volunteers reported in Kwan et al. (1974) and analyzed subsequently in Davidian & Giltinan (1995) and Pinheiro & Bates (2000). Each subject received an intravenous dose, $\mathrm{D} = 25mg \cdot kg^{-1}$. Then, indomethacin was labeled by a radioisotope $^{14}\mathrm{C}$, and the concentrations were observed at $n = 11$ times, using a radioactive counting device. Using a compartment model function, the concentration at time t is given by

$$f(t;\boldsymbol{\beta}) = \frac{\mathrm{D}(\kappa_1^* - \kappa_{21})}{V(\kappa_1^* - \kappa_2^*)} \exp(-\kappa_1^* t) + \frac{\mathrm{D}(\kappa_{21} - \kappa_2^*)}{V(\kappa_1^* - \kappa_2^*)} \exp(-\kappa_2^* t), \tag{10}$$

where $\boldsymbol{\beta} = (V, \kappa_{12}, \kappa_{21}, \kappa_e)^{\mathrm{T}}$, V is the volume of distribution, κ_{12} and κ_{21} are the transfer rate constants, and κ_e is the elimination rate constant. Here, $\kappa_1^* + \kappa_2^* = \kappa_e + \kappa_{12} + \kappa_{21}$, $\kappa_1^* \kappa_2^* = \kappa_e \kappa_{21}$, and $\kappa_1^* > \kappa_2^*$. This original function of Eq.(10) is derived from ordinary differential equations (Gibaldi & Perrier, 1982).

In practice, pharmacokinetic investigators often apply the parameterization for Eq.(10) to facilitate the compartment model:

$$f(t;\boldsymbol{\beta}^*) = \beta_1^* \exp(-\beta_2^* t) + \beta_3^* \exp(-\beta_4^* t), \tag{11}$$

where $\boldsymbol{\beta}^* = (\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*)^{\mathrm{T}}$. $\boldsymbol{\beta}^*$ has no direct pharmacokinetic interpretation. Furthermore, since the parameter $\boldsymbol{\beta}^*$ is positive, the following function can also be utilized:

$$f(t;\boldsymbol{\beta}^{**}) = \exp(\beta_1^{**}) \exp(-\exp(\beta_2^{**}) t) + \exp(\beta_3^{**}) \exp(-\exp(\beta_4^{**}) t), \tag{12}$$

where $\boldsymbol{\beta}^{**} = (\beta_1^{**}, \beta_2^{**}, \beta_3^{**}, \beta_4^{**})^{\mathrm{T}} = (\ln \beta_1^*, \ln \beta_2^*, \ln \beta_3^*, \ln \beta_4^*)^{\mathrm{T}}$.

The parameterized function of Eq.(12) was utilized in Davidian & Giltinan (1995), and the heteroscedastic model of Eq.(1) was applied with the variance function $g(t;\boldsymbol{\beta},\theta) = [f(t;\boldsymbol{\beta})]^{\theta}$ and $\boldsymbol{\Gamma}(\boldsymbol{\alpha}) = \boldsymbol{I}_n$ for an subject (No.5)'s concentration-time profile. To illustrate, we also focus on this subject, but of course data relating to other subjects could also have been considered. The nonlinearity underlying their models that are expressed in terms of compartment model functions, namely, the original function Eq.(10), the parameterized function Eq.(11) and the parameterized function Eq.(12), are assessed by using the relative curvature measure, and especially the performances of the parameterizations are evaluated. The log-likelihood is calculated by substituting the GLS-MPL estimates into Eq.(2). The results are shown in Table 2.

As shown in Table 2, the estimates of θ and $\tilde{\sigma}$, the log-likelihood values, and the values of the intrinsic curvature measure for all the functions of the compartment model coincides with one another. For fitting each of the functions of the compartment model, the value of the parameter-effects curvature measure is larger than that of the corresponding intrinsic curvature measure. All the RMS values of the intrinsic and parameter effects curvature measure are greater than $1/\sqrt{F_{4,11-4,1-0.05}} = 0.493$. This suggests that intrinsic nonlinearity as well as that due to parameter-effects might be quite large, and that the linear approximation and the results of inference based on it (for example, the parameter estimate and its standard error given here) might not be reliable. To reduce the intrinsic nonlinearity, we could reconsider the design of sampling times of blood drug concentration data, by using concepts of the optimal design. On the other hand, to reduce the

Table 2: Results for the indomethacin data.

Original parameterization: Eq.(10)	Estimate	Standard error
γ	6.788	1.154
κ_e	1.377	0.213
κ_{12}	1.100	0.169
κ_{21}	0.335	0.085
σ	0.131	
θ	0.819	
γ_{RMS}^{PE}		4.177
γ_{RMS}^{IN}		0.615
Parameterization 1: Eq.(11)	Estimate	Standard error
β_1^*	3.445	0.609
β_2^*	2.639	0.364
β_3^*	0.240	0.056
β_4^*	0.175	0.046
σ	0.131	
θ	0.819	
γ_{RMS}^{PE}		2.737
γ_{RMS}^{IN}		0.615
Parameterization 2: Eq.(12)	Estimate	Standard error
β_1^{**}	1.237	0.177
β_2^{**}	0.970	0.138
β_3^{**}	−1.428	0.232
β_4^{**}	−1.742	0.263
σ	0.131	
θ	0.819	
γ_{RMS}^{PE}		2.929
γ_{RMS}^{IN}		0.615
Log-likelihood		21.383

parameter-effects nonlinearity, we can find that the parameterization may be useful, as also shown in the results. Although all the values of the parameter-effects curvature measures are greater than 0.493, as compared with that of Eq.(10) the RMS parameter-effects curvature measure values for the parameterized functions of Eq.(11) and (12) are relatively small, and the difference between them are small (2.737 versus 2.929). Therefoer, parameterization 1 or 2 could be valid from the viewpoint of the linear approximation.

4. Discussion

Statistical inference in the compartment model is based on the linear approximation. In this article, the validity of linearization was investigated using the relative curvature measure, and especially the parameterization checking was focused on. We can recommend that pharmacokinetists or pharmacokinetic investigators check whether both of results of fitting the compartment models based on the linear approximation and applied parameterizations are reliable, by using the presented approach. A large value of the relative

curvature measure indicates that inference based on the linear approximation might not be valid, and that results obtained from the linear approximation, e.g., the parameter estimate, standard error, confidence region, or hypothesis test results, could be suspect. Furthermore, the comparison among the values of the parameter-effects curvature measure for the applied parameterizations provides a suggestion about suitable parameterizations. In other words, a suitable parameterization could reduce parameter-effects nonlinearity, thus improving the reliability of the results obtained from the linear approximation. As shown in our results, it is noted that a typical parameterization adopted for pragmatic reasons, for example, log-parameterization, was not always appropriate. To select the optimal parameterization in any data sets, cosequently, it will be necessary to compare the values of the parameter-effects curvature measure among the parameterizations used in various data sets. In order to understand the parameter-effects nonlinearity visually, it will be also so helpful to draw contour plots based on both likelihood ratio and linear approximation for the pair of parameters and to compare them (see Bates & Watts (1988) and Seber & Wild (1989)), and find out the relation between their plot's deviation and the values of the relative curvature measure.

To reduce intrinsic nonlinearity, we can reconsider the design of sampling times of blood drug concentration data, by using concepts of the optimal design. Daimon & Goto (2007b) have proposed the curvature adjusted design of sampling times. In addition, our results suggest that a recommended way to find intrinsic curvatures under various parameterizations is to refit the model with a different parameterization, then to recalculate derivatives curvatures based on new estimates. In fact, in Bates & Watts (1988), it is recommended to appropriately adjust the curvature obtained from the original fit. In other words, values for the relative curvature measure can vary considerably, depending on both parameter values and sampling times.

The linearization is also used to fit population pharmacokinetic/pharmacodynamic models. It would be interesting to know the implications of a curvature on parameter estimates, but the relative curvature measure for the mixed effects model has not been developed yet and it will take considerable effort. Thus in this article we have applied the relative curvature measure to only the fixed effects model, but our results could have provided a guideline for the parameterizations in the nonlinear mixed effects model.

REFERENCES

Aarons, L., Toon, S., & Rowland, M. (1987). Validation of assay methodology used in pharmacokinetic studies. *Journal of Pharmacological Methods* **17**, 337–346.

Bates, D.M., & Watts, D.G. (1980). Relative curvature measures of nonlinearity (with discussion). *Journal of the Royal Statistical Society, Series B* **42**, 1–25.

Bates, D.M., & Watts, D.G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Annals of Statistics* **9**, 1152–1167.

Bates, D.M., & Watts, D.G. (1988). *Nonlinear regression: analysis and its applications.* Wiley: New York.

Beal, S.L., & Sheiner, L.B. (1988). Heteroscedastic nonlinear regression. *Technometrics* **30**, 327–338.

Boeckmann, A.J., Sheiner, L.B., & Beal, S.L. (1994). *NONMEM user's guide, Part V, introductory guide.* NONMEM Project Group, University of California, San Francisco.

Daimon, T., & Goto, M. (2007a). Assessing nonlinearity in compartment models via the relative curvature measure. *Journal of Pharmacokinetics and Pharmacodynamics* **34**, 207–227.

Daimon, T., & Goto, M. (2007b). Curvature-adjusted optimal design of sampling times for the inference of pharmacokinetic compartment models. *Statistics in Medicine* **26**, 2799–2812.

Davidian, M., & Carroll, R.J. (1987). Variance function estimation. *Journal of the American Statistical Association* **50**, 74–82.

Davidian, M., & Giltinan, D. (1995). *Nonlinear models for repeated measurement data.* Chapman & Hall: London.

