Table II. Probabilities of a sensitive subpopulation finding with the fixed upper limit $\eta^* = 0.80$ and the upper and lower probability cutoffs $\pi^* = 0.70$ and $\pi^*_{stop} = 0.20$ when the total sample size $N = 500$.

			Early stopping						
Scenario	Pattern	Method	First	Second	$\mathcal{P}_{none}$	$\mathcal{P}_4$	$\mathcal{P}_{3-4}$	$\mathcal{P}_{2-4}$	$\mathcal{P}_{all}$
(1)	1	S-A	0.04	0.05	**0.80**	0.04	0.05	0.05	0.05
		R-M	0.62	0.18	**0.99**	0.00	0.00	0.00	0.00
	2	S-A	0.04	0.04	**0.78**	0.03	0.08	0.09	0.03
		R-M	0.64	0.17	**0.99**	0.00	0.00	0.00	0.00
	3	S-A	0.04	0.04	**0.77**	0.08	0.03	0.03	0.09
		R-M	0.60	0.19	**0.99**	0.00	0.00	0.00	0.00
	4	S-A	0.03	0.03	**0.72**	0.01	0.04	0.07	0.17
		R-M	0.66	0.16	**1.00**	0.00	0.00	0.00	0.00
	5	S-A	0.03	0.03	**0.70**	0.15	0.09	0.04	0.02
		R-M	0.54	0.21	**0.98**	0.01	0.00	0.00	0.00
(2)	1	S-A	0.00	0.00	0.00	0.13	**0.54**	0.28	0.05
		R-M	0.00	0.00	0.01	0.15	**0.51**	0.23	0.10
	2	S-A	0.00	0.00	0.00	0.19	**0.48**	0.30	0.03
		R-M	0.00	0.00	0.00	0.24	**0.49**	0.23	0.04
	3	S-A	0.00	0.00	0.00	0.08	**0.55**	0.28	0.09
		R-M	0.01	0.00	0.02	0.09	**0.49**	0.22	0.18
	4	S-A	0.00	0.00	0.00	0.09	**0.50**	0.25	0.17
		R-M	0.00	0.00	0.00	0.11	**0.38**	0.16	0.34
	5	S-A	0.00	0.00	0.04	0.15	**0.49**	0.30	0.02
		R-M	0.06	0.02	0.17	0.16	**0.41**	0.23	0.03
(3)	1	S-A	0.00	0.00	0.00	0.04	0.15	**0.76**	0.05
		R-M	0.00	0.00	0.00	0.06	0.09	**0.71**	0.15
	2	S-A	0.00	0.00	0.00	0.09	0.21	**0.68**	0.03
		R-M	0.00	0.00	0.00	0.13	0.16	**0.64**	0.07
	3	S-A	0.00	0.00	0.00	0.01	0.11	**0.79**	0.09
		R-M	0.00	0.00	0.00	0.03	0.05	**0.62**	0.31
	4	S-A	0.00	0.00	0.00	0.04	0.20	**0.59**	0.17
		R-M	0.00	0.00	0.00	0.05	0.07	**0.33**	0.55
	5	S-A	0.00	0.00	0.00	0.04	0.11	**0.83**	0.02
		R-M	0.02	0.00	0.04	0.05	0.07	**0.80**	0.04
(4)	1	S-A	0.00	0.00	0.00	0.04	**0.86**	0.06	0.05
		R-M	0.00	0.00	0.00	0.04	**0.92**	0.02	0.02
	2	S-A	0.00	0.00	0.00	0.11	**0.78**	0.09	0.03
		R-M	0.00	0.00	0.00	0.13	**0.82**	0.03	0.02
	3	S-A	0.00	0.00	0.00	0.01	**0.87**	0.03	0.09
		R-M	0.00	0.00	0.00	0.01	**0.96**	0.01	0.02
	4	S-A	0.00	0.00	0.00	0.02	**0.74**	0.07	0.17
		R-M	0.00	0.00	0.00	0.03	**0.81**	0.05	0.12
	5	S-A	0.00	0.00	0.01	0.10	**0.83**	0.04	0.02
		R-M	0.04	0.01	0.07	0.09	**0.82**	0.02	0.01
(5)	1	S-A	0.00	0.00	0.04	**0.80**	0.05	0.05	0.05
		R-M	0.07	0.02	0.14	**0.82**	0.03	0.00	0.01
	2	S-A	0.00	0.00	0.01	**0.79**	0.08	0.09	0.03
		R-M	0.03	0.01	0.05	**0.87**	0.06	0.01	0.01
	3	S-A	0.00	0.00	0.10	**0.75**	0.03	0.03	0.09
		R-M	0.16	0.05	0.32	**0.66**	0.01	0.00	0.01
	4	S-A	0.00	0.00	0.00	**0.72**	0.04	0.07	0.17
		R-M	0.01	0.00	0.02	**0.94**	0.02	0.00	0.02
	5	S-A	0.01	0.01	0.29	**0.57**	0.09	0.04	0.01
		R-M	0.29	0.14	0.67	**0.31**	0.01	0.00	0.00

The probabilities of identifying (i) none of the four subgroups, (ii) subgroup 4 only, (iii) subgroups 3 and 4, (iv) subgroups 2–4, and (v) all the four subgroups are shown in $\mathcal{P}_{none}$, $\mathcal{P}_4$, $\mathcal{P}_{3-4}$, $\mathcal{P}_{2-4}$, and $\mathcal{P}_{all}$, respectively. The probabilities of early stopping at the first and second interim analyses, which are included in $\mathcal{P}_{none}$, are also separately shown. The probability values of correct identification are indicated in boldface.

R-M, regression model; S-A, subgroup analysis.

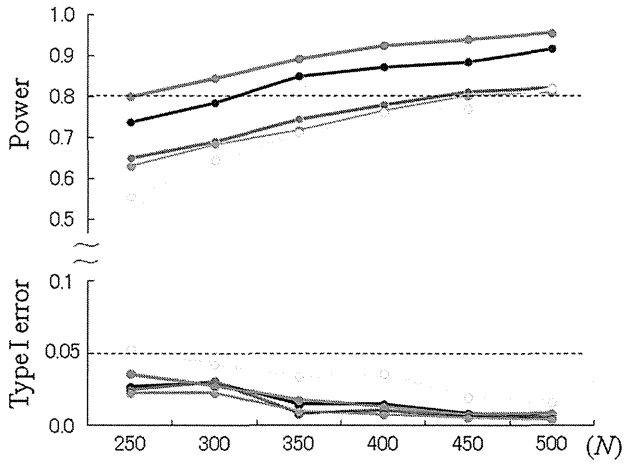

Figure 1. Type I error rates (lower circles) and power values (upper circles) provided by regression model method for the six sample sizes ($N = 250, 300, 350, 400, 450$, and 500) under the five patterns of subpopulation proportions; patterns 1: black, 2: blue, 3: red, 4: green, and 5: yellow. In this investigation, the power is evaluated by the probability of correctly identifying subgroups 3 and 4 under scenario 4. The fixed design parameters $\eta^* = 0.80$, $\pi^* = 0.70$, and $\pi^*_{stop} = 0.20$ are used.

5. Discussion

We have proposed a Bayesian approach with two alternative methods to identify a sensitive subpopulation in the setting of a randomized phase II clinical trial. Taking the simulation results into account, the R-M method may be recommended as the primary choice. The limitations of our proposed approach include the following:

(a) the requirement of a large sample size for a phase II trial,
(b) the inadequate study monitoring,
(c) the monotonicity assumption for hazard ratios of PFS for biomarker subgroups,
(d) the requirement that a specific quantitative biomarker for sensitivity be established in advance, and
(e) lack of experience using our proposed method in an actual clinical trial.

Considering the feasibility of patient enrollment, the projected sample size $N = 300$ may be the upper limit in a clinical trial of second-line therapies for HCC. $N = 300$ may be achievable by enrolling, for instance, 25 patients per month for one year in a multinational trial setting. In some cases, however, it may be unrealistic to enroll such a large number of patients into a phase II trial because of the associated development costs. If we can successfully identify a sensitive subpopulation, however, the required sample size might be minimized in a subsequent phase III trial of an enriched patient population, thereby optimizing the total sample size for the entire clinical development of a new agent. In the phase II trial design, we considered early termination of the entire trial only. Because the trial is still in phase II, it may be highly recommended to monitor the safety of the new treatment. For example, a safety criterion to monitor the probability of toxicity in each subgroup, such as $p(prob(\text{Tox})_g > \eta^*_{Tox} \mid \mathcal{D}) > \pi^*_{stop,Tox}$, where η^*_{Tox} represents an acceptable toxicity level, may be useful. In addition, the efficacy and futility rules for stopping subgroups that we mentioned in Section 3.3 may help reduce the expected sample size of the phase II trial. This should be evaluated in future works. Our study design was based completely on a monotonic change in treatment efficacy for biomarker subgroups. However, such a monotonicity assumption does not necessarily work in all cases. If data observed in the phase II trial indicates a non-monotonic change, such as 'V-shape', the S-A method modified to select the subgroup with the highest value of $p(\lambda_g < \eta^* \mid \mathcal{D}_g)$ may work better than the R-M method. Otherwise, we may need to develop an alternative method based on an isotonic regression model with the pool-adjacent-violator algorithm [23].

In this paper, we focused on identifying a sensitive subpopulation of patients in a randomized phase II trial to develop a new molecular-targeted anticancer agent. It may be useful to incorporate our proposed approach into a seamless phase II/III study design in order to maximize the probability of its successful development, an issue that will be examined in future works.

Acknowledgements

We thank Dr. Richard Simon for his helpful comments and useful suggestions. Satoshi Morita's work was supported in part by a Grant-in-Aid for Scientific Research C-24500345 from the Ministry of Health, Labour, and Welfare of Japan and by the nonprofit organization Epidemiological and Clinical Research Information Network. We thank the associate editor and the referees for their thoughtful and constructive comments and suggestions.

