fukuoka 810-0014, Japan.
E-mail: kurochan@dd.iij4u.or.jp

Abbreviations:
CSP-HOR, Comprehensive support project for health outcome research; CSPOR, Comprehensive support project for oncological research of breast cancer; EORTC QLQ, European organization for research and treatment of cancer quality of life questionnaire; EQ-5D, EuroQoL 5 dimension; FACT, Functional assessment of cancer therapy; GOG-Ntx, Gynecologic oncology group-neurotoxicity; HR-QOL, Health-related quality of life; HOR, Health outcome research; N-SAS BC, National surgical adjuvant study of breast cancer; PNEF, Physician neurotoxicity examination form; PNQ, Patient neurotoxicity questionnaire; PRO, Patient-reported outcome; QOL, Quality of life; QOL-ACD, Quality of life questionnaire for cancer patients treated with anticancer drugs

Concept and Methods of HR-QOL Assessment

Concept of HR-QOL

Quality of life (QOL) is essentially a subjective, personal and multidimensional concept, determined from the viewpoint of the patient. The fundamental domains include physical, functional, psychological or mental, social, and spiritual aspects, but most of these domains are interrelated and cannot be clearly separated, and are influenced by a person's experiences, beliefs, expectations, and

perceptions. Among them, the physical, functional, psychological, and social aspects are usually referred to as HR-QOL, and so far health outcome research (HOR) in cancer patients has focused on the assessment of these domains with the use of profile (questionnaire)-type QOL instruments. Measurements of spiritual aspects of QOL are difficult and there is some concern as to whether medical intervention can bring improvements in this area[3].

Purpose of HR-QOL Measurement

The clinical studies on HR-QOL can be categorized as basic studies or clinical application studies[4]. The former aim to investigate the reliability and validity of instruments, to develop the instruments for HR-QOL assessment, and to cross-culturally validate the instruments after translation into various languages, while the latter includes observational or interventional studies evaluating the effect and net balance of new therapeutic strategies on HR-QOL. To provide the information that patients require when treatment decision are being made, detailed data about HR-QOL needs to be collected using well-validated instruments in an era when patients are meant to be offered more opportunities to be partners with their health-care providers in therapeutical decision-making.

On the other hand, the measurement of an individual patient's HR-QOL can be used in clinical practice to help decision making, to facilitate detection of unexpected physical or psychological problems, to monitor disease and treatment over time, and thus improve the delivery of medical care. This approach might provide useful information to care providers, and facilitate communication[5,6]. However, so far a number of practical, methodological and attitudinal barriers have limited the use of patient-based measures of health within routine practice[7,8]. For example, many questionnaires are lengthy and repeated measures may be a burden for both patients and physicians. There are also logistic and financial barriers to data collection, and their analysis in an understandable manner with prompt feedback. Moreover, the effectiveness of the formal provision of information about perceived health status to the clinician in practice remains unclear. In fact, in a systematic review by Espallargues *et al.*[8], prompt feedback led to an increase in the diagnosis of conditions and the use of health services and referrals, while the health status of the patients was similar to that of patients whose physicians were not provided with such information. Thus, there is still need for a more through evaluation of this type of intervention.

Approach to HR-QOL Assessment

HR-QOL is a measurement of health status filtered by the subjective perceptions and expectations of the individual[9]. In this sense, HR-QOL measurement represents a radical realignment between the objective and subjective elements of clinical medicine[10]. However, each domain of HR-QOL has many components, and there is an almost infinite number of states of health, all with differing qualities, and all quite independent of longevity[1]. Moreover, since expectations regarding health and the ability to cope with limitation and disability can greatly affect a person's perception of health and satisfaction with life, two people with the same health status may have different qualities of life. Therefore, the construct can be defined in a number of ways, and a consensus of what the term means has not been reached. Thus a challenge for clinicians and researchers has been knowing their patients' QOL[11].

