the Rth analysis $(\hat{\delta}_{1R}, \hat{\delta}_{2R})$ or the assumed mean differences during trial planning $(\tilde{\delta}_1, \tilde{\delta}_2)$. We consider the conditional power based on $(\delta_1^*, \delta_2^*) = (\hat{\delta}_{1R}, \hat{\delta}_{2R})$, which allows evaluation of behavior of power independent of $(\tilde{\delta}_1, \tilde{\delta}_2)$.

When recalculating the sample size, three options are possible: (i) only allowing an increase in the sample size, (ii) only allowing a decrease in the sample size, and (iii) allowing an increase or decrease in sample size. For all the cases, we assign Z'_{km} and n'_m instead of Z_{km} and n_m in the conditional powers (5) and (6) for the conditional power with sample size recalculation. Consider the rule for determining the recalculated sample size n'_L, when the sample size may be increased only, which is

$$n'_L = \begin{cases} n_L, & \text{if } CP \geqslant 1-\beta \text{ or } \min(\hat{\delta}_{1R}, \hat{\delta}_{2R}) \leqslant 0, \\ \min\left(n''_L, \lambda n_L\right), & \text{otherwise,} \end{cases}$$

where n''_L is the smallest integer $n'_L (> n_R)$, where the conditional power achieves the desired power $1-\beta$. When the sample size may be decreased only, the recalculated sample size n'_L is

$$n'_L = \begin{cases} n''_L, & \text{if } CP > 1-\beta, \\ n_L, & \text{otherwise.} \end{cases}$$

When the sample size may be increased or decreased, the recalculated sample size n'_L is

$$n'_L = \begin{cases} n''_L, & \text{if } CP > 1-\beta, \\ n_L, & \text{if } CP = 1-\beta \text{ or } \min\left(\hat{\delta}_{1R}, \hat{\delta}_{2R}\right) \leqslant 0, \\ \min\left(n''_L, \lambda n_L\right), & \text{otherwise.} \end{cases}$$

4.2. *Simulation study*

A simulation study was performed to evaluate the impact of sample size recalculation based on DF-1 and DF-2 on the power and Type I error rate. We consider group-sequential designs with a single interim, that is, one interim and one final analyses, and with multiple interims, that is, three interims and one final analyses. In addition, we discuss the three options of (i) only decreasing the sample size, (ii) only increasing the sample size, and (iii) increasing or decreasing the sample size, based upon the observed intervention's effect. The planned MSS per intervention group is calculated to detect the joint difference for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where $(\tilde{\delta}_1, \tilde{\delta}_2) = (0.2, 0.2)$, $\sigma_1^2 = \sigma_2^2 = 1^2$ and the correlation is assumed to be known correlation at the design stage, that is, $\rho = 0.0, 0.3, 0.5$, and 0.8. For the evaluation of the Type I error rate, the two pairs of the mean differences $(\delta_1, \delta_2) = (0.0, 0.0)$ and (0.0, 0.2) are considered under H_0. For the designs with a single interim, the timing of the interim analysis for sample size recalculation is evaluated at 0.25, 0.50, and 0.75 of information time. For designs with multiple interims, one sample size recalculation is considered, and the timing is evaluated at the first, second, and third of interim analysis. The critical values are determined by the OF boundary for both endpoints with the LD alpha-spending method, with equal information space. The upper limit of the recalculated sample size is set to $n'_2 = \lambda n_2$ with $\lambda = 1.5$. The number of replications for the simulation is set to 1,000,000 for the evaluation of the Type I error rate and 100,000 replications for the power. These number of replications for the simulation was determined based on the precision, where a sample size of 1,000,000 provides a two-sided 95% confidence interval with a width equal to 0.001 when the proportion is 0.025, and a total number of replications of 100,000 provides a two-sided 95% confidence interval with a width equal to 0.005 when the proportion is 0.80.

Suppose that the sample size recalculation is based on the interim estimates of (δ_1, δ_2). Note that the value of correlation assumed at the design stage is retained for the sample size recalculation, that is, without updating based on observed correlation at the interim as the correlation is a nuisance parameter in hypothesis testing. All results are summarized in Tables S1–S4 in the Supporting information. As there are no significant differences between DF-1 and DF-2 with respect to the Type I error rates and empirical powers, we limit the discussion to the behavior of the Type I error rates and power for DF-1.

Figure 3 illustrates how the Type I error rates and powers behave as a function of the correlation, the timing of the interim analysis for sample size recalculation, and the sample size recalculation options for DF-1 in the single-interim case. In all three recalculation options, the Type I error rates increase as the correlation increases, but they do not exceed the targeted 2.5%. There is no practical difference in

Copyright © 2014 John Wiley & Sons, Ltd.

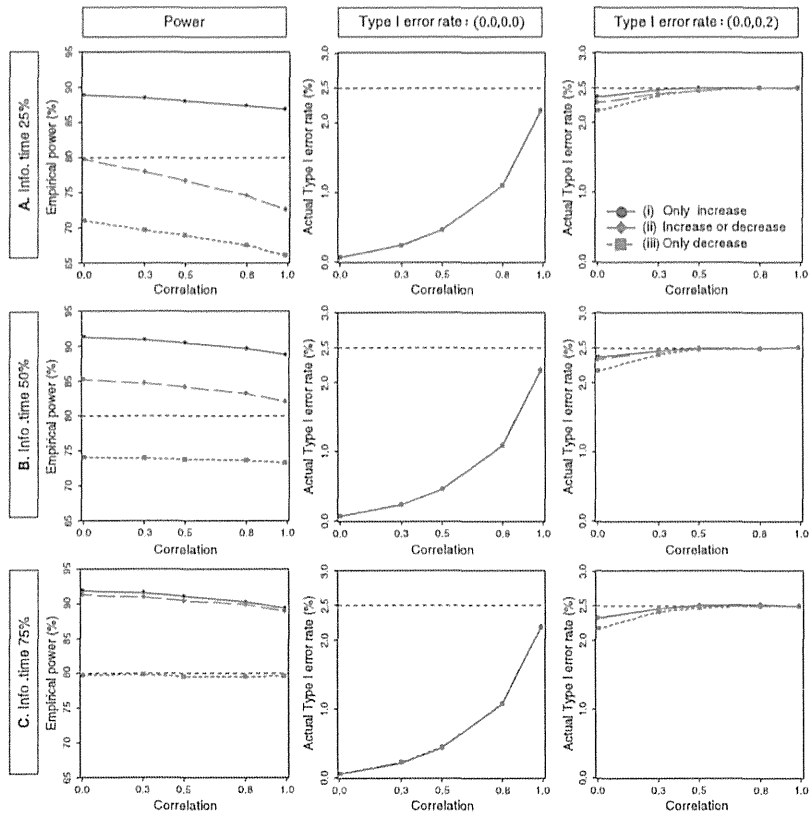

Figure 3. Behavior of the power and Type I error rate as a function of the correlation with sample size recalculation in two-stage group-sequential designs, where the information times of 0.25, 0.50, and 0.75 were selected as the timing of the sample size recalculation. The planned MSS per intervention group is calculated to detect the joint difference for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where one interim and one final analysis are to be performed. The critical values are determined by the OF boundary for both endpoints, with the LD alpha-spending method. The upper limit of recalculation sample size is $n_2' = \lambda n_2$ with $\lambda = 1.5$. The number of replications for simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power (DF-1).

the behavior of the Type I error rates depending on the timing of the interim analysis for sample size recalculation. On the other hand, for the behavior of the power, when only allowing an increase in the sample size, the empirical powers are higher than the desired power of 80% in all of the three timings of sample size recalculation, although the power is slightly decreased with higher correlation. When allowing an increase or a decrease in the sample size, if the timing for sample size recalculation is at 25% information time, then the empirical power is lower than the desired power of 80%, especially with higher correlation. However, if the timing for sample size recalculation is 50% or 75%, then the empirical powers are higher than in all three timings of sample size recalculation. When only allowing a decrease in the sample size, if the timing for the sample size recalculation is at 25% or 50% information time, then the empirical powers are always lower than the desired power, especially with higher correlation. If the timing for sample size recalculation is at 75% information time, then the empirical power is almost achieved at the desired power of 80%.

Figure 4 illustrates how the Type I error rates and powers behave as a function of the correlation, the timing of the interim analysis for sample size recalculation, and the sample size recalculation options for DF-1, in the multiple-interim case. The results are similar to those in the single-interim case; when only allowing an increase in the sample size, compared with the desired power of 80%, the empirical powers are improved in all of the three timings for the sample size recalculation, but the empirical power is much lower than the desired power if the sample size recalculation is conducted early in the study, especially when allowing a decrease in the sample size.

Copyright © 2014 John Wiley & Sons, Ltd.

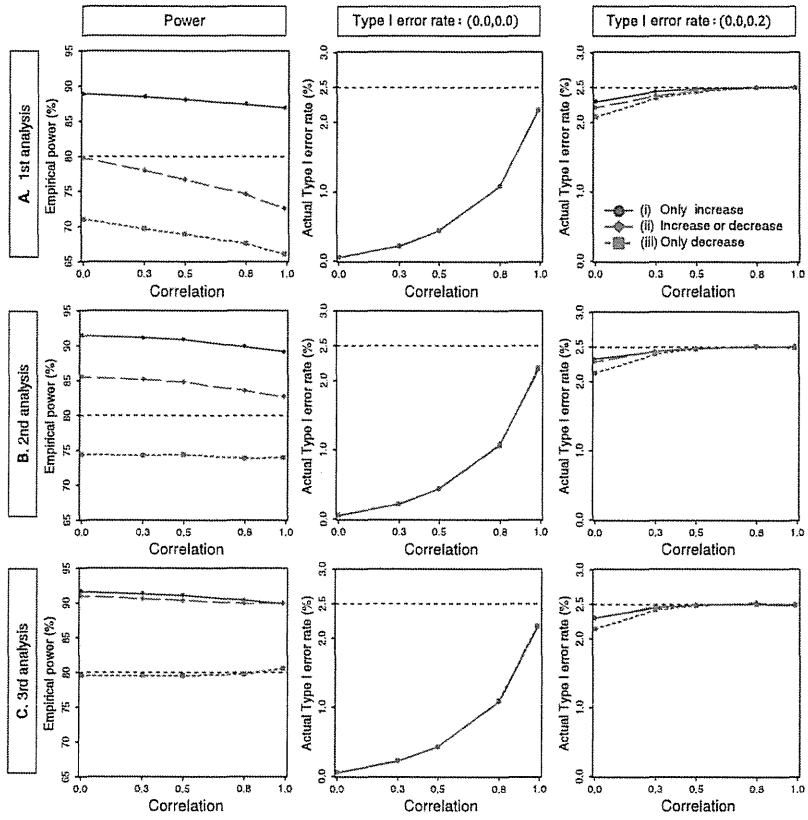

Figure 4. Behavior of the power and Type I error rate as a function of the correlation with sample size recalculation in four-stage group-sequential designs, where the first, second, and third interim points were selected as the timing of the sample size recalculation. The planned MSS per intervention group is calculated to detect the joint difference for two endpoints with the overall power of 80% at the one-sided significance level of 2.5%, where three interims and one final analysis are to be performed. The critical values are determined by the OF boundary for both endpoints, with the LD alpha-spending method. The upper limit of recalculation sample size is $n_2' = \lambda n_2$ with $\lambda = 1.5$. The number of replications for simulation is set to 1,000,000 for evaluation of the Type I error rate and 100,000 replications for the power (DF-1).