References

1. Simon R, Maitournam A. Evaluating the efficiency of targeted designs for randomized clinical trials. *Clinical Cancer Research* 2004; **10**:6759–6763.
2. Seymour L, Ivy SP, Sargent D, Spriggs D, Baker L, Rubinstein L, Ratain MJ, Le Blanc M, Stewart D, Crowley J, Groshen S, Humphrey JS, West P, Berry D. The design of phase II clinical trials testing cancer therapeutics: consensus recommendations from the clinical trial design task force of the national cancer institute investigational drug steering committee. *Clinical Cancer Research* 2010; **16**:1764–1769.
3. McShane LM, Hunsberger S, Adjei AA. Effective incorporation of biomarkers into phase II trials. *Clinical Cancer Research* 2009; **15**:1898–1905.
4. Dancey JE, Dobbin KK, Groshen S, Jessup JM, Hruszkewycz AH, Koehler M, Parchment R, Ratain MJ, Shankar LK, Stadler WM, True LD, Gravell A, Grever MR, Biomarkers Task Force of the NCI Investigational Drug Steering Committee. Guidelines for the development and incorporation of biomarker studies in early clinical trials of novel agents. *Clinical Cancer Research* 2010; **16**:1745–1755.
5. Parmar MK, Barthel FM, Sydes M, Langley R, Kaplan R, Eisenhauer E, Brady M, James N, Bookman MA, Swart AM, Qian W, Royston P. Speeding up the evaluation of new agents in cancer. *Journal of the National Cancer Institute* 2008; **100**:1204–1214.
6. Mandrekar SJ, Sargent DJ. Clinical trial designs for predictive biomarker validation: one size does not fit all. *Journal Of Biopharmaceutical Statistics* 2009; **19**:530–542.
7. Buyse M, Michiels S, Sargent DJ, Grothey A, Matheson A, de Gramont A. Integrating biomarkers in clinical trials. *Expert Review of Molecular Diagnostics* 2011; **11**:171–182.
8. Baselga J. Herceptin alone or in combination with chemotherapy in the treatment of HER2-positive metastatic breast cancer: pivotal trials. *Oncology* 2001; **61**(Suppl 2):14–21.
9. Amado RG, Wolf M, Peeters M, Van Cutsem E, Siena S, Freeman DJ, Juan T, Sikorski R, Suggs S, Radinsky R, Patterson SD, Chang DD. Wild-type KRAS is required for panitumumab efficacy in patients with metastatic colorectal cancer. *Journal of Clinical Oncology* 2008; **26**:1626–1634.
10. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *Journal of the National Cancer Institute* 2007; **99**:1036–1043.
11. Yin G. *Clinical Trial Design: Bayesian and Frequentist Adaptive Methods*. Wiley: Hoboken, 2012.
12. Wang SJ, O'Neill RT, Hung HM. Approaches to evaluation of treatment effect in randomized clinical trials with genomic subset. *Pharmaceutical Statistics* 2007; **6**:227–244.
13. Brannath W, Zuber E, Branson M, Bretz F, Gallo P, Posch M, Racine-Poon A. Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology. *Statistics in Medicine* 2009; **28**:1445–1463.
14. Eickhoff JC, Kim K, Beach J, Kolesar JM, Gee JR. A Bayesian adaptive design with biomarkers for targeted therapies. *Clinical Trials* 2010; **7**:546–556.
15. Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. *Pharmaceutical Statistics* 2011; **10**:347–356.
16. Korn EL, Arbuck SG, Pluda JM, Simon R, Kaplan RS, Christian MC. Clinical trial designs for cytostatic agents: are new approaches needed? *Journal of Clinical Oncology* 2001; **19**:265–272.
17. Sinha D, Ibrahim JG, Chen MH. A Bayesian justification of Cox's partial likelihood. *Biometrika* 2003; **90**:629–641.
18. Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. In *Encyclopedia of Biostatistics*, Armitage P, Colton T (eds). John Wiley and Sons: Chichester, 2005; 352–366.
19. Gilks W, Richardson S, Spiegelhalter D. *Markov Chain Monte Carlo in Practice*. Chapman & Hall: London, 1996.
20. US Food and Drug Administration (USFDA). *Guidance on Enrichment Strategies for Clinical Trials to Support Approval of Human Drugs and Biological Products*. US FDA: Rockville, MD, 2012.
21. Llovet JM, Pena CE, Lathia CD, Shan M, Meinhardt G, Bruix J, SHARP Investigators Study Group. Plasma biomarkers as predictors of outcome in patients with advanced hepatocellular carcinoma. *Clinical Cancer Research* 2012; **18**:2290–2300.
22. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, de Oliveira AC, Santoro A, Raoul JL, Forner A, Schwartz M, Porta C, Zeuzem S, Bolondi L, Greten TF, Galle PR, Seitz JF, Borbath I, Haussinger D, Giannaris T, Shan M, Moscovici M, Voliotis D, Bruix J, SHARP Investigators Study Group. Sorafenib in advanced hepatocellular carcinoma. *New England Journal of Medicine* 2008; **359**:378–390.
23. Yuan Y, Yin G. Dose–response curve estimation: a semiparametric mixture approach. *Biometrics* 2011; **67**:1543–1554.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site.

International Regulatory Science

DIA DEVELOP INNOVATE ADVANCE

Therapeutic Innovation & Regulatory Science
2014, Vol. 48(5) 644-650

DOI: 10.1177/2168479014524582
tirs.sagepub.com

Exploring Ethnic Differences in Toxicity in Early-Phase Clinical Trials for Oncology Drugs

Takashi Ogura, MD[1,2], Satoshi Morita, PhD[2], Kan Yonemori, MD[3], Takahiro Nonaka, PhD[1], and Tsutomu Urano, PhD[4,5]

Abstract

During oncology drug development, it is important that ethnic differences are evaluated to determine the optimal dose and administration schedule in a new region based on the clinical data from other regions. The objective of this study was to explore the possibility of detecting ethnic differences in toxicity during early-phase clinical trials. Data were reviewed from phase I clinical trials for new drug applications conducted in Japan and Western countries. The maximum tolerated doses (MTDs), recommended phase II doses (RP2Ds), and approved doses in Japan were compared with those in Western countries. There were 4 of 28 drugs eligible for analysis that showed differences in MTDs or RP2Ds between Japanese and Western patients. Differences in MTDs or RP2Ds in 2 phase I trials were associated with ethnic differences in toxicity. It may be worthwhile to evaluate ethnic differences in toxicity during early-phase clinical trials for oncology drugs.

Keywords

ethnic differences, maximum tolerated dose, oncology drugs, phase I trials

Introduction

Differences in the dosage and dose regimen of some drugs among regions have been pointed out, although they cannot definitely be attributed to ethnic differences.[1,2] Examination of ethnic differences is important while planning and conducting global clinical trials and determining whether clinical data from other countries or regions are applicable for clinical development in new countries or regions.

In the evaluation of ethnic differences during drug development, endogenous factors such as race, sex, and genetic polymorphisms and exogenous factors including socioeconomic factors and health care environments should be considered.[3] Examples of ethnic differences with known causes include those due to genetic polymorphisms in enzymes involved in drug metabolism and the ethnic differences in the distribution of these polymorphisms. In the development of S-1, differences in the distribution of the CYP2A6 polymorphism between Japanese and Western individuals caused different toxicity profiles, leading to differences in the maximum tolerated dose (MTD) and the recommended dose for subsequent clinical trials.[4] For irinotecan, variations in the distribution of the UGT1A1*6 and *28 polymorphisms by ethnicity resulted in different metabolism profiles, which resulted in different levels of toxicity.[5] An example of ethnic differences of unknown cause is the difference in the incidence of interstitial lung disease (ILD) with the use of gefitinib and bortezomib. The incidence of ILD is higher in Japanese patients than in Western counterparts.[6–8] Ethnic differences in safety often pose a serious problem in the development of oncology drugs with narrow therapeutic windows.

If ethnic differences in the incidence of serious adverse events can be predicted early in drug development in a new

[1] Office of New Drug V, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan
[2] Department of Biostatistics and Epidemiology, Yokohama City University Graduate School of Medicine, Yokohama, Japan
[3] Breast and Medical Oncology Division, National Cancer Center Hospital, Tokyo, Japan
[4] Office of Vaccines and Blood Products, Pharmaceuticals and Medical Devices Agency, Tokyo, Japan
[5] Yokohama City University Graduate School of Medicine, Yokohama, Japan

Submitted 12-Dec-2013; accepted 28-Jan-2014

Corresponding Author:
Takashi Ogura, Office of New Drug V, Pharmaceuticals and Medical Devices Agency, Shin-Kasumigaseki Building, 3-3-2, Kasumigaseki, chiyoda-ku, Tokyo, 100-0013 Japan.
Email: ogura-takashi@pmda.go.jp

region or country, it could be determined early on whether clinical data in other regions or countries can be used or whether additional data are necessary, and then clinical development would proceed more appropriately. For example, the development of erlotinib, which targets EGFR in the same manner as gefitinib, was based on information of an ethnic difference with a similar drug—that is, a higher incidence of ILD in Japanese patients with the use of gefitinib. Since this higher incidence was recognized in Japan, studies evaluating safety in Japanese persons were conducted during the development of erlotinib.[9,10] In addition, postmarketing data collection for erlotinib focused on the occurrence of ILD.[11] The clinical development of drugs in new countries or regions will proceed more appropriately if the extent of ethnic differences can be evaluated in an exploratory manner during phase I clinical trials that are first conducted in the residents of the new country or region, in addition to referring to the data on similar drugs.

In the present study, we examined the MTD in phase I clinical trials and the recommended phase II doses (RP2D) and approved doses of new oncology drugs to evaluate whether or not ethnic differences in toxicity can be detected in early-phase clinical trials in new countries or regions.

Methods

We reviewed the data from phase I clinical trials for new drug applications conducted in Japan and Western countries that had been reviewed by the Pharmaceutical and Medical Devices Agency (PMDA) and approved by the Japanese Ministry of Health, Labour, and Welfare between September 1999 and March 2011. Specifically, we examined the PMDA review reports—the documents submitted by the application sponsors, which have been publicly released on the websites of the PMDA[12]—and the published study reports to compare the MTD (or the maximum administered dose, if MTD was not reached) and the RP2D for the Japanese population and that in the US and Europe. The definitions of the terms in this study were as follows: MTD was the lowest dose level at which more than 33% of patients experience dose-limiting toxicity (DLT). RP2D was one dose level below the MTD.

To evaluate ethnic differences between Japanese and Western populations, we compared the approved doses of all drugs according to the prescribe information shown on the website of the regulatory agencies in each region,[13–15] and we retrospectively analyzed the safety profile and frequency of adverse events of all drugs based on the published study reports when differences in MTD or RP2D were identified.

To assess the adequacy of phase I clinical trial design for detecting any differences in toxicity, we compared the dose escalation methods and reasons for stopping dose escalation in the Japanese trials with those conducted in the US and Europe.

No statistical comparisons were made because of the retrospective nature of this analysis.

Results

Between 1999 and 2011, a total of 97 oncology drugs were approved in Japan. Among them, 39 drugs with novel active ingredients were approved. The following drugs were excluded from this study: 4 drugs that had not been approved in the US and Europe (miriplatin, tamibarotene, talaporfin, amrubicin); 3 hormonal drugs (letrozole, exemestane, anastrozole); 2 drugs for which phase I clinical trials were not conducted in Japan (thalidomide, nelarabine); 1 drug for which dose escalation studies were not conducted in the US and Europe (azacitidine); and 1 drug used with different supportive therapies between Japan and the US and Europe (pemetrexed). Thus, 28 drugs were examined in this study.

Drugs With Differences in MTD, RP2D, and Approved Doses Between Japanese and Western Populations

Differences in MTD or RP2D between Japanese and Western populations were observed for 4 of 28 drugs: temsirolimus (with differences only in MTD) and capecitabine, fludarabine, and topotecan (with differences in both MTD and RP2D). Among the drugs with differences in MTD or RP2D, fludarabine and topotecan had different approved dosages and dose regimens. These differences and details of DLT are shown in Table 1. For the drugs without differences in MTD or RP2D, there was also no differences in the approved dosage and dose regimen.

Safety Profiles of the Drugs With Differences in MTD, RP2D, and Approved Doses

The incidence of adverse events with capecitabine—including pigmentation, diarrhea, increased aspartate aminotransferase level, and elevated bilirubin level—was different between Japanese and non-Japanese patients (Table 2). For temsirolimus, a higher incidence of stomatitis and ILD was observed in Japanese persons than in non-Japanese persons (Table 3). The safety profile of topotecan and fludarabine could not be compared owing to the lack of studies conducted using the same dose regimens in Japan as in the US or Europe. However, the occurrence rate of hematologic toxicity with topotecan in Japanese patients is the same as in European patients despite using their different doses, suggesting that there are differences in the occurrence rate of hematologic toxicity between Japanese and European patients (Table 4). Besides, a higher incidence of hematologic toxicity was observed with fludarabine at lower doses in Japanese people than in US people. The

Table 1. MTD, RP2D, approved dose, and DLT of drugs with different toxicity profiles between Japanese and Western populations found in phase I trials.

