A technique for the quantitative evaluation of dose distributions

Daniel A. Low,[a] William B. Harms, Sasa Mutic, and James A. Purdy
Mallinckrodt Institute of Radiology, Division of Radiation Oncology, 510 South Kingshighway Blvd., St. Louis, Missouri 63110

(Received 9 June 1997; accepted for publication 2 March 1998)

The commissioning of a three-dimensional treatment planning system requires comparisons of measured and calculated dose distributions. Techniques have been developed to facilitate quantitative comparisons, including superimposed isodoses, dose-difference, and distance-to-agreement (DTA) distributions. The criterion for acceptable calculation performance is generally defined as a tolerance of the dose and DTA in regions of low and high dose gradients, respectively. The dose difference and DTA distributions complement each other in their useful regions. A composite distribution has recently been developed that presents the dose difference in regions that fail both dose-difference and DTA comparison criteria. Although the composite distribution identifies locations where the calculation fails the preselected criteria, no numerical quality measure is provided for display or analysis. A technique is developed to unify dose distribution comparisons using the acceptance criteria. The measure of acceptability is the multidimensional distance between the measurement and calculation points in both the dose and the physical distance, scaled as a fraction of the acceptance criteria. In a space composed of dose and spatial coordinates, the acceptance criteria form an ellipsoid surface, the major axis scales of which are determined by individual acceptance criteria and the center of which is located at the measurement point in question. When the calculated dose distribution surface passes through the ellipsoid, the calculation passes the acceptance test for the measurement point. The minimum radial distance between the measurement point and the calculation points (expressed as a surface in the dose–distance space) is termed the γ index. Regions where $\gamma > 1$ correspond to locations where the calculation does not meet the acceptance criteria. The determination of γ throughout the measured dose distribution provides a presentation that quantitatively indicates the calculation accuracy. Examples of a 6 MV beam penumbra are used to illustrate the γ index. [S0094-2405(98)01905-1]

Key words: three-dimensional treatment planning, quality assurance, distance-to-agreement, dose distribution analysis, radiation therapy

I. INTRODUCTION

The commissioning of treatment planning systems routinely requires the comparison of measured and calculated dose distributions. The physicist first determines a set of irradiation conditions for which the treatment planning system is to be evaluated. Measured (often planar) dose distributions are obtained for these geometries, and the corresponding isodose distributions are subsequently displayed or printed.

The qualitative evaluation of the treatment planning system calculation is made by superimposing the isodose distributions, either using software tools or by hand using printed isodose distributions and a light box. This evaluation highlights areas of significant disagreement, but a more quantitative assessment may be needed for final system approval.

Quantitative evaluation methods directly compare the measured and calculated dose distribution values. Van Dyk *et al.*[1] describe the quality assurance procedures of treatment planning systems and subdivide the dose distribution comparisons into regions of high and low dose gradients, each with a different acceptance criterion. In low gradient regions, the doses are compared directly, with an acceptance tolerance placed on the difference between the measured and calculated doses. A dose-difference distribution can be displayed[2–5] that identifies the regions where the calculated dose distributions disagree with measurement. In high dose gradient regions (assuming that the spatial extent of the region is sufficiently large), a small spatial error, either in the calculation or the measurement, results in a large dose difference between measurement and calculation. Dose differences in high dose gradient regions may therefore be relatively unimportant, and the concept of a distance-to-agreement (DTA) distribution is used to determine the acceptability of the dose calculation.[6–9] The DTA is the distance between a measured data point and the nearest point in the calculated dose distribution that exhibits the same dose. The dose-difference and DTA evaluations complement each other when used as determinants of dose distribution calculation quality.

A composite analysis developed by Harms *et al.*[7,10] and based on the concepts of Shiu *et al.*[11] as applied by Cheng *et al.*,[12] uses a pass–fail criterion of both the dose difference and DTA. Each measured point is evaluated to determine if both the dose difference and DTA exceed the selected tolerances (e.g., 3% and 3 mm, respectively). Points that fail both criteria are identified on a composite distribution. Because