These results suggest incorporating the uncertainty of the estimates at the interim into the sample size recalculation is important. The power is much lower than desired power if the sample size recalculation is conducted early in the study, especially when allowing for a decrease in the sample size.

5. Summary and discussion

The determination of sample size and the evaluation of power are fundamental and critical elements in the design of a clinical trial. If a sample size is too small, then important effects may not be detected, while a sample size that is too large is wasteful of resources and unethically puts more participants at risk than necessary. Recently, many clinical trials are designed with more than one endpoint considered as co-primary. As with trials involving a single primary endpoint, designing such trials to include interim analyses (i.e., with repeated testing) may provide efficiencies by detecting trends prior to planned completion of the trial. It may also be prudent to evaluate design assumptions at the interim and potentially make design adjustments (i.e., sample size recalculation) if design assumptions were dramatically inaccurate. However, such design complexities create challenges in the evaluation of power and the calculation of sample size during trial design.

We discuss group-sequential designs with co-primary endpoints. We derive the power and sample size methods under two decision-making frameworks: (i) designing the trial to detect the test intervention's superiority for the two endpoints simultaneously (i.e., at the same interim timepoint of the trial) (DF-1) and (ii) designing the trial to detect superiority for the two endpoints at any interim timepoint (i.e., not

Copyright © 2014 John Wiley & Sons, Ltd.

necessarily simultaneously) (DF-2). The former is simpler while the latter is more flexible and may be useful when the endpoint is very invasive or expensive, as it allows for stopping the measurement of any endpoint upon which superiority has been demonstrated. We evaluate the behavior of sample size with varying design elements and provide an example to illustrate the methods. We also discuss sample size recalculation using CHW statistics and evaluate the impact on the power and Type I error rate. Although DF-2 will provide a slightly smaller sample size than DF-1, there is modest difference between the two. However, if the endpoint is very invasive and thus stopping measurement may be ethically desirable, there is a benefit of using DF-2 as DF-2 offers the option of stopping measurement of an endpoint for which superiority has been demonstrated. However, stopping measurement on one endpoint could also create operational challenges in study conduct and patient monitoring. The timing of the sample size recalculation should also be carefully considered as the power does not reach desired levels if the sample size recalculation is carried out early in the study when considering a decrease in the sample size.

There are other practical issues and extensions to consider when designing a group-sequential clinical trial with co-primary endpoints. They include the following: how the value of correlation should be selected at the planning and interim, evaluating futility or efficacy and futility simultaneously, other endpoint scales, and other inferential goals. We discuss each of these issues.

There are two important questions regarding the choice of the correlation in sample size calculations. One is whether the observed correlation from external or pilot data should be utilized or whether correlation is assumed to be zero. The other is whether the sample size should be recalculated based on the observed correlation at the interim. Incorporating the observed correlation at the planning or interim may affect the Type I error rate and power. Our experience suggests that when standardized effect sizes are unequal between the endpoints, the power is not improved with higher correlation. With unequal standardized effect sizes, incorporating the correlation into the sample size calculation at planning or interim may have no advantage [25, 26]. Further investigation will be required to assess how the choice of the correlation impacts the operation characteristics of the design.

Because the main objective of the paper is to provide the fundamental foundation in group-sequential designs for co-primary endpoints, our discussion is restricted to a superiority clinical trial comparing two interventions based on two continuous endpoints. The study design allows for early stopping when larger intervention differences are observed, that is, rejecting a null hypothesis only. However, this work provides a foundation for designing clinical trials with other design features. In addition to this fundamental situation, the method discussed here can be straightforwardly extended to other situations such as evaluating futility (rejecting the alternative hypothesis) or evaluating both efficacy and futility.

Time-to-event outcomes are common in oncology, cardiovascular, and infectious disease clinical trials. The method for continuous endpoints described in the paper may not be directly extended to time-to-event endpoints. When considering a trial with two time-to-event outcomes as co-primary with a plan for using the logrank test to compare two interventions in a group-sequential design, information for the two endpoints may accumulate at different rates. This creates challenges when designing trials, that is, the amount of information for the endpoints may be different at any particular interim timepoint of the trial. Further investigation is required to assess this issue.

Although our primary interest is *co-primary* endpoints, these results provide a fundamental foundation to other inferential goals, for example, designing a trial to detect an effect on *at least one* endpoint. Many authors have proposed methods for the *at least one* endpoint goal in fixed sample size designs, for example, a weighted Bonferroni procedure, the prospective alpha allocation scheme method, the adaptive alpha allocation approach, the Bonferroni-type parametric procedure, and the fallback-type parametric procedure (e.g., see [4, 27, 28]). In addition, several authors have discussed an extension of methods to the group-sequential designs with an inferential goal of *at least one* endpoint [29–32]. For example, Tang and Geller [30] discuss a method based on closed testing procedures, and Tamhane *et al.* [31, 32] discuss sample size methods in two-stage group-sequential designs based on the gatekeeping procedures with hierarchically ordered multiple endpoints.

Appendix

A.1. Power calculation

The power (1) for DF-1 can be calculated by partitioning the set in (1) into mutually exclusive subsets and taking the sum of their probabilities as follows:

Copyright © 2014 John Wiley & Sons, Ltd.

$$1-\beta = \Pr\left[\bigcup_{l=1}^{L}\{A_{1l}\cap A_{2l}\}\,|\mathrm{H}_1\right] = \Pr\left[A_{11}\cap A_{21}\,|\mathrm{H}_1\right] + \sum_{l=2}^{L}\Pr\left[\bigcap_{l'=1}^{l-1}\{\bar{A}_{1l'}\cup\bar{A}_{2l'}\}\cap\{A_{1l}\cap A_{2l}\}\middle|\,\mathrm{H}_1\right], \tag{A1}$$

where $A_{kl} = \{Z_{kl} > c_{kl}\}$ and $\bar{A}_{kl} = \{Z_{kl} \leqslant c_{kl}\}(k = 1,2;\ l = 1,\ldots,L)$. The probability of $\{\bar{A}_{1l'}\cup\bar{A}_{2l'}\}$ can be written as $\Pr[\bar{A}_{1l'}\cup\bar{A}_{2l'}] = \Pr[\tilde{A}^1_{l'}] + \Pr[\tilde{A}^2_{l'}] + \Pr[\tilde{A}^3_{l'}]$, where $\tilde{A}^1_{l'} = \{\bar{A}_{1l'}\cap A_{2l'}\}$, $\tilde{A}^2_{l'} = \{A_{1l'}\cap\bar{A}_{2l'}\}$ and $\tilde{A}^3_{l'} = \{\bar{A}_{1l'}\cap\bar{A}_{2l'}\}(l' = 1,\ldots,L-1)$. Similarly, the probability of the union of $\{\bar{A}_{1l'}\cup\bar{A}_{2l'}\}$ can be written by the sum of the probabilities of the unions composed of $\tilde{A}^1_{l'}$, $\tilde{A}^2_{l'}$ and $\tilde{A}^3_{l'}$. Then, the second term of the right-hand side in (A.1) can be rewritten as

$$\sum_{l=2}^{L}\Pr\left[\bigcap_{l'=1}^{l-1}\{\bar{A}_{1l'}\cup\bar{A}_{2l'}\}\cap\{A_{1l}\cap A_{2l}\}\middle|\,\mathrm{H}_1\right] = \sum_{l=2}^{L}\left(\sum_{h_1=1}^{3}\cdots\sum_{h_{l-1}=1}^{3}\Pr\left[\left\{\bigcap_{l'=1}^{l-1}\tilde{A}^{h_{l'}}_{l'}\right\}\cap\{A_{1l}\cap A_{2l}\}\middle|\,\mathrm{H}_1\right]\right).$$

The probability of $\tilde{A}^1_{l'}$ is calculated by a bivariate normal integral as follows:

$$\Pr\left[\tilde{A}^1_{l'}\right] = \int_{-\infty}^{c_{1l'}}\int_{c_{2l'}}^{\infty} f_2(z_{1l'}, z_{2l'})\,dz_{2l'}dz_{1l'},$$

where $f_2(z_{1l'}, z_{2l'})$ is the density function of the joint distribution of $(Z_{1l'}, Z_{2l'})$ with the means and the covariance matrix given in Section 2.1. The probabilities of $\tilde{A}^2_{l'}$, $\tilde{A}^3_{l'}$ and $\{A_{1l'}\cap A_{2l'}\}$ are calculated similarly. Then, the probability of the union composed of $\tilde{A}^1_{l'}$, $\tilde{A}^2_{l'}$, $\tilde{A}^3_{l'}$ and $\{A_{1l'}\cap A_{2l'}\}$ is calculated by a multivariate normal integral and the power is the sum of $(3^L-1)/2$ multivariate normal integrals. For details of the computation related to multivariate normal, please see [33].

For illustration, we provide the case of $L = 2$ and $r = r_1 = r_2$. In this case, the power can be rewritten as

$$\begin{aligned}1-\beta =& \Pr\left[A_{11}\cap A_{21}\,|\mathrm{H}_1\right] + \sum_{h_1=1}^{3}\Pr\left[\tilde{A}^{h_1}_1\cap\{A_{12}\cap A_{22}\}\,|\mathrm{H}_1\right]\\ =& \int_{c_{11}}^{\infty}\int_{c_{21}}^{\infty} f_2(z_{11}, z_{21})dz_{21}dz_{11} + \int_{-\infty}^{c_{11}}\int_{c_{21}}^{\infty}\int_{c_{12}}^{\infty}\int_{c_{22}}^{\infty} f_4(z_{11}, z_{21}, z_{12}, z_{22})\,dz_{22}dz_{12}dz_{21}dz_{11}\\ &+ \int_{c_{11}}^{\infty}\int_{-\infty}^{c_{21}}\int_{c_{12}}^{\infty}\int_{c_{22}}^{\infty} f_4(z_{11}, z_{21}, z_{12}, z_{22})dz_{22}dz_{12}dz_{21}dz_{11}\\ &+ \int_{-\infty}^{c_{11}}\int_{-\infty}^{c_{21}}\int_{c_{12}}^{\infty}\int_{c_{22}}^{\infty} f_4(z_{11}, z_{21}, z_{12}, z_{22})\,dz_{22}dz_{12}dz_{21}dz_{11},\end{aligned}$$

where $f_2(z_{11}, z_{21})$ is the density function of the bivariate normal distribution of $\boldsymbol{Z}_2 = (Z_{11}, Z_{21})^{\mathrm{T}}$, which is given by

$$f_2(\boldsymbol{Z}_2) = \frac{1}{2\pi\,|\boldsymbol{\Sigma}_2|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{Z}_2-\boldsymbol{\mu}_2)^{\mathrm{T}}\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{Z}_2-\boldsymbol{\mu}_2)\right],\ -\infty < z_{11}, z_{21} < \infty$$

with mean vector $\boldsymbol{\mu}_2 = \sqrt{rn_1/(1+r)}(\delta_1/\sigma_1, \delta_2/\sigma_2)^{\mathrm{T}}$ and correlation matrix

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 1^2 & (r\rho_{\mathrm{T}}+\rho_{\mathrm{C}})/(1+r)\\ (r\rho_{\mathrm{T}}+\rho_{\mathrm{C}})/(1+r) & 1^2\end{pmatrix}.$$

and $f_4(z_{11}, z_{21}, z_{12}, z_{22})$ is the density function of the tetra-variate normal distribution of $\boldsymbol{Z}_4 = (Z_{11}, Z_{21}, Z_{12}, Z_{22})^{\mathrm{T}}$ given by

$$f_4(\boldsymbol{Z}_4) = \frac{1}{(2\pi)^2\,|\boldsymbol{\Sigma}_4|^{1/2}}\exp\left[-\frac{1}{2}(\boldsymbol{Z}_4-\boldsymbol{\mu}_4)^{\mathrm{T}}\boldsymbol{\Sigma}_4^{-1}(\boldsymbol{Z}_4-\boldsymbol{\mu}_4)\right],\ -\infty < z_{11}, z_{21}, z_{12}, z_{22} < \infty$$

Copyright © 2014 John Wiley & Sons, Ltd.

with mean vector $\mu_4 = \sqrt{(1+r)/r}\left(\sqrt{n_1}\delta_1/\sigma_1, \sqrt{n_1}\delta_2/\sigma_2, \sqrt{n_2}\delta_1/\sigma_1, \sqrt{n_2}\delta_2/\sigma_2\right)^{\mathrm{T}}$ and correlation matrix

$$\Sigma_4 = \begin{pmatrix} \Sigma_2 & \sqrt{n_1/n_2}\Sigma_2 \\ \sqrt{n_1/n_2}\Sigma_2 & \Sigma_2 \end{pmatrix},$$

where Σ_4 is positive definite matrix under $|\rho_{\mathrm{T}}|, |\rho_{\mathrm{C}}| < 1$ and $n_1 \neq n_2$ as $|\Sigma_4| = |\Sigma_2|^2(1-n_1/n_2)^2$.