Drug: Region	MTD or MAD[a]	RP2D	Approved Dose	DLT
Capecitabine				
US	1657 mg/m^2/d; daily	1331 mg/m^2/d; daily	2500 mg/m^2/d; days 1-14 every 3 wk	Hand-foot syndrome, diarrhea, nausea, vomiting, vertigo, dehydration, abdominal pain, dyspnea, venous thrombosis, thrombocytopenia
Europe (UK, NLD)	1657 mg/m^2/d; days 1-14 every 3 wk	2510 mg/m^2/d; days 1-14 every 3 wk	2500 mg/m^2/d; days 1-14 every 3 wk	Hand-foot syndrome, diarrhea, nausea, vomiting, stomatitis, abdominal pain, neutropenia, leucopenia, thrombocytopenia, neutropenia with sepsis
Japan	2510 mg/m^2/d; daily	1657 mg/m^2/d; days 1-21 every 4 wk	2500 mg/m^2/d; days 1-14 every 3 wk[b]	Hemorrhagic gastric ulcer, skin toxicity
Fludarabine				
US	40 mg/m^2/d; days 1-5 every 4 wk	25 mg/m^2/d; days 1-5 every 4 wk for patients without prior therapy[c]	25 mg/m^2/d; days 1-5 every 4 wk	Granulocytopenia, thrombocytopenia
Japan	25 mg/m^2/d	20 mg/m^2/d	20 mg/m^2/d; days 1-5 every 4 wk	Neutropenia, thrombocytopenia
Topotecan				
US	2.5 mg/m^2/d; days 1-5 every 3 wk	*Initial dose:* 1.5 mg/m^2/d; days 1-5 every 3 wk *2nd dose:* 2.0 mg/m^2/d; days 1-5 every 3 wk	1.5 mg/m^2/d; days 1-5 every 3 wk	Neutropenia, febrile neutropenia
Europe (NLD, DNK)	1.5 mg/m^2/d; days 1-5 every 3 wk	1.5 mg/m^2/d; days 1-5 every 3 wk	1.5 mg/m^2/d; days 1-5 every 3 wk	Neutropenia, leukopenia
Japan	1.5 mg/m^2/d; days 1-5 every 3 wk	1.2 mg/m^2/d; days 1-5 every 3 wk	1.0 mg/m^2/d; days 1-5 every 3 wk (maximum dose: 1.5 mg/m^2/d)	Neutropenia, leukopenia
Temsirolimus				
Europe	Not reached (220 mg/m^2)	Not determined	25 mg	Stomatitis, asthenia
Japan	45 mg/m^2	15 mg/m^2	25 mg	Diarrhea, stomatitis

DLT, dose-limiting toxicity; DNK; Denmark; MAD, maximum administered dose; MTD, maximum tolerated dose; NLD, Netherlands; RP2D, recommended phase 2 dose.

[a]If MTD was not reached, MAD was given.

[b]There was a difference in dosage and dose regimen at the time of the first approval application in Japan (2 wk of administration followed by 1 wk without administration in the US and Europe and 3 wk of administration followed by 1 wk without administration in Japan). However, additional clinical studies were conducted in Japan, resulting in the approval of the same dosage and dose regimens as those approved in the US and Europe.

[c]The RP2D was 18 mg/m^2/d for patients with prior chemotherapy or radiotherapy.

incidence of neutropenia was 69% in Japanese people and 18% in US people (Table 5).

Dose Escalation Methods and Reasons for Stopping Dose Escalation

According to the PMDA review reports, for the 28 drugs examined, 78 dose escalation studies were conducted, which consisted of 32 studies in Japanese patients and 46 in European and American patients.

The dose was increased in a 3 + 3 design in 31 of 32 studies in Japanese patients and another design in the remaining study. In the 46 studies in Western persons, the dose was increased in a 3 + 3 design in 37 studies, with a continual reassessment method in 2 studies and other designs in 7 studies (Table 6).

In the 32 studies with Japanese participants, the reason for discontinuation of dose escalation was toxicity in 8 studies, confirmation of the tolerability of the overseas recommended dose in 20 studies, and other in 4 studies. In the 46 studies with Western participants, the reason was toxicity in 24 studies, consideration of pharmacokinetics in 3 studies, achievement of the dose expected to block the target in 3 studies, and other in 16 studies (Table 7).

Table 2. Incidence of treatment-related adverse events with capecitabine (2500 mg/m^2 for 14 d, every 3 wk), No. (%).

	JO15951 (Japan)[20,21] (n = 60)		SO14695 (US, Canada, Mexico, and Brazil)[20,22] (n = 299)		SO14796 (Europe, Australia, New Zealand, Taiwan, and Israel)[20,23] (n = 297)	
	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4
Hand-foot syndrome	44 (73)	8 (13)	175 (59)	54 (18)	143 (48)	48 (16)
Pigmentation	23 (38)	0	3 (1)	0	7 (2)	0
Diarrhea	21 (35)	1 (2)	148 (50)	46 (15)	136 (46)	32 (11)
Nausea	21 (35)	0	121 (41)	10 (3)	104 (35)	5 (2)
Vomiting	9 (15)	0	92 (31)	11 (4)	47 (16)	6 (2)
Appetite loss	20 (33)	3 (5)	66 (22)	3 (1)	37 (13)	2 (1)
Stomatitis	21 (35)	0	81 (27)	9 (3)	37 (13)	2 (1)
Increased AST level	43 (72)	6 (10)	110 (37)	2 (1)	130 (44)	3 (1)
Elevated bilirubin level	40 (67)	20 (33)	123 (41)	52 (17)	162 (55)	84 (28)
Decreased lymphocyte count[a]	33 (55)	5 (8)	276 (92)	117 (39)	276 (93)	103 (35)

[a]Evaluation criterion was different between Japan and other countries.

Table 3. Incidence of adverse events with temsirolimus (25 mg), No. (%).

	2217-AP (Asia)[24–26] (n = 76)		2217-AP (Japanese Patients)[25,26] (n = 14)		304-WW (US, Europe, Australia, Canada, Asia-Pacific, Africa, and South America)[25–28] (n = 208)	
	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4
Stomatitis	42 (55)	4 (5)	11 (79)	1 (7)	45 (22)	2 (1)
Diarrhea	17 (22)	2 (3)	3 (21)	1 (7)	56 (27)	3 (1)
Skin toxicity	55 (72)	2 (3)	14 (100)	0	142 (68)	11 (5)
Hyperglycemia	10 (13)	1 (1)	4 (29)	0	6 (3)	1 (1)
Increased creatinine level	19 (25)	1 (1)	2 (14)	0	25 (12)	4 (2)
Pneumonitis	12 (16)	2 (3)	5 (36)	1 (7)	4 (2)	2 (1)
Pneumonitis (independent review)	42 (59)[a]	NR	8 (57)[b]	NR	52 (29)[c]	NR

NR, not reported.
[a]Chest computed tomographic (CT) images of 71 evaluable patients were read by an independent advisory board.
[b]Chest CT images of 14 evaluable patients were read by an independent advisory board.
[c]Chest CT images of 178 evaluable patients were read by an independent blinded review.

Table 4. Incidence of adverse events with topotecan, No. (%).

	Early Phase II (Japan),[29] 1.2 mg/m^2 (n = 97)		Late Phase II (Japan),[29] 1.0 mg/m^2 (n = 96)		Phase II (Europe),[29,30] 1.5 mg/m^2 (n = 100)	
	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4
Leukopenia	96 (99)	68 (70)	50 (100)	63 (66)	100 (100)	90 (90)
Neutropenia	82 (99)	74 (89)	50 (100)	81 (84)	100 (100)	96 (96)
Anemia	92 (95)	56 (58)	48 (96)	44 (46)	100 (100)	29 (29)
Thrombocytopenia	80 (83)	42 (43)	45 (90)	40 (42)	100 (100)	55 (55)

Discussion

In the present study, 2 cytotoxic drugs—fludarabine and topotecan—showed hematologic toxicity in phase I trials. This eventually led to different doses of these drugs being approved in Japan and in the US and Europe. We cannot confirm ethnic differences from the results of the phase I trials, as the number of patients was limited. However, this finding suggests a hypothesis that the differences in MTD or RP2D in early clinical trials may be associated with ethnic differences in toxicity. Therefore, when we found the differences in MTD or RP2D,

Table 5. Incidence of adverse events with fludarabine, No. (%).

	Phase II (Japan),[31] 20 mg/m^2/d (n = 26)		Phase I/II (MDAH, US),[31] 20, 25, or 30 mg/m^2/d[a] (n = 101)	
	All Grades	Grades 3 and 4	All Grades	Grades 3 and 4
Leukopenia	8 (31)	4 (15)	NR	NR
Neutropenia	18 (69)	14 (54)	18 (18)	2 (2)
Anemia	7 (27)	5 (19)	15 (15)	3 (3)
Thrombocytopenia	13 (50)	3 (12)	8 (8)	3 (3)
Pancytopenia	NR	NR	2 (2)	1 (1)
Bone marrow suppression	NR	NR	6 (6)	2 (2)
Other hematologic toxicity	NR	NR	6 (6)	1 (1)
Red blood cell count decreased	7 (27)	NR	NR	NR
Lymphocyte count decreased	6 (23)	NR	NR	NR

MDAH, MD Anderson Hospital, Houston, Texas, USA; NR, not reported.
[a]Intrapatient dose escalation was permitted.

Table 6. Dose escalation study design.

Design	Japan (n = 32)		Western Countries (n = 46)	
	No. of Trials	%	No. of Trials	%
3 + 3 design	31	96	37	80
Continual reassessment method	0	0	2	4
Other	1	3	7	15

Table 7. Reason for stopping clinical trials.

Reason	Japan (n = 32)		Western Countries (n = 46)	
	No. of Trials	%	No. of Trials	%
Toxicity	8	25	24	52
Study objective met[a]	20	63	0	
Pharmacokinetics	0		3	7
Target inhibition	0		3	7
Other	4	13	16	35

[a]Almost all studies had the objective of evaluating tolerability of the dosage approved for Western populations.

we might need to collect additional data, including pharmacokinetics, genetic polymorphism, and other ethnic factors.

It is unclear why the approved doses of both fludarabine and topotecan were different in 2 regions. These drugs had a DLT, which was hematologic toxicity. However, the other drugs with hematologic toxicity as the DLT did not have different approved doses.

In the pharmacokinetic study of fludarabine, the area under the curve of plasma 2F-ara-A, which is the active metabolite of fludarabine phosphate, was similar between the Japanese and American patients.[16] Although the distribution of the common variant alleles of *CYP* genes is known to vary among different ethnic populations,[17] in an in vitro study, 3H-2F-ara-A was not metabolized by *CYP3A4* and *CYP1A2*.

Topotecan is a topoisomerase I inhibitor, which is a water-soluble derivative of camptothecin. The pharmacokinetic parameters with topotecan—C_{max}, area under the curve, and $T_{1/2}$ levels in the plasma—were not different between Japanese and Western patients. Human liver microsomal metabolism of topotecan and its metabolite was not affected by *CYP1A2*, *CYP2A6*, *CYP2C8/9*, *CYP2C19*, *CYP2D6*, *CYP2E1*, *CYP3A4*, and *CYP4A*.[17]

Yet, capecitabine and temsirolimus showed no differences in approved dosages and dose regimens, although both MTD and RP2D were different between Japan and the US and Europe. For temsirolimus, although gastrointestinal toxicity such as diarrhea and stomatitis was caused by a considerable disparity in MTD between Japan and Europe, the safety profile of this drug in later clinical trials showed a difference in the incidence of not only stomatitis but also ILD.

The reason for such discrepancies in ethnic differences between earlier and later clinical trials is unknown. Note that observed ethnic differences in early clinical trials can be attributed to patient-level differences because these studies were conducted with the limited number of patients. Examination of study designs in the present study showed that 20 of 32 phase I clinical trials in new regions (62.5%) did not employ a dose escalation design to determine a region-specific MTD but rather attempted to confirm tolerability of the doses recommended in the regions where the drugs were previously developed. Only 8 studies (25%) specifically evaluated the development of toxicity to determine MTD. Therefore, we speculate that even if tolerability in new regions is similar to that in the previously approved regions, dosages and dose regimens in new regions would not be sufficiently evaluated.