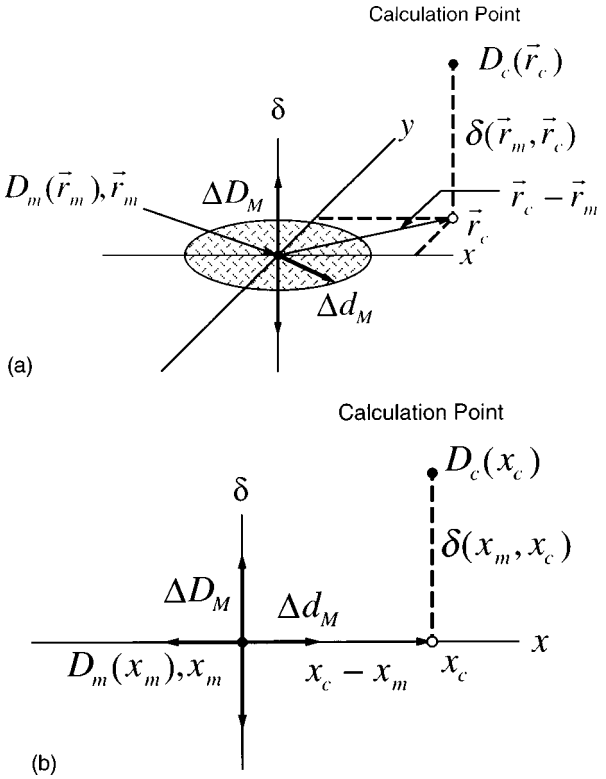

FIG. 1. Geometric representation of dose distribution evaluation criteria for the dose-difference and distance-to-agreement tests. (a) Two-dimensional representation. (b) One-dimensional representation.

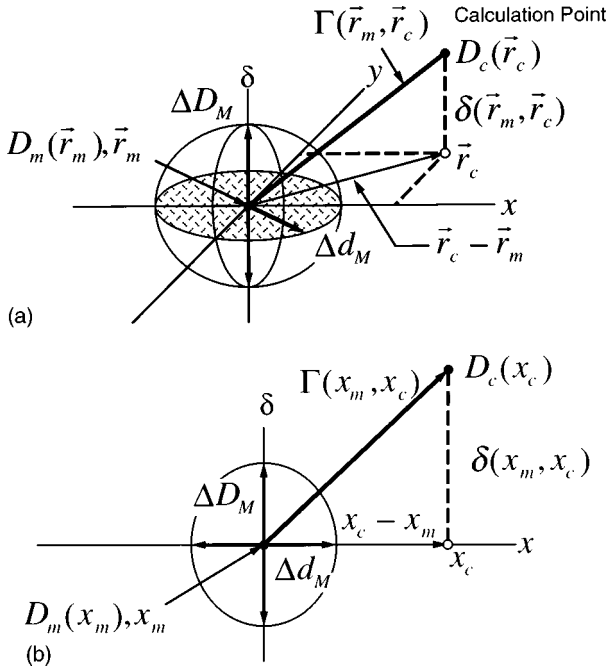

FIG. 2. Geometric representation of dose distribution evaluation criteria using the combined ellipsoidal dose-difference and distance-to-agreement tests. (a) Two-dimensional representation. (b) One-dimensional representation.

the composite distribution is a binary distribution, it does not lend itself to a convenient display. Therefore, by convention, the quantity displayed in the composite distribution is the dose difference. While the composite distribution highlights regions of disagreement, the display of the dose difference may accentuate the impression of failure in high dose gradient regions. An additional limitation to this technique is that there is no unique numerical index that enables the presentation and analysis of a distribution that measures the calculation quality.

An extension of the isodose comparison tools is presented that simultaneously incorporates the dose and distance criteria. It provides a numerical quality index that serves as a measure of disagreement in the regions that fail the acceptance criteria and indicates the calculation quality in regions that pass. Unlike the existing composite distribution, the index can be presented in a graphical form to enable a rapid and efficient evaluation of the algorithm quality by the physicist. An implicit assumption is made that once the passing criteria are selected, the dose-difference and distance-to-agreement analyses have equivalent significance when determining calculation quality.

II. METHODS AND MATERIALS

A. Evaluation methods

The method presented here uses a comparison between measured and calculated dose distributions. The measurement is used as the reference information, and the calculated distribution is queried for comparison. The mathematical formalism describing the dose-difference, DTA, and composite distributions has been presented by Harms *et al.*[7,10] but is summarized here for clarity. The dose-difference criterion is ΔD_M, and the DTA criterion is Δd_M. The passing criteria shown for the examples are $\Delta D_M = 3\%$ and $\Delta d_M = 3$ mm based on our internal clinical standards for photon beams.