Gibaldi, M., & Perrier, D. (1982). *Pharmacokinetics.* Marcel Dekker: New York.

Giltinan, D.M., & Ruppert, D. (1989). Fitting heteroscedastic regression models to individual pharmacokinetic data and using standard statistical software. *Journal of Pharmacokinetics and Biopharmaceutics* **14**, 601–614.

Kwan, K.C., Breault, G.O., Umbenhauer, E.R., Mcmahon, F.G., & Duggan, D.E. (1976). Kinetics of indomethacin absorption, elimination, and enterohepatic circulation in man. *Journal of Pharmacokinetics and Biopharmaceutics* **4**, 255–280.

Oberg, A., & Davidian, M. Estimating data transformations in nonlinear mixed effects models. *Biometrics* **56**, 65–72.

Pinheiro, J.C., & Bates, D.M. (2000). *Mixed-effects models in S and S-PLUS.* Springer: New York.

Ratkowsky, D.A. (1983). *Nonlinear regression modeling.* Marcel Dekker: New York.

Seber, G.A.F., & Wild, C.J. (1989). *Nonlinear regression.* Wiley: New York.

Sheiner, L.B., & Beal, S.L. (1985). Pharmacokinetic parameter estimates from several least squares procedures: Superiority of extended least squares. *Journal of Pharmacokinetics and Biopharmaceutics* **13**, 185–201.

Vonesh, E.F., & Chinchilli, V.M. (1997). *Linear and nonlinear models for the analysis of repeated measurements.* Marcel Dekker: New York.

Wei, B.C. (1998). *Exponential family nonlinear models.* Springer: Singapore.

Yuh, L., Beal, S., Davidian, M., Harrison, F., Hester, A., Kowalski, K., Vonesh, E., & Wolfinger, R. (1994). Population pharmacokinetic/pharmacodynamic methodology and applications: a bibliography. *Biometrics* **50**, 566–575.

Appendix. Calculation for the relative curvature measure

The relative curvature measure is calculated by using the following algorithm (see Bates & Watts (1980, 1988) and Seber & Wild (1989)):

Step.1: Calculate the $n \times p$ first-derivative matrix of the function $\boldsymbol{f}^*(t;\boldsymbol{\beta})$ in Eq.(4), namely, $\dot{\mathbf{F}}^* = \Delta_{\boldsymbol{\beta}} \boldsymbol{f}^*(t;\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = \left[\left(\frac{\partial f^*(t_j;\boldsymbol{\beta})}{\partial \beta_k}\right)\Big|_{\beta_k=\hat{\beta}_k}\right] = \left[\left(\dot{f}^*_{jk}\right)\Big|_{\beta_k=\hat{\beta}_k}\right]$, $j = 1,\ldots,n,\ k = 1,\ldots,p$, evaluated at the parameter estimate $\hat{\boldsymbol{\beta}}$, and the $n \times p \times p$ second derivative array, namely, $\ddot{\mathbf{F}}^* = \nabla_{\boldsymbol{\beta}} \boldsymbol{f}^*(t;\boldsymbol{\beta})|_{\beta_k=\hat{\beta}_k,\beta_l=\hat{\beta}_l} = \left[\left(\frac{\partial f^*(t_j;\boldsymbol{\beta})}{\partial \beta_k \partial \beta_l}\right)\Big|_{\beta_k=\hat{\beta}_k,\beta_l=\hat{\beta}_l}\right] = \left[\left(\ddot{f}^*_{jkl}\right)\Big|_{\beta_k=\hat{\beta}_k,\beta_l=\hat{\beta}_l}\right]$, $j = 1,\ldots,n,\ k = 1,\ldots,p,\ l = 1,\ldots,p$. In addition, calculate $\ddot{\mathbf{F}}^{*PE}$ and $\ddot{\mathbf{F}}^{*IN}$. $\ddot{\mathbf{F}}^{*PE} = \left[\ddot{\mathbf{F}}^{*PE}_j\right] = \left[\left(\ddot{f}^{*PE}_{jkl}\right)\right]$, $j = 1,\ldots,n,\ k = 1,\ldots,p,\ l = 1,\ldots,p$ is the $n \times p \times p$ array, and $\ddot{\boldsymbol{f}}^{*PE}_{kl} = (\ddot{f}^{PE}_{1kl},\ldots,\ddot{f}^{*PE}_{nkl})^{\mathrm{T}}$ is the $n \times 1$ vector,

given by $\ddot{\boldsymbol{f}}_{kl}^{*PE} = \dot{\mathbf{F}}^{*}\left(\dot{\mathbf{F}}^{*\mathrm{T}}\dot{\mathbf{F}}^{*}\right)^{-1}\dot{\mathbf{F}}^{*\mathrm{T}}\ddot{\boldsymbol{f}}_{kl}^{*}$, where $\ddot{\boldsymbol{f}}_{kl}^{*} = (\ddot{f}_{1kl}^{*}, \ldots, \ddot{f}_{nkl}^{*})^{\mathrm{T}}$ On the other hand, $\ddot{\mathbf{F}}^{*IN} = \left[\ddot{\mathbf{F}}_{j}^{*IN}\right] = \left[\left(\ddot{f}_{jkl}^{*IN}\right)\right]$, $j = 1, \ldots, n$, $k = 1, \ldots, p$, $l = 1, \ldots, p$ is the $n \times p \times p$ array, and $\ddot{\boldsymbol{f}}_{kl}^{*IN} = (\ddot{f}_{1kl}^{*IN}, \ldots, \ddot{f}_{nkl}^{*IN})^{\mathrm{T}}$ is the $n \times 1$ vector, given by $\ddot{\boldsymbol{f}}_{kl}^{*IN} = \left\{\mathbf{I}_n - \dot{\mathbf{F}}^{*}\left(\dot{\mathbf{F}}^{*\mathrm{T}}\dot{\mathbf{F}}^{*}\right)^{-1}\dot{\mathbf{F}}^{*\mathrm{T}}\right\}\ddot{\boldsymbol{f}}_{kl}^{*}$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Step.2: Calculate the QR decomposition of $\dot{\mathbf{F}}^{*}$, that is,

$$\dot{\mathbf{F}}^{*} = \mathbf{Q}\mathbf{R}_1 = \left[\mathbf{Q}_p, \mathbf{Q}_{n-p}\right]\begin{bmatrix}\mathbf{R}_{11}\\ \mathbf{O}\end{bmatrix} = \mathbf{Q}_p\mathbf{R}_{11},$$

where $\mathbf{Q}$ is the $n \times n$ orthogonal matrix which rotates the axes of the sample space, however, $\mathbf{Q}^{\mathrm{T}}\mathbf{Q} = \mathbf{I}_n$, $\mathbf{R}_1$ is the $n \times p$ matrix, $\mathbf{O}$ is the $(n-p) \times p$ zero matrix, $\mathbf{Q}_p$ is the $n \times p$ matrix whose columns form a basis for the plane tangent to the expectation surface, $\mathbf{Q}_{n-p}$ is the $n \times (n-p)$ matrix, whose columns form a basis for vectors normal to the tangent plane, and $\mathbf{R}_{11}$ is the nonsingular $p \times p$ upper triangular matrix. Since $\mathbf{R}_{11}^{\mathrm{T}}\mathbf{R}_{11}$ is the Cholesky decomposition of $\dot{\mathbf{F}}^{*\mathrm{T}}\dot{\mathbf{F}}^{*}$, $\mathbf{R}_{11}$ is unique if its diagonal elements are all positive or all negative. If $\mathbf{R}_{11}$ is unique, $\mathbf{Q}_p$ is also unique. In fact, $\mathbf{Q}_p = \dot{\mathbf{F}}^{*}\mathbf{R}_{11}^{-1}$ (On the other hand, $\mathbf{Q}_{n-p}$ is not unique). Then, $\dot{\mathbf{F}}^{*}\boldsymbol{h} = \mathbf{Q}_p\mathbf{R}_{11}\boldsymbol{h} = \mathbf{Q}_p\boldsymbol{d}$, where $\boldsymbol{d} = \mathbf{R}_{11}\boldsymbol{h}$, $p \times 1$ vector. That is, $\mathbf{Q}_p$ in the new coordinate system plays the same role as the first derivative matrix $\dot{\mathbf{F}}^{*}$ in the original coordinate system.

Step.3: Scale the elements of the matrix $\dot{\mathbf{F}}^{*}$ and array $\ddot{\mathbf{F}}^{*}$ by $\rho = \hat{\sigma}\sqrt{p}$.