Considering these issues, one approach might be to apply the techniques of qualitative research. This is an interpretative, naturalistic approach that involves making sense of phenomenon in terms of the meanings people bring to them; to understand thoughts, feelings and experiences of individuals, focusing on direct face-to-face knowledge of patients as human beings coping with their condition and treatment in any setting. This approach sets out without a hypothesis, does not use a control group, a fixed design or any minimum number of subjects, does not attempt objectivity and does not rely on statistical analysis. Thus, qualitative research provides a different type of evidence from quantitative research[12], and can help clinicians to understand the patient's experience of being diagnosed and treated for breast cancer as an individual.

On the other hand, quantitative research as a method holds that both the natural and social sciences strive for testable and confirmable theories that explain phenomena by showing how they are derived from theoretical assumptions, and attempts to reduce social reality to a variable in the same manner as physical reality. This approach places numerical values on HR-QOL to produce a single item score, a scale or a composite index,

and requires psychometric properties such as reliability, validity and responsiveness to change to tightly control the variable in question to see how other variables are influenced. As it is often hard to establish a cause-and-effect relationship on QOL issues for individual patients, conclusions are usually drawn for a group of patients, not an individual.

Instruments used in HR-QOL Assessment

Instruments to measure HR-QOL are commonly classified as generic instruments and disease- or condition-specific instruments. The former are generally applied to assess the QOL of people who suffer no daily illness or patients with benign disease. They are also used occasionally to assess cancer survivors who have gone for a long period without a cancer recurrence. Generic instruments are further classified as those that allow calculation of utility, and the health profile type. The advantage of these instruments is that they allow comparison of QOL beyond the confines of the disease or condition. However, these generic instruments usually lack information specific to the disease or treatment.

Disease- or condition-specific instruments are no different from generic ones in terms of including the main domains of HR-QOL, but the contents are more closely matched to each disease. For example, cancer-specific instruments commonly include items on loss of appetite, weight loss, nausea, and hair loss. In addition, subscales or modules for additional concerns which include questions specifically for the treatment or type of disease have been developed as options to be used with the general scale or core questionnaire in disease- or condition-specific instruments. Using these instruments, it is easy to obtain clinically useful information, while it is virtually impossible to compare QOL not specifically related to the disease or condition, and bias in the measured domains of QOL cannot be denied.

General Principles in HR-QOL Measurement in Clinical Application Studies

In clinical application studies, it is fundamentally important to adopt a questionnaire for which the reliability and validity have been confirmed, to express both the subjective and qualitative concept of HR-QOL in an objective and quantitative way that meets the patient's true needs, and to obtain high-quality information about HR-QOL. In addition, as a true measure of HR-QOL needs to assess not only the occurrence of or ability to experience a given symptom or phenomenon, but also satisfaction and concern about the area[13], more detailed questions regarding treatment-related symptoms, level of satisfaction with treatment, body image and sexual function can be included to obtain instructive information.

Moreover, considering the recent increase in the opportunities for international collaboration in clinical trials, it is also important to adopt the instruments after rigorous translation work if they have been developed in a foreign language. For the cross-cultural validation, the first key matter to compare differences between countries and cultures is the translation of materials to accurately convey meaning in the language of each country. At present, several QOL instruments developed in Western countries such as the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire (EORTC QLQ), the Functional Assessment of Cancer Therapy (FACT) scale and the EuroQol 5 Dimension (EQ-5D) are available in Japanese[14-16]. Table 1 lists the currently available HR-QOL instruments for breast cancer currently available in Japanese[3, 4, 17]. Of them, the Japanese version of the EORTC QLQ-C30, FACT-G and FACT-B, and the Quality of Life Questionnaire for Cancer Patients Treated with Anticancer Drugs (QOL-ACD) have been verified in terms of reliability and validity[3, 17, 18]. The others, such as the EORTC QLQ-BR23 and FACT-Taxane, are now being verified for reliability and validity.