Figure 1(a) shows a diagrammatic representation of the composite analysis tool for two-dimensional dose distribution evaluations. In this and all subsequent cases, the evaluation presented is for a single measurement point $\mathbf{r}_m$, lying at the origin of the figure, and for clinical evaluations, the comparisons are repeated for all measurement points. Two of the axes (x and y) represent the spatial location $\mathbf{r}_c$ of the calculated distribution relative to the measured point. The third axis (δ) represents the difference between the measured $[D_m(\mathbf{r}_m)]$ and calculated $[D_c(\mathbf{r}_c)]$ doses. The DTA criterion, Δd_M, is represented by a disk in the $\mathbf{r}_m - \mathbf{r}_c$ plane with a radius equal to Δd_M. If the calculated distribution surface, $D_c(\mathbf{r}_c)$, intersects the disk, the DTA is within the acceptance criterion, and the calculated distribution passes the DTA test at that point. The vertical line represents the dose-difference test; its length is $2\Delta D_M$. If the calculated distribution surface crosses the line $[|D_c(\mathbf{r}_m) - D_m(\mathbf{r}_m)| \leq \Delta D_M]$, the calculated distribution passes the dose-difference test at the measurement point. Figure 1(b) shows the one-dimensional analog to Fig. 1(a).

Figure 2 shows a representation of a method for determining an acceptance criterion that simultaneously considers the dose difference and DTA. An ellipsoid is selected as the

surface representing the acceptance criterion. The equation defining the surface is

$$1=\sqrt{\frac{r^2(\mathbf{r}_m,\mathbf{r})}{\Delta d_M^2}+\frac{\delta^2(\mathbf{r}_m,\mathbf{r})}{\Delta D_M^2}}, \tag{1}$$

where

$$r(\mathbf{r}_m,\mathbf{r})=|\mathbf{r}-\mathbf{r}_m| \tag{2}$$

and

$$\delta(\mathbf{r}_m,\mathbf{r})=D(\mathbf{r})-D_m(\mathbf{r}_m) \tag{3}$$

is the dose difference at the position $\mathbf{r}_m$.

If any portion of the $D_c(\mathbf{r}_c)$ surface intersects the ellipsoid defined by Eq. (1), the calculation passes at $\mathbf{r}_m$.

Defining the acceptance criteria not just along the δ axis and in the $\mathbf{r}_c-\mathbf{r}_m$ plane allows for a more general comparison between calculation and measurement than does the traditional composite evaluation. The quantity on the right-hand side of Eq. (1) can be used to identify a quality index γ at each point in the evaluation plane $\mathbf{r}_c-\mathbf{r}_m$ for the measurement point $\mathbf{r}_m$,

$$\gamma(\mathbf{r}_m)=\min\{\Gamma(\mathbf{r}_m,\mathbf{r}_c)\}\forall\{\mathbf{r}_c\}, \tag{4}$$

where

$$\Gamma(\mathbf{r}_m,\mathbf{r}_c)=\sqrt{\frac{r^2(\mathbf{r}_m,\mathbf{r}_c)}{\Delta d_M^2}+\frac{\delta^2(\mathbf{r}_m,\mathbf{r}_c)}{\Delta D_M^2}}, \tag{5}$$

$$r(\mathbf{r}_m,\mathbf{r}_c)=|\mathbf{r}_c-\mathbf{r}_m|, \tag{6}$$

and

$$\delta(\mathbf{r}_m,\mathbf{r}_c)=D_c(\mathbf{r}_c)-D_m(\mathbf{r}_m) \tag{7}$$

is the difference between dose values on the calculated and measured distributions, respectively. The pass–fail criteria therefore become

$$\gamma(\mathbf{r}_m)\leqslant 1,\ \text{calculation passes},$$
$$\gamma(\mathbf{r}_m)>1,\ \text{calculation fails}. \tag{8}$$

An important feature of this method is that in the final assessment of the dose distribution quality, the value of $\gamma(\mathbf{r}_m)$ can be displayed in an iso-γ distribution. The regions where $\gamma(\mathbf{r}_m)$ is greater than but nearly unity will be apparent relative to the regions of more significant disagreement.

