Appendix 2 Dental Examiner Calibration Procedure

I. Calibration of Examiners

I.0 Introduction

Periodontal determination of plaque (PI), gingival inflammation (GI), gingival bleeding (BOP), pocket probing depths (PD) and cemento-enamel junction (CEJ) measurements are customarily used in clinical studies to measure the disease status of periodontitis and to measure effects of periodontal therapy. The measurement of these periodontal parameters poses several limitations and technique-sensitive sources of variability. Probe penetration and depth may vary with the degree of inflammation, probing force, angulation, position and instrument tip diameter. Other confounding factors include patient discomfort, accuracy of probe markings, anatomical differences in tooth crown and roots, and technique variability within and between examiners. Thus, all of these measurements are important response variables for inclusion in periodontal clinical trials.

Studies evaluating the efficacy of therapeutic interventions for periodontal disease require stringent control of measurement error. In studies involving more than one examiner, use of these assessment parameters must be based on acceptable levels of inter-examiner variability. The levels of intra- and inter-examiner reliability should be determined for each examiner and center involved in a clinical trial by the performance of a calibration session. A training exercise with a standard examiner (SE) will help to standardize to a common measurement technique and to a common definition of measurement parameters. An interactive calibration session will enable the investigators to quantitate the measurement variability among and between examiners and enable the optimization of the measurement process. Prior to the start of any clinical trial investigating a periodontal therapeutic, a calibration session will be conducted according to the guidelines set forth in this document.

I.1 Summary

The purpose of this examiner calibration session will be to quantify intra- and inter-examiner reliability of measuring periodontal soft and hard tissue parameters. All examiners will participate, and examiner calibration will be conducted at the University of North Carolina School of Dentistry. A sample of persons with a range of periodontal health (AAP types I-III) will be recruited to serve as calibration subjects. Each examiner will assess PI, GI, BOP and PD and CEJ using the manual University of North Carolina probe (UNC-12).

I.2 Structure of the Calibration Sessions

Within three weeks of the initiation of a study, six to nine adult calibration subjects (18 years of age or older) with a range of periodontal health (types I-III) per examiner will be recruited for the first portion of the calibration session. This portion will span three consecutive days at the University of North Carolina. All clinical examiners assigned to the study will participate. Prior to the clinical procedures, all examiners will complete a review seminar during which the trial objectives, measurement parameters, techniques and schedule will be discussed. Prior to the calibration, all Centers will provide examiners with training for Cardio-pulmonary Resuscitation and Ethics of Human Subjects Research.

Prior to arrival in Chapel Hill, examiners will review the study manual. The training session will begin with a review seminar. The standard examiner (SE) will go over the criteria being used for each aspect of the examination and will answer questions. The structure of the calibration session will be described. The seminar will also include a review of the clinical procedure manual, use of the specific indices, instrumentation, case scenario slides, and demonstration models. A review of the use of the direct data entry system (DDES) will also be conducted. During the calibration session, data will be entered directly into the computer to allow for a comparison of data calls. Thus, examiners will have an opportunity to
become familiar with the data entry program during the calibration. In addition, laboratory procedures for the collection, shipping, and handling of biological specimens will be addressed.

At the UNC School of Dentistry General and Oral (GO) Health clinic, each calibration subject will be asked to read and sign an informed consent form. A review and update of the medical history will be completed and vital signs recorded prior to the start of the clinical portion of the calibration session.

One practice calibration subject will be examined by each examiner with the SE observing and providing direct and immediate feedback on examiner technique. Examiners will practice taking a plaque sample using the standard protocol described in the manual. After the practice calibration subject has been examined, an opportunity is provided for discussion of the practice session.

**DAY 1**
Since the exclusion criteria of the study require a determination of hopeless periodontally involved teeth or severely fractured, carious crowns, examiners will be calibrated on methods for determining whether a participant should be included in the study. Six subjects with carious lesions will be examined to apply inclusion criteria and clinical tests for determining study eligibility as described in the manual.

Calibration on PI and GI is especially difficult when multiple examiners are calibrated using the same patients. The act of probing may disrupt the plaque and may alter the degree of gingival inflammation present. To minimize this source of error, the PI will be assessed on six half mouths with every examiner paired at least once with each other examiner and six times with the standard examiner (SE). The plaque scores will be determined by visual inspection so that subsequent examiners will be observing the same conditions. All examiners will simultaneously assess GI for six calibration subjects. The standard examiner will sweep the anterior surfaces of the gingival margins from canine to canine on the maxillary and mandibular arches. Examiners will record their findings for the GI in writing so as not to influence other examiners; the information will later be entered into the DDES by the recorder.

At the end of the clinical session, examiners will be given an opportunity to discuss questions and clinical issues that occurred during the afternoon session.

**DAY 2**
Each examiner will perform a periodontal examination on six out of eight calibration subjects (half-mouth) using a dental chair and light. A recorder will assign each examiner a code for identification in the DDES. An oral soft tissue examination, including oral cancer screening, will be conducted, and present teeth identified. The standard examiner will collect PI, GI, BOP, PD and CEJ for all calibration subjects (full-mouth). Accordingly, each center examiner will perform PI, GI, BOP, PD and CEJ measurements per site (i.e., six sites per tooth) for all calibration patients. He or she will measure PD and CEJ (PI and GI already having been measured) at all applicable sites followed by BOP.