Step.4: Reparameterize γ_h^{PE} and γ_h^{IN} in Eq.(5) by $\boldsymbol{\phi} = \mathbf{R}_{11}\boldsymbol{\beta}$ and multiply their denominators and numerators by $\mathbf{Q}^{\mathrm{T}}$. We can ensure that $\|\boldsymbol{d}\| = 1$, since γ_h^{PE} and γ_h^{IN} depend on the direction of $\boldsymbol{d}$ and $\boldsymbol{h}$, but do not depend on their sizes. That is, since $\left\|\mathbf{Q}^{\mathrm{T}}\dot{\mathbf{F}}_{\phi}^{*}\mathbf{R}_{11}^{-1}\boldsymbol{d}\right\| = 1$,

$$\gamma_d^{PE} = \frac{\left\|\mathbf{Q}^{\mathrm{T}}\boldsymbol{h}^{\mathrm{T}}\ddot{\mathbf{F}}^{*PE}\boldsymbol{h}\right\|}{\left\|\mathbf{Q}^{\mathrm{T}}\dot{\mathbf{F}}^{*}\boldsymbol{h}\right\|^2} = \frac{\left\|\mathbf{Q}^{\mathrm{T}}\boldsymbol{d}^{\mathrm{T}}\left(\mathbf{R}_{11}^{-1}\right)^{\mathrm{T}}\ddot{\mathbf{F}}^{*PE}\mathbf{R}_{11}^{-1}\boldsymbol{d}\right\|}{\left\|\mathbf{Q}^{\mathrm{T}}\dot{\mathbf{F}}^{*}\mathbf{R}_{11}^{-1}\boldsymbol{d}\right\|^2} = \left\|\boldsymbol{d}^{\mathrm{T}}\ddot{\mathbf{F}}_{**}^{*PE}\boldsymbol{d}\right\|$$

and

$$\gamma_d^{IN} = \frac{\left\|\mathbf{Q}^{\mathrm{T}}\boldsymbol{h}^{\mathrm{T}}\ddot{\mathbf{F}}^{*IN}\boldsymbol{h}\right\|}{\left\|\mathbf{Q}^{\mathrm{T}}\dot{\mathbf{F}}^{*}\boldsymbol{h}\right\|^2} = \frac{\left\|\mathbf{Q}^{\mathrm{T}}\boldsymbol{d}^{\mathrm{T}}\left(\mathbf{R}_{11}^{-1}\right)^{\mathrm{T}}\ddot{\mathbf{F}}^{*IN}\mathbf{R}_{11}^{-1}\boldsymbol{d}\right\|}{\left\|\mathbf{Q}^{\mathrm{T}}\dot{\mathbf{F}}^{*}\mathbf{R}_{11}^{-1}\boldsymbol{d}\right\|^2} = \left\|\boldsymbol{d}^{\mathrm{T}}\ddot{\mathbf{F}}_{**}^{*IN}\boldsymbol{d}\right\|, \tag{13}$$

where $\ddot{\mathbf{F}}_{**}^{*PE}$ is the $p \times p \times p$ parameter-effects curvature array, given by $\ddot{\mathbf{F}}_{**}^{*PE} = \left(\mathbf{R}_{11}^{-1}\right)^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\ddot{\mathbf{F}}^{*PE}\mathbf{R}_{11}^{-1}$ and $\ddot{\mathbf{F}}_{**}^{*IN}$ is the $p' \times p \times p$ intrinsic curvature array (p' is at most $p(p+1)/2$), given by $\ddot{\mathbf{F}}_{**}^{*IN} = \left(\mathbf{R}_{11}^{-1}\right)^{\mathrm{T}}\mathbf{Q}^{\mathrm{T}}\ddot{\mathbf{F}}^{*IN}\mathbf{R}_{11}^{-1}$. However, it is noted that the following relation in the expansion of Eq.(13) is satisfied:

$$\mathbf{Q}^{\mathrm{T}}\ddot{\mathbf{F}}^{*} = \mathbf{Q}^{\mathrm{T}}\begin{bmatrix}\ddot{\mathbf{F}}^{*PE}\\ \ddot{\mathbf{F}}^{*IN}\end{bmatrix} = \begin{bmatrix}\mathbf{Q}_p^{\mathrm{T}}\ddot{\mathbf{F}}^{*PE}\\ \mathbf{Q}_{n-p}^{\mathrm{T}}\ddot{\mathbf{F}}^{*IN}\end{bmatrix} = \begin{bmatrix}\ddot{\mathbf{F}}_{**}^{*PE}\\ \ddot{\mathbf{F}}_{**}^{*IN}\end{bmatrix} = \ddot{\mathbf{F}}_{**}^{*},$$

where $\ddot{\mathbf{F}}^{*}_{**}$ is the $(p+p') \times p \times p$ curvature array.

The root mean square (RMS) parameter-effects curvature $\gamma^{PE}_{\mathrm{RMS}}$ and the RMS intrinsic curvature $\gamma^{IN}_{\mathrm{RMS}}$ is computed by, respectively,

$$\gamma^{PE}_{\mathrm{RMS}} = \sqrt{\frac{1}{p(p+2)} \sum_{j=1}^{p} \left[2 \sum_{k=1}^{p} \sum_{l=1}^{p} \left(\ddot{f}^{*}_{**,jkl} \right)^2 + \left(\sum_{k=1}^{p} \ddot{f}^{*}_{**,jkk} \right)^2 \right]}$$

and

$$\gamma^{IN}_{\mathrm{RMS}} = \sqrt{\frac{1}{p(p+2)} \sum_{j=p+1}^{p^*} \left[2 \sum_{k=1}^{p} \sum_{l=1}^{p} \left(\ddot{f}^{*}_{**,jkl} \right)^2 + \left(\sum_{k=1}^{p} \ddot{f}^{*}_{**,jkk} \right)^2 \right]},$$

where $\ddot{f}^{*}_{**,jkl}$ is the element of the curvature array $\ddot{\mathbf{F}}^{*}_{**}$ and $\ddot{f}^{*}_{**,jkk}$ is the diagonal element $p \times p$ matrix in the j th face or slice of the curvature array $\ddot{\mathbf{F}}^{*}_{**}$. The value of p^* lies between $p+1$ to at most $p(p+3)/2$.

(Received April 17 2007, Revised October 22 2007)

Communications in Statistics—Theory and Methods, 37: 508–519, 2008

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920701669777

Taylor & Francis
Taylor & Francis Group

Regression Analysis

The Mean Squared Error Optimum Design Criterion for Parameter Estimation in Nonlinear Regression Models

TAKASHI DAIMON[1] AND MASASHI GOTO[2]

[1]Division of Biostatistics, Department of Drug Evaluation and Informatics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan
[2]Non Profit Organization Biostatistical Research Association, Osaka, Japan

In the context of nonlinear regression models, we propose an optimal experimental design criterion for estimating the parameters that account for the intrinsic and parameter-effects nonlinearity. The optimal design criterion proposed in this article minimizes the determinant of the mean squared error matrix of the parameter estimator that is quadratically approximated using the curvature array. The design criterion reduces to the D-optimal design criterion if there are no intrinsic and parameter-effects nonlinearity in the model, and depends on the scale parameter estimator and on the reparameterization used. Some examples, using a well known nonlinear kinetics model, demonstrate the application of the proposed criterion to nonsequential design of experiments as compared with the D-optimal criterion.

Keywords *D*-optimality; Intrinsic or parameter-effects nonlinearity; Locally optimal designs; Quadratic approximation; Reparameterization.

Mathematics Subject Classification Primary 62K05, 65C60, 65D17; Secondary 46N55, 90C30, 94C30.

1. Introduction

It is common knowledge that a well designed experiment has some advantages including the ability to provide valid inferences as well as cost-effectiveness. In the context of nonlinear models, optimum experimental designs depend on unknown parameter values and are only locally optimum as the model is nonlinear with respect to the parameter, whereas in linear models, such designs are independent of parameter values (see Atkinson, 1982; Silvey, 1980; St. John and Draper, 1975,

Received October 13, 2006; Accepted May 11, 2007
Address correspondence to Takashi Daimon, Division of Biostatistics, Department of Drug Evaluation and Informatics, School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan; E-mail: daimon@u-shizuoka-ken.ac.jp

for general reviews). Therefore, for the nonlinear case, in particular, our interest often focuses on the design of experiments that can provide accurate or precise parameter estimation, and thus a number of nonlinear experimental design criteria have been proposed (for general reviews, Atkinson, 1996; Atkinson and Donev, 1992; Chaudhuri and Mykland, 1993; Cochran, 1973; Ford et al., 1989; Melas, 2006, etc.; in particular, an excellent and comprehensive review for general experimental designs has been presented from a historical viewpoint by Atkinson and Bailey, 2001). In addition, recent developments and applications are found in Atkinson (2003, 2005), Dette and Biedermann (2003), Müller and Pázman (2003), Dette et al. (2003, 2004, 2005), Han and Chaloner (2004), Uciński and Bogacka (2005), etc. However, most works on the design of experiments for nonlinear models follow the *D*-optimal design criterion proposed in the pioneering paper of Box and Lucas (1959). *D*-optimality minimizes the determinant of the information matrix, at a value close to or a preliminary guess for a true value of the parameter. Other interpretations of this criterion are that it minimizes the volume of linear-approximation joint confidence regions for the parameters and that it maximizes the posterior density for the parameter estimator if a noninformative prior distribution is used. Two other criteria that have a statistical interpretation in terms of the information matrix are *A*- and *E*-optimality (Atkinson and Donev, 1992). In *A*-optimality, the trace of the information matrix, or the average variance of the parameter estimates, is minimized. In *E*-optimality, the variance of the least well estimated contrast is minimized. An advantage of *D*-optimality is that the optimum designs for quantitative variables do not depend upon their scale. Linear transformations leave the *D*-optimum design unchanged, which is not in general the case for *A*- and *E*-optimum designs. In principle this is a serious drawback to the other two criteria: it seems undesirable that an optimum design should depend upon units of a variable.

Cochran (1973) noted the asymptotic nature of the *D*-optimal design criterion and invited studies of its small sample performance. Some researchers have responded to this invitation; Box (1971) derived an approximation to the bias of the least squares estimators and suggested designing experiments to minimize this bias of the estimator of a parameter of interest. Clarke (1980) presented an improved formula for the variance–covariance matrix of the parameter estimator and recommended selecting experimental designs to minimize the mean squared error of the estimator of one parameter. Bates and Watts (1981) proposed selecting designs to minimize the parameter-effects curvature to simplify inference procedures. Hamilton and Watts (1985) developed a quadratic approximation to the volume of the parameter inference region based on a curvature array. O'Brien (1992) used the quadratic approximation due to Hamilton and Watts (1985) to find designs which have additional design points. Hughes-Oliver (1998) took the quadratic approximation approach for nonlinear models with correlated errors under the limited scenario. Daimon and Goto (2003) proposed the curvature-adjusted criterion for optimal design of sampling times for the inference of the pharmacokinetic compartment models.