B. Evaluation distributions

An examination of the dose distribution evaluation methods is made using the penumbra of a $10\times10\ \text{cm}^2$, 6 MV photon beam. For convenience and presentation clarity, the evaluations are shown for one-dimensional dose distributions and the data fit to the equation

$$D(x)=\eta\{T+(1-T)(A\ \text{erf}[B_1(x_0-x)]+(1-A)\text{erf}[B_2(x_0-x)])\}+D_1, \tag{9}$$

where $D(x)$ is the dose at position x, A, B_1, and B_2 are fitting coefficients, and T is the transmission of the collimator. The error function erf(x) is the integral of the normalized Gaussian distribution and ranges from a value of 0 at $x=-\infty$ to 1 at $x=\infty$. The remaining parameter values (η, x_0, and D_1) are explained below.

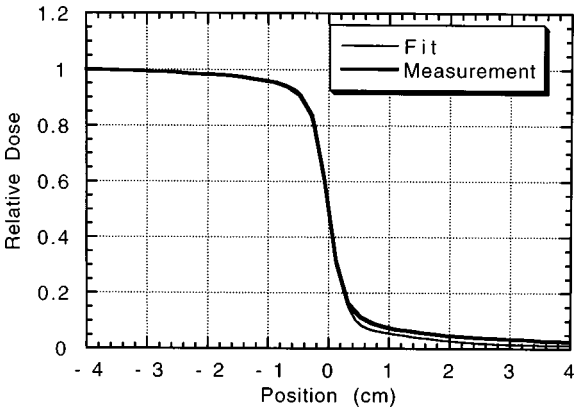

FIG. 3. Measured penumbra and the fit using Eq. (9) for a $10\times10\ \text{cm}^2$, 6 MV photon beam.

To investigate the utility of the quality index $\gamma(\mathbf{r}_m)$, two curves, identified as calculated and measured, are required. The curve labeled measured uses the parameters $\eta=1$, $x_0=0$, and $D_1=0$, yielding a penumbra with the collimator edge at $x=0$. The utility of the γ index is investigated by adjusting the parameters η, x_0, and D_1, to mimic specific clinical situations, including a normalization error (η), position shift (x_0, which models the position of the collimator edge, equal to the 50% dose position when $T=0$), and dosimetric shift (D_1). The dose-difference, DTA, and γ index are determined for these cases.

III. RESULTS

Figure 3 shows the comparison of the measured penumbra and fit using Eq. (9). The fit matches the measured data within 3% over the range of -4 to 4 cm. The fitting parameters are

$$\begin{aligned} &A=0.173,\\ &B_1=0.456\ \text{cm}^{-1},\\ &B_2=2.892\ \text{cm}^{-1},\\ &T=0.010. \end{aligned} \tag{10}$$

The measured and calculated dose distributions for a position shift of 0.25 cm ($x_0=0.25$ cm for the calculated dose distribution) are shown in Fig. 4(a). The dose-difference and DTA are also shown. The dose difference is multiplied by 10 such that the position on the graph of the dose-difference criterion (3%) appears at the same location as the criterion for the DTA (3 mm). The passing criterion is shown as a thick dot–dashed horizontal line at 0.3.

The DTA is a constant 0.25 cm, as expected. The dose difference rises rapidly at the penumbra, peaking beyond the scope of the graph, with a value of 0.33. Because the DTA is always less than 0.3 cm, there are no regions that fail the composite evaluation.