For accurate HR-QOL assessment, one should take into account several sources of bias, which may influence the results of the analysis in oncological research. These include not only the medical status such as disease progression and existing treatments but also age, marital status, occupation, level of education, and family income. In addition, one should make every effort to minimize missing data, and the cause of missing data should be identified before analyzing the data collected, as the analytical technique will differ according to the reason for the missing data[19]. Missing data can be broadly classified qualitatively as missing completely at random, missing at random, missing not at random, and missing data tend to increase particularly in HR-QOL assessments in advanced cancer patients. Therefore, it is important to consider these issues as fully as possible

Table 1. Currently Available Japanese Versions of Health-Related Quality of Life Instruments and Psychological Measures for Breast Cancer

Health-related QOL instruments
Generic instruments (Health profile type)
Medical Outcome Study Short Form (SF-36)
World Health Organization Quality of Life Assessment questionnaire (WHO/QOL-26)
Disease/condition-specific instruments for cancer patients
European Organization for Research and Treatment of Cancer Quality of life Questionnaire (EORTC QLQ)
I: Core questionnaire: C30
II: Modules specific to tumor site, treatment modality, or a QL dimension: EORTC QLQ-BR23 (Breast cancer), CIPN20 (Chemotherapy-induced peripheral neuropathy)
Functional Assessment of Cancer Therapy (FACT) scale
I: General instrument: FACT-G
II: Subscales for cancer-, symptom-, or treatment-specific concerns: FACT-B (Breast cancer), An (Anemia), F (Fatigue), Taxane (Taxane toxicity), GOG/Ntx (Gynecologic Oncology Group-Nerotoxicity), Sp (spirituality), ES (Endocrine-related symptoms), Pal (Palliative care)
Quality of Life Questionnaire for Cancer Patients Treated with Anticancer Drugs (QOL-ACD)
I: General instrument: QOL-ACD
II: Subscale for breast cancer: QOL-ACD-B
Psychological measures
The Center for Epidemiologic Studies Depression Scale (CES-D)
General Health Questionnaire (GHQ)
Hospital Anxiety and Depression Scale (HADS)
Profile of Mood States (POMS)
Self-rating Depression Scale (SDS)
State-Trait Anxiety Inventory (STAI)

before the start of the survey. This is especially true in the longitudinal study, as the treatment baseline is an essential survey point.

Current Status of Clinical Application Studies of HR-QOL Assessment in Breast Cancer

Recent studies evaluating treatment and intervention outcomes in breast cancer are increasingly including some measurement of HR-QOL[20]. According to the systematic review of Shimozuma *et al.*[4], the most common setting of clinical application studies in breast cancer has been postoperative, followed by advanced or recurrent disease. There have been few studies focusing on medical examination, prevention, or terminal stage breast cancer. As for interventions, the most common were chemotherapy and/or endocrine therapy, followed by surgery, nursing/care/counseling/rehabilitation, and radiotherapy and medical testing.

It is beyond the scope of this paper to provide a full review of these studies. However, it should be noted that some adverse events that affect HR-QOL are often underestimated or unrecognized by clinicians if assessed using several questionnaires[21], indicating that the manner in which all the adverse events associated with drugs are collected is not always reliable. It is also important to note that, in recent years, there has been a trend toward increasing HR-QOL and other patient-reported outcomes (PROs) claims in regulatory documents for the approval of new pharmaceutical products by the European Medicines Agency [22]. Docetaxel was the first case in which HR-QOL data appeared in product regulatory documents registered at the European Medicines Agency. The impact of docetaxel on HR-QOL was measured by EORTC QLQ in a large phase III study of locally advanced or metastatic breast cancer in which docetaxel was used in combination with doxorubicin, and the HR-QOL results were discussed both under the efficacy and safety sections of the Scientific Discussion as well as in the summary of product characteristics document, stating "In both arms, QOL measured by EORTC questionnaires was comparable and stable during treatment and follow-up." This experience with HR-QOL in the authorization process has served as a baseline for a review of the use of HR-QOL in regulatory submission in European between 1995 and 2003. However, it has been stated that health

Table 2. Summary of Clinical Studies of Comprehensive Support Project for Health Outcome Research Assessing HR-QOL in Breast Cancer