In this paper, we develop an optimum experimental design criterion that minimizes the "determinant" of the mean squared error matrix of the parameter estimator, which is quadratically approximated by the curvature array. In Sec. 2, the formula for the mean squared error optimum design criterion is presented. In Sec. 3, some applications of the design criterion to static designs are demonstrated with a typical nonlinear model, as compared with the *D*-optimal design criterion.

2. The Formula for the Mean Squared Error Optimum Experimental Design Criterion

Consider the nonlinear regression model

$$Y_j = f(x_j; \beta) + \varepsilon_j; \quad j = 1, \ldots, n, \tag{1}$$

where Y_j is the jth response, x_j is the $q \times 1$ vector of the controllable or design variables for the jth response, given by $x_j = (x_{j1}, \ldots, x_{jq})^T$, $f(x_j; \beta)$ is a known expectation function that depends nonlinearly on the p-vector of unknown parameters, given by $\beta = (\beta_1, \ldots, \beta_p)^T$, ε_j is the error that is usually assumed to be independently, identically, normally distributed with mean 0 and constant variance σ^2 as $N(0, \sigma^2)$, and n is the number of observations or experiments. The vector representation of (1) is

$$Y = f(x; \beta) + \varepsilon, \tag{2}$$

where Y is the $n \times 1$ random vector of the observations given by $Y = (Y_1, \ldots, Y_n)^T$, $f(x; \beta) = (f(x_1; \beta), \ldots, f(x_n; \beta))^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, and ε is assumed to be n dimensional multivariate normal distributed with mean vector 0 and variance–covariance matrix $\sigma^2 I_n$ as $MN_n(0, \sigma^2 I_n)$, where I_n is the $n \times n$ identity matrix. The maximum likelihood (least squares) parameter estimator $\hat{\beta}$ of β minimizes the sum of squares $S(\beta) = (Y - f(x; \beta))^T (Y - f(x; \beta)) = \sum_{j=1}^{n} (Y_j - f(x_j; \beta))^2$.

The estimation of β and σ in (2) is carried out based on the linearized approximation about $\hat{\beta}$:

$$f(x; \beta) \approx f(x; \hat{\beta}) + \dot{F}(\beta - \hat{\beta}), \tag{3}$$

where $\dot{F}$ is the $n \times p$ first derivative matrix with respect to β evaluated at $\hat{\beta}$ and is represented as $\dot{F} = \Delta_\beta f(x; \beta)|_{\beta=\hat{\beta}} = \left[\left(\frac{\partial f(x_j;\beta)}{\partial \beta_k}\right)\Big|_{\beta_k=\hat{\beta}_k}\right]$, $j = 1, \ldots, n$, $k = 1, \ldots, p$.

Chernoff (1953) introduced locally optimum designs in which the design may depend on the unknown values of the parameters and proposed the design criterion that minimizes the trace of $(\dot{F}^T \dot{F})^{-1}$. A disadvantage of this criterion is that the optimal designs are not invariant under changes of scale of the parameters. Box and Lucas (1959), in some nonlinear models arising in chemical or biological kinetics, proposed the D-optimal designs that minimize the determinant of $(\dot{F}^T \dot{F})^{-1}$ and are invariant under scale changes of the parameters. However, it should be noted that the above mentioned design criteria based on the linearized approximation in (3) are applicable to a large sample on the premise that the function of the nonlinear model is not misspecified and that the behavior about $\hat{\beta}$ is approximately or locally close to that in the linear model. If the linearized approximation is inappropriate for a given model-design combination, then $\hat{\beta}$ can be very seriously biased and the asymptotic results of the inference be invalid (see Box, 1971; Clarke, 1980; Hougaard, 1985). Therefore, as pointed out by Cochran (1973), it is necessary to develop an experimental design criterion that relies on an accurate higher-order approximation, rather than the linearized approximation. In addition, it is required for the designs to yield precise and accurate estimates of the parameters in the inference of nonlinear models.

As the validity of the linearized approximation quite strongly depends on the nonlinearity underlying the expectation function $f(x; \beta)$ of the nonlinear model, evaluated at parameter estimators, as pointed out by Hamilton and Watts (1985) and Hughes-Oliver (1998), it is sensible that we utilize the relative curvature measure of the nonlinearity developed by Bates and Watts (1980) to incorporate the higher-order approximation, and develop the curvature optimum design criterion.

The local behavior of the expectation surface of $f(x; \beta)$ can be described by the parameter-effects curvature array $\ddot{\mathrm{A}}^{PE}$ and the intrinsic curvature array $\ddot{\mathrm{A}}^{IN}$ (see Bates and Watts, 1988; Seber and Wild, 1989, and the Appendix). The parameter-effects curvature array measures the degree of nonlinearity depending on the parameters in the model. If each element of the curvature array is so small, it means that straight parallel equispaced lines in the parameter space map into ones in the expectation surface, as they do in the tangent plane. The parameter-effects nonlinearity can be decreased by suitable reparameterizations. On the other hand, the intrinsic curvature array measures the degree of nonlinearity inherent to the model itself. If each element of the intrinsic curvature array is small, it means that the expectation surface can be locally replaced by the tangent plane. But the intrinsic curvature cannot be reduced by reparameterizations.

In (2), the mean squared error of $\ddot{\beta}$ can be approximated by using the curvature array as follows (for details of the derivation and proof, see Clarke, 1980 and the Section 4.7 of Seber and Wild, 1989):

$$MSE(\hat{\beta}) = \sigma^2 \mathrm{R}_{11}^{-1}\left(\mathrm{I}_p + \sigma^2\left(\mathrm{L}^{IN} + \frac{1}{2}\mathrm{M}^{PE} + \frac{1}{4}\mathrm{N}^{PE}\right)\right)(\mathrm{R}_{11}^{-1})^{\mathrm{T}}, \tag{4}$$

where I_p is the $p \times p$ identity matrix, $\mathrm{L}^{IN} = \sum_{j=1}^{n-p}\left(\ddot{\mathrm{A}}_j^{IN}\right)^2$, where $\ddot{\mathrm{A}}_j^{IN}$ is the jth face of $\ddot{\mathrm{A}}^{IN}$ (for the definition of the term "face," see Bates and Watts, 1988; Seber and Wild, 1989, and the Appendix), $\mathrm{M}^{PE} = \left[\left(\mathrm{trace}\left\{\ddot{\mathrm{A}}_j^{PE}\ddot{\mathrm{A}}_{j^*}^{PE}\right\}\right)\right]$, $j, j^* = 1, \ldots, p$, and $\mathrm{N}^{PE} = \left[\left(\mathrm{trace}\left\{\ddot{\mathrm{A}}_j^{PE}\right\}\mathrm{trace}\left\{\ddot{\mathrm{A}}_{j^*}^{PE}\right\}\right)\right]$, $j, j^* = 1, \ldots, p$, are the $p \times p$ matrices, composed of the function of $\ddot{\mathrm{A}}_j^{PE}$, where $\ddot{\mathrm{A}}_j^{PE}$ is the jth face of $\ddot{\mathrm{A}}^{PE}$. The term including N^{PE} on the right hand of (4) is associated with the estimate of the bias of $\hat{\beta}$, and the remaining terms are associated with the estimate of the variance of $\hat{\beta}$.

Here, we can consider the mean squared error optimum experimental design criterion, *MSE-*D_{opt} that selects the designs, $x_1, \ldots, x_n$ to minimize the determinant of (4):

$$\left|\left(\dot{\mathrm{F}}^{\mathrm{T}}\dot{\mathrm{F}}\right)^{-1}\right|\left|\mathrm{I}_p + \sigma^2\left(\mathrm{L}^{IN} + \frac{1}{2}\mathrm{M}^{PE} + \frac{1}{4}\mathrm{N}^{PE}\right)\right|, \tag{5}$$

where the design region is the half-line $[0, \infty]$. For $n = p$, there is no intrinsic curvature ($\ddot{\mathrm{A}}^{IN} = 0$), and the expectation surface is flat, coinciding with the tangent plane. Note that if there is no curvature ($\ddot{\mathrm{A}}^{IN} = \ddot{\mathrm{A}}^{PE} = 0$) or $\sigma^2 = 0$, the *MSE-*D_{opt} reduces to the D-optimal design criterion, D_{opt} to minimize the determinant of the inverse of the asymptotic Fisher's information matrix of $\hat{\beta}$, namely $\left|\left(\dot{\mathrm{F}}^{\mathrm{T}}\dot{\mathrm{F}}\right)^{-1}\right|$. It should be noted that the *MSE-*D_{opt} also assumes that the nonlinear regression function $f(x; \beta)$ is known in advance from previous experiments, theoretical constructs, or other considerations. In addition, the *MSE-*D_{opt} requires that unknown true values β^* and σ^*, of β and σ in (4) are replaced by *a priori* initial

parameter values, $\tilde{\beta}$ and $\tilde{\sigma}$, respectively. We use a quasi-Newton algorithm to select the designs that minimize (5). The FORTRAN program codes for the *MSE-D*$_{opt}$ are available from the author upon request.

An advantage of the proposed *MSE-D*$_{opt}$ is that it could enables us to conduct sensitivity analysis with respect to the nonlinearity underlying a compartment model, leading to checking the D_{opt} designs or other designs or providing additional designs (see O'Brien, 1992). In particular, the *MSE-D*$_{opt}$ could be useful tool for such a check when the nonlinearity underlying the model function is quite high.

