APPENDIX B

Definition of Consensus

Survey data sent to Delphi panelists were evaluated using a 9-point Likert scale. The consensus definition was based on the RAND/UCLA Appropriateness method (Fitch et al., 2001). The scale was ranked with one, meaning "totally disagree" or "harm outweighed the expected benefit," and nine, meaning "totally agree" or "benefit outweighs the expected harm" (Jones & Hunter, 1995, p.311; Fitch et al., 2001, p. 4). The consensus was defined as a score of 7 to 9 as "Agreement," scores of 4 to 6 were considered "Uncertain," and scores of 1 to 3 were considered "Disagreement." If no consensus was established, it was considered "uncertain" (Cho et al., 2019; Fitch et al., 2001; Jones & Hunter, 1995; Lee et al., 2020).

Defining the level of consensus was based on the RAND algorithm (Figure 2) (Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019). Cho et al. (2019), Franco-Sadud et al. (2019), Scheeren et al. (2019), and Soni et al. (2019) describe the terms "Perfect consensus," "Very good consensus," "Good consensus," "Some consensus," and "No consensus" to provide an in-depth understanding of the level of consensus as described during the RAND algorithm. The term "Perfect consensus" describes 100 percent of participants rating the statement 7, 8, or 9. "Very good consensus" describes "median and middle 50% of respondents are found at one integer, or 80% of respondents are within one integer of the median" (Cho et al., 2019, p. E8; Franco-Sadud et al., 2019, p. E4; Soni et al., 2019, p. E3). "Good consensus" is described as "50% of respondents are within one integer of the median or 80% of the respondents are within two integers of the median" (Cho et al., 2019, p. E8; Franco-Sadud et al., 2019, p. E4; Soni et al., 2019, p. E3). "Some consensus" is described as "50% of respondents are within two integers of the median or 80% of respondents are within three integers of the median" (Cho et al., 2019, p. E8; Franco-Sadud et al., 2019, p. E4; Soni et al., 2019, p. E3). "No consensus" indicates "all other responses" or "any median with disagreement" (Cho et al., 2019, p. E8; Franco-Sadud et al., 2019, p. E4; Soni et al., 2019, p. E3).

APPENDIX B-2: RAND Algorithm



(Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019)

Cho et al. (2019). Franco-Sadud et al. (2019), Scheeren et al. (2019), and Soni et al. (2019) further describe the degree of consensus and the strength of recommendation. The description elaborates on the previous definitions of the level of consensus and aligns them with the strength of recommendation. The strength of the recommendation was based on a modification of the Grade guidelines (Guyatt et al., 2011). Cho et al. (2019), Franco-Sadud et al. (2019) Scheeren et al. (2019), and Soni et al. (2019) used a modification of the Grade guidelines by using the terms as noted in the RAND Algorithm such as "Strong Recommendation," "Conditional/Weak Recommendation," and "No Recommendation" in place of the terms "High, Moderate, Low and Very Low" (Guyatt et al., 2011).

The modified Grade Method, as discussed by Cho et al. (2019), Franco-Sadud et al. (2019), Scheeren et al. (2019), and Soni et al. (2019), is based on the appropriateness and degree of consensus. Strong recommendations are based on the degree of consensus is at least good, and the median score is not in the undermined middle zone (the median is not in the four to six-zone; therefore, it is either in the seven to nine-zone or the one to three-zone) (Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019). Therefore, a strong recommendation can have either two categories: "Strong With" or "Strong Against." The "Strong

With" category is categorized as a median of seven to nine, and the "Strong Against" category is categorized as one to three.

Weak recommendations are based on the degree of consensus is "some consensus" with any median score or median score of four to six with any degree of consensus (Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019). Therefore, a "Weak Recommendation" has two categories: "Weak With" and "Weak Against." The "Weak With" category is defined as the middle 50% of the interquartile range is equal to four to nine. The "Weak Against" is defined as the middle 50% of the interquartile range is equal to one or less than four (Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019).

"Conditional Recommendations," which are categorized alongside "Weak Recommendations," were categorized as 70 to 80% of the participants agreeing on a recommendation/statement (Cho et al., 2019; Franco-Sadud et al., 2019; Scheeren et al., 2019; Soni et al., 2019).

Completing the first round was the first phase to determine consensus. Due to a lack of consensus on specific questions, those questions were carried over into the second round. During the second round, survey questions from the first round that did not meet consensus were modified based on feedback from panelist-free discussion boxes. Those questions that did not meet consensus during the second round based on the RAND/UCLA Appropriateness method (Fitch et al., 2001) were explored during the qualitative interview phase of this study.

Analysis

Assessing the degree of agreement and disagreement amongst Delphi panelists, the surveyed results underwent analysis using central tendencies (means, medians) and levels of dispersion (standard deviations and interquartile ranges) to assess the degree of variability between the surveyed responses (Hasson et al., 2000; Lee et al., 2020). Additionally, means, medians, standard deviations, and interquartile ranges were used to compare proportion data between rounds to assess the overall acceptance rate of the surveyed data (Jones & Hunter, 1995; Lee et al., 2020).

The study data were collected by the author of this article and entered into a Microsoft Office Excel for Mac version 16.41 (Microsoft et al.) Spreadsheet for data analysis (Cho et al., 2019; Fitch et al., 2001; Jones & Hunter, 1995; Lee et al., 2020). Once analyzed, the study data provided central tendencies and levels of dispersion to assess the level and degree of consensus for participant responses. The central tendencies expressed in this analysis are expressed as medians, and the dispersions are expressed as interquartile ranges. The consensus was based on the medians, and the level of dispersion was expressed using interquartile ranges. A participants' response sheet was provided for each round, with the final data displayed after the second round.

Analysis of the data used three zones/regions: an Agreement zone/region (median 7 through 9), an Uncertain zone/region (median 4 through 6), and a Disagreement zone/region (median 1 through 3). The median establishes where 50 percent of the votes were cased. Establishing a consensus requires a minimum of 70 percent scoring of a statement within a specific zone/region. Therefore, if 30 percent of the votes are outside a particular "zone/region," there is no consensus. A statement with a median score of seven or higher would be classified within the Agreement zone/region because 50 percent of the votes were categorized between

seven and nine. Despite being classified in the Agreement zone, there would be disagreement about the statement if 30 percent or more participants did not score "7, 8, or 9".