Study	Clinical setting	Study design	Intervention	Instruments
HOR 01-2	Postoperative adjuvant hormone therapy for post-menopausal patients	Longitudinal study (non-controlled study)	Tamoxifen, anastrozole	FACT-G, FACT-ES, Kupperman Menopausal Index, VAS for hot flash, QOL-ACD-B, Original scale
HOR 02	Metastatic breast cancer	Longitudinal study (non-controlled study)	Weekly paclitaxel	FACT-G, FACT-Taxane, PNQ, PNEF
HOR 04-1*	Under treatment	Longitudinal study (non-controlled study)	Chemotherapy, hormone therapy, radiation	FACT-An
HOR 06	Preoperative chemotherapy in operable breast cancer	Longitudinal study (non-controlled study)	Anthracycline and taxane	WAIS-R, HAM-D, POMS, EORTC QLQ-C30, EORTC QLQ-BR23, SF-8
HOR 07	Under treatment	Longitudinal study (non-controlled study)	Taxane	PNQ
N-SAS BC 02	Postoperative adjuvant chemotherapy	RCT	AC followed by taxane vs taxane	FACT-G, FACT-B, FACT-Taxane, EQ-5D, PNQ, PNEF
N-SAS BC 03	Postoperative adjuvant hormone therapy	RCT	Tamoxifen vs tamoxifen followed by anastrozole	FACT-G, FACT-B, FACT-ES, CES-D
N-SAS BC 04 (TEAM-Japan)	Postoperative adjuvant hormone therapy	RCT	Tamoxifen vs exemestane vs anastrozole	FACT-G, FACT-B, FACT-ES, CES-D
SELECT BC	Metastatic or recurrent breast cancer, hormone refractory and HER-2 negative	RCT	Taxane vs TS-1 as first line chemotherapy	EORTC QLQ-C30, PNQ, EQ-5D, PHRS-SCL (SF), CES-D

*Various types of cancer were included.
Abbreviations: AC, doxorubicin and cyclophosphamide; An, anemia; B, breast; BR23; Breast Cacner Module 23; C30, Core Questionnaire 30; CES-D, The Center for Epidemiologic Studies Depression Scale; EORTC QLQ, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire; EQ-5D, EuroQol 5 Dimension; ES, endocrine-related symptoms; FACT, Functional Assessment of Cancer Therapy; G, general; HAM-D, Hamilton Depression Rating Scale; HER-2, human epidermal growth factor-2; HOR, Health outcome research; N-SAS BC, National Surgical Adjuvant Study of Breast Cancer; PHRS-SCL (SF), Public Health Research Foundation Stress Checklist (Short Form); PNEF, Patient Neurotoxicity Examination Form; PNQ, Patient Neurotoxicity Questionnaire; POMS, Profile of Mood States; QOL-ACD-B, Quality of Life Questionnaire for Cancer Patient Treated with Anticancer Drugs; RCT, randomized control trial; SF-8, Short Form 8-item of SF-36; WAIS-R, Wechsler Adult Intelligence Scale-Revised.

researchers need to better justify the inclusion of these outcomes in clinical trials and highlight the added value of PRO data for more efficient utilization of HR-QOL and PROs.

Under these circumstances, the Comprehensive Support Project for Health Outcome Research (CSP-HOR) in the Public Health Research Foundation has conducted several health outcome studies in healthy individuals as well as in patients with various types of cancer in Japan. The aims of these studies are to verify the reliability and validity of existing instruments, to assess the effect of interventions on HR-QOL and to develop instruments. For example, the HOR 01 has developed a scale for endocrine-related symptoms in post-menopausal patients with breast cancer undergoing endocrine treatment (Table 2). HOR 02 aims to assess the effects of weekly paclitaxel (80-100 mg/m^2 per week) on HR-QOL in patients with advanced or metastatic breast cancer. The measures being used in that study are FACT-G, FACT-Taxane including FACT-Gynecologic Oncology Group-Neurotoxicity (GOG/Ntx), Patient Neurotoxicity Questionnaire (PNQ), and the Physician Neurotoxicity Examination Form (PNEF)[23], and the primary endpoint is to evaluate the frequency and severity of neurotoxicity caused by weekly paclitaxel[24]. As a secondary endpoint, HR-QOL and the feasibility, reliability and validity of these questionnaires will be evaluated.