3. Applications of the Mean Squared Error Optimum Experimental Design Criterion

In this section, the *MSE-D*$_{opt}$ is applied to one of the most well known models for nonlinear experimental designs – the intermediate product model (see Atkinson and Donev, 1992; Bates and Watts, 1988; Box and Lucas, 1959; Hamilton and Watts, 1985; Seber and Wild, 1989, etc.). The model function describes the concentration of the intermediate substance created from two consecutive irreversible first-order biological or chemical kinetics; the model function can be expressed as follows:

$$f(x;\beta) = \frac{\beta_1}{\beta_1 - \beta_2}[\exp(-\beta_2 x) - \exp(-\beta_1 x)], \quad \beta_1, \beta_2, x > 0, \tag{6}$$

where $\beta = (\beta_1, \beta_2)^T$ and $p = 2$; x denotes a controllable variable. We investigate the effects of n, $\tilde{\sigma}^2$ and the parameterization.

3.1. *Changing n and* $\tilde{\sigma}^2$

The effects of changing the initial scale parameter value $\tilde{\sigma}^2$ and the number of experiments n on the *MSE-D*$_{opt}$ are investigated. Box and Lucas (1959) used the initial parameter value $\tilde{\beta} = (0.7, 0.2)^T$ in (6) and found that the local D_{opt} two-point design was $x_D = (1.26, 6.86)^T$; these were used as the starting values. Here, $n = 2, \ldots, 9$ and $\tilde{\sigma}^2 = 0$, 0.05, 0.10, 0.15, 0.20, 0.25. However, for $n > 2$, starting designs were randomly sampled from (1.26, 6.86); for example, for $n = 3$, they were randomly selected out from (1.26, 6.86) as one combination such as (1.26, 1.26, 6.86), (1.26, 6.86, 6.86), etc. Table 1 shows the *MSE-D*$_{opt}$ designs for different combinations of n and $\tilde{\sigma}^2$. For $\tilde{\sigma}^2 = 0$, the *MSE-D*$_{opt}$ design coincides with the D_{opt} design. For $n > 2$, irrespective of $\tilde{\sigma}^2$, in every starting design, the procedure for the minimization of (5) terminated at a replicated two-point design; for example, for $n = 4$ and $\tilde{\sigma}^2 = 0.1$, the designs were selected as (1.201, 1.201, 6.646, 6.646), and for $n = 5$ and $\tilde{\sigma}^2 = 0.15$, they were selected as (1.175, 1.175, 1.175, 6.465, 6.465), (1.175, 1.175, 6.465, 6.465, 6.465), etc. That is to say, like the D_{opt} design, in the *MSE-D*$_{opt}$, it would be equal for the even number of designs and for the odd number of designs, with the choice the 'extra' observation could be taken at either design point, but the *MSE-D*$_{opt}$ can yield the selection of designs adjusted, depending on n and $\tilde{\sigma}^2$, for the D_{opt}, which can always do (1.229, 6.858); for example, for $n = 4$, (1.229, 1.229, 6.858, 6.858), and for $n = 5$, (1.229, 1.229, 6.858, 6.858, 6.858), (1.229, 1.229, 1.229, 6.858, 6.858), etc. As shown in these results, the fact that replications of the optimal p-point design are optimal or near-optimal (see Atkinson and Donev, 1992) would be applied to the *MSE-D*$_{opt}$.

Table 1
Design points selected by the *MSE-D_{opt}* for various combinations of n and $\tilde{\sigma}^2$

$\tilde{\sigma}^2$	$n=2$	$n=3$	$n=4$	$n=5$
0	(1.229,6.858)	(1.229,6.858)	(1.229,6.858)	(1.229,6.858)
0.05	(1.134,6.673)	(1.216,6.750)	(1.222,6.799)	(1.222,6.804)
0.1	(1.049,6.378)	(1.182,6.490)	(1.201,6.646)	(1.203,6.659)
0.15	(0.984,6.087)	(1.138,6.203)	(1.172,6.448)	(1.175,6.465)
0.2	(0.932,5.851)	(1.091,5.975)	(1.138,6.245)	(1.144,6.265)
0.25	(0.889,5.677)	(1.047,5.816)	(1.104,6.062)	(1.110,6.086)
$\tilde{\sigma}^2$	$n=6$	$n=7$	$n=8$	$n=9$
0	(1.229,6.858)	(1.229,6.858)	(1.229,6.858)	(1.229,6.858)
0.05	(1.223,6.815)	(1.225,6.821)	(1.225,6.821)	(1.226,6.832)
0.1	(1.208,6.711)	(1.211,6.719)	(1.211,6.722)	(1.216,6.758)
0.15	(1.188,6.560)	(1.193,6.575)	(1.192,6.575)	(1.202,6.652)
0.2	(1.163,6.393)	(1.170,6.453)	(1.168,6.440)	(1.184,6.499)
0.25	(1.136,6.229)	(1.140,6.248)	(1.154,6.338)	(1.163,6.359)

In addition, local minima were found in some cases, in particular, for large n ($n = 7$, 8, or 9) and $\tilde{\sigma}^2$ ($\tilde{\sigma}^2 = 0.15$, 0.20, or 0.25); therefore, a smaller value for (5) was sometimes obtained by starting with a replicated design that is different from x_D. As shown in these results, we can exploit the fact that replications of the optimal p-point designs for general models are optimal or near-optimal (see Atkinson and Hunter, 1968; Box, 1968, 1970), and the *MSE-D_{opt}* can be applied for a replicated p-point design.

Figure 1 shows the behavior of the design points selected by the *MSE-D_{opt}* for x_1 and x_2 with a continuous change in $\tilde{\sigma}^2$ for each value of n. For each n, both the design points selected by the *MSE-D_{opt}* decrease with an increase in $\tilde{\sigma}^2$; the design point for x_1 decreases slowly, whereas the design point for x_2 decreases remarkably. This is because the *MSE-D_{opt}* accounts for the underlying nonlinearity in the model with an increase in $\tilde{\sigma}^2$. As n is larger, the decrease of the design point for x_2 becomes smaller. Each curve of the design points for x_1 and x_2 tends to approach a specialized one. This behavior may result from the effect of the intrinsic nonlinearity relative to the parameter-effects nonlinearity with an increase in n.

Figure 2 shows the relative percentage efficiency of the D_{opt} to the *MSE-D_{opt}* for $n = 2$, which is calculated by $100 \times [(\text{Minimum of (5) for } \tilde{\sigma}^2)/(\text{Minimum of (5) for } \tilde{\sigma}^2 = 0)]^{1/p}$, where the power factor accounts for the number of parameters.

As shown in Figure 2, the relative efficiency decreases in proportion with an increase in $\tilde{\sigma}^2 > 0.05$.

3.2. *Changing $\tilde{\beta}$*

For $n = 2$, the *MSE-D_{opt}* two-point design for $\tilde{\sigma}^2 = 0.05$ is compared with the local D_{opt} design, for a range of parameter values $\tilde{\beta}$. The initial parameter values are

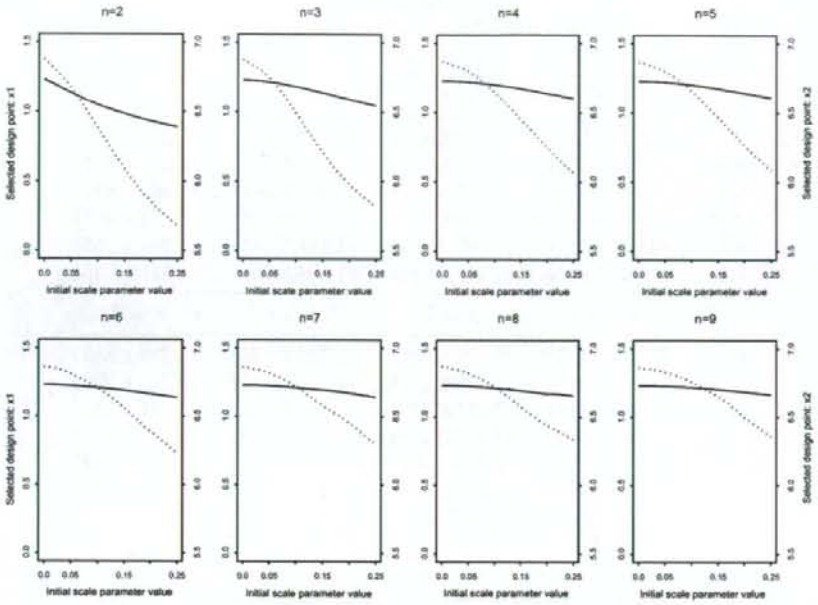

Figure 1. Changes in the design points due to changes in n and $\tilde{\sigma}^2$; the solid and dotted lines represent the design points for x_1 and x_2, respectively.

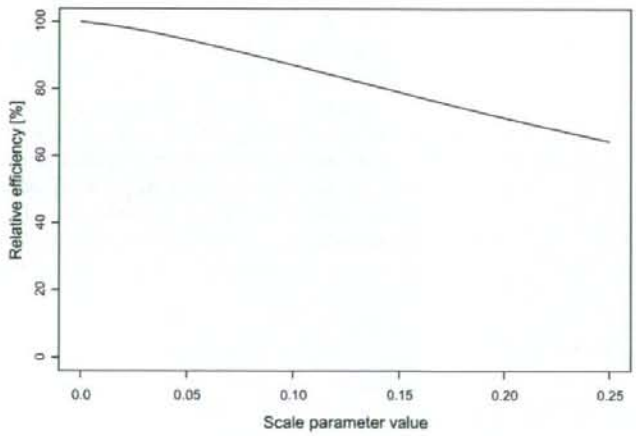

Figure 2. Efficiency of the D_{opt} relative to the $MSE\text{-}D_{opt}$; $n = 2$.

