# Medevac Interrater Reliability Study

**Final Report**

Medevac Interrater Study Team

2025-12-04

## Table of contents

**1 Study Overview** **1**
1.1 Table 1: Study Composition . . . 1
1.2 Figure 1: Response Distribution by Physician . . . 3
1.3 Overall Results . . . 3
1.4 Tables 2.1-2.3: Summary Statistics by Individual Vignette . . . 4
1.5 Figure 2: Agreement and Confidence by Subclass . . . 10
1.6 Tables 3.1-3.3: Summary Statistics by Individual Vignette (High Confidence Only, n=20) . . . 10
1.6.1 Table 3.1: One Correct Answer (High Confidence Only) . . . 10
1.6.2 Table 3.2: Two Reasonable Options (High Confidence Only) . . . 12
1.6.3 Table 3.3: Highly Ambiguous (High Confidence Only) . . . 14
1.7 Figure 3: Agreement and Confidence by Subclass (High Confidence Only, n=20) 16
1.8 Tables 4.1-4.3: Summary Statistics by Individual Vignette (Experienced Respondents Only, n=13) . . . 16
1.8.1 Table 4.1: One Correct Answer (Experienced Respondents Only) . . . 16
1.8.2 Table 4.2: Two Reasonable Options (Experienced Respondents Only) . . . 18
1.8.3 Table 4.3: Highly Ambiguous (Experienced Respondents Only) . . . 20
1.9 Figure 4: Agreement and Confidence by Subclass (Experienced Respondents Only, n=13) . . . 22

# 1 Study Overview

## 1.1 Table 1: Study Composition

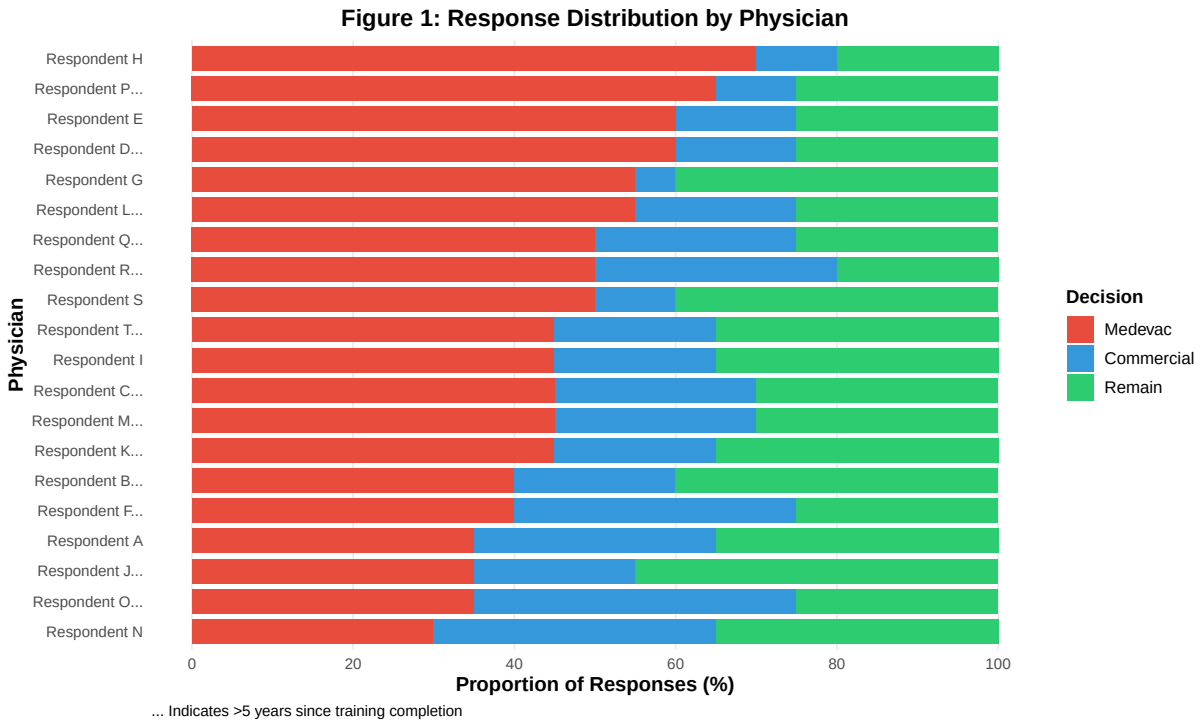
This table provides a comprehensive overview of the study design, organized into two sections: Vignettes and Respondents.