Concurrently, CSP-HOR has supported the health outcome studies in the Comprehensive Support Project for Oncological Research of Breast Cancer (CSPOR-BC). Among them, the National Surgical Adjuvant Study of Breast Cancer (N-SAS BC) 02 is planned to investigate HR-QOL as a secondary endpoint along with survival time, adverse events and cost. This randomized controlled trial compares four cycles of doxorubicin and cyclophosphamide (AC) followed by four cycles of taxanes (docetaxel or paclitaxel), and eight cycles of either taxane, and the comparison of docetaxel and paclitaxel is planned in a two-by-two factorial design. The measures used in this study are FACT-G, FACT-B, FACT-Taxane and EuroQol 5 Dimension (EQ-5D). For analysis of documented patient neurotoxicity symptoms and functional assessment, PNQ and a PNEF are also included. Thus, several health outcome studies are now investigating HR-QOL of medical treatment using several criteria in Japan.

Conclusion

This review has outlined the concept and use of HR-QOL assessment in the medical treatment of breast cancer. As HR-QOL is essentially a subjective, personal concept determined by the viewpoint of the patient, it is fundamentally important to understand the concept and use of HR-QOL assessment, to express both the subjective and qualitative concept of HR-QOL in an objective and quantitative way that meets the patient's true needs, and also to obtain high-quality information about HR-QOL. This will enable us to provide the information that patients require when treatment decisions are being made in an era when patients are meant to be offered more opportunities to be partners with their health-care providers in decision-making.

Acknowledgement

The authors thank Tomoko Moriya for providing information about the health outcome studies and N-SAS BC conducted by CSP-HOR.

References

1) Testa MA, Simonson DC: Assesment of quality-of-life outcomes. *N Engl J Med* 334:835-840, 1996.
2) Constitution of the World Health Organization. World Health Organization. Handbook of basic documents. 5th ed, Palais des Nations, Geneva, p3-20, 1952.
3) Shimozuma K: Quality of life assessment. *Breast Cancer* 9:100-106, 2002.
4) Shimozuma K, Okamoto T, Katsumata N, Koike M, Tanaka K, Osumi S, Saito M, Shikama N, Watanabe T, Mitsumori M, Yamauchi C, Hisashige For The Task Force Of The Japanese Breast Cancer Society For 'The Development Of Guidelines For Quality Of Life Assessment Studies Of Breast Cancer Patients A: Systematic Overview of Quality of Life Studies for Breast Cancer. *Breast Cancer* 9:196-202, 2002.
5) Deyo RA, Carter WB: Strategies for improving and expanding the application of health status measures in clinical settings. A researcher-developer viewpoint. *Med Care* 30:MS176-186; discussion MS196-209, 1992.
6) Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, Selby PJ: Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol* 22:714-724, 2004.
7) Greenhalgh J, Meadows K: The effectiveness of the use of patient-based measures of health in routine practice in improving the process and outcomes of patient care: a literature review. *J Eval Clin Pract* 5:401-416, 1999.
8) Espallargues M, Valderas JM, Alonso J: Provision of feedback on perceived health status to health care professionals: a systematic review of its impact. *Med Care* 38:175-186, 2000.
9) Testa MA: Interpretation of quality-of-life outcomes: issues that affect magnitude and meaning. *Med Care* 38:II166-174, 2000.
10) Sullivan M: The new subjective medicine: taking the patient's point of view on health care and health. *Soc Sci Med* 56:1595-1604, 2003.
11) Okamoto T, Shimozuma K, Katsumata N, Koike M, Hisashige A, Tanaka K, Ohsumi S, Saito M, Shikama N, Mitsumori M, Yamauchi C, Watanabe T: Measuring quality of life in patients with breast cancer: a systematic review of reliable and valid instruments available in Japan. *Breast Cancer* 10:204-213, 2003.
12) Morse JM: Qualitative research is not a modification of quantitative research. *Qual Health Res* 15:1003-1005, 2005.
13) Bonomi AE, Patrick DL, Bushnell DM, Martin M: Quality of life measurement: will we ever be satisfied? *J Clin Epidemiol* 53:19-23, 2000.
14) Kobayashi K, Takeda F, Teramukai S, Gotoh I, Sakai H, Yoneda S, Noguchi Y, Ogasawara H, Yoshida K: A cross-validation of the European Organization for Research and Treatment of Cancer QLQ-C30 (EORTC QLQ-C30) for Japanese with lung cancer. *Eur J Cancer* 34:810-815, 1998.
15) Shimozuma K, Ohashi Y, Yoshimura K: Reliability and validity of the Japanese version of the Functional Assessment of Cancer Therapy-Breast (FACT-B) quality-of-life instrument; Women's Health Outcome Study (WHOS)-01. *Quality Life Res* 9:287, 2000.
16) Tsuchiya A, Ikeda S, Ikegami N, Nishimura S, Sakai I, Fukuda T, Hamashima C, Hisashige A, Tamura M: Estimating an EQ-5D population value set: the case of Japan. *Health Econ* 11:341-353, 2002.
17) Ohsumi. S, Shimozuma. K: Breast cancer treatment and quality of life of patients. *Jpn J Breast Cancer* 18:113-120, 2003.
18) Kurihara M, Shimizu H, Tsuboi K, Kobayashi K,