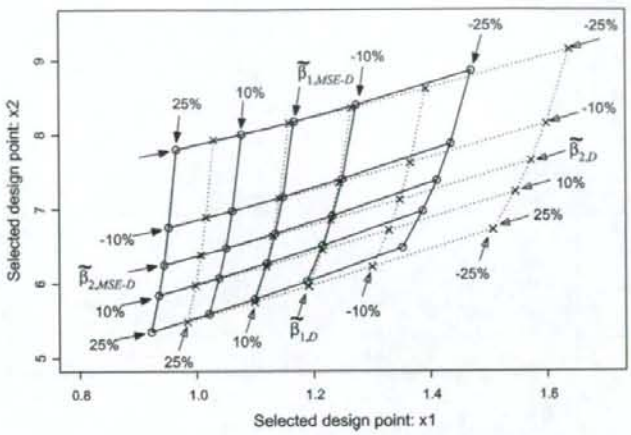

Figure 3. Changes in the design points due to changes in $\tilde{\beta}$; the solid and dotted lines represent the design points for $MSE\text{-}D_{opt}$ and D_{opt}, respectively.

on a grid that was generated by changing each value of $\tilde{\beta} = (0.7, 0.2)^{\mathrm{T}}$ by $\pm 10\%$ and $\pm 25\%$. The locations of the optimal designs are shown in Figure 3. For both the criteria, increasing the initial values of β_1 and β_2 reduces the optimal values for x_1 and x_2. An increase in $\tilde{\beta}_1$ causes a greater relative decrease in x_1 than in x_2, whereas an increase in $\tilde{\beta}_2$ causes a greater relative decrease in x_2 than in x_1. For every combination of the initial values, the optimal values of both x_1 and x_2 are smaller for the $MSE\text{-}D_{opt}$ for $\tilde{\sigma}^2 = 0.05$ than for the D_{opt}. Overall, the $MSE\text{-}D_{opt}$ is slightly less sensitive to changes in the initial parameter values than the D_{opt}. It is evident that the $MSE\text{-}D_{opt}$ is less influenced by incorrect initial values than the D_{opt}.

3.3. *Changing the Parameterization*

The $MSE\text{-}D_{opt}$ contains the terms M^{PE} and N^{IN}, as shown in (5); thus, it depends on the parameterization applied to the expectation function. On the other hand, the D_{opt} does not show such dependency. Here, we can consider typical three parameterizations used for (5):

a. $\phi_1 = \beta_1, \phi_2 = \beta_2/\beta_1$;
b. $\phi_1 = \log \beta_1, \phi_2 = \log \beta_2$; and
c. $\phi_1 = [\log(\beta_1/\beta_2)]/(\beta_1 - \beta_2), \phi_2 = \exp(-\beta_2\phi_1)$,

where the parameters in case c correspond to the maximum $\mathrm{E}(Y)$ $(=\phi_2)$ and to the time of occurrence $-x = \phi$. To illustrate the effects of these parameterizations for $n = 2$, the $MSE\text{-}D_{opt}$ designs were selected with initial parameter values corresponding to $\tilde{\beta} = (0.7, 0.2)^{\mathrm{T}}$ and with an increase in $\tilde{\sigma}^2$, ranging from 0 to 0.25.

Graphs of the design points selected by the $MSE\text{-}D_{opt}$ are plotted in Figure 4. An increase in $\tilde{\sigma}^2$ causes the $MSE\text{-}D_{opt}$ design to move away from the D_{opt} $(\tilde{\sigma}^2 = 0)$ designs in different directions, depending on the parameterization. This dependence on the parameterization of $\tilde{\sigma}^2$ is the least for case b. The design for case a is more

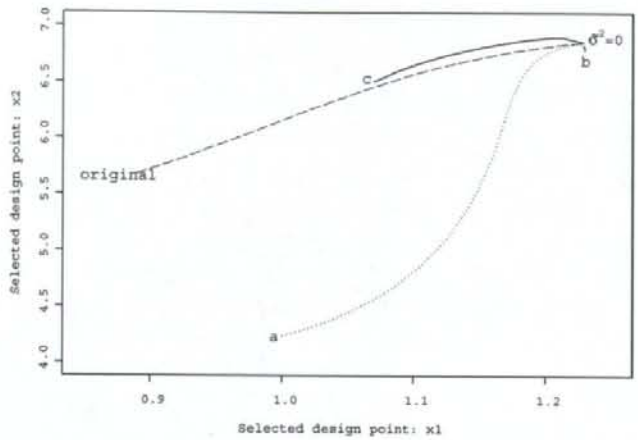

Figure 4. Changes in the design points due to changes in parameterization.

remarkably decreased than that for the original. On the other hand, for case c, the design point for x_2 decreases with a slight increase in $\tilde{\sigma}^2$ when compared with that for $\tilde{\sigma}^2 = 0$; in contrast, the design point for x_1 is decreased.

4. Discussion

We have proposed the mean squared error optimum design criterion, *MSE-*D_{opt}, and evaluated its performance as compared with that of the *D*-optimal design criterion, D_{opt}. Similar to D_{opt} designs, *MSE-*D_{opt} designs are influenced by the initial parameter values $\tilde{\beta}$, based on the premise that the true expectation function can be specified *a priori*. The example shows that the *MSE-*D_{opt} is slightly less sensitive to changes in the initial parameter values than is the D_{opt}. Consequently, inaccurate specification of the initial parameter values may lead to a smaller loss in efficiency for the *MSE-*D_{opt} than for the D_{opt}. It is possible that the *MSE-*D_{opt} is not very sensitive to the variation of the initial parameter values. In this article, we assume that the parameters are fixed, and our interest is limited to locally optimum designs; however, construction of the *MSE-*D_{opt} in the framework of the other approaches including sequential, minimax, and Bayesian designs will be possible (see Atkinson, 2003, 2005; Dette and Biedermann, 2003; Dette et al., 2003, 2004, 2005; Han and Chaloner, 2004; Müller and Pázman, 2003, etc.). For example, as pointed out by Atkinson and Bailey (2001), we could assume that the parameters are taken to be sampled from the prior once, for all experimental units, and consider an extension of the *MSE-*D_{opt} that takes account of the variation or uncertainty with regard to the parameters in the Bayesian approach or in a framework of random effects nonlinear regression models. Mukhopadhyay and Haines (1995) investigated Bayesian *D*-optimal designs for the exponential growth model. Mentré et al. (1997) presented *D*-optimal designs for random effects nonlinear models with applications to pharmacokinetics/toxicokinetics.

Another common feature of the *MSE-*D_{opt} and D_{opt} is the occurrence of replications. As pointed out by Hamilton and Watts (1985) and other authors, a

design consisting only of replicates at the p points provides no information about the lack of fit. However, as shown in the examples, at least, with regard to the *MSE-*D_{opt}, an increase in the sample size can lead to taking into account the intrinsic nonlinearity and adjusting the designs although they are replicated.

The major features that distinguish the *MSE-*D_{opt} from the D_{opt} are its sensitivity to the scale parameter value, the number of experiments, and the parameterization. The examples have shown that these three factors can cause the location of the *MSE-*D_{opt} design to be quite different from the D_{opt}. In contrast, the D-optimal criterion is unaffected by any of these factors because the linearized approximation of the model fails to take into account the intrinsic and parameter-effects nonlinearity.

With an increase in $\tilde{\sigma}^2$, the gain in efficiency of the *MSE-*D_{opt} is more than that of the D_{opt}. A simple way to determine whether a given initial value of the scale parameter is too large for the D_{opt} to be safely used, has not been found; however, the *MSE-*D_{opt} takes into consideration the degree of the nonlinearity in the model and produces reasonable results for a wider range of scale parameter value. The *MSE-*D_{opt} would be helpful when the nonlinearity in the model is quite high.

Appendix A

Curvature array. The relative curvature measure is given by assigning its role to the parameter-effects curvature and intrinsic curvature, to describe the local behavior of the expectation surface of $f(x; \beta)$. Following the notation in Bates and Watts (1988) and Seber and Wild (1989), the parameter-effects curvature array $\ddot{\mathrm{A}}^{PE}$ and the intrinsic curvature array $\ddot{\mathrm{A}}^{IN}$ are respectively defined, in (2), by

$$\ddot{\mathrm{A}}^{PE} = [\mathrm{Q}_p^{\mathrm{T}}][(\mathrm{R}_{11}^{-1})^{\mathrm{T}}\ddot{\mathrm{F}}(\mathrm{R}_{11}^{-1})] \quad \text{and} \quad \ddot{\mathrm{A}}^{IN} = [\mathrm{Q}_{n-p}^{\mathrm{T}}][(\mathrm{R}_{11}^{-1})^{\mathrm{T}}\ddot{\mathrm{F}}(\mathrm{R}_{11}^{-1})], \tag{A.7}$$

where $\ddot{\mathrm{F}}$ is the $n \times p \times p$ three-dimensional array of second derivatives of $f(x; \beta)$ with respect to β evaluated at $\hat{\beta}$, given by $\ddot{\mathrm{F}} = \nabla_\beta f(x; \beta)|_{\beta=\hat{\beta}} = \{(\frac{\partial f(x_j;\beta)}{\partial \beta_k \partial \beta_l}\big|_{\beta=\hat{\beta}})\}$, $j = 1, \ldots, n, k = 1, \ldots, p, l = 1, \ldots, p$, and Q_p, Q_{n-p}, and R_{11} are obtained from the QR decomposition of $\dot{\mathrm{F}}$:

$$\dot{\mathrm{F}} = \mathrm{QR}_1 = \left(\underbrace{\mathrm{Q}_p}_{n\times p} \mid \underbrace{\mathrm{Q}_{n-p}}_{n\times(n-p)} \right) \begin{pmatrix} \mathrm{R}_{11} \\ \mathrm{O} \end{pmatrix} \begin{matrix} \}p \times p \\ \}(n-p) \times p \end{matrix} = \mathrm{Q}_p \mathrm{R}_{11},$$

where $(\cdot|\cdot)$ denotes a block-partitioning of the matrix, Q is an $n \times n$ orthogonal matrix that plays a role of rotating the axes of the sample space and so $\mathrm{Q}^{\mathrm{T}}\mathrm{Q} = \mathrm{I}_n$ (I_n is the $n \times n$ identity matrix), R_1 is an $n \times p$ matrix, Q_p is an $n \times p$ matrix, whose columns form a basis for the tangent plane parallel to the expectation surface, Q_{n-p} is an $n \times (n-p)$ matrix, whose columns form a basis for vectors perpendicular to the tangent plane, R_{11} is a nonsingular $p \times p$ upper triangular matrix, and O is an $(n-p) \times p$ zero matrix. Since $\mathrm{R}_{11}^{\mathrm{T}}\mathrm{R}_{11}$ is the Cholesky decomposition of $\dot{\mathrm{F}}^{\mathrm{T}}\dot{\mathrm{F}}$, R_{11} is unique if its diagonal elements are all positive or all negative.