Table 1: Study Composition: Vignettes, Respondents, and Response Metrics

| | |
|---|---|
| ****VIGNETTES**** | |
| **Total Vignettes** | **20** |
| ***Class A (One Correct Answer)*** | ***8*** |
| Medevac | 4 |
| Commercial | 2 |
| Remain | 2 |
| Class B (Two Reasonable Options) | 9 |
| ***Medevac or Commercial*** | ***2*** |
| Commercial or Remain | 4 |
| Medevac or Remain | 3 |
| Class C (Highly Ambiguous) | 3 |
| ***All Options Reasonable*** | ***3*** |
| ****RESPONDENTS**** | |
| Total Physicians | 20 |
| By Affiliation: | |
| ANMC | 10 |
| Maniilaq | 8 |
| Other | 2 |
| Years Since Training: | |
| Mean (SD) | 9.6 (7.6) |
| Median | 8 |
| Range | 1-30 |
| ****RESPONSES**** | |
| Total Responses | 400 |

**Note:** This table summarizes the complete study composition with 20 unique clinical vignettes evaluated by 20 physicians (mean experience: 9.6 years since training completion), generating 400 total responses.

## 1.2 Figure 1: Response Distribution by Physician




---

## 1.3 Overall Results

Mean Agreement %: 74.5%

Median Agreement %: 75%

Agreement Range: 45% - 100%

Mean Confidence: 7.8

Mean Entropy: 0.786

Vignettes with >=75% Agreement: 13

Vignettes with <50% Agreement: 3

---

## 1.4 Tables 2.1-2.3: Summary Statistics by Individual Vignette

These tables show detailed statistics for each vignette organized by class, with response distributions presented as percentage and count (% (n=)).

::: .cell ::: .cell-output-display

Table 2: Table 2.1: Class A - One Correct Answer

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **Medevac** | **1** | 100.0% (n=20) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100** | 1.000 | Perfect |
| **Medevac** | **4** | 95.0% (n=19) | 0.0% (n=0) | 5.0% (n=1) | Medevac | **95** | 0.889 | Very Good |
| **Medevac** | **13** | 100.0% (n=20) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100** | 1.000 | Perfect |
| **Medevac** | **19** | 100.0% (n=20) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100** | 1.000 | Perfect |
| **Commercial** | **9** | 0.0% (n=0) | 75.0% (n=15) | 25.0% (n=5) | Commercial | **75** | 0.368 | Fair |
| **Commercial** | **15** | 65.0% (n=13) | 30.0% (n=6) | 5.0% (n=1) | Medevac | **65** | 0.326 | Fair |
| **Remain** | **8** | 5.0% (n=1) | 15.0% (n=3) | 80.0% (n=16) | Remain | **80** | 0.576 | Moderate |
| **Remain** | **11** | 0.0% (n=0) | 25.0% (n=5) | 75.0% (n=15) | Remain | **75** | 0.368 | Fair |

::: :::

