Innovation/Impact:
This study examines the reproducibility of the TG119 test criteria in quantifying the overall accuracy of an IMRT system. Although patient-specific IMRT QA is widely performed, there is no recognized standard. Furthermore, even for the most commonly used QA tests, such as gamma analysis of composite films and per-field measurements using a diode array, there is no recognized standard passing criteria. TG119 addresses these needs by providing a method to assess the accuracy of a given IMRT system. Institutions interested in utilizing TG119 to likely be concerned with reliability of the test protocol as it stands, especially when comparing their results to those of the medical physics community.

Introduction:
TG119 describes a test protocol that can be performed by any institution to quantify the accuracy of their IMRT system. The TG119 test consists of several standardized IMRT and non-IMRT test cases that the institution should plan, deliver, and analyze, while closely conforming to planning, delivery, and analysis techniques used in the report. Delivery and measurement are to take the form of per-field measurements using film or a diode array, and composite measurements, using film and an ionization chamber. Analysis results are to be quantified as an accuracy assessment, presented as a confidence limit, which can be used to evaluate the quality of commissioning and can be compared to a benchmark established by a collective group of institutions. The determined confidence limit for a given IMRT QA test establishes a rational acceptance criteria for evaluating patient specific IMRT QA for their system.

We evaluated the reproducibility of the TG119 accuracy assessment by performing four independent tests on a single IMRT system.

Methods and Materials:
1. Planning
Planning utilized the TG119 test suite structures and CT study, in addition to a local phantom CT study for plan delivery. One AP:PA plan (Test P1) and one Bands plan (Test P2) were developed using several fields of symmetric or asymmetric jaws. Four planners created each of the five IMRT plans (Tests I1-15) following the planning parameters described in TG119, including dose goals and beam angles for each plan. All plans were developed with the same isocenter coordinate values to eliminate setup variation between plans during delivery. All plans were developed using iPlan RT Dose 4.1 treatment planning software (BrainLAB) for a Varian 6EX accelerator utilizing a step-and-shoot delivery technique. These 20 IMRT plans will be referred to as the primary set, and constitute four complete sets for minimum completion of the TG119 test.
Each plan was transferred to MOSAIQ v. 1.60Q3 record and verify system (IMPAC Medical Systems Inc) for delivery on a Varian 6EX with Millennium 120-leaf MLCs (Varian Medical Systems, Palo Alto, CA) using 4DTC version 8. Calculated per-field dose planes, point-dose values, and composite dose planes were collected for comparison to respective measurement data at locations prescribed by the TG119 instructions.

In order to investigate the effects of planer optimization technique, five planners each created a plan following the C shape easy goals (Test I4), referred to as the secondary set. Unlike the primary set of plans, planners were forced to use identical user-selectable planning parameters (i.e. number of segments, calculation grid, beamlet size, etc.) which were selected to closely match those of planner 3 from the primary set.

2. Measurements and Analysis
Following the TG119 guidelines, both per-field measurements and composite measurements were performed for each plan. Per-field measurements for the primary set were collected using a MAPCHECK model 1175 diode array device (Sun Nuclear Corporation, Melbourne, Fl.) and analyzed using MAPCHECK 4.01.01 software. Measurements were performed in five separate sessions over consecutive days, with each session measuring all plans of one type (i.e. Mock Prostate, Multitarget, etc.). Analysis of each field utilized the same MAPCHECK software options selected in the TG119 study, namely gamma criteria of 3% dose and 3mm distance to agreement, absolute dose, 10% threshold, Van Dyk percentage difference, and applied measurement uncertainty selected.

Composite film and chamber measurements for the primary set and secondary set were completed across two sessions, one for each set. The number of treatment sessions was minimized in order to reduce session-dependent variation, such as phantom setup or film quality, between compared measurements. Both measurement sessions were performed by one individual to reduce the likelihood of user dependence during delivery. AP:PA test, Bands test, and a film calibration set were performed immediately prior to each delivery session. All treatment plans were delivered on the local phantom consisting of a 30cm x 30cm x 15cm set of water-equivalent slabs, which permitted ion chamber measurements and planar dose measurements on coronal planes. Since each plan used the same CT sets and isocenter values, minimal phantom positioning between measurements was required. Chamber measurements were obtained with an Exradin A1SL ionization chamber, with a sensitive volume of 0.056 cm³. Composite film measurements were obtained utilizing Kodak EDR2 film. After each composite film was exposed, the film was immediately developed and digitized for later analysis. Machine output, treatment room temperature, and atmospheric pressure were monitored throughout each session in order to confirm consistent delivery conditions. Delivery of the 20 plans was ordered in groups by type (i.e. Mock Prostate, Multitarget, etc.) in order to minimize time and positioning variation within respective plans of a single type.

In addition to the TG119 required measurements, repeat measurements were performed for one IMRT plan (Easy Cshape) and the Bands test to further assess the consistency of the system.