In (A.7) we notice the relation between the term "array" and the operation "$[\cdot]$", square-bracket multiplication (see Seber and Wild, 1989, pp. 142, 691–692). For example, let us consider the $n \times p \times p$ three-dimensional array $\ddot{\mathrm{W}} = \{(\mathrm{w}_{kl})\}$ made up of a $p \times p$ matrix of n-dimensional vectors, $\mathrm{w}_{kl} = (\mathrm{w}_{1kl}, \ldots, \mathrm{w}_{nkl})^{\mathrm{T}}$;

$k = 1, \ldots, p, l = 1, \ldots, p$. If w_{jkl} is the jth element of w_{kl}, then the matrix of j^{th} elements $\ddot{W}_j = [(w_{jkl})]$ is called the jth face of $\ddot{W}$. Now, two types of multiplication are defined. Firstly, if B and C are $p \times p$ matrices, then

$$\ddot{V} = \{(v_{kl})\} = B\ddot{W}C$$

denotes the array with jth face $\ddot{V}_j = B\ddot{W}_jC$, i.e.,

$$v_{kl} = \sum_{k'} \sum_{l'} b_{kk'} w_{k'l'} c_{l'l},$$

where $b_{kk'}$ and $c_{l'l}$ denote the (k, k')th element of B and the (l', l)th element of C, respectively. Secondly, if D is a $q \times n$ matrix, then we define square-bracket multiplication by the equation

$$[D][\ddot{W}] = \{(Dw_{kl})\},$$

where the right-hand side is a $q \times p \times p$ array.

References

Atkinson, A. C. (1982). Developments in the design of experiments. *Rev. Int. Statist. Inst.* 50:161–177.

Atkinson, A. C. (1996). The usefulness of optimum experimental designs (with discussion). *J. R. Statist. Soc. B* 58:59–76.

Atkinson, A. C. (2003). Horwitz's rule, transforming both sides and the design of experiments for mechanistic models. *Appl. Statist.* 52:261–278.

Atkinson, A. C. (2005). Robust optimum designs for transformation of the responses in a multivariate chemical kinetic model. *Technometrics* 47:478–487.

Atkinson, A. C., Hunter, W. G. (1968). The design of experiments for parameter estimation. *Technometrics* 10:271–289.

Atkinson, A. C., Donev, A. N. (1992). *Optimum Experimental Designs*. London: Oxford University Press.

Atkinson, A. C., Bailey, R. A. (2001). One hundred years of the design of experiments on and off the pages of Biometrika. *Biometrika* 88:53–97.

Bates, D. M., Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. R. Statist. Soc. B* 42:1–25.

Bates, D. M., Watts, D. G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Statist.* 9:1152–1167.

Bates, D. M., Watts, D. G. (1988). *Nonlinear Regression: Analysis and Its Applications*. New York: Wiley.

Bates, D. M., Hamilton, D. C., Watts, D. G. (1983). Calculation of intrinsic and parameter-effects curvatures for nonlinear regression models. *Commun. Statist. Theor. Meth.* 12:469–477.

Box, M. J. (1968). The occurrence of replications in optimal designs of experiments to estimate parameters in non-linear models. *J. R. Statist. Soc. B* 30:290–302.

Box, M. J. (1970). Some experiences with a nonlinear experimental design criterion. *Technometrics* 12:569–589.

Box, M. J. (1971). Bias in nonlinear estimation (with discussion). *J. R. Statist. Soc. B* 33:171–201.

Box, G. E. P., Lucas, H. L. (1959). Design of experiments in non-linear situations. *Biometrika* 46:77–90.

Chaudhuri, P., Mykland, P. A. (1993). Nonlinear experiments: Optimal design and inference based on likelihood. *J. Am. Statist. Assoc.* 88:538–546.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Statist.* 24:586–602.

Clarke, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model. *J. R. Statist. Soc. B* 42:227–237.

Cochran, W. G. (1973). Experiments for nonlinear functions. *J. Am. Statist. Assoc.* 44:32–61.

Daimon, T., Goto, M. (2003). Optimum design of sampling times for inference of compartment models. *J. Jpn. Soc. Comp. Stat.* 16:23–38.

Dette, H., Biedermann, S. (2003). Robust and efficient design for the Michaelis–Menten model. *J. Am. Statist. Assoc.* 98:679–686.

Dette, H., Melas, V., Pepelyshev, A., Strigul, N. (2003). Efficient design of experiments in the Monod model. *J. R. Statist. Soc. B* 65:725–742.

Dette, H., Melas, V., Pepelyshev, A. (2004). Optimal designs for a class of nonlinear regression models. *Ann. Statist.* 32:2142–2167.

Dette, H., Melas, V., Wong, W. K. (2005). Optimal design for goodness-of-fit of the Michaelis–Menten enzyme kinetic function. *J. Am. Statist. Assoc.* 100:1370–1381.

Ford, I., Titterington, D. M., Kitsos, C. P. (1989). Recent advances in nonlinear experimental design. *Technometrics* 31:49–60.

Hamilton, D. C., Watts, D. G. (1985). A quadratic design criterion for precise estimation in nonlinear regression models. *Technometrics* 27:241–250.

Han, C., Chaloner, K. (2004). Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. *Biometrics* 60:25–33.

Hougaard, P. (1985). The appropriateness of the asymptotic distribution in a nonlinear regression model in relation to curvature. *J. R. Statist. Soc. B* 47:103–114.

Hughes-Oliver, J. M. (1998). Optimal designs for nonlinear models with correlated errors. In: Flournoy, N., Rosenberger, W. F., Wong, W. K., eds. *New Developments and Applications in Experimental Design*. Institute of Mathematical Statistics. Lecture Notes—Monograph Series No. 34, Hayward, California, pp. 163–174.

Melas, V. B. (2006). *Functional Approach to Optimal Experimental Design* (Lecture Notes in Statistics). New York: Springer.

Mentré, F., Mallet, A., Baccar, D. (1997). Optimal design in random-effects regression models. *Biometrika* 84:429–442.

Mukhopadhyay, S., Haines, L. M. (1995). Bayesian *D*-optimal designs for the exponential growth model. *J. Statist. Planng. Inf.* 44:385–397.

Müller, W. G., Pázman, A. (2003). Measures for designs in experiments with correlated errors. *Biometrika* 84:429–442.

O'Brien, T. E. (1992). A note on quadratic designs for nonlinear regression models. *Biometrika* 79:847–849.

Seber, G. A. F., Wild, C. J. (1989). *Nonlinear Regression*. New York: Wiley.

Silvey, S. D. (1980). *Optimum Design*. London: Chapman and Hall.

St. John, R. C., Draper, N. R. (1975). *D*-optimality for regression designs: a review. *Technometrics* 17:15–23.

Uciński, D., Bogacka, B. (2005). *T*-optimum designs for discrimination between two multiresponse dynamic models. *J. R. Statist. Soc. B* 67:3–18.

British Journal of Cancer (2008) 98, 1039–1045

www.bjcancer.com

Intra-arterial chemoradiotherapy for locally advanced oral cavity cancer: analysis of therapeutic results in 134 cases

N Fuwa*[,1], T Kodaira[1], K Furutani[1], H Tachibana[1], T Nakamura[1], R Nakahara[1], T Tomoda[1], H Inokuchi[1] and T Daimon[2]

[1]Department of Radiation Oncology, Aichi Cancer Center Hospital, Nagoya, Japan; [2]Division of Drug Evaluation & Informatics, Graduate School of Pharmaceutical Sciences, University of Shizuoka, Shizuoka, Japan

The objective of this study was to investigate the therapeutic results of arterial injection therapy via the superficial temporal artery for 134 cases of stages III and IV (M0) oral cavity cancer retrospectively, and to clarify the prognostic factors. We administered intra-arterial chemoradiotherapy by continuous infusion of carboplatin in 65 cases from January 1993 to July 2002. Systemic chemotherapy was performed on 26 cases at the same time. We administered intra-arterial chemoradiotherapy by cisplatin with sodium thiosulphate in 69 cases from October 2002 to December 2006. Systemic chemotherapy was performed on 48 cases at the same time. The 3-year local control rate was 68.6% (T2-3: 77.9%; T4: 51.3%), and the 3-year survival rate was 53.9% (stage III: 62.9%; stage IV: 45.3%). Regarding the results of multivariate analysis of survival rates, age (<65), selective intra-arterial infusion, and the use of cisplatin as an agent for intra-arterial infusion were significant factors. The therapeutic results of intra-arterial chemoradiotherapy via the superficial temporal artery were not inferior to the results of surgery. In particular, the results of arterial injection therapy by cisplatin with sodium thiosulphate were excellent, so we believe that it will be a new therapy for advanced oral cavity cancer.
British Journal of Cancer (2008) **98,** 1039–1045. doi:10.1038/sj.bjc.6604272 www.bjcancer.com
Published online 19 February 2008

Keywords: carcinoma of the oral cavity; intra-arterial chemotherapy; radiation therapy; prognostic factor

As locally advanced oral cavity cancer is difficult to control by radiotherapy, surgery remains the most effective curative therapy (Poulsen *et al*, 1996). In this case, an extended surgery markedly reduces the quality of life, thus affecting the patient's social life. Therefore, the development of effective non-resection therapy is extremely important.