::: .cell ::: .cell-output-display

Table 3: Table 2.2: Class B - Two Reasonable Options

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation | En |
|---|---|---|---|---|---|---|---|---|---|
| **Medevac or Commercial** | **7** | 95.0% (n=19) | 5.0% (n=1) | 0.0% (n=0) | Medevac | **95** | 0.889 | Very Good | |
| **Medevac or Commercial** | **10** | 70.0% (n=14) | 30.0% (n=6) | 0.0% (n=0) | Medevac | **70** | 0.238 | Fair | |
| **Commercial or Remain** | **2** | 20.0% (n=4) | 45.0% (n=9) | 35.0% (n=7) | Commercial | **45** | 0.021 | Slight | |
| **Commercial or Remain** | **5** | 0.0% (n=0) | 75.0% (n=15) | 25.0% (n=5) | Commercial | **75** | 0.368 | Fair | |
| **Commercial or Remain** | **6** | 25.0% (n=5) | 45.0% (n=9) | 30.0% (n=6) | Commercial | **45** | -0.002 | None | |
| **Commercial or Remain** | **14** | 0.0% (n=0) | 20.0% (n=4) | 80.0% (n=16) | Remain | **80** | 0.505 | Moderate | |
| **Medevac or Remain** | **16** | 75.0% (n=15) | 0.0% (n=0) | 25.0% (n=5) | Medevac | **75** | 0.368 | Fair | |
| **Medevac or Remain** | **17** | 50.0% (n=10) | 0.0% (n=0) | 50.0% (n=10) | Medevac | **50** | -0.053 | None | |
| **Medevac or Remain** | **18** | 25.0% (n=5) | 0.0% (n=0) | 75.0% (n=15) | Remain | **75** | 0.368 | Fair | |

::: :::

::: .cell ::: .cell-output-display

Table 4: Table 2.3: Class C - Highly Ambiguous

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **All Options Reasonable** | **3** | 20.0% (n=4) | 0.0% (n=0) | 80.0% (n=16) | Remain | **80** | 0.505 | Moderate |
| **All Options Reasonable** | **12** | 45.0% (n=9) | 40.0% (n=8) | 15.0% (n=3) | Medevac | **45** | 0.065 | Slight |
| **All Options Reasonable** | **20** | 65.0% (n=13) | 25.0% (n=5) | 10.0% (n=2) | Medevac | **65** | 0.289 | Fair |