In the present study, as far as we can determine, the approved doses were the same for drugs without any differences in MTD

or RP2D between Japanese and Western participants in early clinical trials. However, some drugs without any differences in MTD or RP2D demonstrated different toxicity profiles in Japanese participants. For example, there was a higher incidence of ILD with gefitinib and bortezomib in Japanese participants.[6–8]

Two theories have been put forth on why there was a failure to detect ethnic differences in the toxicity profile of gefitinib and bortezomib in early clinical trials and a discrepancy between the toxicity observed in the early versus later clinical trials for temsirolimus and capecitabine. First, less frequent adverse events cannot be detected in clinical studies with a small number of participants. In the present study, the toxicity of the 2 drugs (fludarabine and topotecan) shown to be different among different populations was hematologic toxicity, a relatively frequent adverse event. Conversely, early clinical trials with the small number of participants have only a limited capacity to detect ethnic differences in adverse events with relatively low incidences—for example, ILD. Depending on the properties of the specific drug and those in the same class, it may be more helpful to search for evidence of ethnic differences in later clinical trials. Second, dose escalation design was not strictly followed in the phase I trials when a drug is being studied for a new region. For bortezomib, dose escalation was discontinued because tolerability of the overseas recommended dose was confirmed and a sufficient determination of MTD was not performed.

A potentially more significant problem is the possibility that uncommon but severe adverse events do not surface during the clinical development stage. In Japan, immediately after the launch of gefitinib, ILD associated with the drug's use caused multiple cases of death. Its prescribing information was ultimately revised to raise awareness of the risk of ILD.[18,19] It should be recognized that information collected by early and late clinical trials is not sufficient. We consider it meaningful to collect the data from multinational trials, including early clinical trials, and continue the examination for ethnic differences in a larger number of patients, including postmarketing surveys.

Two limitations of the present study should be considered. One is that it examined only drugs that were eventually approved. Drugs whose development was discontinued, potentially due to ethnic differences detected during clinical development, were not examined. The other is that we could not find the information on the difference of sampling interval for laboratory variables and the criteria in each trial to report the laboratory-related adverse events between Japan and Western countries. As for the adverse event reporting, the slightly abnormal laboratory values tend to be strictly reported as adverse events in Japan, while they did not tend to in Western countries. Although these tendencies could not cause the ethnic differences in severe hematologic toxicity, these points should be noted in the interpretation of the results.

The present study found that phase I clinical trials detected ethnic differences in the toxicity profile of 2 of 28 drugs examined, suggesting that it is important to collect additional data in later clinical trials when MTD or RP2D in a new region is different from that in previously approved regions.

Authors' Note

This study was presented as part of the European Multidisciplinary Cancer Congress, September 23-27, 2011, Stockholm, Sweden. The views expressed are the result of independent work and do not represent the views of the Pharmaceuticals and Medical Devices Agency of Japan.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported in part by a Grant-in-Aid for Scientific Research [C-24500345 to S.M.] from the Ministry of Health, Labour and Welfare of Japan.

References

1. Malinowski HJ, Westelinck A, Sato J, Ong T. Same drug, different dosing: differences in dosing for drugs approved in the United States, Europe, and Japan. *J Clin Pharmacol.* 2008;48:900-908.
2. Arnold FL, Kusama M, Ono S. Exploring differences in drug doses between Japan and Western countries. *Clin Pharmacol Ther.* 2010;87:714-720.
3. ICH. Ethnic factors in the acceptability of foreign clinical data. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E5_R1/Step4/E5_R1__Guideline.pdf. Accessed December 9, 2013.
4. Ajani JA, Faust J, Ikeda K, et al. Phase I pharmacokinetic study of S-1 plus cisplatin in patients with advanced gastric carcinoma. *J Clin Oncol.* 2005;23:6957-6965.
5. Minami H, Sai K, Saeki M, et al. Irinotecan pharmacokinetics/pharmacodynamics and UGT1A genetic polymorphisms in Japanese: roles of UGT1A1*6 and *28. *Pharmacogenet Genomics.* 2007;17:497-504.
6. Kudoh S, Kato H, Nishiwaki Y, et al. Interstitial lung disease in Japanese patients with lung cancer: a cohort and nested case-control study. *Am J Respir Crit Care Med.* 2008;177:1348-1357.
7. Miyakoshi S, Kami M, Yuji K, et al. Severe pulmonary complications in Japanese patients after bortezomib treatment for refractory multiple myeloma. *Blood.* 2006;107:3492-3494.
8. FDA. Label and approval history. http://www.accessdata.fda.gov/drugsatfda_docs/label/2006/021602s008,s009.pdf. Accessed December 9, 2013.
9. Okusaka T, Furuse J, Funakoshi A, et al. Phase II study of erlotinib plus gemcitabine in Japanese patients with unresectable pancreatic cancer. *Cancer Sci.* 2011;102:425-431.

10. Kubota K, Nishiwaki Y, Tamura T, et al. Efficacy and safety of erlotinib monotherapy for Japanese patients with advanced non-small cell lung cancer: a phase II study. *J Thorac Oncol.* 2008; 3:1439-1445.
11. Nakagawa K, Kudoh S, Ohe Y, et al. Postmarketing surveillance study of erlotinib in Japanese patients with non-small-cell lung cancer (NSCLC): an interim analysis of 3488 patients (POLARSTAR). *J Thorac Oncol.* 2012;7:1296-1303.
12. Pharmaceuticals Medical Devices Agency. Approved products [in Japanese]. http://www.info.pmda.go.jp/info/syounin_index.html. Accessed December 9, 2013.
13. European Medicines Agency. European public assessment reports. http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp&mid=WC0b01ac058001d124. Accessed December 9, 2013.
14. FDA. FDA approved drug products. http://www.accessdata.fda.gov/scripts/cder/drugsatfda/. Accessed December 9, 2013.
15. Pharmaceuticals Medical Devices Agency. Package inserts [in Japanese]. http://www.info.pmda.go.jp/info/iyaku_index.html. Accessed December 9, 2013.
16. Pharmaceuticals Medical Devices Agency. Review reports (fludarabine) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/g990914/75repo01.pdf. Accessed December 9, 2013.
17. Zhou SF, Liu JP, Chowbay B. Polymorphism of human cytochrome P450 enzymes and its clinical impact. *Drug Metab Rev.* 2009;41(2):89-295.
18. Pharmaceuticals Medical Devices Agency. The yellow letter / blue letter [in Japanese]. http://www.info.pmda.go.jp/kinkyu_anzen/kinkyu20021015.html. Accessed December 9, 2013.
19. Inoue A, Saijo Y, Maemondo M, et al. Severe acute interstitial pneumonia and gefitinib. *Lancet.* 2003;361(9352):137-139.
20. Pharmaceuticals Medical Devices Agency. Review reports (capecitabine) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/P200700068/45004500_21500AMZ00400_A100_1.pdf. Accessed December 9, 2013.
21. Hyodo I, Shirao K, Doi T, et al. A phase II study of the global dose and schedule of capecitabine in Japanese patients with metastatic colorectal cancer. *Jpn J Clin Oncol.* 2006;36:410-417.
22. Hoff PM, Ansari R, Batist G, et al. Comparison of oral capecitabine versus intravenous fluorouracil plus leucovorin as first-line treatment in 605 patients with metastatic colorectal cancer: results of a randomized phase III study. *J Clin Oncol.* 2001;19: 2282-2292.
23. Van Cutsem E, Twelves C, Cassidy J, et al. Oral capecitabine compared with intravenous fluorouracil plus leucovorin in patients with metastatic colorectal cancer: results of a large phase III study. *J Clin Oncol.* 2001;19:4097-4106.
24. Sun Y, Rha S, Lee SH, et al. Phase II study of the safety and efficacy of temsirolimus in East Asian patients with advanced renal cell carcinoma. *Jpn J Clin Oncol.* 2012;42:836-844.
25. Pharmaceuticals Medical Devices Agency. Published study report (temsirolimus) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/P201000043/index.html. Accessed December 9, 2013.
26. Pharmaceuticals Medical Devices Agency. Review reports (temsirolimus) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/P201000043/67145000_22200AMX00870_A100_1.pdf. Accessed December 9, 2013.
27. Maroto JP, Hudes G, Dutcher JP, et al. Drug-related pneumonitis in patients with advanced renal cell carcinoma treated with temsirolimus. *J Clin Oncol.* 2011;29:1750-1756.
28. Hudes G, Carducci M, Tomczak P, et al. Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *N Engl J Med.* 2007;356:2271-2281.
29. Pharmaceuticals Medical Devices Agency. Published study report (topotecan) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/g001212/index.html. Accessed December 9, 2013.
30. Ardizzoni A, Hansen H, Dombernowsky P, et al. Topotecan, a new active drug in the second-line treatment of small-cell lung cancer: a phase II study in patients with refractory and sensitive disease. The European Organization for Research and Treatment of Cancer Early Clinical Studies Group and New Drug Development Office, and the Lung Cancer Cooperative Group. *J Clin Oncol.* 1997;15: 2090-2096.
31. Pharmaceuticals Medical Devices Agency. Published study report (fludarabine) [in Japanese]. http://www.info.pmda.go.jp/shinyaku/g990914/index.html. Accessed December 9, 2013.

Therapeutic Innovation
& Regulatory Science
2014, Vol. 48(2) 213-219

DOI: 10.1177/2168479013500970
tirs.sagepub.com

A Continual Reassessment Method With Cohort Size Adaptation Based on Bayesian Posterior Probabilities in Phase I Dose-Finding Studies

Tomoyuki Kakizume, MS[1] and Satoshi Morita, PhD[1]

Abstract

In phase I cancer studies, the maximum tolerated dose (MTD) is estimated by gradually increasing dose levels while accumulating safety information. Recently, Bayesian dose-finding methods such as the continual reassessment method (CRM) have gained popularity. Due to the lack of safety information, phase I studies on new drugs must start at doses low enough that efficacy is not expected but safety is certain up to an acceptable level. To reach the MTD with fewer patients, a 2-stage method has been proposed that enrolls only a single patient at each dose level until the first dose-limiting toxicity is observed. If the study drug is less toxic, it may require many cohorts to complete the study and thus may lead to a longer study period. In this paper, the authors propose a new CRM with cohort size adaptation to reduce the number of cohorts without reducing the accuracy of MTD selection. The cohort size is determined based on the Bayesian posterior probabilities computed during a study. Simulation studies show that the proposed method reduced the number of cohorts compared with the 2-stage method while still yielding a comparable probability of selecting the MTD correctly.

Keywords

continual reassessment method, dose-finding, cohort size adaptation, phase I cancer trial, Bayesian posterior probability.

Introduction

The primary objective of a phase I dose-finding study for cancer is to estimate the maximum tolerated dose (MTD). In many cases, the MTD is estimated through the use of rule-based designs, typically the 3+3 design. However, model-based dose-finding methods have gained popularity because of their ability to estimate the MTD more accurately.[1-6] The MTD is defined as the dose with a probability of dose-limiting toxicity (DLT) closest to a given target (eg, 33%); DLT is determined clinically before the start of each study. Model-based methods integrate all DLT information observed during the study and update model parameters sequentially based on the observed data as the cohort progresses. A dose-toxicity model is used to make dose escalation decisions for the next cohort based on the estimated toxicity probabilities.

