



Short communication

Improved regional control and survival with “low Maruyama Index” surgery in gastric cancer: autopsy findings from the Dutch D1-D2 Trial

SCOTT A. HUNDAHL¹, KOEN C.M.J. PEETERS², E. KLEIN KRANENBARG², HENK HARTGRINK²,
and CORNELIS J.H. VAN DE VELDE²

¹Department of Surgery, University of California at Davis, U.S.A.; and VA Northern California Health Care System, Sacramento VA at Mather, SMAT-112, 10535 Hospital Way, Mather, CA 95655-1200, USA

²Department of Surgery, Leiden University Medical Center, Leiden, The Netherlands

Abstract

Based on more than 11 years of follow-up, autopsy-based analysis of recurrence in the Dutch D1-D2 Trial permits meaningful assessment of patterns of failure with respect to the Maruyama Index (MI). We previously reported that a low Maruyama Index was an independent predictor of both overall and disease-specific survival. Autopsy results are available for 441 deaths on study. Distant-only failure (15% vs 13%) was no different between the MI categories, but isolated “regional” failure (8% for MI < 5 group vs 21%) and “regional + distant” failure (19% for MI < 5 group vs 36%) occurred less frequently in the MI < 5 group ($P < 0.001$). We conclude that “low Maruyama Index” surgery enhances regional control and survival but does not alter the occurrence of isolated distant metastases unassociated with regional failure. Our results speak to the substantial survival value of local-regional control in this disease.

Key words Gastric cancer · Lymphadenectomy · Node · Survival · Pattern of failure · Computer · Maruyama Index · Autopsy

Background: the Maruyama program

An alternative means of guiding node dissection in a particular case of gastric cancer was pioneered by Kampschoer, Maruyama, and colleagues in 1989 [1]. A meticulously maintained database of 3843 cases from the National Cancer Center, Tokyo, was computerized and, through use of a simple search routine, cases could be matched on the basis of seven characteristics identifiable pre-operatively (age, sex, Borrmann morphologic type, presumed depth of primary tumor, gastric tumor location, maximum tumor diameter, and histologic

type). Based on recorded experience with actual patients, the likelihood of disease in 16 defined nodal stations around the stomach could be thus estimated. Performance of the so-called Maruyama program has been evaluated in German [2] and Italian populations [3] and found to be fairly accurate. For more than 15 years, the program has been used throughout the world as a tool to augment Japanese D-level guidelines with respect to lymph node dissection. From 1988 to 2000, diskettes of the Maruyama program were disseminated free of charge by the authors. In 2000, a Windows-based version, with expanded case volume and graphical presentation of output, was sold as part of an instructional CD-ROM produced by Springer Electronic Media [4].

Definition of “Maruyama Index of unresected disease” and early results

The Maruyama program has recently been used to assess the impact of varying degrees of lymphadenectomy in gastric cancer trials. Defining the “Maruyama Index of unresected disease” (MI) as the sum of Maruyama program predictions for regional lymph node stations 1–12 left *undissected* by the patient’s surgeon, one of the authors (SH) demonstrated that surgical undertreatment of nodal sites for patients enrolled in the Macdonald Chemoradiation Trial (SWOG 9008 / Intergroup 0116) in the United States, as determined by MI, significantly undermined survival. In this trial, by multivariate analysis MI < 5 was shown to be a strong independent predictor of both disease-free and overall survival [5].

Maruyama Index analysis of the Dutch D1-D2 Trial

In the Dutch D1-D2 Trial, pathology determinations revealed a noncompliance rate (i.e., nontreatment of

Offprint requests to: S.A. Hundahl

Presented at the 79th Annual Meeting of the Japanese Gastric Cancer Association, held in Nagoya, Japan, March 1–3, 2007
Received: April 28, 2007 / Accepted: May 3, 2007

node stations that should have been dissected) of 51% for patients who should have undergone a D-2 dissection [6]. Did such undertreatment influence survival in this trial? A blinded re-analysis of the trial by the authors indeed showed that patients with MI < 5 enjoyed superior survival, and MI was, again, found to be an *independent* predictor of survival [7]. Revised graphs for overall and disease-specific survival, by MI quartiles, are depicted in Fig. 1.

Because surgery is a local-regional treatment, better surgical treatment should be reflected in lower local-regional failure rates. Is low-MI surgery associated with alteration in the pattern of treatment failure and improved local-regional control? The very long follow-up (>11 years) in the Dutch Trial, as well as almost uniform autopsy at time of death for patients in this trial, invites such analysis.