Measured ion chamber values were converted to dose using the AP:PA dose/chamber reading method recommended in TG119. These dose values were compared to their respective calculated doses for each plan and location. Differences are expressed as a percentage ratio of the prescription dose instead of predicted local dose, per TG119 recommendations. Film measurement values were converted to dose using the film calibration data for the matching session. Composite film data was analyzed in Doselab 4.11 with gamma criteria of 3% dose and 3mm distance to agreement using a rectangular region of interest defined by the respective jaw positions for the AP field (gantry=0°). A second analysis was performed, which utilized point normalization on the film that corresponded to a high dose/low gradient area in the calibrated profile. This second analysis relied only on film data, and was independent of ion chamber results.

RESULTS:
1. Planning results
Although the prescribed dose constraints were not all met, each planner produced plans that were deemed reasonably similar by visual inspection of isodose lines and dose-volume histograms. Wide variation of user-selectable planning parameters (dose grid size, number of segments, DVH constraints) existed among plans. Each planner had unique selections of planning parameters, but all plans followed the major guidelines such as beam angles, isocenter point, dose per fraction, etc. as specified in TG119. Figure 1 represents the Dose-Volume Histograms
for each of five planners for the secondary set. The user-selectable planning parameters were fixed in the secondary across all plans except for DVH constraints, with variation only in parameters controlled by the dose optimization, namely leaf positions and monitor units.

2. Per-field measurements

Confidence limits established from field-by-field measurements are displayed in Figure 2, with values of 1.55%, 0%, 0%, and 2.89% for planners 1 through 4 respectively. The values are expressed as a reduction from 100%, as described by TG119. The confidence limits of planners 2 and 3 indicate that no points failed the gamma test throughout all per-field measurements for the two planners (41 fields for each planner). Confidence limit results for planner 1 and planner 4 were both non-zero and unequal, with largest mean variance and standard deviation values from the set attributed to planner 4. Using identical gamma analysis parameters as the TG119 group, analysis of the per-field measurements resulted in relatively low confidence limit values as compared to the average value reported by the TG119 study of 7.0% (93.0% passing). However, the analysis showed inconsistent confidence limit values between data sets attributed to each planner. The worst per-field value across all fields was attributed the Mock head/neck (Test I3) of planner 4, with a gamma pass rate of 96.8%.

3. Ion chamber results

For target and avoidance structure regions, the mean variance of ion chamber results for each planner were within 1.7% of the planned dose, reported as percentage of the prescription dose rather than local dose. Only planners 1 and 4 had any measurements that differed from planned by more than 3%, both reported from the target region of their Easy CShape plans (-3.5% for planner 1 and -5.2% for planner 4). Figure 3 presents the respective confidence limit value for each planner for the primary set. The average deviation for ion chamber comparisons belonging to the secondary set resulted in 0.78%, 0.86%, 0.47%, 1.00%, and 0.98% for planners 1 through 5.

Confidence limit values from ion chamber analysis resulted in some inconsistency among planners. The average confidence limit across all planners was 4.06%, which is slightly lower than the average values reported by the TG119 group. It is of note that ion chamber values for planner 4 resulted in a confidence limit of 5.99%, which is the largest in the study and greater than the average TG119 result.

4. Composite film results

Initial analysis of film measurements utilized ion chamber normalization per TG119. Six repetitions of the Bands test led to a mean gamma passing value of 96.13% with standard deviation of 2.8%. The AP:PA film analysis resulted in 99.77% points passing. Confidence limit values for each planner using chamber normalization are presented in Figure 4. Confidence limit values for all composite film measurements are expressed as a reduction from 100%. Composite film confidence limits averaging over all planes for each planner significantly exceeds the TG119 group average value of 12.4%. Additionally, there remains an inconsistency in confidence limit values amongst planners. Composite film values for planner 4 resulted in an especially poor confidence limit of 56.58% (43.42% passing). It was noted that the worst set of film measurements were attributed to the Mock head/neck plan of planner 4, with gamma passing rates of 75.9% for the target plane and 42.4% for the avoidance structure plane. The lowest per-field measurement of 96.8% passing also belonged to this same plan.

Average deviation from expected values for the secondary set resulted in 20.9%, 19.45%, 13.00%, 21.25%, and 23.30% for planners 1 through 5, indicating that the variations seen in the primary set remained even when the planners were constrained to use the same user-selectable parameters.

An analysis of the primary set composite film data, utilizing a reasonable high dose point on the film as normalization instead of an ion chamber, is presented in Figure 5. With this method, the confidence limits found here are approximately equal to those of TG119.

**DISCUSSION:**

Performing the TG199 protocol on an IMRT system provides confidence limits designed to assess system performance. The confidence limits are determined by the mean deviation and the standard deviation of the measured doses from the planned doses over a test suite of IMRT cases [CL = |mean deviation| + 1.96SD]. TG119 states that if the CL is dominated by the mean deviation, then the system most likely can be improved. All of the confidence limits found, per-field, ion chamber, and film, for all four of our complete TG119 characterizations, were dominated by the standard deviation values. Surprisingly, the confidence limits were significantly different for a single IMRT system when the person constructing the plans changed. This suggests that a single TG119 test does not uniquely characterize an IMRT system.