One of the early model-based designs was the continual reassessment method (CRM) proposed by O'Quigley et al.[1] However, concerns arose regarding the fact that the CRM allowed dose skipping and that its approach to dosing the initial patients in a study was based on a priori dose-toxicity curves.[2] Several modifications were proposed, including treating the first patients at a low starting dose and prohibiting dose escalation to no more than 2 dose levels at a time.[3,4] Goodman et al[3] and Ahn[5] recommended assigning more than 1 patient to each cohort. Many studies using CRM have fixed the size of each cohort to 3 patients. When new drugs are developed, phase I dose-finding studies must start with low doses at which efficacy is not expected but safety is certain up to an acceptable level. Dose levels must be increased gradually because of unknown safety characteristics.[2,3,7] If there is a large gap between the starting dose and the MTD, it may be necessary to treat a relatively large number of patients at suboptimal dose

[1] Department of Biostatistics and Epidemiology, Yokohama City University Graduate School of Medicine, Yokohama, Japan

Submitted 11-Apr-2013; accepted 16-Jul-2013

Corresponding Author:
Tomoyuki Kakizume, Department of Biostatistics and Epidemiology, Yokohama City University Graduate School of Medicine, 4-57 Urafune-cho, Minami-ku, Yokohama 232-0024, Japan.
Email: t106015a@yokohama-cu.ac.jp

levels that are below the MTD. From an ethical perspective, however, it is required that investigators minimize patients treated at ineffective doses while minimizing patients treated at toxic doses.[6]

Moller[7] proposed the restricted CRM (R-CRM), which is a 2-stage design that begins as a rule-based design (first stage), including only a single patient at each dose level, and then switches to a CRM (second stage) once the first DLT is observed. It was shown that the R-CRM reaches the MTD with fewer patients compared with CRM with a fixed cohort size of 3. However, the R-CRM may require many cohorts to complete the study when the study drug is less toxic, because it enrolls only a single patient at each dose level until the first DLT is observed. In this case, there is concern that the study period may be unnecessary long. Estimating the accurate MTD rapidly contributes to accelerate new drug development and leads to possible treatments for patients suffering from cancer.

In this paper, we propose a new CRM with cohort size adaptation that is determined based on the Bayesian posterior probabilities calculated during a study. This Bayesian posterior probability CRM (BPP-CRM) reduces cohort size at doses suggested to be far from MTD or increases cohort size for doses suggested to be near MTD based on its posterior probability, and thereby it reduces the number of cohorts while still yielding a comparable probability of selecting the true MTD. Simulations are used to compare the BPP-CRM with the R-CRM.

In the next section, we summarize the dose-finding method and present the cohort size determination algorithm of the BPP-CRM. Next, we conduct extensive simulation studies to examine the operating characteristics of our proposed method. We close with a brief discussion.

Methods

The second stage of the R-CRM and the BPP-CRM use the same dose-finding process described next. However, each method uses different cohort size determination algorithms. The R-CRM fixes the cohort size to 1 during the first stage and 3 during the second stage. The BPP-CRM adjusts the cohort size based on the Bayesian posterior probabilities, as explained in the section on the cohort size determination rule.

Dose-Finding Process

The CRM is based on a Bayesian parametric model characterized by a model parameter or parameters representing the dose-toxicity relationship.[8] The general idea behind the CRM proposed by O'Quigley et al[1] was that a dose-toxicity relationship would be updated with all available toxicity data using Bayes' theorem and that each patient would be assigned the dose most likely to be the MTD. However, concerns arose regarding the fact that the CRM allowed dose skipping and that

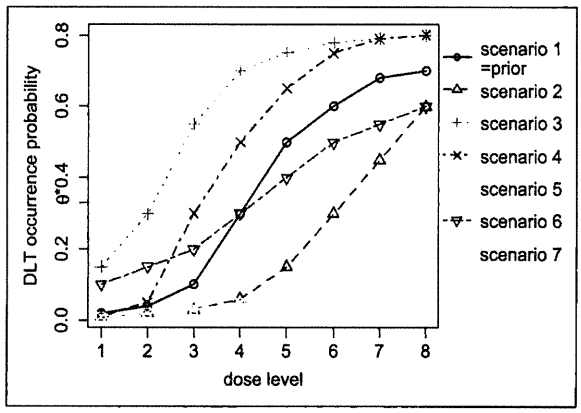

Figure 1. Prior and true dose-limiting toxicity occurrence probabilities.

its approach to dosing the initial patients in a study was based on an a priori dose-toxicity relationship.[2] Several modifications were proposed, including treating the first patients at a low starting dose and prohibiting dose escalation to no more than 2 dose levels at a time.[3,4]

In this paper, the following dose-finding steps are performed (more details are provided in the appendix):

Step 1: Assume a priori dose-toxicity curve and the target probability. This dose-toxicity curve characterizes clinical investigators' uncertainty or knowledge before starting the study and is sometimes based on historical data from previous clinical studies in which identical or similar study treatments were examined.[8]

Step 2: Treat several patients (depending on method) at the assigned dose level and evaluate the occurrence of DLT.

Step 3: Update the dose-toxicity curve with all available DLT data using Bayes' theorem and compute the posterior expected DLT rates at each dose level.

Step 4: Determine the next dose level at which the posterior expected DLT rate is the closest to the target probability. In this regard, dose skipping is prohibited.

Step 5: Repeat steps 2 through 4 until the fixed sample size of 30 is reached, or terminate early in the case of unacceptable toxicity at the lowest dose level.

Cohort Size Determination Rule in BPP-CRM

After determining the next dose level x_i by using the dose-finding method provided in the preceding section and the appendix, the BPP-CRM adjusts the cohort size based on the posterior probability of the DLT rate at x_i given available $(i-1)$ enrolled patients' data $\Omega_{i-1} = \{x_1, x_2, \ldots, x_{i-1}, y_1, y_2, \ldots, y_{i-1}\}$

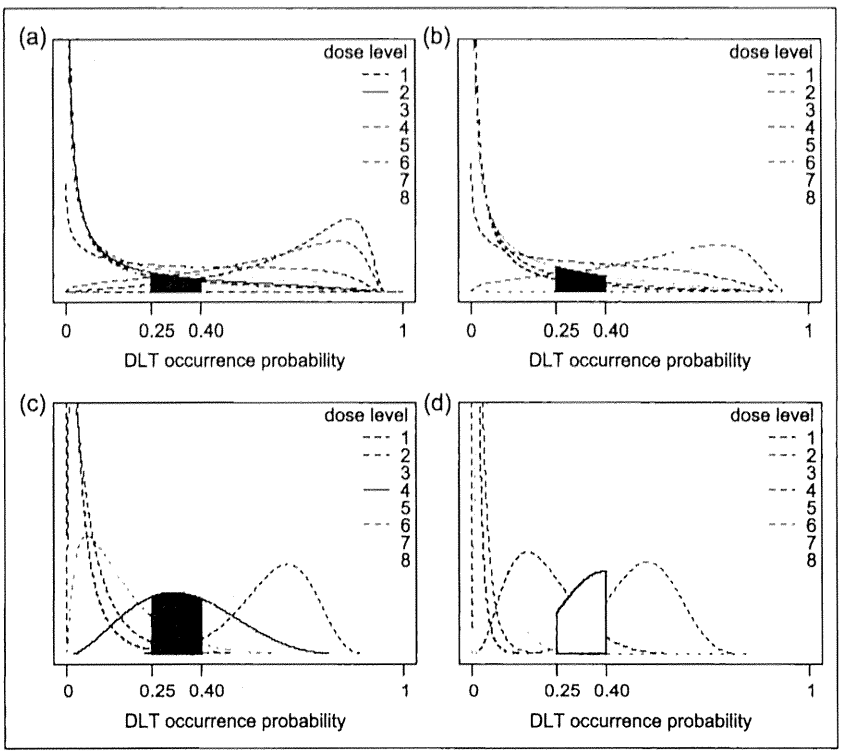

Figure 2. Prior and posterior density functions of the dose-limiting toxicity (DLT) rate estimated at each of the 8 dose levels and probabilities of the DLT rate falling in the target probability at the next dose level (shaded zone) based on toxicity data (a) prior, (b) after the first cohort (1 patient), (c) after the fifth cohort (12 patients), and (d) after the seventh cohort (21 patients).

falling in the target interval [0.25, 0.40], which centers on the target probability: $\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\}$. In other words, $\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\}$ represents the distance between x_i and the MTD. When $\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\}$ is large, x_i is assessed to be near the MTD, and therefore a large cohort size is assigned to reduce the number of cohorts. In contrast, when $\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\}$ is small, x_i is assessed to be suboptimal or highly toxic, and therefore a small cohort size is assigned so as to limit the number of patients who receive suboptimal or highly toxic doses. In summary, the cohort size is calculated by $[\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\}*M] + 1$, where x is the gauss symbol, which is the greatest integer that is $\leq x$, and M is the design parameter. For example, when $M = 10$ and $\Pr\{R(x_i|\Omega_{i-1})\in[0.25, 0.40]\} = 0.28$, the cohort size is $[0.28*10] + 1 = 2 + 1 = 3$. M should be determined based on the prior probability of the DLT rate and the maximum cohort size in simulation under some scenarios. To determine M efficiently, it is useful to find the maximum M first, and M can be finalized based on the maximum cohort size under some scenarios. Maximum M can be determined based on the prior probability of the DLT rate falling in the target interval at the starting dose x_1 $\Pr\{R(x_1)\in[0.25, 0.40]\}$ and appropriate cohort size at first cohort. If the expected cohort size at first cohort is 1, the maximum M can be the largest number that meets $\max\{M|\Pr\{R(x_1)\in[0.25, 0.40]\}*M < 1\}$. It is easy to see that the maximum cohort size should be smaller when a smaller M is used. The final M can be determined based on the maximum cohort size under some scenarios. For example, if the expected cohort size at first cohort is 1 under the setting provided in the next section, M should be ≤ 10 because $\Pr\{R(x_1)\in[0.25, 0.40]\} = 0.096$.

Simulation Studies

Simulation Settings

We ran simulations to compare the operating characteristics of the BPP-CRM with those of the R-CRM. We considered 8 dose levels with the target probability $\theta^* = 0.33$. The patients enrolled in the first cohort were always treated at dose level 2. We used the same dose-toxicity model, dose-finding process, and stopping rule given in the section on dose-finding process and the appendix for both CRMs. For the cohort size determination rule for the BPP-CRM, we set $M = 10$ (see "Example of

Table 1. Simulation results for the BPP-CRM and R-CRM.

		Dose Level									
		1	2	3	4	5	6	7	8	Early Termination	Mean DLT
Scenario 1		0.02	0.04	0.10	**0.30**	0.50	0.60	0.68	0.70		
BPP-CRM	%MTD	0	0	3.5	**75.6**	20.2	0.6	0	0	0	9.0
	#Pats	0.1	1.2	4.7	**17.3**	6.1	0.5	0	0		
R-CRM	%MTD	0	0	3.5	**75.9**	19.9	0.6	0	0	0	9.3
	#Pats	0.1	1.5	4.1	**16.7**	6.5	0.8	0.2	0.1		
Scenario 2		0.01	0.02	0.03	0.06	0.15	**0.30**	0.45	0.60		
BPP-CRM	%MTD	0	0	0	0.5	16.5	**60.2**	20.0	2.8	0	6.8
	#Pats	0	1.1	2.3	3.4	8.6	**10.4**	3.5	0.8		
R-CRM	%MTD	0	0	0	0.6	17.0	**62.3**	18.5	1.7	0	7.5
	#Pats	0.1	1.2	1.4	2.5	8.3	**11.0**	4.2	1.4		
Scenario 3		0.15	**0.30**	0.55	0.70	0.75	0.78	0.79	0.80		
BPP-CRM	%MTD	11.2	**67.6**	21.1	0.1	0	0	0	0	0	10.8
	#Pats	5.8	**14.0**	9.2	0.9	0.1	0	0	0		
R-CRM	%MTD	10.7	**68.7**	20.2	0.2	0	0	0	0	0.2	10.9
	#Pats	5.2	**15.0**	8.5	1.0	0.2	0	0	0		
Scenario 4		0.01	0.05	**0.30**	0.50	0.65	0.75	0.79	0.80		
BPP-CRM	%MTD	0	2.7	**71.8**	25.4	0.2	0	0	0	0	10.3
	#Pats	0.5	2.9	**16.9**	8.9	0.9	0	0	0		
R-CRM	%MTD	0	2.7	**70.8**	26.3	0.3	0	0	0	0	10.4
	#Pats	0.2	3.3	**16.3**	8.9	1.2	0.2	0	0		
Scenario 5		**0.30**	0.50	0.60	0.65	0.69	0.72	0.74	0.75		
BPP-CRM	%MTD	**68.2**	27.4	2.3	0.2	0	0	0	0	2.0	11.9
	#Pats	**16.5**	8.8	3.7	0.6	0	0	0	0		
R-CRM	%MTD	**67.8**	26.6	2.9	0.3	0	0	0	0	2.4	11.9
	#Pats	**16.6**	9.1	2.9	0.7	0.2	0	0	0		
Scenario 6		0.10	0.15	0.20	**0.30**	0.40	0.50	0.55	0.60		
BPP-CRM	%MTD	0	1.7	21.6	**51.0**	22.5	2.9	0.2	0	0	8.4
	#Pats	0.8	2.7	8.3	**11.9**	5.3	1.0	0.1	0		
R-CRM	%MTD	0.1	1.6	22.7	**47.4**	24.4	3.6	0.3	0	0	8.7
	#Pats	0.8	3.1	7.5	**10.5**	5.9	1.6	0.4	0.3		
Scenario 7		0.01	0.02	0.03	0.05	0.08	0.13	0.20	**0.30**		
BPP-CRM	%MTD	0	0	0	0.2	2.5	11.5	23.6	**62.2**	0	4.6
	#Pats	0.1	1.1	2.3	3.0	5.0	6.2	5.6	**6.9**		
R-CRM	%MTD	0	0	0	0.3	2.0	9.2	27.1	**61.4**	0	5.7
	#Pats	0.1	1.2	1.3	1.9	3.3	4.6	6.7	**11.0**		

True DLT rates are presented in the first row of each scenario; MTDs under each scenario are shown in boldface. BPP-CRM, Bayesian posterior probability continual reassessment method (CRM); DLT, dose-limiting toxicity; %MTD, the percentage of times each dose level was selected as the MTD; #Pats, mean number of allocated patients; R-CRM, restricted CRM.