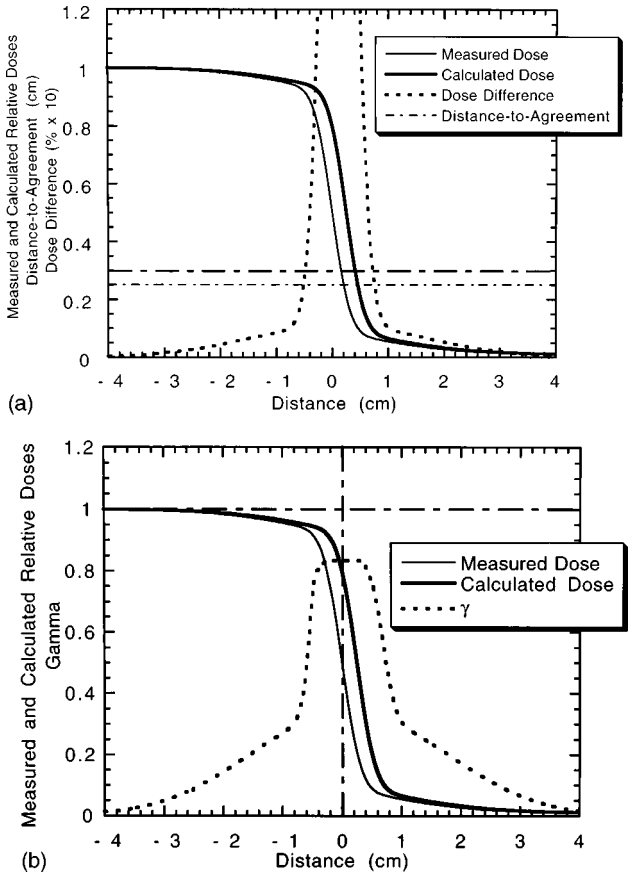

FIG. 4. (a) Superimposed dose, dose-difference, and distance-to-agreement distributions for the penumbra distributions modeling a 0.25 cm spatial shift between calculation and measurement. (b) Superimposed dose, dose-difference, and γ-index distributions for the same conditions as in (a).

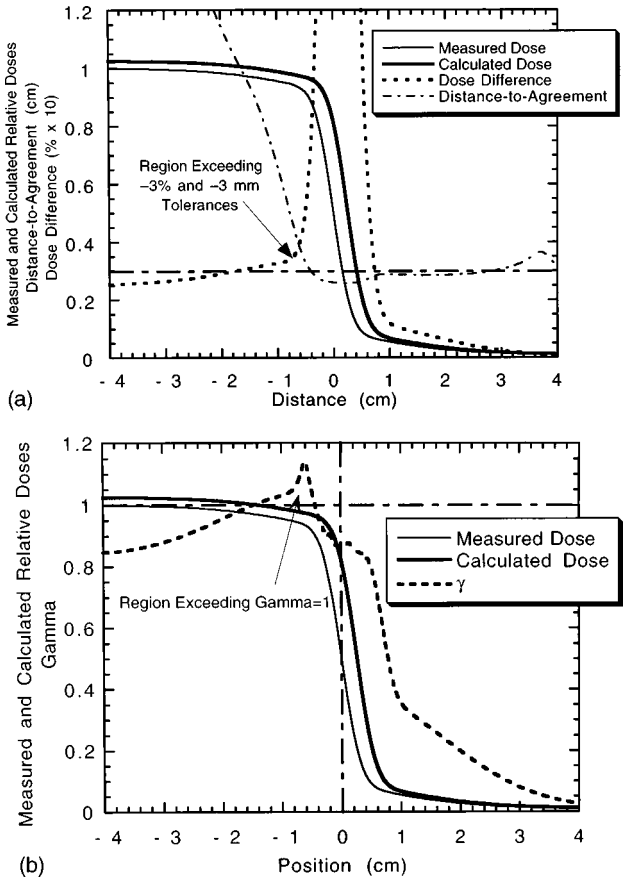

FIG. 5. (a) Superimposed dose, dose-difference, and distance-to-agreement distributions for the penumbra distributions modeling a 0.25 cm spatial shift and 2.5% normalization difference between calculation and measurement. (b) Superimposed dose, dose-difference, and γ-index distributions for the same conditions as in (a).

Figure 4(b) shows the measured and calculated dose distributions, and the γ index, for the same conditions as in Fig. 4(a). The maximum value of γ is 0.83, corresponding to the maximum value of the dose difference. The value of γ in all regions is less than 1.0, indicating that the calculation passes the acceptance tests, consistent with the composite evaluation [Fig. 4(a)].

Figure 5(a) shows the measured and calculated dose distributions and the dose difference and DTA for the penumbra, where there is a normalization difference of 2.5% and a lateral shift of 0.25 cm ($\eta=1.025$, $x_0=0.25$ cm for the calculated dose distribution). The DTA is no longer constant, with large values where the dose distributions differ by 2.5% and the dose gradients are small. For the region $x<-1.50$ cm, the dose difference lies below the passing criterion of 3%, and the calculation passes the test. The dose difference has a similar peak (maximum value$=0.35$) to that in Fig. 4(a). Both the dose difference and DTA lie below the passing criteria for $x>0.74$ cm, and the DTA lies below the passing criterion between $x=-0.43$ cm and $x=0.74$ cm. Both the DTA and dose differences fail the criteria in the region -1.74 cm$<x<-0.43$ cm. The calculation would be said to fail in this region.

Figure 5(b) shows the measured and calculated dose distributions and the γ index for the same conditions as in Fig. 5(a). The γ index exceeds a value of unity in nearly the same region as the DTA and dose differences cross the passing criteria in Fig. 5(a). The failure region lies between -1.53 and -0.44 cm for the γ index, and -1.74 and -0.43 cm for the composite analysis. The maximum value of γ is 1.14, indicating that the calculation fails the acceptance criteria by 14%.