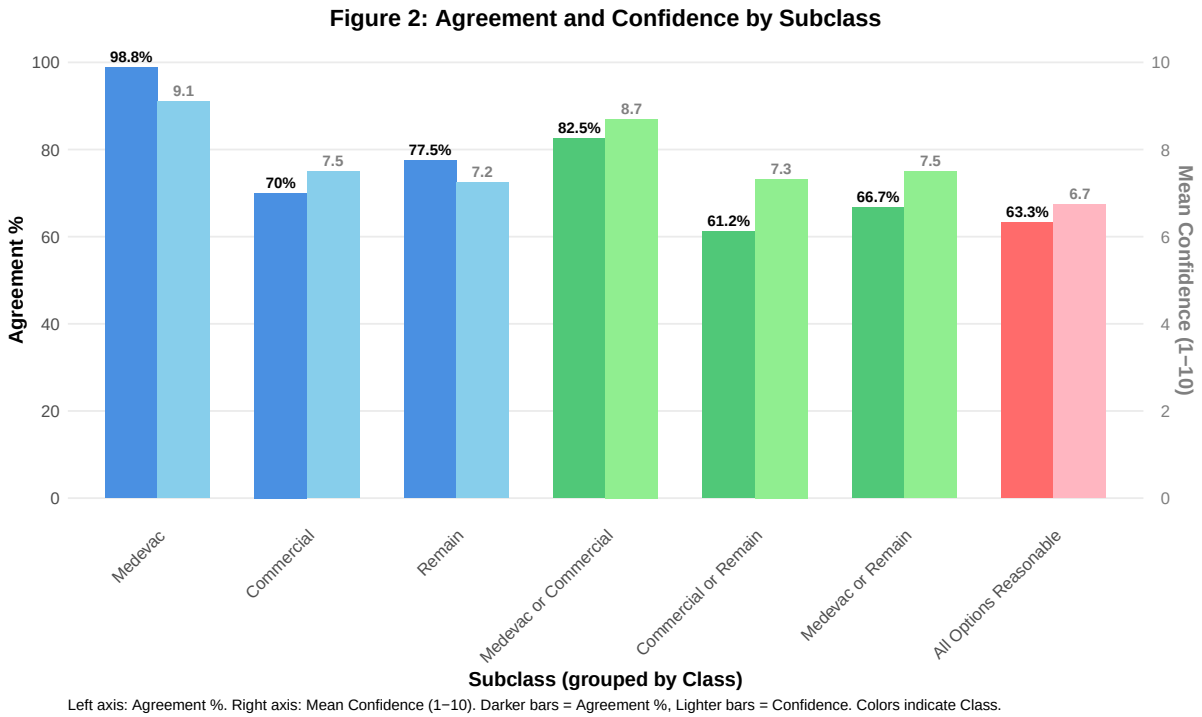


::: :::

**Note:** Each vignette was evaluated by 20 physicians. Agreement % (highlighted in green  75%, orange 50-74%, red <50%) indicates the percentage who selected the modal decision. Entropy measures disagreement (0 = perfect agreement, higher = more disagreement).

---

## 1.5 Figure 2: Agreement and Confidence by Subclass




---

## 1.6 Tables 3.1-3.3: Summary Statistics by Individual Vignette (High Confidence Only, n=20)

These tables show detailed statistics for each vignette, including only responses with confidence scores of 8 or higher.

### 1.6.1 Table 3.1: One Correct Answer (High Confidence Only)

::: .cell ::: .cell-output-display

Table 5: Table 3.1: Class A - One Correct Answer (Confidence >= 8)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **Medevac** | **1** | 100.0% (n=19) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Medevac** | **4** | 100.0% (n=18) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Medevac** | **13** | 100.0% (n=19) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Medevac** | **19** | 100.0% (n=18) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Commercial** | **9** | 0.0% (n=0) | 66.7% (n=8) | 33.3% (n=4) | Commercial | **66.7** | 0.127 | Slight |
| **Commercial** | **15** | 83.3% (n=10) | 16.7% (n=2) | 0.0% (n=0) | Medevac | **83.3** | 0.580 | Moderate |
| **Remain** | **8** | 0.0% (n=0) | 7.7% (n=1) | 92.3% (n=12) | Remain | **92.3** | 0.821 | Very Good |
| **Remain** | **11** | 0.0% (n=0) | 25.0% (n=3) | 75.0% (n=9) | Remain | **75.0** | 0.345 | Fair |

::: :::

### 1.6.2 Table 3.2: Two Reasonable Options (High Confidence Only)

::: .cell ::: .cell-output-display

Table 6: Table 3.2: Class B - Two Reasonable Options (Confidence >= 8)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation | En |
|---|---|---|---|---|---|---|---|---|---|
| **Medevac or Commercial** | **7** | 94.7% (n=18) | 5.3% (n=1) | 0.0% (n=0) | Medevac | **94.7** | 0.883 | Very Good | |
| **Medevac or Commercial** | **10** | 92.3% (n=12) | 7.7% (n=1) | 0.0% (n=0) | Medevac | **92.3** | 0.821 | Very Good | |
| **Commercial or Remain** | **2** | 27.3% (n=3) | 45.5% (n=5) | 27.3% (n=3) | Commercial | **45.5** | -0.046 | None | |
| **Commercial or Remain** | **5** | 0.0% (n=0) | 76.9% (n=10) | 23.1% (n=3) | Commercial | **76.9** | 0.404 | Moderate | |
| **Commercial or Remain** | **6** | 50.0% (n=4) | 12.5% (n=1) | 37.5% (n=3) | Medevac | **50.0** | 0.035 | Slight | |
| **Commercial or Remain** | **14** | 0.0% (n=0) | 0.0% (n=0) | 100.0% (n=13) | Remain | **100.0** | 1.000 | Perfect | |
| **Medevac or Remain** | **16** | 91.7% (n=11) | 0.0% (n=0) | 8.3% (n=1) | Medevac | **91.7** | 0.803 | Very Good | |
| **Medevac or Remain** | **17** | 62.5% (n=5) | 0.0% (n=0) | 37.5% (n=3) | Medevac | **62.5** | -0.008 | None | |
| **Medevac or Remain** | **18** | 18.8% (n=3) | 0.0% (n=0) | 81.2% (n=13) | Remain | **81.2** | 0.533 | Moderate | |