Dose Escalation History" below for the determination of M). Prior DLT rates at dose levels 1 to 8 were estimated as 0.02, 0.04, 0.10, 0.30, 0.50, 0.60, 0.68, and 0.70, respectively. Figure 1 shows the prior DLT rates and true DLT rates under 7 scenarios, covering a very broad range of scenarios that might be true dose-toxicity relationships.

We simulated 5000 trials for each scenario. To investigate the operating characteristics of each design, we calculated the percentage of times each dose level was selected as the MTD, the mean number of patients treated at each dose level, the mean number of DLTs per study, and the mean number of cohorts per study. The accuracy was assessed based on the percentage of trials that identified the true MTD, and the risk control of underdosing or overdosing was assessed based on the mean number of patients treated at doses under or over the MTD. These simulations ran under the assumption that there were enough patients standing by. Under this assumption, the mean number of cohorts was reasonable to assess the duration to estimate the MTD. In addition, the mean number of cohort size at each cohort was calculated to investigate how cohort size changes.

Example of Dose Escalation History

This section describes how to determine the design parameter M and gives 2 examples of dose escalation histories of the BPP-CRM.

Table 2. Summary of the number of cohorts for the BPP-CRM and R-CRM.

Scenario		Mean	SD	Minimum	Maximum	Quantile		
						25%	50%	75%
1	BPP-CRM	9.2	0.40	8	10	9.0	9.0	9.0
	R-CRM	12.9	0.90	11	14	12.0	13.0	14.0
2	BPP-CRM	9.0	0.17	8	14	9.0	9.0	9.0
	R-CRM	13.6	0.72	11	14	14.0	14.0	14.0
3	BPP-CRM	8.5	0.55	3	12	8.0	8.0	9.0
	R-CRM	11.8	0.76	2	14	11.0	12.0	12.0
4	BPP-CRM	9.0	0.62	8	10	9.0	9.0	9.0
	R-CRM	12.4	0.76	11	14	12.0	12.0	13.0
5	BPP-CRM	8.7	1.12	3	17	8.0	9.0	9.0
	R-CRM	11.4	1.29	2	14	11.0	11.0	12.0
6	BPP-CRM	9.0	0.46	8	10	9.0	9.0	9.0
	R-CRM	12.6	1.07	2	14	12.0	12.0	14.0
7	BPP-CRM	9.5	1.10	8	17	9.0	9.0	10.0
	R-CRM	13.7	0.69	11	14	14.0	14.0	14.0

BPP-CRM, Bayesian posterior probability continual reassessment method (CRM); R-CRM, restricted CRM; SD, standard deviation.

As described in the methods section, M can be determined based on the prior probability of DLT rate falling in the target interval at the starting dose d_2. Because the expected cohort size at first cohort is 1 and $\Pr\{R(d_2)\in[0.25, 0.40]\} = 0.096$, M should be ≤ 10. To meet our objective that the number of cohorts should be minimized, we decided to use $M = 10$.

In the BPP-CRM, the cohort size is determined after the dose that is estimated to be the closest to the target probability is selected as the next dose (see the appendix). In the case that no DLT is observed out of 1 patient at first cohort, posterior expected DLT rates at each dose level given available information at first cohort $\Omega_1 = \{d_2,0\}$ are updated, and d_3 is selected as the dose at second cohort by using the method provided in the appendix. In this case, the cohort size at second cohort is $[\Pr\{R(d_3|\Omega_1)\in[0.25, 0.40]\}*10] + 1 = [0.121*10] + 1 = 2$. In the case that no DLT is observed out of 2 patients at second cohort also (ie, $\Omega_2 = \{d_2, d_3, d_3, 0, 0, 0\}$), the next dose is d_4 and the cohort size is $[\Pr\{R(d_4|\Omega_2)\in[0.25, 0.40]\}*10] = [0.194*10] + 1 = 2$.

Through repetition of this process, both dose level and cohort size at the next cohort can be selected sequentially. Figure 2 shows prior and some posterior density functions of the DLT rate at each of the 8 dose levels. The graphs suggest that the density functions of the DLT rate become narrower and that the cohort size, as well as the posterior probability of the DLT rate falling in the target interval at the next dose level, becomes bigger as available DLT data increase.

Simulation Results

Table 1 shows the percentage of times each dose level was selected as the MTD, the mean number of patients treated at each dose level, and the mean number of DLTs per study in each scenario. In all scenarios, both methods selected the true MTD with the same accuracy. In scenarios 1, 4, and 6, in which the true dose-toxicity relationship is almost identical to the prestudy estimates, BPP-CRM showed a tendency to concentrate more patients to the target dose compared with the R-CRM. In scenarios 2 and 7, in which the true dose-toxicity relationship is flatter than the prestudy estimates, the BPP-CRM showed a tendency to concentrate more patients to lower doses compared with the R-CRM. However, in scenarios 3 and 5, in which the true MTD is lower than the prestudy estimate, the BPP-CRM tended to be comparable to the R-CRM.

Table 2 summarizes the number of cohort size in each scenario. In all scenarios, the BPP-CRM reduced from 2 to 4 cohorts compared with the R-CRM.

Figure 3 shows the mean cohort size and cumulative number of patients by cohort in scenario 1. Results in other scenarios are not shown in this paper because they show almost the same tendency. R-CRM and BPP-CRM used the same cohort size of 1 at first cohort. However, while the R-CRM moderately increased the cohort size to 3, the BPP-CRM increased cohort size more rapidly and reached 5.1 at the eighth cohort, leading to a reduction of the number of cohorts.

Discussion

In this paper, we propose a new CRM design with cohort size adaptation to reduce the number of cohorts. The cohort size was determined based on Bayesian posterior probabilities falling to the target interval. The BPP-CRM reduced the number of cohorts by 2 to 4 compared with the R-CRM while still yielding a comparable probability of selecting the true MTD.

In many cases except for rare diseases, because a certain period (eg, 2 months) is required the time when the new cohort

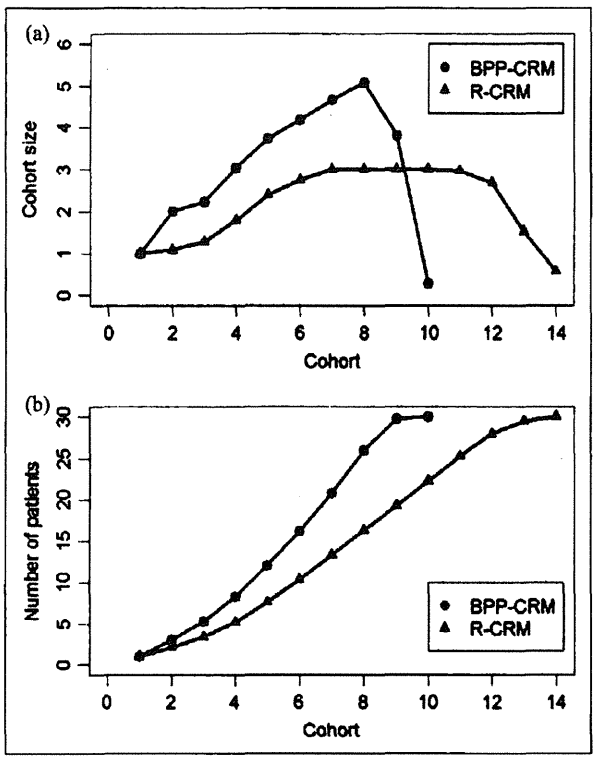

Figure 3. (a) The mean number of cohort size by cohort in scenario 1. (b) Mean cumulative number of patients by cohort in scenario 1. BPP-CRM, Bayesian posterior probability continual reassessment method (CRM); R-CRM, restricted CRM.

opens to the time when the next dose level is finalized, it is expected that BPP-CRM would shorten the study period by 4 to 8 months.

The BPP-CRM can enroll more patients at a single cohort as the study progresses. There may be concerns that this leads to an increased safety risk. Actually, the BPP-CRM showed a tendency to assign more patients to higher doses compared with the R-CRM if the true dose-toxicity relationship was steeper than the prestudy estimates. However, the BPP-CRM showed almost the same DLTs as the R-CRM, suggesting that the BPP-CRM can enroll many patients at a single cohort while ensuring patients' safety.

Further, to control the risk of overdosing, it may be useful to use our proposed design in conjunction with the escalation with overdose control (EWOC) method proposed by Babb et al.[9] We used a binary response for each patient to indicate the presence or absence of DLT. The severity of toxicity was evaluated using multiple grades from 0 to 3 or 4 based on common toxicity criteria for adverse events. It may be desirable to consider the possibility of using these grades to determine the cohort size more adequately.

Appendix

Let d_j ($j = 1, 2, \ldots, K$) denote numerical dose levels, with d_j specified using backward fitting[2] as described below. Let $R(d_j)$ denote the true DLT rate at dose level d_j. The binary response is the indicator $Y_i = 1$ if the ith patient ($i = 1, 2, \ldots, n$) suffers a DLT, 0 if not. In this paper, a 1-parameter logistic regression model,

$$\psi(d_j, a) = \frac{\exp(3 + a \times d_j)}{1 + \exp(3 + a \times d_j)} \quad (1)$$

is assumed for the dose-toxicity working model $\psi(d_j, a) = \Pr(Y_i = 1|d_j, a)$, ($j = 1, 2, \ldots, K$, $i = 1, 2, \ldots, n$), with the assumed prior distribution $g(a)$ for parameter a being Gamma(5,5).[10] We specified numerical dose levels of d_j for $j = 1, 2, \ldots, K$ using backward fitting[2] so that d_j is satisfied with the equation $\psi(d_j|a = \mathrm{E}(a)) = p_j$ ($j = 1, 2, \ldots, K$), where p_j is the prestudy estimate of the proportion of patients who would experience a DLT at dose level j. In addition, we fixed the parameter a to E(a), where E(a) denotes the prior mean of a under $g(a)$, in this case E(a) = 1. For example, if $\{p_1, p_2, \ldots, p_8\} = \{0.02, 0.04, 0.10, 0.30, 0.50, 0.60, 0.68, 0.70\}$, then $\{d_1, d_2, \ldots, d_8\} = \{-6.89, -6.18, -5.20, -3.85, -3.00, -2.59, -2.25, -2.15\}$.