The power (2) for DF-2 can be calculated from two L-variate normal integrals and a $2L$-variate normal integral.

$$\begin{aligned} 1-\beta &= \Pr\left[\left\{\bigcup_{l=1}^{L} A_{1l}\right\} \cap \left\{\bigcup_{l=1}^{L} A_{2l}\right\} \middle| \mathrm{H}_1\right] \\ &= 1 - \left(\Pr\left[\bigcap_{l=1}^{L} \bar{A}_{1l} \middle| \mathrm{H}_1\right] + \Pr\left[\bigcap_{l=1}^{L} \bar{A}_{2l} \middle| \mathrm{H}_1\right] - \Pr\left[\bigcap_{l=1}^{L} \{\bar{A}_{1l} \cap \bar{A}_{2l}\} \middle| \mathrm{H}_1\right]\right). \end{aligned}$$

The power can be calculated similarly as discussed in the power (1) for DF-1.

A.2. ASN calculation

The ASN (3) for DF-1 can be calculated by the sum of multivariate normal integrals

$$\begin{aligned} \mathrm{ASN} &= n_L\left(1 + \sum_{l=1}^{L-1} \Pr\left[\{\bar{A}_{11} \cup \bar{A}_{21}\} \cap \cdots \cap \{\bar{A}_{1l} \cup \bar{A}_{2l}\}\right]\right) \Big/ L \\ &= n_L\left\{1 + \sum_{l=1}^{L-1}\left(\sum_{h_1=1}^{3} \cdots \sum_{h_l=1}^{3} \Pr\left[\bigcap_{l'=1}^{l} \bar{A}_{l'}^{h_{l'}}\right]\right)\right\} \Big/ L. \end{aligned}$$

Similarly, the ASN (4) for DF-2 can be calculated by

$$\begin{aligned} \mathrm{ASN} &= n_L\left(1 + \sum_{l=1}^{L-1} \Pr\left[\{\bar{A}_{11} \cap \cdots \cap \bar{A}_{1l}\} \cup \{\bar{A}_{21} \cap \cdots \cap \bar{A}_{2l}\}\right]\right) \Big/ L \\ &= n_L\left\{1 + \sum_{l=1}^{L-1}\left(\Pr\left[\bigcap_{l'=1}^{l} \bar{A}_{1l'}\right] + \Pr\left[\bigcap_{l'=1}^{l} \bar{A}_{2l'}\right] - \Pr\left[\bigcap_{l'=1}^{l} \{\bar{A}_{1l'} \cap \bar{A}_{2l'}\}\right]\right)\right\} \Big/ L. \end{aligned}$$

A.3. Conditional power

The conditional power (5) for DF-1 is described by

$$\begin{aligned} CP &= \Pr\left[\bigcup_{m=R+1}^{L} \{A_{1m} \cap A_{2m}\} \middle| a_{1R}, a_{2R}\right] \\ &= \Pr\left[A_{1,R+1} \cap A_{2,R+1} \middle| a_{1R}, a_{2R}\right] \\ &\quad + \sum_{m=R+2}^{L} \Pr\left[\bigcap_{m'=R+1}^{m-1} \{\bar{A}_{1m'} \cup \bar{A}_{2m'}\} \cap \{A_{1m} \cap A_{2m}\} \middle| a_{1R}, a_{2R}\right], \end{aligned} \tag{A2}$$

if $Z_{1l} \leqslant c_{1l}$ or $Z_{2l} \leqslant c_{2l}$ for all $l = 1, \ldots, R$, where $A_{km} = \{Z_{km} > c_{km}\}$, $\bar{A}_{km} = \{Z_{km} \leqslant c_{km}\}$ $(k = 1, 2;\ m = R+1, \ldots, L)$ and (a_{1R}, a_{2R}) is a given observed value of (Z_{1R}, Z_{2R}). The second term of the right-hand side in (A2) can be calculated in a similar way to that for the power calculation (Appendix A.1). The conditional distribution of $(Z_{1,R+1}, Z_{2,R+1}, \ldots, Z_{1L}, Z_{2L} | a_{1R}, a_{2R})$ is a multivariate normal with their means $\mathrm{E}[Z_{km} | a_{1R}, a_{2R}] = \sqrt{n_m/2}\delta_k + \sqrt{n_R/n_m}(a_{kR} - \sqrt{n_R/2}\delta_k)$ and covariance given by $\mathrm{cov}[Z_{km}, Z_{k'm'} | a_{1R}, a_{2R}] = (n_{m'} - n_R)/\sqrt{n_m n_{m'}}$ if $k = k'$;

Copyright © 2014 John Wiley & Sons, Ltd.

$(n_{m'} - n_R)\rho / \sqrt{n_m n_{m'}}$ if $k \neq k'$, where $m' \leqslant m = R+1, \ldots, L$. For DF-2, the conditional power (6) can be described as

$$CP = \begin{cases} \Pr\left[\bigcup_{m=R+1}^{L} A_{1m} \,|a_{1R}, a_{2l'}\right] = 1 - \Pr\left[\bigcap_{m=R+1}^{L} \bar{A}_{1m} \,|a_{1R}, a_{2l'}\right] \\ \quad \text{if } Z_{1l} \leqslant c_{1l} \text{ for all } l = 1, \ldots, R \text{ and } Z_{2l'} > c_{2l'} \text{ for some } l' = 1, \ldots, R, \\ \Pr\left[\bigcup_{m=R+1}^{L} A_{2m} \,|a_{2R}, a_{1l'}\right] = 1 - \Pr\left[\bigcap_{m=R+1}^{L} \bar{A}_{2m} \,|a_{2R}, a_{1l'}\right] \\ \quad \text{if } Z_{2l} \leqslant c_{2l} \text{ for all } l = 1, \ldots, R \text{ and } Z_{1l'} > c_{1l'} \text{ for some } l' = 1, \ldots, R, \\ \Pr\left[\left\{\bigcup_{m=R+1}^{L} A_{1m}\right\} \cap \left\{\bigcup_{m=R+1}^{L} A_{2m}\right\} \,|a_{1R}, a_{2R}\right] \\ \quad = 1 - \Pr\left[\bigcap_{m=R+1}^{L} \bar{A}_{1m} \,|a_{1R}, a_{2R}\right] - \Pr\left[\bigcap_{m=R+1}^{L} \bar{A}_{2m} \,|a_{1R}, a_{2R}\right] \\ \qquad + \Pr\left[\bigcap_{m=R+1}^{L} \{\bar{A}_{1m} \cap \bar{A}_{2m}\} \,|a_{1R}, a_{2R}\right] \\ \quad \text{if } Z_{1l} \leqslant c_{1l} \text{ and } Z_{2l} \leqslant c_{2l} \text{ for all } l = 1, \ldots, R, \end{cases}$$

and calculated similarly as discussed in the power for DF-2 (Appendix A.1).

When $R = L - 1$, the conditional power for DF-1 can be rewritten as

$$CP = \Pr[A_{1L} \cap A_{2L} \,|a_{1R}, a_{2R}] = \Phi_2\left(-c_1^*, -c_2^* | \rho\right),$$

where $\Phi_2(\cdot, \cdot|\rho)$ is the cumulative distribution function of the standard bivariate normal distribution with the correlation ρ, and $c_1^* = (c_{1L} - a_{1R}\sqrt{t})/\sqrt{1-t} - \delta_1\sqrt{n_L - n_R}/\sqrt{2}$ and $c_2^* = (c_{2L} - a_{2R}\sqrt{t})/\sqrt{1-t} - \delta_2\sqrt{n_L - n_R}/\sqrt{2}$ with $t = n_R/n_L$. For DF-2, the conditional power can be rewritten as

$$CP = \begin{cases} \Pr[A_{1L} \,|a_{1R}, a_{2l'}] = 1 - \Phi\left(c_1^*\right) \\ \quad \text{if } Z_{1l} \leqslant c_{1l} \text{ for all } l = 1, \ldots, R \text{ and } Z_{2l'} > c_{2l'} \text{ for some } l' = 1, \ldots, R, \\ \Pr[A_{2L} \,|a_{2R}, a_{1l'}] = 1 - \Phi\left(c_2^*\right) \\ \quad \text{if } Z_{2l} \leqslant c_{2l} \text{ for all } l = 1, \ldots, R \text{ and } Z_{1l'} > c_{1l'} \text{ for some } l' = 1, \ldots, R, \\ \Pr[A_{1L} \cap A_{2L} \,|a_{1R}, a_{2R}] = \Phi_2\left(-c_1^*, -c_2^* \,|\rho\right) \\ \quad \text{if } Z_{1l} \leqslant c_{1l} \text{ and } Z_{2l} \leqslant c_{2l} \text{ for all } l = 1, \ldots, R, \end{cases}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standardized normal distribution.

Acknowledgements

The authors are grateful to the two anonymous referees and the associate editor for their valuable suggestions and helpful comments that improved the content and presentation of the paper. We thank Dr. H. M. James Hung and Dr. Sue-Jane Wang for encouraging us in this research with their helpful advice. Research reported in this publication was supported by JSPS KAKENHI under Grant Number 23500348 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Numbers UM1AI104681 and UM1AI068634. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Committee for Medicinal Products for Human Use (CHMP). Guideline on Medicinal Products for the Treatment Alzheimer's Disease and Other Dementias (CPMP/EWP/553/95 Rev.1). EMEA: London, 2008.
2. Offen W, Chuang-Stein C, Dmitrienko A, Littman G, Maca J, Meyerson L, Muirhead R, Stryszak P, Boddy A, Chen K, Copley-Merriman K, Dere W, Givens S, Hall D, Henry D, Jackson JD, Krishen A, Liu T, Ryder S, Sankoh AJ, Wang J, Yeh CH. Multiple co-primary endpoints: medical and statistical solutions. *Drug Information Journal* 2007; **41**:31–46. DOI: 10.1177/009286150704100105.
3. Hung HMJ, Wang SJ. Some controversial multiple testing problems in regulatory applications. *Journal of Biopharmaceutical Statistics* 2009; **19**:1–11. DOI: 10.1080/10543400802541693.
4. Dmitrienko A, Tamhane AC, Bretz F. *Multiple Testing Problems in Pharmaceutical Statistics*. Chapman & Hall: Boca Raton, FL., 2010.
5. Xiong C, Yu K, Gao F, Yan Y, Zhang Z. Power and sample size for clinical trials when efficacy is required in multiple endpoints: application to an Alzheimer's treatment trial. *Clinical Trials* 2005; **2**:387–393. DOI: 10.1191/1740774505cn112oa.
6. Sozu T, Kanou T, Hamada C, Yoshimura I. Power and sample size calculations in clinical trials with multiple primary variables. *Japanese Journal of Biometrics* 2006; **27**:83–96. DOI: 10.5691/jjb.27.83.