Murakami M, Eguchi K, Shimozuma K: Development of quality of life questionnaire in Japan: quality of life assessment of cancer patients receiving chemotherapy. *Psychooncology* 8:355-363, 1999.
19) Ohashi Y, Morita T: Theoretical basis for quality of life assessment: statistics in quality of life assessment. In: Ikegami N, Fukuhara N, Shimozuma K, *et al.* ed, Handbook of quality of life assessment for clinicians, ed, Igaku-Shoin, Tokyo, p19-31, 2001 (in Japanese).
20) Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R: Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ* 324:1417, 2002.
21) Fallowfield L: Acceptance of adjuvant therapy and quality of life issues. *Breast* 14:612-616, 2005.
22) Szende A, Leidy NK, Revicki D: Health-related quality of life and other patient-reported outcomes in the European centralized drug regulatory process: a review of guidance documents and performed authorizations of medicinal products 1995 to 2003. *Value Health* 8:534-548, 2005.
23) Hausheer FH, Schilsky RL, Bain S, Berghorn EJ, Lieberman F: Diagnosis, management, and evaluation of chemotherapy-induced peripheral neuropathy. *Semin Oncol* 33:15-49, 2006.
24) Kuroi K, Shimozuma K: Neurotoxicity of taxanes: symptoms and quality of life assessment. *Breast Cancer* 11:92-99, 2004.

Jpn J Clin Oncol 2007;37(2)96–101
doi:10.1093/jjco/hyl145

The Role of the Outpatient Clinic in Chemotherapy for Patients with Unresectable or Recurrent Gastric Cancer

Kentaro Yamazaki, Narikazu Boku, Kaoru Shibamoto, Hirofumi Yasui, Akira Fukutomi, Takayuki Yoshino, Shuichi Hironaka, Yusuke Onozawa, Yosuke Otake, Noriaki Hasuike, Hiroyuki Matsubayashi, Tetsuya Inui, Yuichiro Yamaguchi and Hiroyuki Ono

Division of Gastrointestinal Oncology and Endoscopy, Shizuoka Cancer Center, Sunto-gun, Shizuoka, Japan

Received July 5, 2006; accepted September 30, 2006; published online February 1, 2007