To determine the ith patient's dose level $x_i \in \{d_1, d_2, \ldots, d_K\}$, the posterior distribution of parameter a is updated based on available ($i - 1$) enrolled patients' data $\Omega_{i-1} = \{x_1, x_2, \ldots, x_{i-1}, y_1, y_2, \ldots, y_{i-1}\}$ using Bayes' rule. The posterior distribution of a is given by

$$p(a|\Omega_{i-1}) = \frac{L(a|\Omega_{i-1})g(a)}{\int L(u|\Omega_{i-1})g(u)du},$$

where the likelihood function $L(a|\Omega_{j-1})$ is given by

$$L(a|\Omega_{i-1}) = \prod_{l=1}^{i-1} \psi(x_l, a)^{y_l}\{1 - \psi(x_l, a)\}^{1-y_l}.$$

Based on the posterior distribution of a, the posterior expected DLT rate at dose level d_i is given by

$$\bar{R}(d_i|\Omega_{i-1}) = \int \psi(d_i, a)p(a|\Omega_{i-1})da. \quad (2)$$

The next dose level x_i is determined based on the criterion

$$\bar{R}(x_i|\Omega_{i-1}) - \theta^*| < |\bar{R}(d_j|\Omega_{i-1}) - \theta^*| < |-\theta^* \quad (i = 1, 2, \ldots, n,\ j = 1, 2, \ldots, K, x_i \le x_{i-1} + 1, x_i \neq d_j) \quad (3)$$

That is, x_i is the dose at which the posterior expected DLT rate given all available data is the closest to the target probability θ^*. In this regard, dose skipping is prohibited.

The fixed sample size of 30 is used based on Thall et al.[11] In addition, the study is terminated early in the case of unacceptable toxicity at the lowest dose level if $\Pr\{R(d_1|\Omega_{i-1}) > \theta_{toxic}\} \ge 0.95$, where θ_{toxic} is the lower limit of the DLT rate considered to be

toxic and in many cases is set at the same value of the target probability θ^*.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: Satoshi Morita's work was supported in part by a Grant-in-Aid for Scientific Research C-24500345 from the Ministry of Health, Labour and Welfare of Japan and by the nonprofit organization Epidemiological and Clinical Research Information Network.

References

1. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics*. 1990;46:33-48.
2. Garrett-Mayer E. The continual reassessment method for dose-finding studies: a tutorial. *Clin Trials*. 2006;3:57-71.
3. Goodman SN, Zahurak ML, Piantadosi S. Some practical improvements in the continual reassessment method for phase I studies. *Stat Med*. 1995;14:1149-1161.
4. Korn EL, Midthune D, Chen TT, Rubinstein LV, Christian MC, Simon RM. A comparison or two phase I trial designs. *Stat Med*. 1994;13:1149-1161.
5. Ahn C. An evaluation of phase I cancer clinical trial designs. *Stat Med*. 1998;17:1537-1549.
6. Ratain MJ, Mick R, Schilsky RL, Siegler M. Statistical and ethical issues in the design and conduct of phase I and II clinical trials of new anticancer agents. *J Natl Cancer Inst*. 1993;85: 1637-1643.
7. Moller S. An extension of the continual reassessment methods using a preliminary up-and-down design in a dose finding study in cancer patients, in order to investigate a greater range of doses. *Stat Med*. 1995;14:911-922.
8. Morita S, Toi M, Saji S, et al. Practical application of the continual reassessment method to a phase I dose-finding trial in advanced breast cancer. *Clin Trials*. 2007;41:691-700.
9. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Stat Med*. 1998; 17:1103-1120.
10. Ishizuka N, Morita S. Practical implementation of the continual reassessment method. In: Crowley J, ed. *Handbook of Statistics in Clinical Oncology*. 2nd ed. New York, NY: CRC Press; 2005:97-116.
11. Thall PF, Lee JJ, Tseng CH, Estey EH. Accrual strategies for phase I trials with delayed patient outcome. *Stat Med*. 1999;18: 1155-1169.

DOI:10.1093/jnci/djt269
Advance Access publication October 9, 2013

BRIEF COMMUNICATION

Progression-Free Survival as a Surrogate for Overall Survival in Advanced/Recurrent Gastric Cancer Trials: A Meta-Analysis

Xavier Paoletti, Koji Oba, Yung-Jue Bang, Harry Bleiberg, Narikazu Boku, Olivier Bouché, Paul Catalano, Nozomu Fuse, Stefan Michiels, Markus Moehler, Satoshi Morita, Yasuo Ohashi, Atsushi Ohtsu, Arnaud Roth, Philippe Rougier, Junichi Sakamoto, Daniel Sargent, Mitsuru Sasako, Kohei Shitara, Peter Thuss-Patience, Eric Van Cutsem, Tomasz Burzykowski, Marc Buyse; on behalf of the GASTRIC group

Manuscript received February 10, 2013; revised July 25, 2013; accepted July 25, 2013.

Correspondence to: Xavier Paoletti, PhD, Bisotatistics Dept / INSERM U900, Institut Curie, 26 rue d'Ulm, 75005 Paris, France (e-mail: xavier.paoletti@curie.net).

The traditional endpoint for assessing efficacy of chemotherapies for advanced/recurrent gastric cancer is overall survival (OS), but OS requires prolonged follow-up. We investigated whether progression-free survival (PFS) is a valid surrogate for OS. Using individual patient data from the GASTRIC meta-analysis, surrogacy of PFS was assessed through the correlation between the endpoints and through the correlation between the treatment effects on the endpoints. External validation of the prediction based on PFS was also evaluated. Individual data from 4069 patients in 20 randomized trials were analyzed. The rank correlation coefficient between PFS and OS was 0.853 (95% confidence interval [CI] = 0.852 to 0.854). The R^2 between treatment effects on PFS and on OS was 0.61 (95% CI = 0.04 to 1.00). Treatment effects on PFS and on OS were only moderately correlated, and we could not confirm the validity of PFS as a surrogate endpoint for OS in advanced/recurrent gastric cancer.

J Natl Cancer Inst;2013;105:1667–1670

The prognosis of patients with advanced or recurrent gastric cancer (AGC) remains poor, with a 1-year median overall survival (OS) for commonly used chemotherapy regimens, consisting of fluoropyrimidine, platinum, taxane or anthracyclines agents (1). The most important issue in the development of agents for AGC is their ability to prolong OS with acceptable toxicity. Even though median postprogression survival ranges from 5 to 10 months, a validated shorter-term surrogate endpoint would likely reduce drug development costs, sample sizes, or the duration of trials aimed at establishing the benefit of new drugs. Progression-free survival (PFS) is commonly used in phase II and phase III trials. It has been evaluated as a surrogate endpoint for OS in several types of cancers (2–4). The ability to predict clinical benefits on OS from earlier benefits on PFS could be useful at all stages of clinical development. Here, we investigate the surrogacy of PFS for OS within the framework of the GASTRIC meta-analysis (5).

Trials were eligible if they were randomized, closed to accrual before the end of 2006, and collected individual patient data on PFS. To explore the correlation between the treatment effects at the trial level, we relied on the comparison between the experimental arms of the trials included in the meta-analysis with their corresponding control arms. We defined as experimental the treatment that contained the larger number of drugs (eg, triple combinations vs double combinations). In case of equal number of drugs, we defined as experimental the treatment that included the newer agent. When two experimental arms were tested in the same trial, we combined their data for the purposes of the analyses. All data were centrally checked for inconsistencies (6).

We used a meta-analytic validation approach (3,4,7). OS was defined as the time from randomization to death from any cause or to the last follow-up. PFS was the time to tumor progression or death from any cause or time to the last follow-up assessment. A detailed description of statistical methods used is provided in the Supplementary Material (available online). For external validation, we applied the identified relation to predict the hazard ratio (HR) for OS (HR_{OS}) from the hazard ratio for PFS (HR_{PFS}) in randomized trials published since 2000 for which we had not obtained the individual patient data. We extracted the summary statistics for both endpoints (8) and compared the predicted value of HR_{OS} to the one reported in the articles. To determine whether surrogacy also applied to other classes of agents, we extended the validation to three published trials of targeted agents (9–11).

Individual data were obtained on 4069 patients from 20 eligible randomized trials (12–30). The characteristics of the trials have been described elsewhere (5). Thirteen trials defined the progression using radiological criteria, whereas seven used both clinical and radiological assessments. Overall and at the trial level, the treatment effect on PFS (HR = 0.79; 95% confidence interval [CI] = 0.74 to 0.85) tended to be larger than on OS (HR = 0.85; 95% CI = 0.79 to 0.92) as shown on the forest plot of Supplementary Figures 1 and 2 (available online).

The individual-level association, as measured by the rank correlation coefficient, was 0.853 (95% CI = 0.852 to 0.854), indicating substantial correlation between PFS and OS for a given patient. The association at the trial level between log HR_{OS} and log HR_{PFS} was only moderate, with a coefficient of determination, R^2, adjusted for the estimation errors (31), of 0.61 (95% CI = 0.04 to 1.00). The large confidence interval reflects the uncertainty around this estimate. The linear regression model that relates the treatment effect on PFS and on OS adjusted for estimations errors was

$$\log(HR_{OS}) = 0.042 + 0.779 \times \log(HR_{PFS})$$

where the standard errors of the intercept and the slope were 0.79 and 0.295, respectively. This is shown as a straight line in Figure 1. The 95% prediction limits indicate the range of effect on OS that can be expected for a given effect on PFS. The moderate predictive accuracy at the trial level is reflected by the large interval width and a surrogate threshold effect of 0.56; hence, one should observe an HR_{PFS} less than 0.56 to predict, with 95% probability, an HR_{OS} less than 1.

Validation on independent literature data (9–11,32–39) is shown in Table 1 and Supplementary Figure 3 (available online). The larger the number of progressions, the more precise the prediction; however, precision is limited by the variability of the regression line. The observed HR_{OS} fell within the prediction interval in all trials, even in trials using humanized monoclonal antibodies [Trastuzumab (10), bevaxizumab (9), matuzumab (11)]. However, in the trial that concluded a statistically significant benefit of trastuzumab on OS (10), the effect on PFS was smaller than the surrogate threshold effect and therefore could not have been used to predict a statistically significant effect on OS.

This is the first study based on individual patient data to evaluate whether PFS is a reasonable surrogate endpoint to use for randomized trials in AGC. Our results show a high correlation of PFS and OS in individual patients but only a modest correlation ($R^2 = 0.61$) between treatment effects on PFS and OS. It is lower than that found in trials of 5-fluorouracil–based therapies for advanced colorectal cancer (4). The correlation was also lower than in the adjuvant setting (40).

Possible limitations that may explain the moderate correlation observed in our analysis include the numerous processes involved in the progression of stomach cancer (eg, local or distant metastasis, peritoneum involvement), the use of clinical and radiological assessments for progression, and the impact of our definition of investigational treatment related to the heterogeneity in chemotherapies considered here; variability in the investigated treatments and in the effects of the treatments is a condition to generalize any results to future trials. Last, patients included in more recent trials received second-line treatments, including crossover (30), which may have diluted the effect of first-line treatment on OS (2). Because not all trials reported the same information at baseline, we could not assess the surrogacy in clinically relevant subset analyses.

All in all, we would not conclude that PFS is an adequate surrogate for OS in AGC. No precise prediction of the effect of a treatment on OS can be reliably drawn from the effect estimated on PFS.