Copyright © 2014 John Wiley & Sons, Ltd.

7. Chuang-Stein C, Stryszak P, Dmitrienko A, Offen W. Challenge of multiple co-primary endpoints: a new approach. *Statistics in Medicine* 2007; **26**:1181–1192. DOI: 10.1002/sim.2604.
8. Eaton ML, Muirhead RJ. On multiple endpoints testing problem. *Journal of Statistical Planning & Inference* 2007; **137**:3416–3429. DOI: 10.1016/j.jspi.2007.03.021.
9. Senn S, Bretz F. Power and sample size when multiple endpoints are considered. *Pharmaceutical Statistics* 2007; **6**:161–170. DOI: 10.1002/pst.301.
10. Kordzakhia G, Siddiqui O, Huque MF. Method of balanced adjustment in testing co-primary endpoints. *Statistics in Medicine* 2010; **29**:2055–2066. DOI: 10.1002/sim.3950.
11. Sozu T, Sugimoto T, Hamasaki T. Sample size determination in clinical trials with multiple co-primary binary endpoints. *Statistics in Medicine* 2010; **29**:2169–2179. DOI: 10.1002/sim.3972.
12. Sozu T, Sugimoto T, Hamasaki T. Sample size determination in superiority clinical trials with multiple co-primary correlated endpoints. *Journal of Biopharmaceutical Statistics* 2011; **21**:650–668. DOI: 10.1080/10543406.2011.551329.
13. Julious S, McIntyre NE. Sample sizes for trials involving multiple correlated must-win comparisons. *Pharmaceutical Statistics* 2012; **11**:177–185. DOI: 10.1002/pst.515.
14. Sugimoto T, Sozu T, Hamasaki T. A convenient formula for sample size calculations in clinical trials with multiple co-primary continuous endpoints. *Pharmaceutical Statistics* 2012; **11**:118–128. DOI: 10.1002/pst.505.
15. Hamasaki T, Sugimoto T, Evans SR, Sozu T. Sample size determination for clinical trials with co-primary outcomes: exponential event times. *Pharmaceutical Statistics* 2013; **12**:28–34. DOI: 10.1002/pst.1545.
16. Sugimoto T, Sozu T, Hamasaki T, Evans SR. A logrank test-based method for sizing clinical trials with two co-primary time-to-event endpoints. *Biostatistics* 2013; **14**:409–421. DOI: 10.1093/biostatistics/kxs057.
17. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika* 1983; **70**:659–663. DOI: 10.1093/biomet/70.3.659.
18. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics* 1979; **35**:549–556. DOI: 10.2307/2530245.
19. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika* 1977; **64**:191–199. DOI: 10.1093/biomet/64.2.191.
20. Glimm E, Maurer W, Bretz F. Hierarchical testing of multiple endpoints in group-sequential trials. *Statistics in Medicine* 2010; **29**:219–228. DOI: 10.1002/sim.3748.
21. Tamhane AC, Mehta CR, Liu L. Testing a primary and secondary endpoint in a group sequential design. *Biometrics* 2010; **66**:1174–1184. DOI: 10.1111/j.1541-0420.2010.01402.x.
22. Green RC, Schneider LS, Amato DA, Beelen AP, Wilcock G, Swabb EA, Zavitz KH, for the Tarenflurbil Phase 3 Study Group. Effect of tarenflurbil on cognitive decline and activities of daily living in patients with mild Alzheimer disease: a randomized controlled trial. *Journal of the American Medical Association* 2009; **302**:2557–2564. DOI: 10.1001/jama.2009.1866.
23. Doraiswamy PM, Bieber F, Kaiser L, Krishnan KR, Reuning-Scherer J, Gulanski B. The Alzheimer's disease assessment scale: patterns and predictors of baseline cognitive performance in multicenter Alzheimer's disease trials. *Neurology* 1997; **48**:1511–1517. DOI: 10.1212/WNL.48.6.1511.
24. Cui L, Hung HMJ, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics* 1999; **55**:853–857. DOI: 10.1111/j.0006-341X.1999.00853.x.
25. Asakura K, Hayashi K, Sugimoto T, Sozu T, Hamasaki T. Sample size evaluation in group sequential designs for clinical trials with two continuous endpoints as co-primary contrasts. *Joint Statistical Meetings 2013*, Montreal, Quebec, Canada, August 3-8, 2013.
26. Hamasaki T, Asakura K, Sugimoto T, Evans SR. Sample size modification in group-sequential clinical trials with two co-primary endpoints. *Proceedings of Joint Meeting of the IASC Satellite Conference and 8th Conference of the Asian Regional Section of the IASC*, Seoul, Korea, August 21-14, 2013; 311–317.
27. Moyé LA. *Multiple Analyses in Clinical Trials*. Springer: New York. NY, 2003.
28. Moyé LA, Baraniuk S. Dependence, hyper-dependence and hypothesis testing in clinical trials. *Contemporary Clinical Trials* 2013; **28**:68–78. DOI: 10.1016/j.cct.2006.05.010.
29. Jennison C, Turnbull BW. Group sequential tests for bivariate response: interim analyses of clinical trials with both efficacy and safety. *Biometrics* 1993; **49**:741–752.
30. Tang DI, Geller NL. Closed testing procedures for group sequential clinical trials with multiple endpoints. *Biometrics* 1999; **55**:1188–1192. DOI: 10.1111/j.0006-341X.1999.01188.x.
31. Tamhane AC, Wu Y, Mehta C. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (I): unknown correlation between the endpoints. *Statistics in Medicine* 2012; **31**:2027–2040. DOI: 10.1002/sim.5372.
32. Tamhane AC, Wu Y, Mehta C. Adaptive extensions of a two-stage group sequential procedure for testing primary and secondary endpoints (II): sample size re-estimation. *Statistics in Medicine* 2012; **31**:2041–2054. DOI: 10.1002/sim.5377.
33. Genz A, Bretz F. *Computation of Multivariate Normal and t Probabilities*. Springer Verlag: Berlin, 2009.

Supporting information

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

Copyright © 2014 John Wiley & Sons, Ltd.

Acta Med. Okayama, 2014
Vol. 68, No. 2, pp. 111-117
Copyright© 2014 by Okayama University Medical School.

Acta Medica Okayama
http://escholarship.lib.okayama-u.ac.jp/amo/

Original Article

Assistant-Based Standardization of Prone Position Thoracoscopic Esophagectomy

Yasuhiro Shirakawa*, Kazuhiro Noma, Naoaki Maeda, Ryoichi Katsube, Shunsuke Tanabe, Toshiaki Ohara, Kazufumi Sakurama, and Toshiyoshi Fujiwara

Department of Gastroenterological Surgery, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama 700-8558, Japan

Thoracoscopic esophagectomy in the prone position (TEPP) might enable solo-surgery in cases requiring resection of the esophagus and the surrounding lymph nodes due to the associated advantages of good exposure of the surgical field and ergonomic considerations for the surgeon. However, no one approach can be for all patients requiring extensive lymphadenectomy. We recently developed an assistant-based procedure to standardize exposure of the surgical field. Patients were divided into 1 of 2 groups: a pre-standardization group (n = 37) and a post-standardization group (n = 28). The thoracoscopic operative time was significantly shorter ($p = 0.0037$) in the post-standardization group (n = 28; 267 ± 31 min) than in the pre-standardization group (n = 37; 301 ± 53 min). Further, learning curve analysis using the moving average method showed stabilization of the thoracoscopic operative time after the standardization. No significant differences were found in the number of mediastinal lymph nodes dissected or intraoperative blood loss between the 2 groups. There were also no significant differences in the complication rate. Assistant-based surgery and standardization of the procedure resulted in a well-exposed and safe surgical field. TEPP decreased the operative time, even in patients requiring extensive lymphadenectomy.

Key words: thoracoscopic esophagectomy, prone position, standardization

Thoracoscopic esophagectomy in the prone position (TEPP) was first reported by Cuschieri *et al.* [1] in 1994 and has gradually grown in popularity. Several reports have described the tolerability and efficacy of this procedure [2-4]. Planivelu *et al.* [5] conducted a study of 130 patients and reported that TEPP resulted in a decreased operative time and a decreased frequency of respiratory complications when compared with other techniques of minimally invasive esophagectomy and open esophagectomy, mainly due to the associated advantages of good exposure of the surgical field and improved ergonomics for the surgeon. Other reports have described that TEPP permits solo-surgery for resection of the esophagus after isolation from the surrounding organs, regardless of the skill of the assistant [4-8].

In Japan, thoracoscopic esophagectomy is conventionally performed in the left lateral decubitus position. Some studies have reported that the traditional Japanese technique of precise mediastinal lymph node dissection is equally or more effective than open surgery via thoracotomy, with an added advantage of a low respiratory complication rate [9, 10]. Several recent studies have also described the advantages of

Received September 30, 2013; accepted December 10, 2013.
*Corresponding author. Phone:+81-86-235-7257; Fax:+81-86-221-8775
E-mail:yasuwr@md.okayama-u.ac.jp (Y. Shirakawa)

mediastinal lymph node dissection using TEPP [3, 4, 6, 11–13]. However, performing extensive and precise dissection consistently (as conducted in Japan), irrespective of the body type of the patient, can be difficult for a single surgeon. This is because the boundaries of the surgical field are limited when a single surgeon is involved. Therefore, we hypothesized that exposure of the surgical field by an assistant is necessary to achieve optimal outcomes for patients undergoing this procedure.

Teaching and university hospitals are increasingly moving towards: (1) decreased dependency on single expert surgeons, (2) increased recruitment of young surgeons who can function as assistants, and (3) development of defined safety measures that can apply to all members of the team, even staff working under short-term contracts who are not otherwise familiar with the facility. In this regard, standardization of procedures involving surgical assistants is important and necessary.

The purpose of this study was to establish and evaluate our new standardized procedure for performing thoracoscopic esophagectomy with patients in the prone position, with particular reference to the work of the surgical assistants.

Materials and Methods

Patients. Thoracoscopic esophagectomy was performed for 65 patients (62 males, 3 females) in the prone position at our facility from June 2011 to September 2012. This group comprised 75.6% of the 86 patients with thoracic esophageal carcinoma who underwent resection at our facility during this time period. The preoperative diagnosis was squamous cell carcinoma in 61 patients, adenocarcinoma in 2 patients, and mixed squamous cell carcinoma combined with neuroendocrine tumor in 2 patients. No patient had a prior history of thoracic surgery. Preoperative adjuvant chemotherapy was administered to 45 patients, and 3 patients underwent salvage surgery following radical chemoradiotherapy. The inclusion criteria for the study were as follows: no pleural adhesions, no T4 or M1 cancer, and no serious impairment in circulatory, respiratory, or liver function. Patients were divided into 1 of 2 groups: a pre-standardization group (n = 37) that underwent surgery before April 2012 and a post-standardization group (n = 28) that underwent surgery after April 2012. Clinical outcomes were compared between these 2 groups.