The same analysis is made for a calculated penumbra distribution with spatial and dosimetric offsets of 0.25 cm ($x_0=0.25$ cm) and 2.5% ($D_1=0.025$), respectively. Figure 6(a) shows the superimposed dose distributions, dose difference, and DTA. As before, the dose difference exhibits a peak value (of 0.35) near the origin. This time, however, the DTA has a single minimum near the origin, with large values in both regions of low dose gradient. There are two regions, lying on each side of the origin, that fail the composite criteria.

The γ index is shown with the superimposed dose distributions in Fig. 6(b). The two regions that fail the criteria are clearly identified where $\gamma>1$. The calculated dose distribution fails, based on the composite analysis, in the regions $-1.81<x<-0.38$ and $0.30<x<2.08$, while the γ index-value test fails between $\gamma-1.66<x<-0.41$ and $0.37<x<1.84$. Once again, the region where the calculation fails

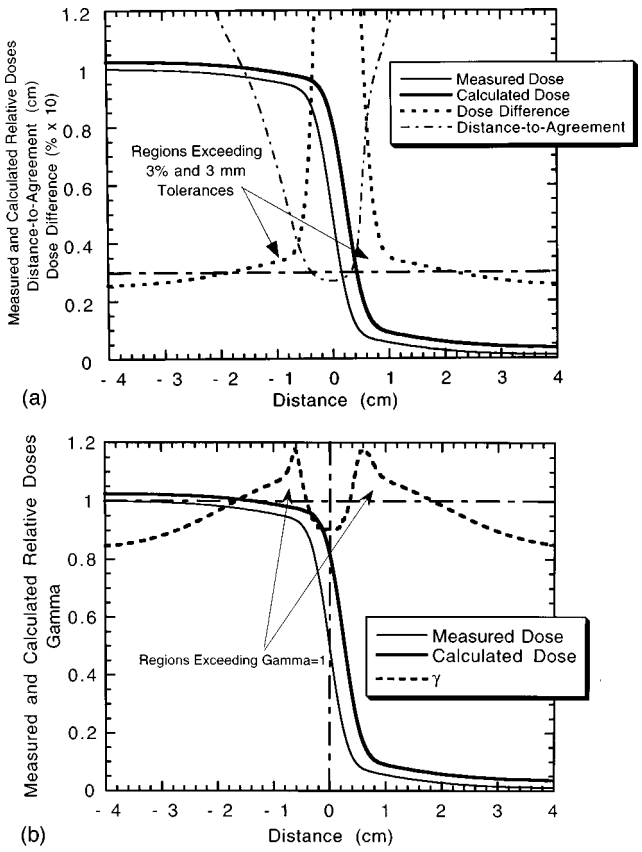

FIG. 6. (a) Superimposed dose, dose-difference, and distance-to-agreement distributions for the penumbra distributions modeling a 0.25 cm spatial shift and 2.5% dose offset between calculation and measurement. (b) Superimposed dose, dose-difference, and γ-index distributions for the same conditions as in (a).

according to the γ index is slightly smaller than the region failing according to the composite distribution. The maximum value of γ is equal to 1.18, indicating that the calculation fails by 18% relative to the selected acceptance criteria.

IV. DISCUSSION AND CONCLUSIONS

An important benefit of this technique is the determination of a quality measurement, termed the γ index, that indicates the difference between the calculation and measurement relative to the acceptance tolerances. A γ-index distribution can be generated and displayed, providing a quantitative assessment of the quality of the calculation, both in regions that pass and fail the acceptance criteria. This has the potential to provide a powerful analysis tool in the quantitative evaluation of three-dimensional treatment planning systems. One example of this is the presentation of a γ-index distribution histogram. This may provide a rapid presentation of the calculation quality and may be especially useful in the comparison of three-dimensional dose data, where direct presentation is not convenient.

While not investigated here, the angle between the δ axis and the vector that defines the quantity $\Gamma(\mathbf{r}_m, \mathbf{r}_c)$ can be used to indicate whether the difference between calculated and measured dose distributions is due to the dose difference or DTA. A graphical representation of the angle should be investigated to determine its clinical utility.