Results: survival and autopsy analysis of patterns of failure

Figure 1 shows revised graphs with survival curves by MI quartile. The best survival is seen when MI is less than 5. With T, N, margin status (R0 or R1), and D-level as co-variables, MI was an *independent* predictor of both overall survival ($P = 0.016$, HR = 1.45, 95% CI 1.07–1.95), and disease-specific survival ($P = 0.010$, HR = 1.72, 95% CI 1.14–2.60). D-level was not a significant predictor [7].

Table 1 depicts the autopsy-based pattern of failure for patients who died. In this trial, the exact date of death is known but the exact date of failure is not. This

precludes separate time-to-failure analysis or failure-based Kaplan-Meier curves. However, detailed analysis of pattern of failure among patients who died is possible. We observed significant differences ($P < 0.001$ by the Kruskal-Wallis test). The proportion of both isolated local-regional failure and local-regional-plus-distant failure was far less for the MI < 5 group (see Table 1). For the MI < 5 cases versus others, the proportion of isolated, distant-only failure was about the same (15% vs 13%, respectively).

Discussion

Short follow-up can confound failure analysis. In the Dutch Trial, all living cases have now been followed for more than 11 years, and we believe that this is long enough to permit meaningful results. In other gastric cancer studies, relative survival curves plateau at 7 to 8 years after treatment [8]. Autopsy findings for Dutch

Table 1. Failure analysis by Maruyama Index (MI), as shown by autopsy

	MI < 5 n (%)	MI ≥ 5 n (%)
Died, no recurrence	44 (59)	110 (30)
Regional recurrence	6 (8)	78 (21)
Regional + distant	14 (19)	130 (36)
Distant only	11 (15)	48 (13)
Total	75 (100)	366 (100)

$P < 0.001$. There is a significant difference in the pattern of failure. See text for details

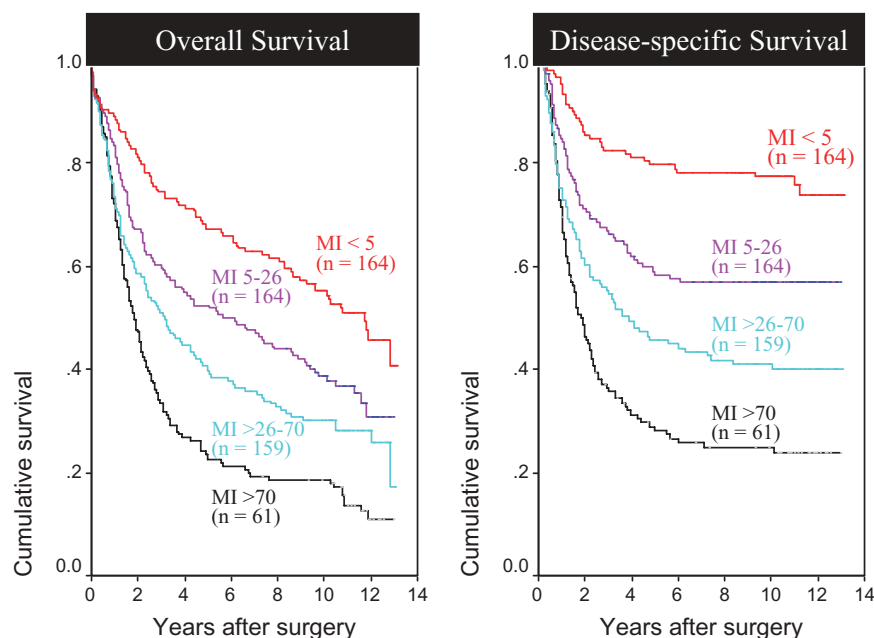


Fig. 1. Overall survival (left) and disease-specific survival (right), from date of surgery, by Maruyama Index (MI) quartiles (log-rank $P < 0.01$ for both graphs). Top line reflects the MI < 5 quartile

Trial cases clearly reveal that low MI cases appear to have less frequent local-regional failure. This is consistent with the therapeutic impact of a local-regional modality such as surgery. In contrast to patients in the Macdonald Chemoradiation Trial [5,9], in which half of the cases received adjuvant chemoradiation, all patients in this cohort were treated by surgery alone, and shorter time to failure can be expected.

The dose–response effect for MI versus survival, as depicted in Fig. 1, and the status of MI as an independent predictor of survival encourage us to view MI as a quantitative measure for the adequacy of lymphadenectomy in a given case of gastric cancer. This conclusion is reinforced by autopsy data indicating that low MI surgery “works” by decreasing regional failure. We believe that MI should henceforth be calculated for every patient entered into an adjuvant therapy trial.