::: :::

### 1.6.3 Table 3.3: Highly Ambiguous (High Confidence Only)

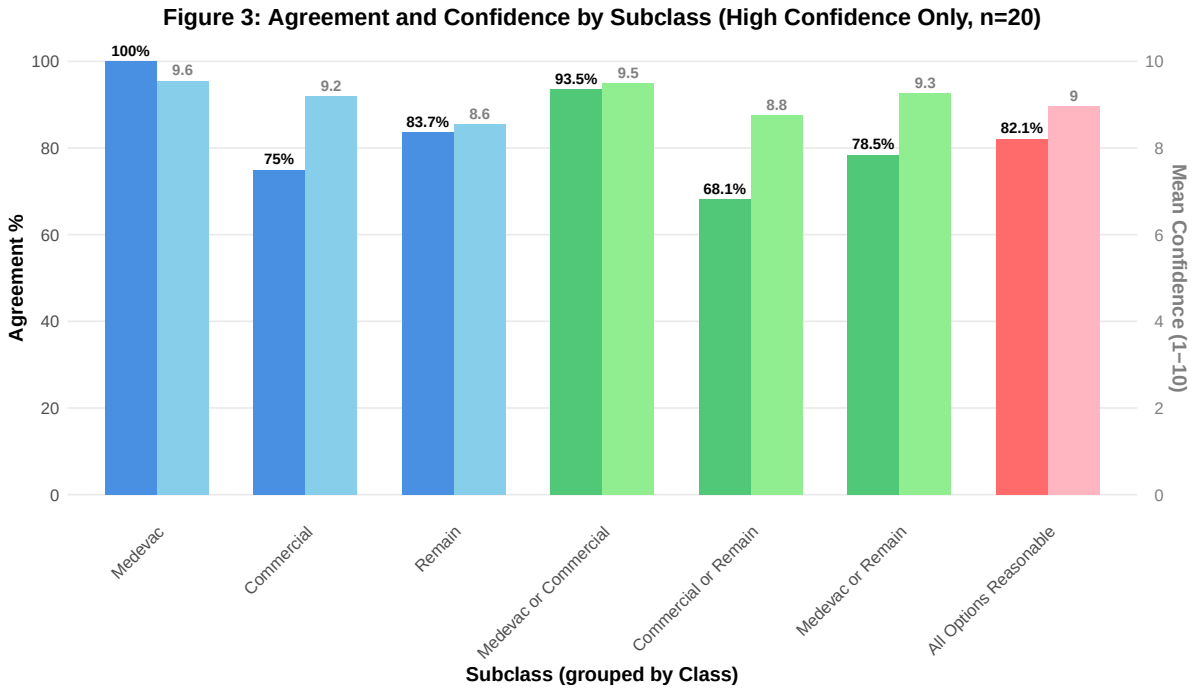
::: .cell ::: .cell-output-display

Table 7: Table 3.3: Class C - Highly Ambiguous (Confidence >= 8)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **All Options Reasonable** | **3** | 28.6% (n=2) | 0.0% (n=0) | 71.4% (n=5) | Remain | **71.4** | 0.195 | Slight |
| **All Options Reasonable** | **12** | 75.0% (n=6) | 25.0% (n=2) | 0.0% (n=0) | Medevac | **75.0** | 0.314 | Fair |
| **All Options Reasonable** | **20** | 100.0% (n=9) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |

::: :::

## 1.7 Figure 3: Agreement and Confidence by Subclass (High Confidence Only, n=20)




Left axis: Agreement %. Right axis: Mean Confidence (1–10). Darker bars = Agreement %, Lighter bars = Confidence. Colors indicate Class. Only includes responses with

## 1.8 Tables 4.1-4.3: Summary Statistics by Individual Vignette (Experienced Respondents Only, n=13)

These tables show detailed statistics for each vignette, including only responses from physicians with more than 5 years since training completion.

