

Supplemental Digital Content 3: Supplemental Text 3. The nominal group technique and consensus results

The existing literature lacks sufficient evidence to effectively guide extubation decisions making for neurosurgical patients. To address this gap, we decided to select model items using the nominal group technique. Previous research has established the efficacy of the nominal group technique in consolidating individual judgments into group decisions regarding the most suitable course of action. This method, which involves voting and discussion, has demonstrated its feasibility and reliability in determining research priorities among clinicians, particularly in scenarios where evidence is scarce. The nominal group technique has been extensively employed for developing guidelines and reaching consensus.

The consensus was established in three steps: 1) Step I: literature review and small-scale clinical consultation to acquire the relevant items; 2) Step II: five rounds of online face-to-face nominal group meetings to acquire the consensus rate for each item; 3) Step III: one round of comments and iterative review to acquire the final items.

After a literature review and small-scale clinical consultation, a list of 10 consciousness and airway protecting function items with 36 different assessment methods was summarized. In the five rounds of online face-to-face nominal group meetings, experts rated each item independently based on a five-point Likert scale ranging from disagree (1 point) to agree (5 points) (Table S1) and gave explanations for their rate. An open discussion was allowed after successive statements from each expert. The consensus rate was defined a priori as the number of experts rated somewhat agree to agree (five-point Likert scale 4-5)/the number of all the experts. Items with a consensus rate over 80% were directly included as final assessments at any round. Items with a consensus rate of less than 50% were excluded at any round. While items with a consensus rate of 50-80% were reconsidered and rated in the next round. If the consensus rate was no more than 80% at round five, the item would be deleted. For

each item, the first question was “Do you think the item need to be assessed” followed by different assessment methods. If the consensus rate for “The item needs to be assessed” was less than 50%, the rate for different assessment methods of the item would not be conducted. The results of five rounds of nominal group meetings were shown below in Table S2. In the comments and iterative review step, the chosen items and assessment criteria were provided to experts who participated in the nominal group meetings through emails. They were invited to adjust to form a final version for five chosen items.

Table S1. Five-point Likert scale

Level	Response
1	Disagrees
2	Somewhat disagree
3	Neither agree nor disagree
4	Somewhat agree
5	Agree

Table S2. Results of five rounds of nominal group meetings

Items	Round 1			Round 2		Round 3		Round 4		Round 5		Final included
	No	Question	Rate*	No	Rate*	No	Rate*	No	Rate*	No	Rate*	
Spontaneous cough	1	Do you think spontaneous cough need to be assessed?	100									√
	1.1	Two categories: Yes/No	59	1.1	53	1.1	29					
	1.2	Two categories: Strong/Poor	71	1.2	76	1.2	82					√
	1.3	Three categories: Strong/Poor/None	53	1.3	12							

	1.4	Four categories: Strong/Moderate/Poor/None	35						
	1.5	Six categories: No cough/Audible movement of air through the airways/Cough audible, but weak/Clearly audible cough/Strong cough/Multiple sequential strong coughs	12						
	1.6	White card test	41						
	1.7	Assessment of cough peak flow	35						
Suctioning cough	2	Do you think suctioning cough need to be assessed?	100					√	
	2.1	Two categories: Yes/No	53	2.1	53	2.1	53	2.1	24
	2.2	Two categories: Strong/Poor	59	2.2	65	2.2	71	2.2	82
	2.3	Three categories: Strong/Poor/None	53	2.3	18				
	2.4	Four categories: Strong/Moderate/Poor/None	18						
	2.5	Six categories: No cough/Audible movement of air through the airways/Cough audible, but weak/Clearly audible cough/Strong cough/Multiple sequential strong coughs	12						
Swallowing	3	Do you think swallowing need to be assessed?	88						√
	3.1	Two categories: Yes/No	53	3.1	47				
	3.2	Two categories: Strong/Poor	65	3.2	82				√
Gag reflex	4	Do you think gag reflex need to be assessed?	41						
	4.1	Two categories: Yes/No	NA						
	4.2	Two categories: Strong/Poor	NA						
	4.3	Four categories: Strong/Moderate/Poor/None	NA						
Tongue protrusion	5	Do you think tongue protrusion need to be assessed?	76						√
	5.1	Two categories: Yes/No	59	5.1	24				
	5.2	Two categories: Strong/Poor	65	5.2	88				√
Sputum quantity	6	Do you think sputum quantity need to be assessed?	35						

	6.1	Two categories: Copious/Few	NA					
	6.2	Four categories: ≥ 3 passes/2 passes/1 pass/None	NA					
	6.3	Quantitative measurement	NA					
Sputum viscosity	7	Do you think sputum viscosity need to be assessed?	35					
	7.1	Three categories: Watery/Frothy/Thick	NA					
	7.2	Four categories: Watery/Frothy/Thick/Tenacious	NA					
Suctioning frequency	8	Do you think suctioning frequency need to be assessed?	47					
	8.1	Two categories: $>2h/\leq 2h$	NA					
	8.2	Three categories: 2-4h/1-2h/ $<1h$	NA					
	8.3	Four categories: $>3h/2-3h/1-2h/\leq 1h$	NA					
Sputum character	9	Do you think sputum character need to be assessed?	41					
	9.1	Four categories: Clean/Tan/Yellow/Green	NA					
Consciousness	10	Do you think consciousness need to be assessed?	100					√
	10.1	GCS: $\geq 10/\leq 10$	65	10.1	53	10.1	29	
	10.2	GCS-M: $\geq 5/\leq 5$	71	10.2	76	10.2	88	√
	10.3	GCS-E: $\geq 3/\leq 3$	47					
	10.4	Open eyes: Yes/No	29					
	10.5	Visual pursuit: Yes/No	35					
	10.6	Grasp hand: Yes/No	53	10.6	18			
	10.7	Wiggle toes: Yes/No	12					
	10.8	Show fingers: Yes/No	29					

*The consensus rate (%) = the number of experts rated somewhat agree to agree (five-point Likert scale 4-5)/the number of all the experts.

†Three colors represent different consensus rates. Items with green grids (consensus rate $\geq 80\%$) were included as final assessments at any round. Items with yellow grids (consensus rate between 50-80%) were rated at the next round. Items with gray grids (consensus rate $\leq 50\%$) were deleted at that round.

‡ NA: not applicable. If the consensus rate for “The item needs to be assessed” was less than 50%, rate for different assessment methods of the item would not be conducted.