Background: Recently, outpatient chemotherapy centers have become popular in Japan. To clarify the actual conditions of outpatient clinics, we surveyed entire clinical courses of chemotherapy in patients with unresectable or recurrent gastric cancer.
Methods: From the medical records of 64 patients with unresectable or recurrent gastric cancer with no prior chemotherapy, we obtained data on overall survival, non-hospitalized survival, the number of and reasons for attendance at the outpatient clinic and hospitalization, and medical conditions at discharge.
Results: The median follow-up time was 520 days, the median survival time was 353 days, and the median non-hospitalized survival time was 282 days. Patients attended the outpatient clinic 1917 times in total; 145 (8%) of these were unplanned visits for accidental disease, disease progression, or toxicity. Patients were hospitalized 291 times in total: 110 (38%) of hospitalizations were unplanned or emergencies because of disease progression or toxicity. Patients were discharged 290 times in total; in 56 of these discharges (19%) unresolved medical problems remained, such as toxicity, total parenteral nutrition, or symptoms related to cancer. Three patients (5%) died from treatment-related leucopenia and thrombocytopenia.
Conclusions: Patients with unresectable and recurrent gastric cancer were treated at outpatient clinics for periods up to 80% longer than the entire clinical course of chemotherapy. However, there were some unplanned or emergency hospitalizations and some patients still experienced medical problems at discharge. The role of the outpatient clinic is very important to chemotherapy for patients with unresectable or recurrent gastric cancer.

Key words: gastric cancer – chemotherapy – outpatient clinic

INTRODUCTION

Gastric cancer is one of the leading causes of death in Japan and throughout the world. Recent progress in diagnostic procedures and surgical treatment has improved the curability of gastric cancer in the resectable stages. However, the prognosis of unresectable or recurrent gastric cancer still remains poor. Randomized trials have demonstrated that fluorouracil (5-FU)-based chemotherapy can improve survival and quality of life (QOL) in patients with unresectable or recurrent gastric cancer compared with best supportive care (1). Although several phase III trials have been conducted for patients with advanced gastric cancer in recent decades, no standard treatment has been established.

However, various novel anti-tumor agents have been developed recently, including irinotecan (CPT-11), oral pyrimidines, taxanes and molecular target agents. Many phase I and II trials have reported on the activities of these new agents, which are used either as single agents or as combination therapy. For the patient, hospitalization deteriorates daily activity, and non-hospitalized survival can thus represent one substantial improvement to QOL. Many of these new drugs, especially oral anti-tumor drugs, can be used in an outpatient setting and may therefore contribute to prolonging the non-hospitalized survival of patients with gastric cancer treated with chemotherapy. While in Japan

For reprints and all correspondence: Kentaro Yamazaki, Division of Gastrointestinal Oncology and Endoscopy, Shizuoka Cancer Center, 1007 Shimonagakubo, Nagaizumi-cho, Sunto-gun, Shizuoka 411-8777, Japan. E-mail: k.yamazaki@scchr.jp

most patients for chemotherapy have received in-hospital treatment, many hospitals have recently been establishing chemotherapy centers, where efforts are made to treat patients on an outpatient basis. The Japan Clinical Oncology Group (JCOG) has adopted non-hospitalized survival time as a secondary endpoint in JCOG9912 (Randomized phase III study of 5-FU continuous infusion versus CPT-11 plus cisplatin versus S-1 in advanced gastric cancer).

In gastric cancer, conditions of patients may deteriorate suddenly as a result of various complications such as peritoneal dissemination, which is usually undetectable by radiological imaging and sometimes causes bowel obstruction, hydronephrosis and obstructive jaundice. It is suggested that management of gastric cancer by chemotherapy at outpatient clinics may be more difficult than other non-digestive malignancies.

Few reports have documented the clinical course from the initiation of treatment to death in patients with gastric cancer treated with chemotherapy, and actual problems in outpatient clinics have scarcely been reported in detail. For example, it is not even known how long the non-hospitalized survival is and what kind of problems are encountered at outpatient clinics during chemotherapy and therefore we are left to conclude that the provision of chemotherapy over a full clinical course for cancer patients is still in its infancy in Japan. In this retrospective study, we surveyed the entire clinical course of patients with unresectable and recurrent gastric cancer treated with chemotherapy to investigate the actual conditions of outpatient clinics during cancer treatment, in order to improve the system in the near future.