Operative procedure. The patient was immobilized in the prone position after endotracheal intubation using a single-lumen endotracheal tube and a bronchial blocker. The surgeon stood on the right of the patient. A 12-mm port was placed at the 9th intercostal space (ICS) at the inferior scapular line, and a thoracoscope was inserted at a 30-degree angle. In addition, a 12-mm port for the assistant was placed at the 3rd ICS at the mid-axillary line, while another 12-mm port for the right hand of the surgeon was placed at the 5th ICS at the posterior axillary line. A 5-mm port for the left hand of the surgeon was also placed at the 7th ICS at the posterior axillary line. Before procedure standardization (*i.e.*, through March 2012), the procedure was conducted with 4 ports. After procedure standardization (from April 2012), an additional 12-mm port for the assistant was placed at the 8th ICS between the mid-axillary line and the posterior axillary line (Fig. 1). A 6-mmHg artificial pneumothorax was induced using left one-lung ventilation and carbon dioxide (CO_2).

The surgical procedure was also standardized with the goal of obtaining a well-exposed surgical field. The standardized procedure is as follows. The surgeon mainly uses the 5th and 7th ICS ports, while the assistant uses the 3rd or 8th port. Exposure of the surgical field by the assistant is important to facilitate the procedure. For middle and lower mediastinal surgeries, the surgical field is exposed, with exclusion of the diaphragm or pericardium, by manipulation using thoraco-cotton (Wyeth Lederle, Tokyo, Japan) from the 8th ICS by the assistant prior to esophageal transection in the initial phase of the surgery (Fig. 2A). Following esophageal transection using autosutures, the pericardium is excluded using thoraco-cotton from the 3rd ICS port while the assistant pulls the esophagus caudally with forceps from the 8th ICS port (2 assistants may be needed at times; Fig. 2B). At this point, it is essential that the contralateral pleura develops into a trapezoid shape in order to prevent the procedure from converting into a left thoracotomy. As the esophageal hiatus is caudally approached by dissection and isolation of the esophagus, the diaphragm is excluded with thoraco-cotton from the 8th ICS port while the assistant pulls the esophagus cranially with forceps from the 3rd ICS port (Fig. 2C). This readily allows for accurate dis-

section of the lymph nodes above the diaphragm. In our cases, exposure with thoraco-cotton or forceps from the 3rd ICS port is also useful while operating on the superior mediastinum, even in patients requiring dissection along both sides of the recurrent laryngeal nerve, and in the subcarinal area (Fig. 2D, E and Fig. 3).

Description and statistical analysis. Clini-

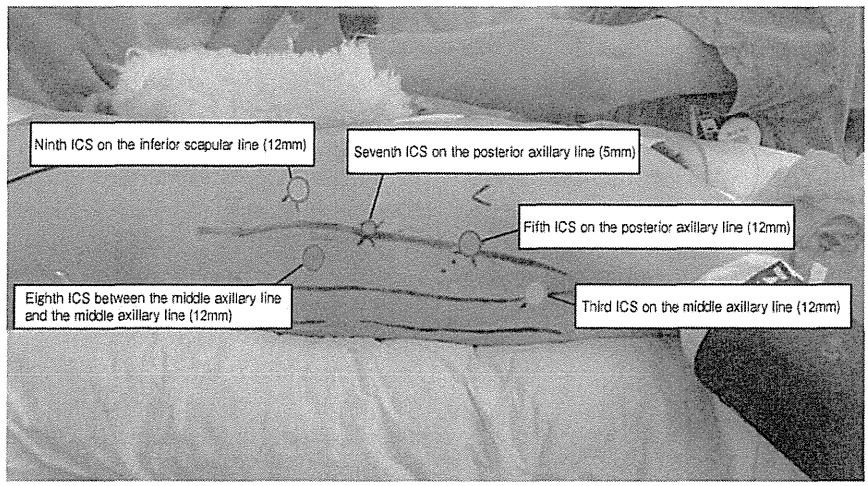

Fig. 1 Port insertion point. Green indicates the camera point, blue the surgeon, orange and red the assistant. The port at the 8th intercostal space was added from April 2012. ICS, intercostal space.

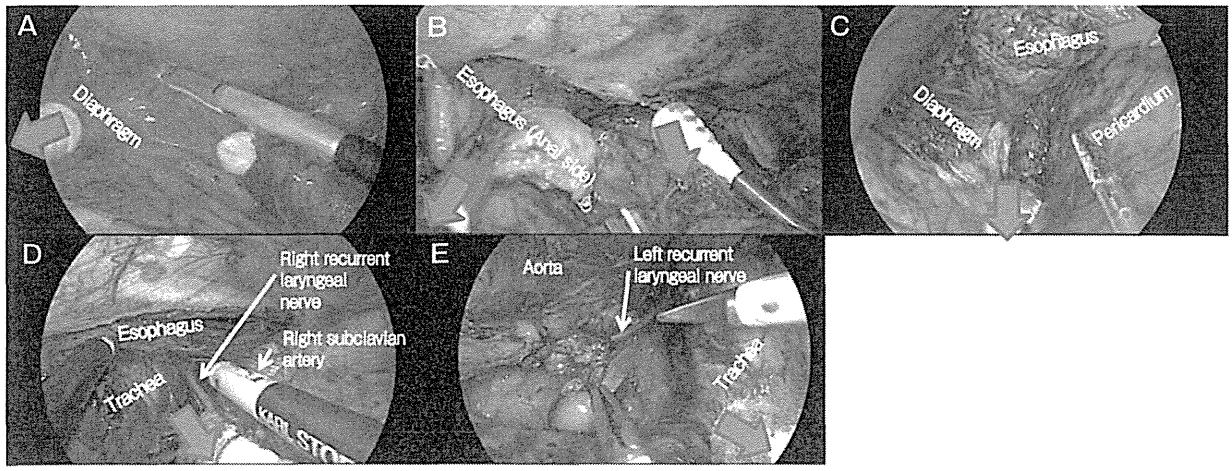

Fig. 2 Exposure of the surgical field by the assistant in thoracoscopic esophagectomy in the prone position. A) In the initial stage of the middle and lower mediastinal procedures, the assistant expands the visual field by excluding the diaphragm by pulling it caudally using thoraco-cotton inserted from the port at the 8th intercostal space (ICS). B) After esophageal transection, the assistant expands the visual field by excluding the pericardium and right main bronchus by pulling them with thoraco-cotton inserted from the port at the 3rd ICS and by pulling the esophagus caudally on the anal side using forceps inserted from the port at the 8th ICS. C) During surgery around the esophageal hiatus, the assistant expands the visual field by pulling the diaphragm ventrally with thoraco-cotton inserted from the port at the 8th ICS and by pulling the esophagus cranially with forceps inserted from the port at the 3rd ICS. D) During lymph node dissection around the right recurrent laryngeal nerve, the assistant expands the visual field by rolling the right subclavian artery cranially with thoraco-cotton inserted from the port at the 3rd ICS. E) During lymph node dissection around the left recurrent laryngeal nerve, the assistant expands the visual field by rolling the trachea to the right-hand side with thoraco-cotton inserted from the port at the 3rd ICS.

copathological factors were noted as per the 10th edition of the General Rules for Esophageal Cancer [14] and the Union for International Cancer Control (UICC) Tumor Nodes Metastasis (TNM) Classification of Malignant Tumors, 7th edition [15]. Postoperative complications were categorized as per the Clavien-Dindo classification [16]. Data are expressed as the mean ± standard deviation. Statistical analysis was conducted using Student's *t*-test, the Chi-square test, or Fisher's exact test to suit the category in question. All analyses were performed using statistical analysis software (JMP version 11; SAS Institute Inc., Cary, NC, USA). The thoracoscopic operative time learning curve was analyzed using the moving average method [17, 18]. Trends in thoracoscopic operative times may be unclear based on changes in individual cases. With the moving average method, using the mean thoracoscopic operative times, the individual changes are removed, and the trends are clarified. In addition, as new data are added, by shifting the mean values, the changes in thoracoscopic operative times are smoothed. A 5-cases moving average was used.

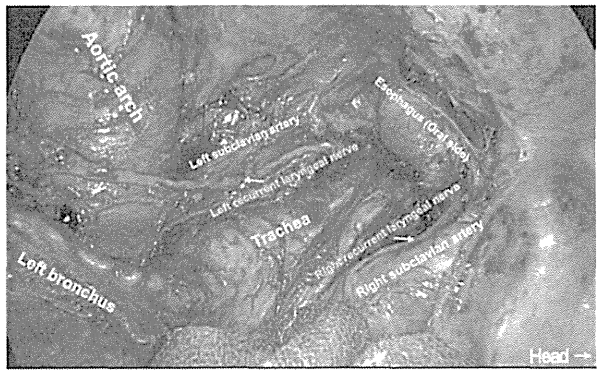

Fig. 3 Thoracoscopic view after competing lymphadenectomy along both sides of the recurrent laryngeal nerve.

Results

There were no significant differences in patient background when comparing the 2 groups (Table 1). The thoracoscopic operative time was significantly shorter ($p = 0.0037$) in the post-standardization group (n = 28; 267 ± 31 min) than in the pre-standardization group (n = 37; 301 ± 53 min) (Table 2). The learning curve analysis using the moving average method showed stabilization of the thoracoscopic operative time after the technique was standardized (Fig. 4). No significant differences were found between the 2 groups in terms of the number of mediastinal lymph nodes dissected or the amount of intraoperative blood loss. None of the patients in either group required conversion to thoracotomy (Table 2). There were no significant differences in the overall complication rate or incidence of respiratory complications or recurrent nerve palsy when comparing the 2 groups (Table 3). However, 1 patient developed chylothorax in the pre-standardization group. There were no mortalities in either group.

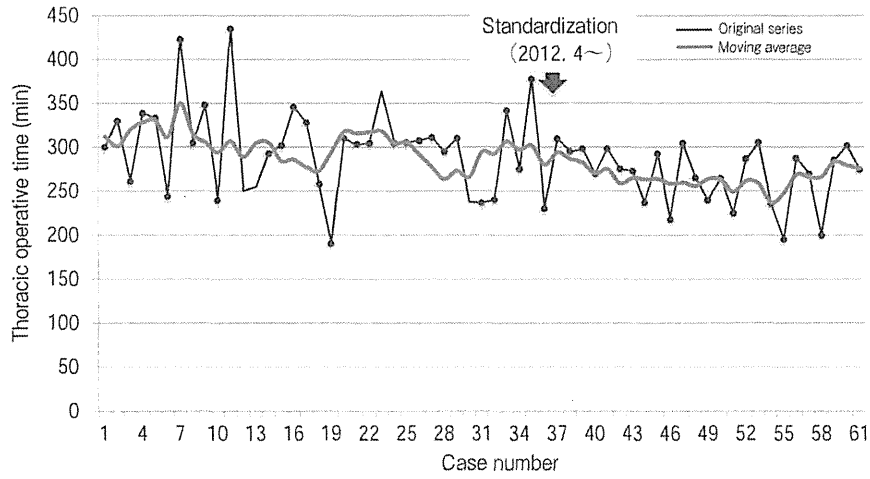

Fig. 4 The moving average method was used to determine changes in thoracoscopic operative time.

Table 1 Clincal characteristics (n=65)

		Pre-standardization group (n = 37)	Post-standardization group (n = 28)	*P*-value
Age (years)		67.3 ± 1.3	64.5 ± 1.6	0.188
Gender	Male	35	27	0.75
	Female	2	1	
Location	Upper	7	4	0.667
	Middle	17	16	
	Lower	13	8	
Histology	SCC	34	27	0.451
	Adeno	1	1	
	NET	2	0	
pTNM Stage	Stage 0	6	3	0.883
	Stage I	8	7	
	Stage II	13	10	
	Stage III	9	6	
	Stage IV	1	2	
Administration of neoadjuvant chemotherapy		24	21	0.381

NET, neuroendocrine tumor.