In 1992, we started chemoradiation therapy, in which continuous arterial infusion therapy with carboplatin was combined with radiotherapy, by selectively inserting a catheter into the target artery through the superficial temporal artery in patients with locally advanced head and neck cancer (Fuwa *et al*, 2000). We started therapy in two courses of systemic chemotherapy combined with intra-arterial chemoradiotherapy by continuous intra-arterial infusion of carboplatin and radiation therapy, in an effort to control metastasis in cervical lymph nodes and distant metastasis in 1997 (Fuwa *et al*, 2007). Furthermore, we changed the agent for intra-arterial infusion from carboplatin to cisplatin in an effort to improve the local control rate in October 2002. This is an improved technique of the Robbins *et al* method (Robbins *et al*, 1994, 2000), whereby an infusion dose of cisplatin was increased by infusing sodium thiosulphate, which is a neutralising agent of cisplatin, from a vein at the time of intra-arterial infusion of cisplatin.

In this article, we analyse the therapeutic results of 134 cases of stages III and IV (M0) oral cavity cancer retrospectively to investigate the prognostic factors, and we inspect the effectiveness of arterial injection therapy via the superficial temporal artery for cases of advanced oral cavity cancer.

*Correspondence: Dr N Fuwa; E-mail: nfuwa@aichi-cc.jp
Received 29 June 2007; revised 17 January 2008; accepted 22 January 2008; published online 19 February 2008

MATERIALS AND METHODS

Patient selection criteria

The subjects met the following criteria: (1) the pathology is squamous cell carcinoma; (2) stage III or higher oral cavity cancer (except carcinoma of the base of tongue) without distant metastasis according to the TNM staging published in 2002; (3) patients in whom the performance status (PS) was evaluated as 0–3 according to the classification described by the Eastern Cooperative Oncology Group; (4) ages ranging from 20 to 89 years; (5) the bone marrow function was maintained (leukocyte count: 3000 mm^{-2} or more, platelet count: 100 000 mm^{-2} or more); (6) patients without severe liver, kidney, heart, or lung dysfunction; (7) untreated patients; (8) patients without active double cancer at the start of treatment, and who had not previously undergone radiotherapy in the head and neck region; and (9) patients from whom written informed consent was obtained.

Treatment schedule, administration of the agent

The treatment schedule was divided into four groups (Figure 1). Continuous arterial injection of carboplatin was performed using a portable electrical pump for 6 weeks in Group 1. Using Calvert's

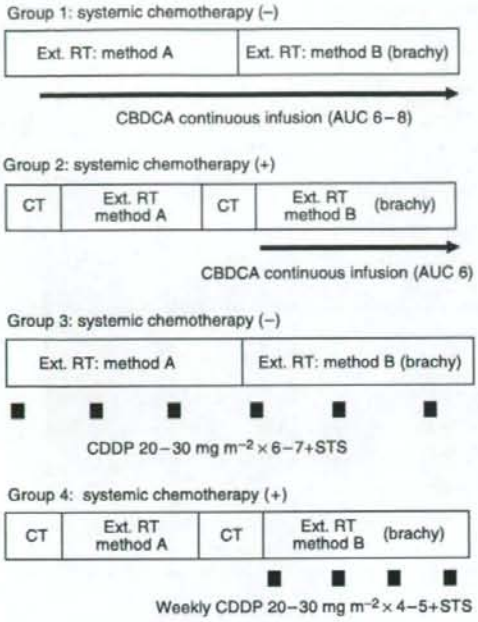

Figure 1 Scheme of the therapy. CT, chemotherapy; Ext RT, external beam radiation therapy. Method A: wide field irradiation, 36–40 Gy/20 fraction. Method B: reduced field irradiation, 26–30 Gy/15 fraction. Brachy, brachytherapy. Alternating therapy involving systemic chemotherapy and radiation therapy is performed. Intra-arterial chemotherapy is initiated after the end of the second course of systemic chemotherapy in the Groups 2 and 4.

formula, the total dose of carboplatin was established as six- to eight-fold the area under the plasma concentration–time curve (AUC) according to both the kidney function and PS.

As mentioned above, we started a new chemoradiation therapy in which continuous arterial infusion therapy with carboplatin was combined with two courses of systemic chemotherapy with 5-fluorouracil and nedaplatin in order to control the neck lymph nodes and distant metastases (Group 2) in 1997. Carboplatin (total dose: AUC 6) was administered continuously in the latter half of radiotherapy after the end of the second course of systemic chemotherapy. The regimen of systemic chemotherapy consisted of continuous intravenous injection of 5-fluorouracil at 700 mg m^{-2} for 5 days (from Day 1 until Day 5) and intravenous drip of nedaplatin at 120 mg m^{-2} over 6 h on Day 6.

To further improve the local control, we modified the procedure described by Robbins *et al* from October 2002. The dose of cisplatin was established as 20 mg m^{-2} when the catheter was inserted into the selected artery, and 30 mg m^{-2} when the catheter was inserted into the external carotid artery. During the arterial injection of cisplatin, a cisplatin-neutralising agent, sodium thiosulphate, at 8–10 g m^{-2} was intravenously administered over 7 h.

When inserting catheters into arteries on both sides, we set the amount of the infused dose of CDDP up to 40 mg m^{-2} in total per week; and to distribute the agent appropriately, we decided the amount of agent distributed from the findings of the MRI.

In patients who were not eligible for systemic chemotherapy, including elderly patients (≥75 years) and those with a poor PS score, cisplatin arterial injection chemotherapy was repeated six to seven times in combination with radiotherapy at 60–70 Gy (Group 3).

In patients in whom systemic chemotherapy was possible, alternating therapy involving systemic chemotherapy and radiation therapy was performed. Arterial injection therapy was repeated four to five times after the end of the second course of systemic chemotherapy (Group 4). The regimen of systemic chemotherapy consisted of the continuous intravenous injection of 5-fluorouracil at 700 mg m^{-2} for 5 days (from Day 1 until Day 5) and intravenous drip of cisplatin at 85 mg m^{-2} over 24 h on Day 6. In patients with a poor renal function (24-h creatinine clearance was 60 ml min^{-1} or less), nedaplatin at 100 mg m^{-2} was administered in place of cisplatin.

Radiation therapy

Radiotherapy was performed five times a week by irradiating 1.8–2 Gy of photon beam in a fraction using a 6 MV linear accelerator. The initial irradiation (irradiation method A) was performed five times a week for 4 weeks at a radiation dose of 1.8–2 Gy (total dose: 36–40 Gy). The latter half of irradiation (irradiation method B) was performed five times a week for 3 weeks at a radiation dose of 2 Gy (total dose: 26–30 Gy) (A + B: 66 Gy).

In the irradiation method A, using the bilateral opposing portal irradiation method, 36–40 Gy in 20 fractions was irradiated between the primary lesion, the middle cervical lymph nodes, and a 2 cm safety margin, whereas 36–40 Gy of photon beam was irradiated between the lower cervical region and the supraclavicular fossa using the anterior single irradiation method.

In irradiation method B, an area involving the tumour site on the initial consultation and a 1 cm safety margin was established as the planned target volume (PTV). The radiation dose for the spinal cord was established as 40 Gy or less. In patients with tongue or oral floor cancer in whom brachytherapy was possible, external irradiation at a radiation dose of approximately 50 Gy or less was combined with brachytherapy using a Cs needle or Au grain.

Arterial injection therapy

As previously reported (Fuwa *et al*, 2000), the anterior ear on the affected side was incised under local anaesthesia to expose the superficial temporal artery. During fluoroscopy, a thin catheter was selectively inserted into the selected artery. When the lesion involved the contralateral side beyond the median line, another catheter was inserted in the contralateral side for bilateral arterial injection. The target artery was the lingual artery in carcinoma of the tongue, the facial artery in carcinomas of the floor of mouth, the buccal mucosa, and lower gingiva, and the maxillary artery in carcinomas of the hard palate and upper gingiva.

When the tumour involved beyond the perfusion area by selected arterial injection, or when severe arteriosclerosis made the selective insertion of a catheter into the selected artery difficult, a catheter was placed in the external carotid artery.

We confirmed that the extent of arterial injection covered the tumour by a pigment, angiography, and MRI from 2001, in which an extremely low dose of contrast medium for MRI was slowly infused via a catheter for arterial injection.

This clinical trial was approved by the Ethics Committee of Aichi Cancer Center Hospital.

Patient assessments

The treatment response was evaluated based on the MRI. The subjects consulted the outpatient clinic at 1-month intervals for 1 year after the end of treatment, at 2- to 3-month intervals in the second and third years of follow-up, and at 3–5 month intervals after 3 years of follow-up. Follow-up MRI was performed at 4- to 6-month intervals for 2 years after the end of treatment, and at 6- to 8-month intervals thereafter. Chest X-rays were performed at 6- to 8-month intervals, and liver CT or echogram was performed every year until 3 years after the end of treatment.