During the clinical sessions, percent agreement, weighted kappa, and intraclass correlation will be calculated for each exam pairing. The SE and the examiners will then review any differences while the patient is still available. Since repeated probing on the same subject can produce different measurements, we will limit each volunteer to three full mouth probing exams. Therefore, the order in which each examiner and the SE will see a subject is randomized.

Whereas each examiner will manually probe half mouths comparable to three adult periodontitis calibration subjects, the SE will probe full mouths on all subjects involved in the calibration session. A calibrated UNC-12 probe will be used to obtain PD and CEJ in millimeters at six periodontal sites (mesiofacial, facial, distofacial, mesiolingual, lingual, distolingual) per tooth. If a reading falls between two millimeter markings, the decision rule will be to record the higher of the two numbers.

Immediately following completion of data collection, scores will be printed and compared for agreement within one millimeter. All disagreement calls will be discussed and the call repeated on the subject in the presence of the SE. For PD and CEJ, disagreement will be defined as a difference of more than 1
At this point, adjustments to the examiners' technique will be made. By the end of the second clinical training day, we usually find that the SE and examiners are in close agreement.

**DAY 3**

Due to the nature of the study, it is also necessary to assess examiner reliability while the subject is in a hospital bed. Thus, for the third day of the calibration sessions, the same subjects will be reexamined by the same examiners at the General Clinical Research Center (GCRC) located in North Carolina Memorial Hospital. Examiners will be equipped with a headlamp, apply the same indices (PD, CEJ, BOP), and examine the same subjects they did the day before. The standard examiner will examine each of the subjects. Immediately following data collection, scores will be printed out and compared for agreement. All disagreement calls will be discussed and the call repeated on the subject in the presence of the SE. Again, adjustments to the examiners' technique will be made when necessary. An opportunity for discussion will be provided. After the clinical data collection is completed, a debriefing session will occur to allow examiners an opportunity to discuss any points that are unclear prior to completing the calibration session.

Examiners will return to their respective field center and will examine 5-10 “trial patients”. They will then ship the examination data and biological specimens to study investigators at the UNC School of Dentistry, Chapel Hill, where the data will be evaluated. When the investigators are satisfied that the examiners can consistently follow the protocol, the examiners will be given written permission to begin examinations in the field. Centralized training sessions will be repeated annually and the SE will visit each field center at least once per year to ensure that the examination, sample collection, storage, and shipping procedures are implemented according to the study protocol.

**1.3 Criteria for agreement**

Statistics will be computed for all parameters for examiner agreement as follows: Percent agreement, weighted kappa, and intra-class correlation coefficient.

**1.4 Data Analysis**

The following table (developed by Dr. Gary Koch) is an example of the type of analysis of agreement. The first three columns provide an indication of intra-examiner reliability. Percent agreement within 1.0 mm will be calculated for the entire group and for the SE, but calculation of these measures for each examiner are based on two repeated measurements and will be unreliable. The mean absolute differences between the two passes can be used to establish a definition of change in attachment level. In the example below, a change in attachment level of 1 mm or more is likely to be a real change. A calculation of the standard deviation between the two passes for all examiners could provide an alternative measure. Again, analyses of individual examiners will be unreliable.

Similar statistics as well as Intra-class correlation’s can be calculated to determine how well all of the examiners agree with the SE (top row of table). These statistics can be calculated for each pass. The Standard Deviation of the Mean Square Error Column provides verification that 95% of the differences are within 1.0 mm. We also can calculate Intra-class Correlation for all examiners compared to the SE.
## Results from Calibration Trial
### Intra- and Inter-examiner Reliability
#### Site Based Analysis

<table>
<thead>
<tr>
<th></th>
<th>Between Pass 1 and Pass 2</th>
<th>Between Each Examiner and Standard Examiner</th>
<th>Among Examiners</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intra-examiner</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>% Agreement within 1.0mm</td>
<td>% Agreement within 1.0mm</td>
<td></td>
</tr>
<tr>
<td></td>
<td>between two passes</td>
<td>between examiner &amp; SE</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean absolute diff between two passes</td>
<td>Mean absolute diff between examiner &amp; SE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Std Dev of Mean Square Error</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ICC ** Each examiner against each other examiner.</td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>96.6</td>
<td>88.5</td>
<td>85.3</td>
</tr>
<tr>
<td>SE</td>
<td>96.4</td>
<td>N/A</td>
<td>not included</td>
</tr>
<tr>
<td>E1</td>
<td>93.8</td>
<td>84.2</td>
<td>85.0</td>
</tr>
<tr>
<td>E2</td>
<td>98.8</td>
<td>88.1</td>
<td>82.9</td>
</tr>
<tr>
<td>E3</td>
<td>95.8</td>
<td>90.8</td>
<td>88.0</td>
</tr>
<tr>
<td>E4</td>
<td>88.9</td>
<td>82.9</td>
<td>82.4</td>
</tr>
<tr>
<td>E5</td>
<td>100.0</td>
<td>90.5</td>
<td>86.2</td>
</tr>
<tr>
<td>.</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>En</td>
<td>95.2</td>
<td>92.9</td>
<td>85.3</td>
</tr>
</tbody>
</table>

SE = Standard Examiner  
N/A = Not applicable for this calculation  
** not shown due to space are the ICC’s for each examiner against each other examiner.

The hypothetical data set provided above is modeled after the actual data obtained from a calibration exercise that was performed for a clinical study conducted by Center for Oral and Systemic Disorders.

**Inter-rater/inter-examiner reliability will be at least .60 or above and intra-rater or intra-examiner reliability will be at least .75 or above.
REFERENCES