### 1.8.1 Table 4.1: One Correct Answer (Experienced Respondents Only)

::: .cell ::: .cell-output-display

Table 8: Table 4.1: Class A - One Correct Answer (>5 Years Experience)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **Medevac** | **1** | 100.0% (n=13) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Medevac** | **4** | 92.3% (n=12) | 0.0% (n=0) | 7.7% (n=1) | Medevac | **92.3** | 0.821 | Very Good |
| **Medevac** | **13** | 100.0% (n=13) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Medevac** | **19** | 100.0% (n=13) | 0.0% (n=0) | 0.0% (n=0) | Medevac | **100.0** | 1.000 | Perfect |
| **Commercial** | **9** | 0.0% (n=0) | 76.9% (n=10) | 23.1% (n=3) | Commercial | **76.9** | 0.404 | Moderate |
| **Commercial** | **15** | 53.8% (n=7) | 38.5% (n=5) | 7.7% (n=1) | Medevac | **53.8** | 0.165 | Slight |
| **Remain** | **8** | 7.7% (n=1) | 23.1% (n=3) | 69.2% (n=9) | Remain | **69.2** | 0.350 | Fair |
| **Remain** | **11** | 0.0% (n=0) | 30.8% (n=4) | 69.2% (n=9) | Remain | **69.2** | 0.196 | Slight |

::: :::

### 1.8.2 Table 4.2: Two Reasonable Options (Experienced Respondents Only)

::: .cell ::: .cell-output-display

Table 9: Table 4.2: Class B - Two Reasonable Options (>5 Years Experience)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation | En |
|---|---|---|---|---|---|---|---|---|---|
| **Medevac or Commercial** | **7** | 92.3% (n=12) | 7.7% (n=1) | 0.0% (n=0) | Medevac | **92.3** | 0.821 | Very Good | |
| **Medevac or Commercial** | **10** | 76.9% (n=10) | 23.1% (n=3) | 0.0% (n=0) | Medevac | **76.9** | 0.404 | Moderate | |
| **Commercial or Remain** | **2** | 23.1% (n=3) | 53.8% (n=7) | 23.1% (n=3) | Commercial | **53.8** | 0.064 | Slight | |
| **Commercial or Remain** | **5** | 0.0% (n=0) | 76.9% (n=10) | 23.1% (n=3) | Commercial | **76.9** | 0.404 | Moderate | |
| **Commercial or Remain** | **6** | 23.1% (n=3) | 53.8% (n=7) | 23.1% (n=3) | Commercial | **53.8** | 0.064 | Slight | |
| **Commercial or Remain** | **14** | 0.0% (n=0) | 30.8% (n=4) | 69.2% (n=9) | Remain | **69.2** | 0.196 | Slight | |
| **Medevac or Remain** | **16** | 69.2% (n=9) | 0.0% (n=0) | 30.8% (n=4) | Medevac | **69.2** | 0.196 | Slight | |
| **Medevac or Remain** | **17** | 46.2% (n=6) | 0.0% (n=0) | 53.8% (n=7) | Remain | **53.8** | -0.071 | None | |
| **Medevac or Remain** | **18** | 23.1% (n=3) | 0.0% (n=0) | 76.9% (n=10) | Remain | **76.9** | 0.404 | Moderate | |

::: :::