Despite desirable per-field and chamber results, confidence limits calculated from composite film measurements were unusually poor. Although these values represent true confidence limits using the TG119 prescribed analysis techniques, it is not immediately apparent that they are truly representative of the quality of the IMRT planning and delivery system. An additional analysis of the primary and secondary film sets were performed with modification to the normalization alone. Rather than normalizing each film to its respective ion chamber reading at a specific point, each film was normalized to a representative point in a high dose/low gradient region of the calculated profile. Although most gamma values improved, the normalization point selected for each film does not necessarily yield the optimal gamma passing value. This analysis method led to confidence limit results that were essentially equal to the TG119 group average. The new analysis set shows less dramatic deviations between planners, but some deviation still exists.

Investigation into the cause of the deviations took two forms: investigation into random deviation, and systematic user-caused deviation. Random deviation was investigated by repeat deliveries of the Bands test, and one IMRT plan. Six repetitions of the Bands test led to a mean gamma passing value of 96.13% with standard deviation of 2.8%. From these results there is an inconsistency with a component of the film system of a few percent. However, the film dosimetry variation appears small when compared to the various planner dependent confidence limits. Unfortunately, this test does not emulate the same conditions of the primary set. Namely, the processing load and time span over which films were processed is less, which may not introduce the same stresses on the film processor. Nevertheless, the primary test was delivered one test case at a time. So that, for any given test case, the plans generated by planners 1 through 4 were all delivered and all films were developed and digitized before moving to the next test case. This method is expected to minimize apparent planner dependence due to drift in film quality alone.

Six repetitions of an IMRT case were performed. This test was performed to investigate if the delivery system was consistent in its IMRT delivery, and the analysis was consistent when given a single case was repeated. It is expected that a consistent system would not introduce apparent variation between consecutively delivered plans, whether identical or otherwise. The Easy CShape (Test I4) of planner 1 was selected for repetition because of the relatively high complexity expected for this type of test, although a comparable plan from any other planner would have sufficed. Analysis of resultant composite films led to a mean 97.23% gamma pass rate over both target and avoidance structure films, with a standard deviation of 1.83%. Ion chamber measurements were also repeated simultaneously with mean deviation of 0.06%, and a standard deviation of 0.94%. The repeated measurements indicate that random deviation in sequential plan delivery or measurement techniques...
did not lead to the variation found between planners. Again, this test does not perfectly emulate the stress placed upon the system during the delivery of the primary set.

It is hypothesized that the deviation of confidence limit among planners may have been systematically caused by user selected planning parameters. It is possible that a single planning parameter or combination of planning parameters may result in less accurate dose calculations when compared with measured dose delivery. The planning parameters used in the primary set of plans varied greatly amongst each planner, with some variation even within each planner’s set of five plans. Surprisingly, the secondary set of plans, which had uniform planning parameters amongst all planners, also resulted in deviation between planners across both target and avoidance structure planes. The number of measurements is too low to establish comparable confidence limits, but the trend in variation of the secondary set is similar to the trend for confidence limits for the primary set. For the composite film analysis planner 3 had the smallest average variation of 13.00% (87.00% passing) and planner 4 with larger variation of 21.25% (78.75% passing).

Retrospective analysis of the influence of planning parameters to ion chamber or film deviation was performed using a large set of variables. These parameters were obtained from the ‘Parameters for Physicist’ report provided by the iPlan treatment planning software. The variables were the number of segments used, total monitor units per plan, dose rate, calculation grid size, beamlet size, number of overlay beamlets, leaf sequence factor, percent above conformal monitor units for each field, and planning margin. Additionally, each plan was scored according to adherence to planning goals, where satisfying more dose constraints may indicate a delivery that is more difficult for the machine to perform. No significant correlation between any single parameter and analysis result has been found.

TG119 presents a procedure for assessing the accuracy of an IMRT system, providing a rational basis for establishing IMRT QA confidence limits. Our study indicates that performing the TG119 test a single time does not accurately establish confidence limits for the overall system quality.

Figure 1. Plot of dose-volume histogram result for the secondary set of plans. All plans followed the Easy Cshape goals and had identical user-selectable parameters.
Figure 2. Primary set confidence limit results for each planner from pre-field comparisons utilizing a diode array (MAPCHECK). The dashed line represents the average confidence limit value of 7.0% established as a benchmark by the TG119 group.

Figure 3. Primary set confidence limit results for each planner from chamber results. Each confidence limit represents all measured ion chamber values for each planner, including both target and avoidance structure locations. The dashed line represents the average confidence limit value 4.6% established as a benchmark by the TG119 group.

Figure 4. Confidence limit results of composite film plane results belonging to the primary set. Analysis utilized chamber point normalization. Each confidence limit represents the all measured film planes for each planner, including both target and avoidance structure locations. The dashed line represents the average confidence limit value 12.4% established as a benchmark by the TG119 group.

Figure 5. Confidence limit results for each planner composite film plane results. Analysis utilized film normalization to a reasonable high dose point, rather than chamber results. Each confidence limit represents all measured film planes for each planner, including both target and avoidance structure locations. The dashed line represents the average confidence limit value 12.4% established as a benchmark by the TG119 group.