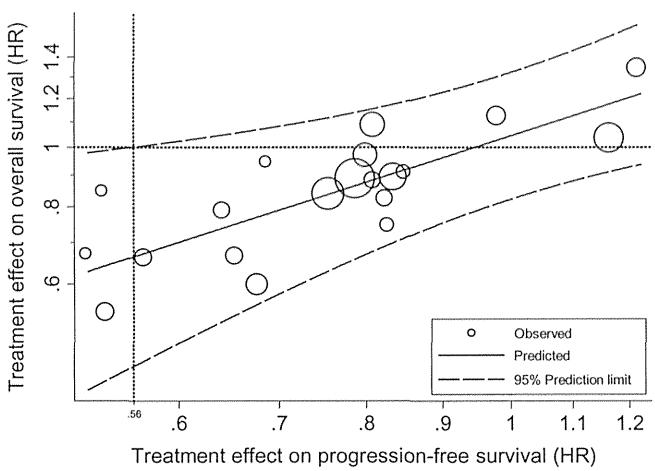

Figure 1. Trial-level association between treatment effects. Log scale was used for the x and y axes; the **horizontal line** (**circles**) corresponds to the hazard ratio (HR) on overall survival of 1, which indicates the absence of effect on the overall survival. At the crossing point, the vertical line corresponds to the minimum amount of effect on PFS that will predict a hazard ratio on OS below 1 with 95% probability. This indicates the surrogate threshold effect.

Table 1. Observed and predicted treatment effect on overall survival, based on the observed treatment effect on progression-free survival*

Trial label	Trial	Observed HR_{PFS} (95% CI)	Observed HR_{OS} (95% CI)	Predicted HR_{OS} (95% CI)
A	Jeung et al. (36)	0.63 (0.38 to 1.05)	0.56 (0.35 to 0.88)	0.73 (0.46 to 1.04)
B	AIO (33)	0.67 (0.43 to 1.04)	0.82 (0.47,1.45)	0.76 (0.53 to 1.07)
C	ToGA (10)	0.71 (0.59 to 0.85)	0.74 (0.60 to 0.91)	0.80 (0.58 to 1.09)
D	AVAGAST (9)	0.80 (0.68 to 0.93)	0.87 (0.73 to 1.03)	0.88 (0.76 to 1.14)
E	Kang et al. (35)	0.80 (0.63 to 1.03)	0.85 (0.64 to 1.13)	0.88 (0.76 to 1.14)
F	Park et al. (38)	0.86 (0.54 to 1.37)	0.96 (0.60 to 1.52)	0.93 (0.71 to 1.18)
G	REAL (a)† (34)	0.92 (0.80 to 1.04)	0.92 (0.80 to 1.10)	0.98 (0.77 to 1.22)
H	REAL (b) (34)	0.92 (0.81 to 1.05)	0.86 (0.80 to 0.99)	0.98 (0.77 to 1.22)
I	Ross et al. (39)	0.95 (0.80 to 1.08)	0.91 (0.76 to 1.04)	1.00 (0.79 to 1.29)
J	FLAGS (32)	0.99 (0.86 to 1.14)	0.92 (0.80 to 1.05)	1.03 (0.81 to 1.31)
K	Rao et al. (11)	1.13 (0.63 to 2.01)	1.02 (0.61 to 1.70)	1.14 (0.89 to 1.46)
L	Moehler et al. (37)	1.14 (0.59 to 2.21)	0.77 (0.51 to 1.17)	1.15 (0.90 to 1.48)

* HR = hazard ratio; PFS = progression-free survival; CI = confidence interval; OS = overall survival.

† This trial was designed as a factorial 2 × 2 plan to test two comparisons: a platinum comparison (a) and a fluoropyrimidine comparison (b).

References

1. Oba K, Paoletti X, Bang YJ, et al. Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis. *Eur J Cancer*. 2013;49(7):1565–1577.
2. Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst*. 2009;101(23):1642–1649.
3. Burzykowski T, Buyse M, Piccart-Gebhart MJ, et al. Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol*. 2008;26(12):1987–1992.
4. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol*. 2007;25(33):5218–5224.
5. GASTRIC. Role of chemotherapy for advanced/recurrent gastric cancer: an individual-patient-data meta-analysis. *Eur J Cancer*. 2013; 49(7):1565–1577
6. Stewart LA, Clarke MJ. Practical methodology of meta-analyses (overviews) using updated individual patient data. Cochrane Working Group. *Stat Med*. 1995;14(19):2057–2079.
7. Burzykowski T, Molenberghs G, Buyse M, Geys H. Validation of surrogate endpoints in multiple randomized clinical trials with failure time endpoints. *J R Stat Soc Series C Appl Stat*. 2001; 50(4): 405–422.
8. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med*. 1998;17(24):2815–2834.
9. Ohtsu A, Shah MA, Van Cutsem E, et al. Bevacizumab in combination with chemotherapy as first-line therapy in advanced gastric cancer: a randomized, double-blind, placebo-controlled phase III study. *J Clin Oncol*.; 2011; 29(30):3968–3976.
10. Bang YJ, Van Cutsem E, Feyereislova A, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet*. 2010;376(9742):687–697.
11. Rao S, Starling N, Cunningham D, et al. Matuzumab plus epirubicin, cisplatin and capecitabine (ECX) compared with epirubicin, cisplatin and capecitabine alone as first-line treatment in patients with advanced oesophago-gastric cancer: a randomised, multicentre open-label phase II study. *Ann Oncol*.; 2010; 21(11):2213–2219.
12. Douglass HO Jr, Lavin PT, Goudsmit A, Klaassen DJ, Paul AR. An Eastern Cooperative Oncology Group evaluation of combinations of methyl-CCNU, mitomycin C, adriamycin, and 5-fluorouracil in advanced measurable gastric cancer (EST 2277). *J Clin Oncol*. 1984;2(12):1372–1381.
13. Cullinan SA, Moertel CG, Fleming TR, et al. A comparison of three chemotherapeutic regimens in the treatment of advanced pancreatic and gastric carcinoma. Fluorouracil vs fluorouracil and doxorubicin vs fluorouracil, doxorubicin, and mitomycin. *JAMA*. 1985;253(14):2061–2067.
14. Wils JA, Klein HO, Wagener DJ, et al. Sequential high-dose methotrexate and fluorouracil combined with doxorubicin—a step ahead in the treatment of advanced gastric cancer: a trial of the European Organization for Research and Treatment of Cancer Gastrointestinal Tract Cooperative Group. *J Clin Oncol*. 1991;9(5):827–831.
15. Kyoto Research Group for Chemotherapy of Gastric Aancer. A randomized, comparative study of combination chemotherapies in advanced gastric cancer: 5-fluorouracil and cisplatin (FP) versus 5-fluorouracil, cisplatin, and 4'-epirubicin (FPEPIR). *Anticancer Res*. 1992;12(6B):1983–1988.
16. Cullinan SA, Moertel CG, Wieand HS, et al. Controlled evaluation of three drug combination regimens versus fluorouracil alone for the therapy of advanced gastric cancer. North Central Cancer Treatment Group. *J Clin Oncol*. 1994;12(2):412–416.
17. Glimelius B, Ekstrom K, Hoffman K, et al. Randomized comparison between chemotherapy plus best supportive care with best supportive care in advanced gastric cancer. *Ann Oncol*. 1997;8(2):163–168.
18. Vanhoefer U, Rougier P, Wilke H, et al. Final results of a randomized phase III trial of sequential high-dose methotrexate, fluorouracil, and doxorubicin versus etoposide, leucovorin, and fluorouracil versus infusional fluorouracil and cisplatin in advanced gastric cancer: a trial of the European Organization for Research and Treatment of Cancer Gastrointestinal Tract Cancer Cooperative Group. *J Clin Oncol*. 2000;18(14):2648–2657.
19. Kim TW. A prospective randomized phase III trial of 5-fluorouracil and cusplatin (FP) versus epirubicin, cisplatin, and 5-FU (ECF) in the treatment of patients with previously untreated advanced gastric cancer (AGC). *Eur J Cancer*. 2001; 47(3)S314.
20. Ohtsu A, Shimada Y, Shirao K, et al. Randomized phase III trial of fluorouracil alone versus fluorouracil plus cisplatin versus uracil and tegafur plus mitomycin in patients with unresectable, advanced gastric cancer: the Japan Clinical Oncology Group Study (JCOG9205). *J Clin Oncol*. 2003;21(1):54–59.
21. Koizumi W, Fukuyama Y, Fukuda T, et al. Randomized phase II study comparing mitomycin, cisplatin plus doxifluridine with cisplatin plus doxifluridine in advanced unresectable gastric cancer. *Anticancer Res*. 2004;24(4):2465–2470.
22. Pozzo C, Barone C, Szanto J, et al. Irinotecan in combination with 5-fluorouracil and folinic acid or with cisplatin in patients with advanced gastric or esophageal-gastric junction adenocarcinoma: results of a randomized phase II study. *Ann Oncol*. 2004;15(12):1773–1781.
23. Bouche O, Ychou M, Burtin P, et al. Adjuvant chemotherapy with 5-fluorouracil and cisplatin compared with surgery alone for gastric cancer: 7-year results of the FFCD randomized phase III trial (8801). *Ann Oncol*. 2005;16(9):1488–1497.
24. Ajani JA, Fodor MB, Tjulandin SA, et al. Phase II multi-institutional randomized trial of docetaxel plus cisplatin with or without fluorouracil in patients with untreated, advanced gastric, or gastroesophageal adenocarcinoma. *J Clin Oncol*. 2005;23(24):5660–5667.
25. Moehler M, Eimermacher A, Siebler J, et al. Randomised phase II evaluation of irinotecan plus high-dose 5-fluorouracil and leucovorin (ILF) vs 5-fluorouracil, leucovorin, and etoposide (ELF) in untreated metastatic gastric cancer. *Br J Cancer*. 2005;92(12):2122–2128.
26. Thuss-Patience PC, Kretzschmar A, Repp M, et al. Docetaxel and continuous-infusion fluorouracil versus epirubicin, cisplatin, and fluorouracil for advanced gastric adenocarcinoma: a randomized phase II study. *J Clin Oncol*. 2005;23(3):494–501.
27. Van Cutsem E, Moiseyenko VM, Tjulandin S, et al. Phase III study of docetaxel and cisplatin plus fluorouracil compared with cisplatin and fluorouracil as first-line therapy for advanced gastric cancer: a report of the V325 Study Group. *J Clin Oncol*. 2006;24(31):4991–4997.
28. Roth AD, Fazio N, Stupp R, et al. Docetaxel, cisplatin, and fluorouracil; docetaxel and cisplatin; and epirubicin, cisplatin, and fluorouracil as systemic treatment for advanced gastric carcinoma: a randomized phase II trial of the Swiss Group for Clinical Cancer Research. *J Clin Oncol*. 2007;25(22):3217–3223.
29. Dank M, Zaluski J, Barone C, et al. Randomized phase III study comparing irinotecan combined with 5-fluorouracil and folinic acid to cisplatin combined with 5-fluorouracil in chemotherapy naive patients with advanced adenocarcinoma of the stomach or esophagogastric junction. *Ann Oncol*. 2008;19(8):1450–1457.
30. Boku N, Yamamoto S, Fukuda H, et al. Fluorouracil versus combination of irinotecan plus cisplatin versus S-1 in metastatic gastric cancer: a randomised phase 3 study. *Lancet Oncol*. 2009;10(11):1063–1069.
31. Burzykowski T, Molenberghs G, Buyse M. *The Evaluation of Surrogate Endpoints*. New York: Springer; 2006.
32. Ajani JA, Rodriguez W, Bodoky G, et al. Multicenter phase III comparison of cisplatin/S-1 with cisplatin/infusional fluorouracil in advanced gastric or gastroesophageal adenocarcinoma study: the FLAGS trial. *J Clin Oncol*. 2010;28(9):1547–1553.
33. Al-Batran SE, Hartmann JT, Probst S, et al. Phase III trial in metastatic gastroesophageal adenocarcinoma with fluorouracil, leucovorin plus either oxaliplatin or cisplatin: a study of the Arbeitsgemeinschaft Internistische Onkologie. *J Clin Oncol*. 2008;26(9):1435–1442.
34. Cunningham D, Starling N, Rao S, et al. Capecitabine and oxaliplatin for advanced esophagogastric cancer. *N Engl J Med*. 2008;358(1):36–46.
35. Kang YK, Kang WK, Shin DB, et al. Capecitabine/cisplatin versus 5-fluorouracil/