PATIENTS AND METHODS

Patient Selection

From 199 patients with unresectable and recurrent gastric cancer receiving chemotherapy at the Shizuoka Cancer Center between September 2002 and March 2004, we selected patients who fulfilled the following eligibility criteria listed in JCOG9912: (i) histologically proven unresectable or recurrent adenocarcinoma of the stomach, except for patients whose unresectable cancer was limited to class V by cytological examination of the abdominal cavity or with no visible tumor; (ii) no prior chemotherapy; (iii) adequate oral intake without nutritional support; (iv) no severe peritoneal dissemination associated with massive ascites or remarkable findings detected by barium enema; (v) age between 20 and 75 years; (vi) Eastern Cooperative Oncology Group (ECOG) performance status (PS) of 2 or better; (vii) no massive pleural effusion; (viii) no other active malignancies; (ix) adequate bone marrow (white blood count 3000–12 000/μl, platelets ≥100 000/μl), renal (creatinine: ≤1.5 mg/dl), and hepatic functions (aspartate aminotransferase ≤99 IU/l, alanine aminotransferase ≤99 IU/l, bilirubin ≤2.0 mg/dl); (x) no other serious medical complications; (xi) no symptomatic brain metastasis; and (xii) written informed consent for chemotherapy.

Treatment Schedule

All chemotherapy regimens were approved in clinical practice to treat patients with gastric cancer by the Clinical Practice Review Committee of the Shizuoka Cancer Center. All patients provided informed consent before chemotherapy was initiated and the chemotherapy continued until tumor progression, unacceptable toxicity, or patient's refusal to continue. For each patient, the chemotherapy regimen was selected according to the patient and physician's choice for the first-line treatment. Treatment was generally performed by the following schedule: (i) S-1 alone: S-1 (40 mg/m^2 per day, orally twice daily) on days 1–28 every 6 weeks (2,3); (ii) S-1 and cisplatin (CDDP): S-1 (40 mg/m^2 per day, orally twice daily) on days 1–21 and CDDP (70 mg/m^2, intravenously) on day 8 every 5 weeks (4); (iii) sequential methotrexate (MTX) and 5-fluorouracil (5-FU): weekly administration of MTX (100 mg/m^2, bolus) followed by 5-FU (600 mg/m^2, bolus) at 3-h intervals, calcium leucovorin (10 mg/m^2, orally or intravenously) administered six times every 6 h starting 24 h after MTX (5), (iv) CPT-11 and CDDP: CPT-11 (70 mg/m^2, intravenously) on days 1 and 15, and CDDP (80 mg/m^2, intravenously) on day 1 every 4 weeks (6); (v) 5-FU continuous infusion (5-FU c.i.): 5-FU (800 mg/m^2, continuous infusion) on days 1–5 every 4 weeks (7); (vi) weekly paclitaxel (w-PTX): weekly administration of PTX (80 mg/m^2, intravenously) for 3 weeks every 4 weeks (8); (vii) CPT-11 and mitomycin C (MMC): CPT-11 (150 mg/m^2, intravenously) and MMC (5 mg/m^2, bolus) every 2 weeks (9); (viii) 5-FU and isovorin (l-LV): weekly administration of 5-FU (600 mg/m^2, bolus) and l-LV (250 mg/m^2, 2-h infusion) for 6 weeks every 8 weeks (10); (ix) 5-FU and CDDP: 5-FU (800 mg/m^2, continuous infusion) on days 1–5 and CDDP (80 mg/m^2, intravenously) on day 1 every 4 weeks (7); (x) CDDP injected intraperitoneally (11); (xi) CPT-11 alone: CPT-11 (150 mg/m^2, intravenously) every 2 weeks (12,13); (xii) CDDP and etoposide (VP-16): CDDP (80 mg/m^2, intravenously) on day 1 and VP-16 (100 mg/m^2, intravenously) on day 1 every 3 weeks (14); (xiii) hepatic arterial infusion (HAI): 5-fluorouracil (333 mg/m^2) each week, epirubicin (30 mg/m^2) once every 4 weeks and mitomycin-C (2.7 mg/m^2) once every 2 weeks administered by HAI (15); (xiv) MMC alone: weekly administration of MMC (5 mg/m^2, bolus). Dose and schedule were modified according to each patient's medical condition and any toxicities observed in the previous courses.

Evaluation and Statistical Analysis

The overall survival time was calculated from the date of the first administration of chemotherapy of the first-line treatment to the date of death by any causes, or to the last date of confirmed survival. The non-hospitalized survival time was