UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF VIRGINIA
Norfolk Division

IN RE PEANUT FARMERS ANTITRUST LITIGATION	Civil Action No. 2:19-cv-00463 **Honorable Raymond A. Jackson** **Honorable Lawrence R. Leonard**

VALIDATION PROTOCOL ORDER

This Validation Protocol Order provides the validation procedures referred to in ¶ VI(C)(5) of the Court's ESI Protocol (ECF No. 92). This Validation Protocol Order shall govern the parties in the above-captioned case whether they currently are involved or become so in the future, and any actions that may later be deemed related to or consolidated with this case.

A producing party's review process should incorporate quality-control and quality-assurance procedures to ensure a reasonable production consistent with the requirements of Federal Rule of Civil Procedure 26(g). Once a producing party reasonably believes that it has produced or identified for production substantially all responsive non-privileged documents, it shall conduct validation according to the sampling protocol described below and in Paragraph I (hereinafter collectively referred to as "Validation Protocol").

This Validation Protocol shall apply to the review process regardless of whether technology-assisted review or exhaustive manual review was used by the producing party.

A. The Document Collection ("Collection") is defined as including all documents identified for review for responsiveness and/or privilege following the application of keywords or other culling criteria. This Validation Protocol assumes that the completeness or adequacy of the Collection has already been established. For the purposes of Plaintiffs' validation requirements

under this Validation Protocol, the Collection refers to the combined set of documents from all named representatives rather than to each individual named representative.

B. The Collection shall be partitioned into the following two Subcollections:

1. Documents designated by the review as responsive to at least one Request for Production, including any privileged documents, but not including family members of responsive documents, unless those family members are deemed to be responsive in their own right ("Subcollection B(1)");

2. Documents designated by the review as non-responsive to any of the Requests for Production ("Subcollection B(2)").

C. A sample shall be drawn consisting of the following:

1. 300 documents selected at random from Subcollection B(1) ("Subsample C(1)");

2. A sample of documents selected at random from Subcollection B(2) ("Subsample C(2)"); the size of Subsample C(2) will be determined after completing the review of Subsample C(1) and will be chosen so as to ensure that the margin of error associated with the recall estimate (*see* Paragraph I) does not exceed +/- 3% (calculated at a level of 95% statistical confidence); if the specified margin of error cannot be guaranteed with a sample of 2,500 documents or fewer, the size of Subsample C(2) will be 2,500 documents.

D. Should a producing party believe that the sample sizes specified in Paragraph C would be disproportionate or unduly burdensome under the circumstances, that party shall promptly raise the issue with the requesting party. To the extent a dispute remains concerning the

sample sizes to be used after good faith negotiations have occurred, either party may request the assistance of the Court in resolving such dispute.

E. The Subsamples C(1) and C(2) shall be combined into a single "Validation Sample," with no indication of the Subcollection from which the documents were derived or how they were previously coded. The Validation Sample shall be reviewed and coded by a subject matter expert ("SME") who is knowledgeable about the subject matter of the litigation. This should be an attorney who is familiar with the RFPs and the issues in the case. During the course of the review of the Validation Sample, the SME shall not be provided with information on the Subcollection or Subsample from which any document was derived or the prior coding of any document. The intent of this requirement is to ensure that the review of the Validation Sample is blind; it does not preclude a party from selecting as SMEs attorneys who may have had prior involvement in the original review process.

F. Once the coding in Paragraph E has been completed, the producing party shall prepare a table listing each of the documents in Subsamples C(1) and C(2). For each document, the table shall include:

1. the Bates number of the document (for documents produced), or a control/identification number (for non-produced documents);

2. the Subsample from which the document came (*i.e.*, C(1) or C(2));

3. the SME's responsiveness coding for the document (*i.e.*, responsive or non-responsive);

4. the SME's privilege coding for the document (*i.e.*, privileged or not privileged). If the document is coded as non-responsive, a privilege determination need

not be made. All documents in the samples coded as privileged shall be included on the producing party's privilege log.

5. for Plaintiffs, the name of the named representative associated with the document.

G. The following items shall be provided to the requesting party:

1. the table described in Paragraph F;

2. a copy of any responsive, non-privileged document in Subsamples C(1) and C(2) that was not previously produced or identified for production to the requesting party;

3. the statistics and recall estimate detailed in Paragraph I to this Validation Protocol.

H. Once the requesting party has received and has had an opportunity to review the items described in Paragraph G and Paragraph I, the parties shall meet and confer to determine whether or not the parties agree that the recall estimate, and the quantity and nature of the responsive documents identified through the sampling process, indicate that the review is substantially complete. If the recall estimate and the samples indicate that Subcollection B(2) still contains a substantial number of non-marginal, non-duplicative responsive documents as compared to Subcollection B(1), the review and quality assurance process shall continue, and the validation process shall be repeated, as warranted. If the parties are unable to agree on whether the review is substantially complete, or on whether the validation process must be repeated, the parties may request the assistance of the Court in resolving the dispute.

I. An estimate of recall shall be computed to inform the decision-making process described in Paragraph G of the Validation Protocol; however, the absolute number in its own right shall not be dispositive of whether or not a review is substantially complete. Also of

concern is the novelty and materiality (or conversely, the duplicative or marginal nature) of any responsive documents identified in Subsample C(2). The estimate of recall shall be derived as described below. It should be noted that, when conducted by an SME pursuant to Paragraph E of the Validation Protocol, a recall estimate on the order of 70% to 80% is consistent with, but not the sole indicator of, an adequate (*i.e.*, high-quality) review. A recall estimate somewhat lower than this does not necessarily indicate that a review is inadequate, nor does a recall in this range or higher necessarily indicate that a review is adequate; the final determination also will depend on the quantity and nature of the documents that were missed by the review process.

Recall Estimation Method:[1]

The number of responsive documents found ≈ the size of Subcollection B(1) × the number of responsive docs in Subsample C(1) ÷ 300.

The number of responsive documents coded incorrectly ≈ the size of Subcollection B(2) × the number of responsive documents in Subsample C(2) ÷ the size of Subsample C(2).

Estimated recall ≈ the number of responsive documents found ÷ (the number of responsive documents found + the number of responsive documents coded incorrectly).

The Clerk is **REQUESTED** to deliver a copy of this Order to all counsel of record.

It is so **ORDERED**.

LAWRENCE R. LEONARD
UNITED STATES DISTRICT MAGISTRATE JUDGE

[1] The margins of error (or confidence intervals) associated with the recall estimate can be calculated according to methods discussed in the relevant academic literature (*e.g.*, Webber, W. (2013), Approximate recall confidence intervals. *ACM Transactions on Information Systems (TOIS)*, 31(1), 1-33).