### 1.8.3 Table 4.3: Highly Ambiguous (Experienced Respondents Only)

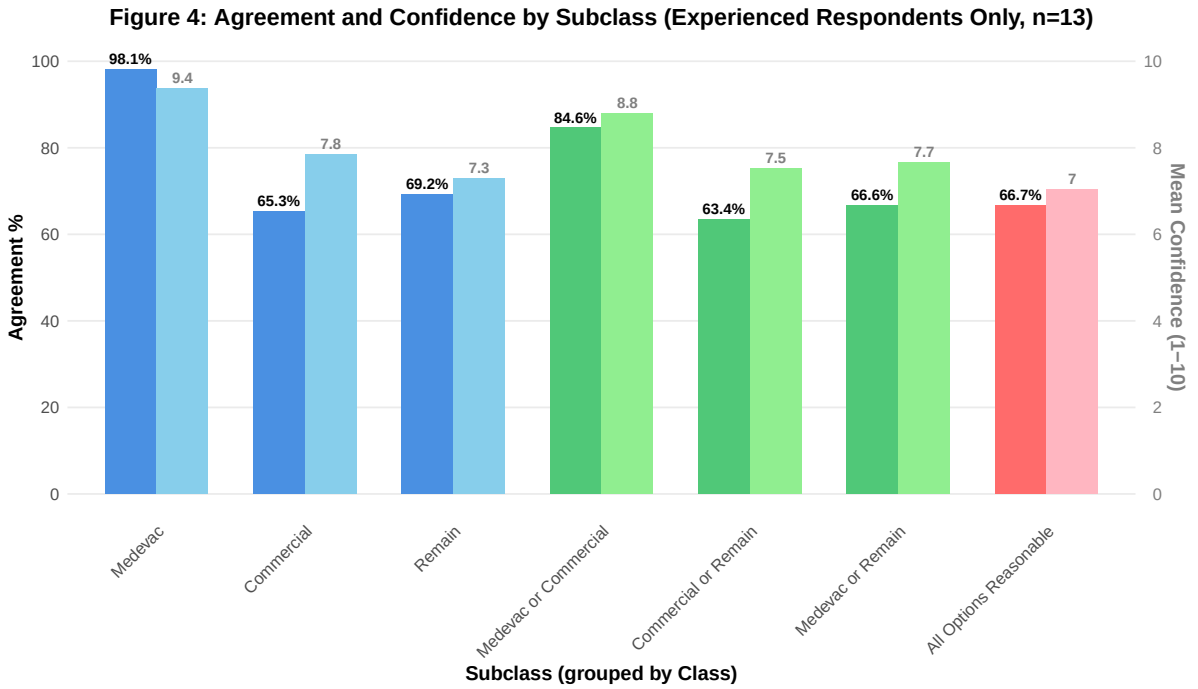
::: .cell ::: .cell-output-display

Table 10: Table 4.3: Class C - Highly Ambiguous (>5 Years Experience)

| Subclass | Question Number | Medevac | Commercial | Remain | Modal Decision | Agreement % | Gwet's AC1 | Interpretation |
|---|---|---|---|---|---|---|---|---|
| **All Options Reasonable** | **3** | 15.4% (n=2) | 0.0% (n=0) | 84.6% (n=11) | Remain | **84.6** | 0.619 | Good |
| **All Options Reasonable** | **12** | 46.2% (n=6) | 38.5% (n=5) | 15.4% (n=2) | Medevac | **46.2** | 0.037 | Slight |
| **All Options Reasonable** | **20** | 69.2% (n=9) | 15.4% (n=2) | 15.4% (n=2) | Medevac | **69.2** | 0.328 | Fair |

::: :::

## 1.9 Figure 4: Agreement and Confidence by Subclass (Experienced Respondents Only, n=13)




Left axis: Agreement %. Right axis: Mean Confidence (1−10). Darker bars = Agreement %, Lighter bars = Confidence. Colors indicate Class. Only includes responses from