Table 2 Clinical outcomes of thoracoscopic esophagectomy in the prone position

	Pre-standardization group (n = 37)	Post-standardization group (n = 28)	*P*-value
Conversion to thoracotomy	0	0	1.000
Thoracic operative time (min)	301 ± 53	267 ± 31	0.004
Estimated blood loss (g)	184 ± 190	171 ± 129	0.757
Number of dissected mediastinal lymph nodes	33.9 ± 12.1	33.6 ± 12.6	0.911

Table 3 Complications after thoracoscopic esophagectomy in the prone position

	Pre-standardization group (n = 37)	Post-standardization group (n = 28)	*P*-value
Morbidity	5	3	0.734
Pneumonia (Grade IIIa-IVa)	0	0	1.000
Recurrent laryngeal nerve palsy (Grade I-II)	4	3	0.878
Chylothorax (Grade IIIa-IIIb)	1	0	0.407
Operative mortality	0	0	1.000
Hospital mortality	0	0	1.000

Complications are described on the Clavien-Dindo classification [16].

Discussion

Conventional open esophagectomy is a highly invasive procedure [19-21]. A key disadvantage of this procedure is the size of the incision involved in the thoracotomy, laparotomy and cervical incision. In fact, some studies have reported that these factors are associated with systemic inflammatory response syndrome [22, 23]. The overall operative mortality after open esophagectomy is approximately 10%, and the rate of serious complications is also high [24]. Respiratory complications (*e.g.*, respiratory failure,

pneumonia) are the most common, and patients with these complications have a 20% rate of operative mortality [25].

Thoracoscopic esophagectomy was introduced as a minimally invasive option to minimize postoperative complications and operative mortality. Initially, thoracoscopic esophagectomy was performed in the left lateral decubitus position [9, 10, 26], and the survival benefits were the same as those seen in response to open procedures [3, 9, 10]. In addition, when the left lateral decubitus position was used during thoracoscopic surgery, the operative time was almost the same as when it was used during open surgery via thoracotomy [10]. However, a special team composed of 3 experts (*i.e.*, a surgeon, an assistant, and an endoscopist) is required to perform this procedure smoothly. Therefore, thoracoscopic esophagectomy has not been widely utilized.

In 1994, Cuschieri *et al.* [1] first described the use of TEPP in a series of 6 patients and reported that there were no differences in the postoperative course between these patients and 20 patients who underwent open surgery via thoracotomy in the left lateral decubitus position during the same period. The same investigators also suggested that the prone position was practically and physiologically superior to the left lateral decubitus position. In 2006, Planivelu *et al.* [5] described their experience with 130 patients undergoing TEPP and reported that this procedure was associated with a shorter operative time, excellent exposure of the surgical field, improved ergonomics for the surgeon, and a lower rate of respiratory complications.

Open or thoracoscopic procedures in the left lateral decubitus position have been performed in Japan over the previous decade. Some recent reports suggest that the prone position is advantageous for esophagectomy, as it is associated with an excellent surgical view [4, 6]. However, there have been no reports in Japan showing that TEPP is superior in terms of operative time over open surgery via thoracotomy or thoracoscopic surgery in the left lateral decubitus position. The thoracic operative time in the prone position might be decreased for dissections involving en-bloc resection of the esophagus and surrounding lymph nodes; however, while this procedure is widely performed in Western countries, it is quite different from the approach used in Japan. When thorough dissection of the mediastinal lymph nodes (including those in the upper mediastinum) is undertaken as recommended in Japan, an average of approximately 5 h is required for patients in the prone position [4, 6]. Also, in this position the difficulty of mediastinal lymph node dissection differs according to the patient's body type. It is very difficult to constantly maintain an optimal surgical field without the use of an assistant, particularly in the case of a narrow space between vertebral bodies and the trachea in the superior mediastinum or in the case of a raised dome of the diaphragm in the inferior mediastinum. It is important to conduct thorough lymph node dissection at these sites, because these are common sites of lymph node metastases. To summarize, it is difficult to maintain a short thoracic operative time when using conventional methods, since it is difficult to achieve an optimal visual field and to perform extensive and precise mediastinal lymph node dissection.

Better exposure of the surgical field through the use of a surgical assistant during surgery may help to improve outcomes. In the present study, we used a protocol to standardize the assistant-based exposure of the surgical field by organizing the procedure in each section and by adding an assistant port in the 8th ICS. The standardization of this assisted procedure resulted in a significant decrease in thoracic operative time without compromising the quality of dissection. Standardization also resulted in a smaller variation in thoracic operative time. Furthermore, this protocol is playing an important role in the development of the next generation of surgeons at our institution and will likely lead to increased satisfaction by the assistants, since it allows them to actively participate in procedures. Finally, none of the patients in this study complained of increased postoperative pain related to the use of an additional port.

Despite standardization, the operative time tended to be longer with TEPP than that of open surgery via thoracotomy or thoracoscopic esophagectomy in the left lateral decubitus position. Further, the incidence of recurrent nerve palsy was also slightly higher with our procedure than with open surgery via thoracotomy [10], and most of these cases were unilateral left-sided palsies. Although we have generally avoided using energy devices during lymph node dissection at that site in order to avoid heat damage, the handling and traction of en block lymph nodes may be issues in

cases of palsy. With the greater visibility of the superior mediastinum in response to this assistant-based exposure, there may be a tendency to dissect up to the nerve, which might contribute to the higher incidence of nerve palsy. Recurrent nerve palsy is a cause of both aspiration and dysphagia and is a major factor contributing to decreased postoperative quality of life. Therefore, further studies are needed to determine how to reduce the incidence of this complication.

In conclusion, TEPP performed by a single surgeon is still technically difficult, mainly because a suboptimal surgical field can otherwise impede thorough dissection of the mediastinal lymph nodes and performance of esophagectomy. The present study demonstrated that assistant-based surgery with the addition of an assistant port and standardization of the procedures involved resulted in greater safety and better outcomes in TEPP.

References

1. Cuschieri A: Thoracoscopic subtotal oesophagectomy. Endosc Surg Allied Technol (1994) 2: 21-25.
2. Tsujimoto H, Takahata R, Nomura S, Yaguchi Y, Kumano I, Matsumoto Y, Yoshida K, Horiguchi H, Hiraki S, Ono S, Yamamoto J and Hase K: Video-assisted thoracoscopic surgery for esophageal cancer attenuates postoperative systemic responses and pulmonary complications. Surgery (2012) 151: 667-673.
3. Petri R, Zuccolo M, Brizzolari M, Rossit L, Rosignoli A, Durastante V, Petrin G, De Cecchis L and Sorrentino M: Minimally invasive esophagectomy: thoracoscopic esophageal mobilization for esophageal cancer with the patient in prone position. Surg Endosc (2012) 26: 1102-1107.
4. Ozawa S, Ito E, Kazuno A, Chino O, Nakui M, Yamamoto S, Shimada H and Makuuchi H: Thoracoscopic esophagectomy while in a prone position for esophageal cancer: a preceding anterior approach method. Surg Endosc (2013) 27: 40-47.
5. Palanivelu C, Prakash A, Senthilkumar R, Senthilnathan P, Parthasarathi R, Rajan PS and Venkatachlam S: Minimally invasive esophagectomy: thoracoscopic mobilization of the esophagus and mediastinal lymphadenectomy in prone position--experience of 130 patients. J Am Coll Surg (2006) 203: 7-16.
6. Noshiro H, Iwasaki H, Kobayashi K, Uchiyama A, Miyasaka Y, Masatsugu T, Koike K and Miyazaki K: Lymphadenectomy along the left recurrent laryngeal nerve by a minimally invasive esophagectomy in the prone position for thoracic esophageal cancer. Surg Endosc (2010) 24: 2965-2973.
7. Fabian T, Martin J, Katigbak M, McKelvey AA and Federico JA: Thoracoscopic esophageal mobilization during minimally invasive esophagectomy: a head-to-head comparison of prone versus decubitus positions. Surg Endosc (2008) 22: 2485-2491.
8. Fabian T, McKelvey AA, Kent MS and Federico JA: Prone thoracoscopic esophageal mobilization for minimally invasive esophagectomy. Surg Endosc (2007) 21: 1667-1670.
9. Akaishi T, Kaneda I, Higuchi N, Kuriya Y, Kuramoto J, Toyoda T and Wakabayashi A: Thoracoscopic en bloc total esophagectomy with radical mediastinal lymphadenectomy. J Thorac Cardiovasc Surg (1996) 112: 1533-1541.
10. Osugi H, Takemura M, Higashino M, Takada N, Lee S and Kinoshita H: A comparison of video-assisted thoracoscopic oesophagectomy and radical lymph node dissection for squamous cell cancer of the oesophagus with open operation. Br J Surg (2003) 90: 108-113.
11. Shen Y, Zhang Y, Tan L, Feng M, Wang H, Khan MA, Liang M and Wang Q: Extensive mediastinal lymphadenectomy during minimally invasive esophagectomy: optimal results from a single center. J Gastrointest Surg (2012) 16: 715-721.
12. Feng M, Shen Y, Wang H, Tan L, Zhang Y, Khan MA and Wang Q: Thoracolaparoscopic esophagectomy: is the prone position a safe alternative to the decubitus position? J Am Coll Surg (2012) 214: 838-844.
13. Yatabe T, Kitagawa H, Yamashita K, Hanazaki K and Yokoyama M: Comparison of the perioperative outcome of esophagectomy by thoracoscopy in the prone position with that of thoracotomy in the lateral decubitus position. Surg Today (2012) 43: 386-391.
14. Japan Esophageal Society: Japanese classification of esophageal cancer, 10th Ed. Kanehara & Co. Tokyo (2008).
15. Sobin LH, Gospodarowicz MK and Wittekind C: TNM classification of malignant tumors (UICC international union against cancer), 7th Ed., Wiley-Blackwell, Oxford (2009).
16. Dindo D, Demartines N and Clavien PA: Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. Ann Surg (2004) 240: 205-213.
17. Diggle P: Time series: a biostatistical introduction. Oxford University Press, London (1990).
18. Kayano H, Okuda J, Tanaka K, Kondo K and Tanigawa N: Evaluation of the learning curve in laparoscopic low anterior resection for rectal cancer. Surg Endosc (2011) 25: 2972-2979.
19. Akiyama H, Tsurumaru M, Udagawa H and Kajiyama Y: Radical lymph node dissection for cancer of the thoracic esophagus. Ann Surg (1994) 220: 364-373.
20. Fujita H, Kakegawa T, Yamana H, Shima I, Toh Y, Tomita Y, Fujii T, Yamasaki K, Higaki K, Noake T, Ishibashi N and Mizutani K: Mortality and morbidity rates, postoperative course, quality of life, and prognosis after extended radical lymphadenectomy for esophageal cancer. Comparison of three-field lymphadenectomy with two-field lymphadenectomy. Ann Surg (1995) 222: 654-662.
21. Watson A: Operable esophageal cancer: current results from the West. World J Surg (1994) 18: 361-366.
22. Ni NC, Redmond HP: Cell response to surgery. Arch Sur (2006) 141: 1132-1140.
23. Tsujimoto H, Ono S, Takahata R, Hiraki S, Yaguchi Y, Kumano I, Matsumoto Y, Yoshida K, Aiko S, Ichikura T, Yamamoto J and Hase K: Systemic inflammatory response syndrome as a predictor of anastomotic leakage after esophagectomy. Surg Today (2012) 42: 141-146.
24. Jamieson GG, Mathew G, Ludemann R, Wayman J, Myers JC and Devitt PG: Postoperative mortality following oesophagectomy and problems in reporting its rate. Br J Surg (2004) 91: 943-947.
25. Atkins BZ, Shah AS, Hutcheson KA, Mangum JH, Peppas TN, Harpole DH Jr and D'Amico TA: Reducing hospital morbidity and mortality following esophagectomy. Ann Thorac Surg (2004) 78: 1170-1176.
26. Cuschieri A, Shimi S and Banting S: Endoscopic oesophagectomy through a right thoracoscopic approach. J R Coll Surg Edinb (1992) 37: 7-11.