Can we meaningfully enhance the survival of patients with gastric cancer by customizing their surgery through use of the Maruyama program? It seems reasonable to assume that this is true. However, despite compelling observations from re-analyzed data from two trials, we must caution that this has yet to be shown in a prospective, randomized trial dedicated to the topic. As seen in previous reports, surgical compliance with the assigned surgical procedure is an issue, as is contamination [6,10]. For a trial involving a surgical procedure, in which survival is an endpoint, pre-operative screening, pre-operative preparation, post-procedure reconstruction, the management of postoperative complications, and other factors that can impact on in-hospital mortality must be controlled, or at least minimized. The difficulty of standardizing such things and ensuring quality control makes conduct of multi-center surgical trials particularly challenging [11]. Thus, successful prospective application of Maruyama program information in a multi-center trial may be more difficult in practice than suggested by this retrospective analysis. Currently, prospective validation of the concept is incorporated in the Dutch CRITICS trial, which has already opened for accrual.

Conclusion

Failure analysis for gastric cancer patients in the Dutch D1-D2 Trial supports our previous multivariate analysis showing favorable impact for low Maruyama Index

surgery. The Maruyama Index may be viewed as a quantitative measure for the adequacy of lymphadenectomy. Both local-regional control and survival appear enhanced when this index is low.

Acknowledgments This work was supported by grants from the Dutch Health Insurance Funds Council, The Netherlands Cancer Foundation, the EORTC, and the U.C. Davis Department of Surgery Pathfinder Fund.

References

1. Kampschoer GH, Maruyama K, van de Velde CJ, Sasako M, Kinoshita T, Okabayashi K. Computer analysis in making preoperative decisions: a rational approach to lymph node dissection in gastric cancer patients. *Br J Surg* 1989;76(9):905–8.
2. Bollschweiler E, Boettcher K, Hoelscher AH, Sasako M, Kinoshita T, Maruyama K, et al. Preoperative assessment of lymph node metastases in patients with gastric cancer: evaluation of the Maruyama computer program. *Br J Surg* 1992;79(2):156–60.
3. Guadagni S, de Manzoni G, Catarci M, Valenti M, Amicucci G, De Bernardinis G, et al. Evaluation of the Maruyama computer program accuracy for preoperative estimation of lymph node metastases from gastric cancer. *World J Surg* 2000;24(12):1550–8.
4. Siewert JR, Kelsen D, Maruyama K, Feussner H, Omote K, Etter M, et al. Gastric cancer diagnosis and treatment — an interactive training program. 1st ed. Heidelberg: Springer Electronic Media; 2000.
5. Hundahl SA, Macdonald JS, Benedetti J, Fitzsimmons T. Surgical treatment variation in a prospective, randomized trial of chemoradiotherapy in gastric cancer: the effect of undertreatment. *Ann Surg Oncol* 2002;9(3):278–86.
6. Bonenkamp JJ, Hermans J, Sasako M, van de Velde CJ. Extended lymph-node dissection for gastric cancer. Dutch Gastric Cancer Group. *N Engl J Med* 1999;340(12):908–14.
7. Peeters KCMJ, Hundahl SA, Kranenbarg EK, Hartgrink H, van de Velde CJH. “Low-Maruyama-Index” surgery for gastric cancer — a blinded re-analysis of the Dutch D1-D2 Trial. *World J Surg* 2005;29:1576–84.
8. Hundahl SA, Phillips JL, Menck HR. The National Cancer Data Base Report on poor survival of U.S. gastric carcinoma patients treated with gastrectomy: 5th ed. American Joint Committee on Cancer Staging, proximal disease, and the “different disease” hypothesis. *Cancer* 2000;88(4):921–32.
9. Macdonald JS, Smalley SR, Benedetti J, Hundahl SA, Estes NC, Stemmermann GN, et al. Chemoradiotherapy after surgery compared with surgery alone for adenocarcinoma of the stomach or gastroesophageal junction. *N Engl J Med* 2001;345(10):725–30.
10. Bunt AM, Hermans J, Boon MC, van de Velde CJ, Sasako M, Fleuren GJ, et al. Evaluation of the extent of lymphadenectomy in a randomized trial of Western- versus Japanese-type surgery in gastric cancer. *J Clin Oncol* 1994;12(2):417–22.
11. Hundahl SA. Surgical quality control in gastric cancer trials. *Surg Oncol Clin N Am* 2002;11(2):445–58.