While an ellipse was selected as the determinant of the pass–fail criteria, the selection was based principally on the convenience of description and coding. Other shapes could be used, such as a right cylinder or cone, based on the user's dose-distribution test specifications. The relative importance of the dose difference and distance-to-agreement tests could also be adjusted by suitable selection of the pass–fail criteria surface. In addition, the dose-difference criteria could be generalized to a spatially dependent function to account for regions where dose differences are more critical.

This manuscript serves as a description of the γ index and comparisons of actual measured and calculated data are not conducted. To enable evaluations of treatment planning system data, a two-dimensional application of the γ index is being developed using the software platform developed by Harms *et al.*[7,10] that will supplement the existing analysis tools. The index will be used to evaluate the dose calculation algorithm of three-dimensional treatment planning systems, including an intensity modulated planning and delivery system. Once this tool has been developed, experience will be gained in the appropriate use of the tool with actual clinical data.

Additional studies will be conducted to determine the tool's usefulness under clinical conditions. A study of the index reliability and sensitivity to noise and further evaluation of its clinical utility will be conducted and presented in a future manuscript. The effects of data smoothing and other data manipulations will also be evaluated.

[a]Address all correspondence to Daniel A. Low, Ph.D., Division of Radiation Oncology, Mallinckrodt Institute of Radiology, 510 South Kingshighway Blvd., St. Louis, Missouri 63110. Electronic mail: low@castor.wustl.edu

[1]J. Van Dyk, R. B. Barnett, J. E. Cygler, and P. C. Shragge, ''Commissioning and quality assurance of treatment planning computers,'' Int. J. Radiat. Oncol., Biol., Phys. **26**, 261–273 (1993).

[2]E. Mah, J. Antolak, J. W. Scrimger, and J. J. Battista, ''Experimental evaluation of a 2D and 3D electron pencil beam algorithm,'' Phys. Med. Biol. **34**, 1179–1194 (1989).

[3]B. A. Fraass, M. K. Martel, and D. L. McShan, ''Tools for dose calculation verification and QA for conformal therapy treatment techniques,'' *XIth International Conference on the Use of Computers in Radiation Therapy* (Medical Physics Publishing, Madison, WI, 1994).

[4]B. A. Fraass and D. McShan, ''Three-dimensional photon beam treatment planning,'' in *Radiation Therapy Physics*, edited by A. R. Smith (Springer-Verlag, New York, 1995).

[5]B. A. Fraass, ''Quality assurance for 3D treatment planning,'' *Teletherapy: Present and Future* (Advanced Medical Publishing, Vancouver, British Columbia, 1996).

[6]K. R. Hogstrom, M. D. Mills, J. A. Meyer, J. R. Palta, D. E. Mellenberg, R. T. Meoz, and R. S. Fields, ''Dosimetric evaluation of a pencil-beam algorithm for electrons employing a two-dimensional heterogeneity correction,'' Int. J. Radiat. Oncol., Biol., Phys. **10**, 561–569 (1984).

[7]W. B. Harms, D. A. Low, J. A. Purdy, and J. W. Wong, ''A software tool to quantitatively evaluate 3D dose calculation algorithms,'' submitted to Med. Phys.

[8]H. Dahlin, I. L. Lamm, T. Landberg, S. Levernes, and N. Ulso, ''User requirements on CT based computerized dose planning systems in radiotherapy,'' Acta Radiol.: Oncol. **22**, 397–415 (1983).

[9]*ICRU Report 42. Use of Computers in External Beam Radiotherapy Procedures with High-Energy Photons and Electrons* [International Commission on Radiation Units and Measurements (ICRU), Bethesda, 1987].
[10]W. B. Harms, D. A. Low, J. A. Purdy, and J. W. Wong, ''A quantitative software analysis tool for verifying 3D dose-calculation programs (abstract),'' Int. J. Radiat. Oncol., Biol., Phys. **30**, 187 (1994).
[11]A. S. Shiu, S. Tung, K. R. Hogstrom, J. W. Wong, R. L. Gerber, W. B. Harms, J. A. Purdy, R. K. Ten Haken, D. L. McShan, and B. A. Fraass, ''Verification data for electron beam dose algorithms,'' Med. Phys. **19**, 623–636 (1992).
[12]A. Cheng, W. B. Harms, R. L. Gerber, J. W. Wong, and J. A. Purdy, ''Systematic verification of a three-dimensional electron beam dose calculation algorithm,'' Med. Phys. **23**, 685–693 (1996).