Percutaneous Radiofrequency Ablation for Pulmonary Metastases from Esophageal Cancer: Retrospective Evaluation of 21 Patients

Yusuke Matsui, MD, Takao Hiraki, MD, Hideo Gobara, MD, Hiroyasu Fujiwara, MD, Toshihiro Iguchi, MD, Yasuhiro Shirakawa, MD, Toshiyoshi Fujiwara, MD, Shinichi Toyooka, MD, and Susumu Kanazawa, MD

ABSTRACT

Purpose: To evaluate retrospectively outcomes after radiofrequency (RF) ablation for pulmonary metastases from esophageal cancer.

Materials and Methods: This study included 21 consecutive patients who met inclusion criteria (all men; mean age, 66.0 y) and had pulmonary metastases from esophageal cancer. There were 31 tumors (mean size, 1.7 cm) that were treated with 27 planned ablation sessions. At the initial RF ablation sessions, 3 patients had viable extrapulmonary recurrences, and 18 patients had viable recurrences confined to the lung. Primary study endpoints included patient survival and the determination of prognostic factors. Secondary endpoints included local efficacy and safety of the treatment. The log-rank test was used to identify prognostic factors. Adverse events were evaluated according to the National Cancer Institute Common Terminology Criteria for Adverse Events Version 4.0.

Results: Median follow-up duration after the initial RF ablation was 22.4 months (range, 6.2–76.1 mo). Estimated overall survival rates were 85.7% at 1 year, 54.8% at 2 years, and 38.4% at 3 years after the initial RF ablation session. The presence of viable extrapulmonary recurrences at the initial RF ablation session was an unfavorable prognostic factor ($P < .001$). Local tumor progression was observed in 25.8% (8 of 31) of tumors and occurred 2.6–10.0 months (median, 4.8 mo) after RF ablation. Grade 3 adverse events occurred in 7.4% (2 of 27) of sessions, including pleural effusion requiring chest tube placement and pneumoderma requiring surgical intervention. No grade 4 or greater adverse events occurred.

Conclusions: RF ablation is a promising treatment option for patients with pulmonary metastases from esophageal cancer.

Esophageal cancer is the eighth most frequently occurring cancer, accounting for nearly 500,000 new cases per year, and it is the sixth most common cause of cancer-related death worldwide (1). At the time of presentation, > 30% of patients with esophageal cancer have metastatic disease (2). Even after curative esophagectomy, distant recurrence occurs in 13%–24% of patients with esophageal cancer (3–5). The lung is one of the most common sites of metastasis from esophageal cancer (3,5). Esophageal cancer with distant metastasis is associated with poor survival outcomes, including a 5-year survival rate of 3.5% (2). Although the use of local therapy for pulmonary metastasis is controversial, several studies have investigated the outcomes of surgery for pulmonary metastasis from esophageal cancer in selected patients. These studies found median or mean survival periods of 24–29 months (6–9), which appears favorable compared with survival outcomes after chemotherapy (10–13).

Radiofrequency (RF) ablation is a widely used local therapy for malignant tumors in various organs. The long-term outcomes of RF ablation for pulmonary metastasis remain to be determined. However, promising short-term and midterm survivals have been observed after RF ablation for pulmonary metastasis from various primary lesions, including colorectal cancer, hepatocellular carcinoma, renal cell carcinoma, and sarcoma (14–20). Patients with pulmonary metastases from esophageal cancer may also be candidates for RF ablation (21). However, there is a paucity of data on

From the Departments of Radiology (Y.M., T.H., H.G., H.F., T.I., S.K.), Gastroenterological Surgery (Y.S., T.F.), and General Thoracic Surgery (S.T.), Okayama University Medical School, 2-5-1 Shikatacho, Kita-ku, Okayama 700-8558, Japan. Received April 21, 2014; final revision received June 21, 2014; accepted June 27, 2014. Address correspondence to Y.M.; E-mail: wckyh140@yahoo.co.jp

From the SIR 2014 Annual Meeting.

None of the authors have identified a conflict of interest.

© SIR, 2014

J Vasc Interv Radiol 2014; 25:1566–1572

http://dx.doi.org/10.1016/j.jvir.2014.06.030

the use of this therapy for pulmonary metastases from esophageal cancer. The purpose of this study was to evaluate retrospectively the outcomes of RF ablation for pulmonary metastases from esophageal cancer.

MATERIALS AND METHODS

RF ablation of lung tumors was performed after receiving informed consent from the patient and approval from the institutional review board. The institutional review board also approved this retrospective study.

Study Population

From April 2002 to March 2012, 26 patients with pulmonary metastases from esophageal cancer underwent lung RF ablation at our institution. The inclusion criteria for this study were as follows: (*i*) patients were medically unable to undergo surgery or refused to undergo surgery, (*ii*) all of the pulmonary metastases that were demonstrated on computed tomography (CT) in the initial RF ablation session were treated, and (*iii*) the follow-up period was > 6 months. The exclusion criteria were as follows: (*i*) intent to treat pulmonary metastases for palliation or cytoreduction or (*ii*) follow-up period < 6 months.

Based on the above-mentioned criteria, 4 patients who underwent RF ablation for palliation or cytoreduction and 1 patient whose follow-up period was < 6 months were excluded; the remaining 21 consecutive patients (all men; mean age, 66.0 y; age range, 44–85 y) were selected for the present study. Patient characteristics are listed in **Table 1**. Although the diagnosis of pulmonary metastases from esophageal cancer was mainly based on the results of serial CT studies, the diagnosis was confirmed with histologic examination in nine patients. ^{18}F-fluorodeoxyglucose positron emission tomography was performed in 14 patients. At the time of the initial RF ablation session, patients had one (n = 14), two (n = 4), three (n = 2), or four (n = 1) pulmonary metastases. One patient who had three tumors underwent RF ablation for two tumors. RF ablation for the remaining tumor was not feasible because the CT equipment was in the process of replacement, and stereotactic radiation therapy was performed as an alternative. There were 31 metastases treated with 27 RF ablation sessions. The mean size of the 31 tumors was 1.7 cm ± 0.9 (SD) (range, 0.6–4.1 cm). The mean size of the largest tumor per patient was 1.9 cm ± 0.8 (SD) (range, 0.7–4.1 cm).

In 12 of the 21 patients, there was no history of extrapulmonary recurrences until the initial RF ablation session. The other nine patients already had histories of extrapulmonary recurrences at the time of the initial RF ablation session. In two of these nine patients, extrapulmonary recurrences (abdominal or supraclavicular lymph node metastasis) had developed and had been treated with radiation therapy before the manifestation of pulmonary metastasis. In the other seven patients, extrapulmonary recurrences developed concurrently with pulmonary metastasis. Four of these seven patients had neck or mediastinal lymph node metastases and underwent radiation therapy or surgical resection for these extrapulmonary lesions before the initial RF ablation sessions. The other three patients had viable extrapulmonary recurrences at the time of the initial RF ablation sessions: one patient had local recurrence of the primary lesion, and two patients had neck lymph node metastasis. We undertook lung RF ablation in these cases because we hoped that control of lung metastases might contribute to patients' survival. The local recurrence of

Table 1. Characteristics of 21 Patients

Age (y)	
Mean (SD, range)	66.0 (8.8, 44–85)
Gender	
Male	21 (100.0%)
Female	0 (0.0%)
Histology of primary lesions	
Squamous cell carcinoma	20 (95.2%)
Basaloid squamous cell carcinoma	1 (4.8%)
Stage at primary therapy*	
I	2 (9.5%)
II	2 (9.5%)
III	13 (61.9%)
IVA/B	4 (19.0%)
Disease-free interval (mo)	
Median (range)	7.6 (1.9–29.0)
Time to pulmonary metastases from primary therapy (mo)	
Median (range)	9.0 (1.9–29.0)
Number of pulmonary metastases at the time of initial RF ablation sessions	
Single	14 (66.7%)
Multiple	7 (33.3%)
Largest size of ablated tumors (cm)	
Mean (SD, range)	1.9 (0.8, 0.7–4.1)
History of extrapulmonary recurrences	
Yes	9 (42.9%)
No	12 (57.1%)
Viable extrapulmonary recurrences at the time of initial RF ablation sessions	
Yes	3 (14.3%)
No	18 (85.7%)
Chemotherapy before RF ablation	
Yes	8 (38.1%)
No	13 (61.9%)
Chemotherapy after RF ablation	
Yes	15 (71.4%)
No	5 (23.8%)
Unknown	1 (4.8%)

RF = radiofrequency.
*Stages were classified based on the Japanese classification of esophageal cancer.

the primary lesion was surgically treated, and the neck lymph node metastases were treated with radiation after lung RF ablation. A history of pulmonary metastasectomy was present in 2 of the 21 patients before they underwent RF ablation.

Lung RF Ablation Techniques

Details of the RF ablation techniques that are used at our institution have been described previously in the literature (22). In each RF ablation session, one or two tumors in the unilateral lung were treated. When a patient had bilateral metastases, the lesions in each lung were treated at different sessions.

After local anesthetic had been administered along with conscious sedation, the electrode was percutaneously introduced into the tumor under CT fluoroscopy guidance and connected to a generator (CC-1; Covidien, Mansfield, Massachusetts; RF 2000 or RF 3000; Boston Scientific, Natick, Massachusetts). The electrodes used for ablation included a multitined expandable electrode with a 2-cm (seven sessions), 3-cm (seven sessions), or 3.5-cm (two sessions) diameter array (LeVeen; Boston Scientific) and a single internally cooled electrode with a 1-cm (two sessions), 2-cm (six sessions), or 3-cm (three sessions) noninsulated tip (Cool-tip; Covidien). Multitined expandable electrodes were preferred based on data that have shown a significantly better local efficacy compared with the efficacy of internally cooled electrodes (23). However, in the early years of performing RF ablation at our institution, internally cooled electrodes were used. In addition, internally cooled electrodes were used in cases in which a multitined expandable electrode was deemed unsuitable because of the tumor location and other relevant factors. When using the Covidien device, RF energy was applied with an impedance control algorithm for 12 minutes during internal cooling of the electrode. When using the Boston Scientific device, the energy was applied until the impedance showed a rapid increase or an automatic shut-off at 15 minutes; this was repeated once at each site. If necessary, multiple overlapping ablations were performed to obtain the ablative margin. In four procedures for the tumors close to pleura, an artificial pneumothorax was created for relief from severe pain because artificial pneumothorax has previously been demonstrated to be efficacious for pain relief (24).

Immediately after the procedure, CT scans of the entire lung were performed to assess the procedure. A posteroanterior upright chest radiograph was obtained 3 hours later and again the following morning.

Follow-up Examination

Whenever possible, the patients were followed 1, 3, 6, 9, and 12 months after the procedure and at 6-month intervals thereafter. Unless contraindicated, a chest CT scan with contrast medium was performed to assess the outcomes of RF ablation at each follow-up examination. ^{18}F-fluorodeoxyglucose positron emission tomography was also performed in 11 patients, although it was not used routinely.

The exact definition of local tumor progression has been described previously in the literature (22). Briefly, local progression of the tumor was indicated by the appearance of an irregular, scattered, nodular, or eccentric enhancement focus in the ablation zone or when the ablation zone was circumferentially enlarged with contrast enhancement. A new pulmonary tumor outside the ablation zone was considered to be an intrapulmonary de novo recurrence.

Study Endpoints

The primary study endpoints were patient survival and the determination of prognostic factors. To identify prognostic factors, a wide range of variables were analyzed, including age ($\geq$ 65 y or $<$ 65 y), disease-free interval (time from primary therapy to manifestation of any recurrence; $\geq$ 1 y or $<$ 1 y), time from primary therapy to pulmonary metastases ($\geq$ 1 y or $<$ 1 y), number of pulmonary metastases at the time of the initial RF ablation sessions (single or multiple), history of extrapulmonary recurrences, presence of viable extrapulmonary recurrences at the time of the initial RF ablation sessions, chemotherapy before RF ablation, and chemotherapy after RF ablation.

Secondary endpoints included local efficacy and safety of the treatment. To evaluate local efficacy, the presence of local tumor progression was determined on every follow-up CT scan using the diagnostic criteria specified earlier. Adverse events were evaluated according to the National Cancer Institute Common Terminology Criteria for Adverse Events Version 4.0 (25).

Statistical Analysis

Overall and disease-specific survival rates were estimated using the Kaplan-Meier method, from the initial RF ablation for pulmonary metastases to deaths from any cause and deaths related to esophageal cancer, respectively. The log-rank test was used to compare the survival rates of the two groups associated with each of the above-mentioned variables. If there was a statistically significant difference in survival between the two groups, the associated variable was considered to be a prognostic factor. For all analyses, P values $<$ .05 were considered to indicate a statistically significant difference. Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 22 (IBM Corp, Armonk, New York).

RESULTS

Patient Survival and Prognostic Factors

The median duration of follow-up after the initial RF ablation session was 22.4 months (range, 6.2–76.1 mo) for all 21 patients; it was 47.9 months (range, 14.3–76.1

mo) for 5 surviving patients. Survival outcomes of the 21 patients are summarized in Figure 1. Three patients who had viable extrapulmonary recurrences at the time of the initial RF ablation sessions died as a result of the progression of intrapulmonary or extrapulmonary recurrences within 2 years after RF ablation. Among the patients without viable extrapulmonary recurrences at the time of initial RF ablation sessions, 10 patients had experienced extrapulmonary recurrences after RF ablation and died of esophageal cancer. Seven of these patients also had developed intrapulmonary recurrences (de novo recurrence or local tumor progression). Three patients had not developed any recurrences after RF ablation and died of other causes (hypopharyngeal cancer [n = 2] or pneumonia [n = 1]) > 3 years after RF ablation. Among the five patients who survived the follow-up period, one patient showed intrapulmonary de novo recurrence after RF ablation, another developed extrapulmonary recurrences after RF ablation, and the remaining three had not developed any recurrence.

The estimated overall survival rates were 85.7% (95% confidence interval, 100%, 70.7%) at 1 year, 54.8% (95% confidence interval, 76.9%, 32.7%) at 2 years, and 38.4% (95% confidence interval, 60.3%, 16.4%) at 3 years after initial RF ablation session (Fig 2). The estimated disease-specific survival rates up to 3 years were identical to the overall survival rates. The median survival time was 24.5 months. The results of the univariate analyses of potential prognostic factors are presented in Table 2. The presence of viable extrapulmonary recurrences at the time of the initial RF ablation sessions was determined to be a negative prognostic factor ($P < .001$). None of the other variables was associated with a significant survival difference after RF ablation.

Local Efficacy

Local tumor progression was observed in 8 (25.8%) of the 31 tumors and occurred 2.6–10.0 months (median, 4.8 mo) after the initial RF ablation procedure. Three locally progressing tumors were treated with RF ablation; thereafter, local progression was observed again in one tumor. Two other locally progressing tumors were surgically resected, and another was treated with stereotactic radiation therapy. The remaining two tumors were either treated with chemotherapy or observed without any treatment. During the follow-up period, the proportion of local progression after all local therapy, including salvage surgery, stereotactic radiation therapy, and repeat RF ablation, was 9.7% (3 of 31 tumors).

Adverse Events

Adverse events are summarized in Table 3. One or more complications occurred after 16 (59.3%) of the 27 sessions. Pneumothorax occurred after 10 (37.0%) of the 27 sessions. Two (7.4%) pneumothorax cases required chest tube placement (grade 2), and the others were managed with observation only (grade 1). Intractable pneumoderma with pulmonary fistula occurred after two (7.4%) sessions: one case was surgically treated (grade 3), and the other was treated with placement of a drainage tube (grade 2). In the latter case, bronchiolitis obliterans organizing pneumonia–like reactive pneumonitis also developed after the procedure (grade 2); this case was previously described in the literature (26). Pleural effusion was encountered after eight (29.6%) sessions, one (3.7%) of which required chest tube placement (grade 3). Hemothorax occurred in two (7.4%) patients, both of whom were managed with observation only (grade 1). Consequently, grade 3 adverse events occurred in 2 (7.4%) of the 27 sessions. There were no adverse events of grade 4 or greater.

DISCUSSION

Palliative chemotherapy is the standard treatment for patients who have esophageal cancer with distant meta-

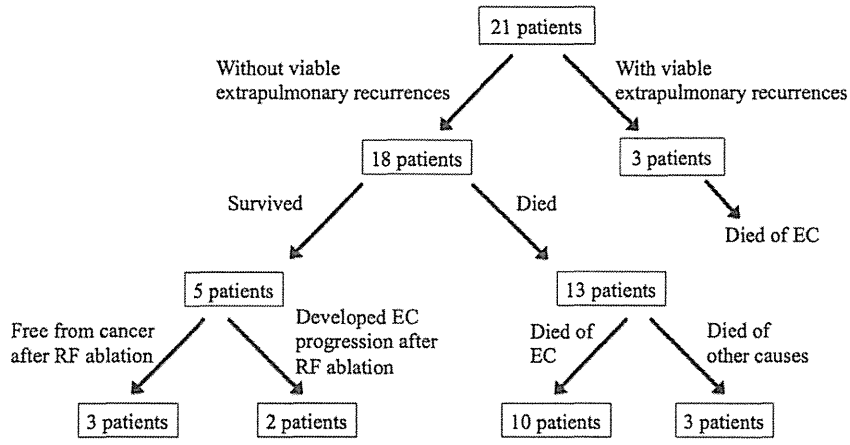

Figure 1. Summary of the survival outcomes of 21 patients with esophageal cancer (EC) who underwent RF ablation for pulmonary metastases.

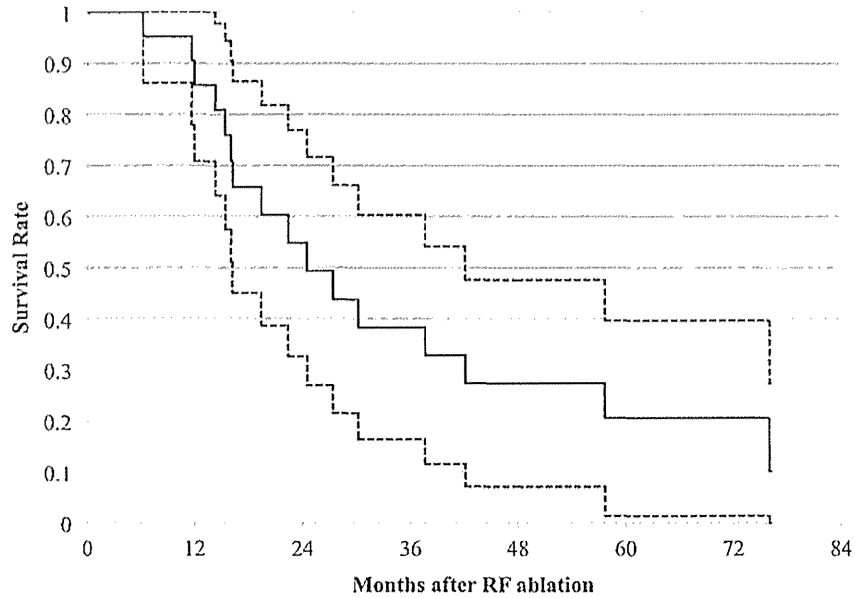

Figure 2. Overall survival after RF ablation for pulmonary metastases from esophageal cancer. Dotted lines indicate 95% confidence bands.

stasis. The combination of cisplatin and 5-fluorouracil has been widely used for metastatic esophageal squamous cell carcinoma, but various alternative regimens have been attempted more recently, including the use of oxaliplatin, nedaplatin, paclitaxel, capecitabine, and leucovorin (10–13). However, the median overall survival times remained approximately 8–15 months in the studies that have examined these new regimens (10–13). There is a need for alternative therapeutic options that can prolong patient survival.

Some researchers have discussed the use of surgical metastasectomy for pulmonary metastases from esophageal cancer. In a multiinstitution retrospective study that included 49 patients, Shiono et al (6) found a 29.6% overall 5-year survival rate and a 27-month median survival time after metastasectomy for pulmonary metastases from esophageal cancer. Ichikawa et al (7) also retrospectively studied 23 patients who underwent metastasectomy for pulmonary metastases from esophageal cancer, estimating that 1-, 3-, and 5-year survival rates were 73.9%, 43.5%, and 43.5% with a median survival time of approximately 29 months. Kozu et al (8) also observed a median survival time of 29 months after pulmonary metastasectomy in five patients. Similarly, Chen et al (9) found a mean survival time of 24 months after pulmonary metastasectomy in five patients. Takemura et al (27) showed that four of five patients who underwent surgical resection for solitary metastasis survived 32–124 months after metastasectomy.

When evaluating survival data after pulmonary metastasectomy in the above-mentioned studies, patient selection bias should be considered. Most of the patients included in these studies had limited numbers of metastases, and their general condition was good enough to undergo surgical treatment. Nonetheless, these reports of relatively favorable outcomes after pulmonary metastasectomy encouraged us to use local therapies for pulmonary metastases from esophageal cancer. The survival outcomes after lung RF ablation that were observed in the present study population appeared comparable to the reported survival outcomes after surgical metastasectomy. In addition, RF ablation offers some advantages, such as low invasiveness and minimal impairment of pulmonary function (28). A disadvantage of the procedure is that RF ablation may result in a substantial incidence of local progression. In the present study, local progression was observed within 1 year after the first RF ablation session in approximately a quarter of ablated tumors; the local efficacy resembled data obtained from prior examinations of primary lung cancer and metastases from colorectal cancer (20,29). However, repetition of RF ablation is possible after the detection of local progression, which may contribute to improvements in local tumor control (30).

To our knowledge, the outcomes of RF ablation for pulmonary metastases from esophageal cancer have been evaluated retrospectively by only one prior study (21). The authors assessed 10 patients with 17 pulmonary metastases from esophageal cancer. Their selection criteria restricted the study to cases with no extrapulmonary metastases and no more than three tumors. They observed that the feasibility of RF ablation was 100%, local tumor control for 1 year was 83%, and the estimated 1- and 2-year overall survival rates after RF ablation were 77.8% and 62.2%, respectively (21). The survival results from the current study are comparable to those reported.

Stereotactic radiation therapy is another alternative local treatment option for pulmonary metastases. We do not have definite criteria for selecting either RF ablation