Physician Medevac Propensity Analysis

Comprehensive Analysis of Clinical Appropriateness and Practice Variation

Medevac Interrater Study Team

2025-12-05

Table of contents

1 Executive Summary **1**

2 Table 1: Sample Characteristics and Study Design **2**

3 Figure 1: Response Distribution by Physician **4**

4 Table 2: Descriptive Statistics by Vignette **4**

5 Figure 2: Appropriateness Gradient **6**

6 Primary Analysis **6**
6.1 Mixed-Effects Model: Vignette Type + Confidence + Experience 6
6.2 Table 3: Model Results - All Fixed Effects 6
6.3 Table 4: Model Results - Variance Components 7
6.4 Figure 3: Physician-Level Variation 8
6.5 Figure 4: Appropriateness Calibration by Physician 9

7 Key Findings **10**
7.1 Methodological Strengths 10
7.2 Implications 10

1 Executive Summary

This report examines physician variation in medevac decision-making through the lens of **clinical appropriateness**, with explicit modeling of **confidence** and **years of experience**.

We categorize all 20 vignettes by **vignette type** (appropriateness of the medevac option):

- **Medevac Only Correct** (n=4): Medevac is the single best answer
- **Medevac One of Two Correct** (n=5): Medevac is appropriate, along with one other option
- **Ambiguous (All Plausible)** (n=3): All three options are reasonable
- **Medevac One of Two Wrong** (n=4): Medevac is inappropriate; two other options are reasonable
- **Medevac Only Wrong** (n=4): Medevac is the single worst answer

Statistical Model: Binary mixed-effects logistic regression - Outcome: Chose medevac vs. not - Fixed effects: Vignette type + Confidence (centered) + Experience (centered) - Random effects: Physician + Vignette - No interaction terms (for simplicity)

Key Findings:

1. **Strong appropriateness gradient**
    - Odds of choosing medevac increase systematically from "Only Wrong" to "Only Correct"
    - Strong evidence of appropriate clinical reasoning
2. **Confidence and experience effects**
    - Confidence significantly predicts medevac choice
    - Experience shows positive trend
    - Together explain meaningful physician variance
3. **Substantial residual variation persists**
    - Even after accounting for appropriateness, confidence, and experience
    - Suggests additional unmeasured factors influence decision-making

---

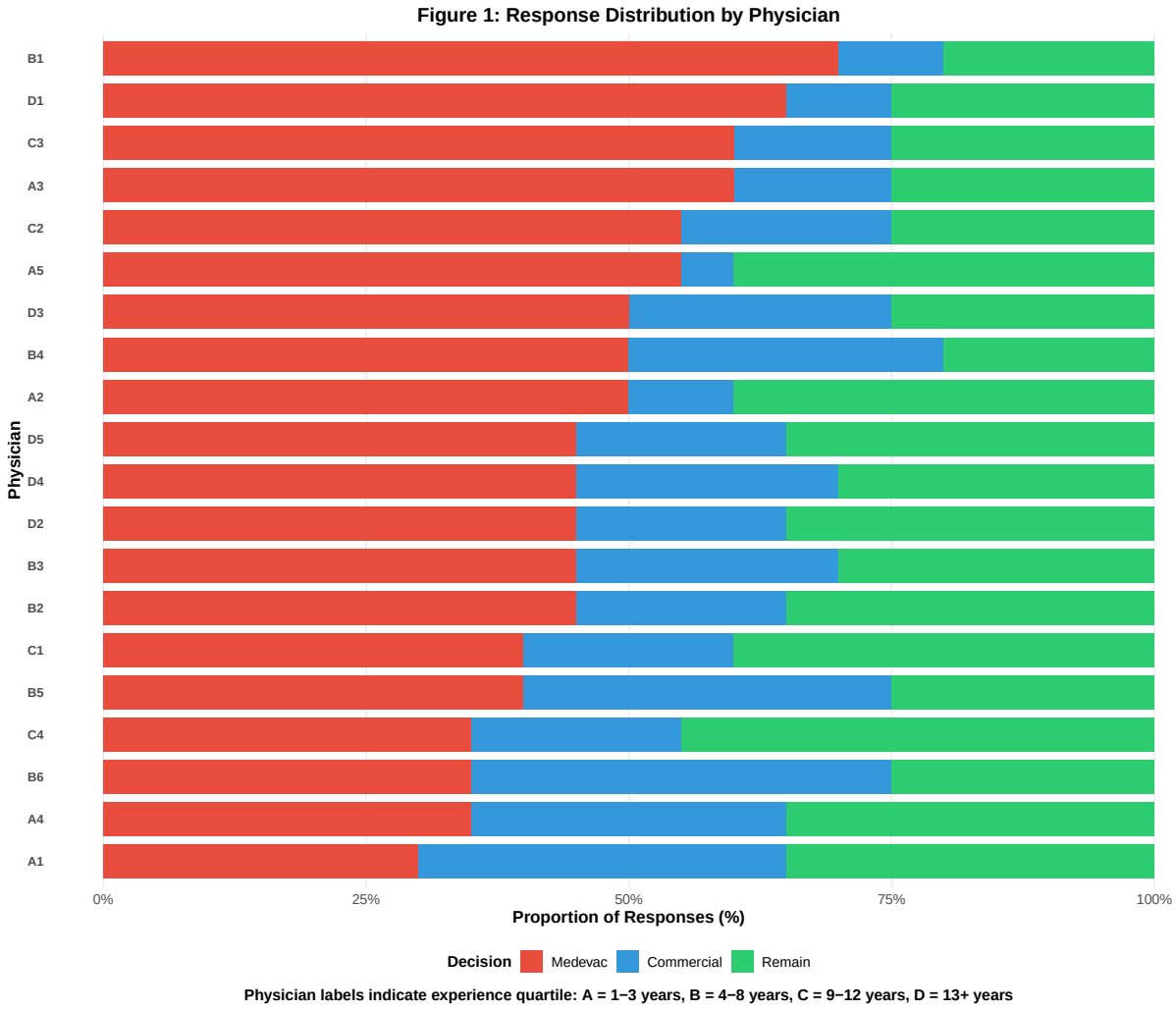
2 Table 1: Sample Characteristics and Study Design

Table 1: Table 1: Sample Characteristics and Study Design

Characteristic	Value
Vignettes	
Total Vignettes	20
Medevac Only Wrong	4
Medevac One of Two Wrong	4
Ambiguous (All Plausible)	3

Medevac One of Two Correct	5
Medevac Only Correct	4
Physicians	
Total Physicians	20
Responses	
Total Responses	400
Years in Practice	
Median [IQR]	8 [4 - 13]
Range	1 - 30 years
Confidence (1-10 scale)	
Median [IQR]	8 [6 - 10]
Range	1 - 10
Medevac Selection	
Overall % Choosing Medevac	47.7%
Range Across Physicians	30.0% - 70.0%

3 Figure 1: Response Distribution by Physician



---

4 Table 2: Descriptive Statistics by Vignette

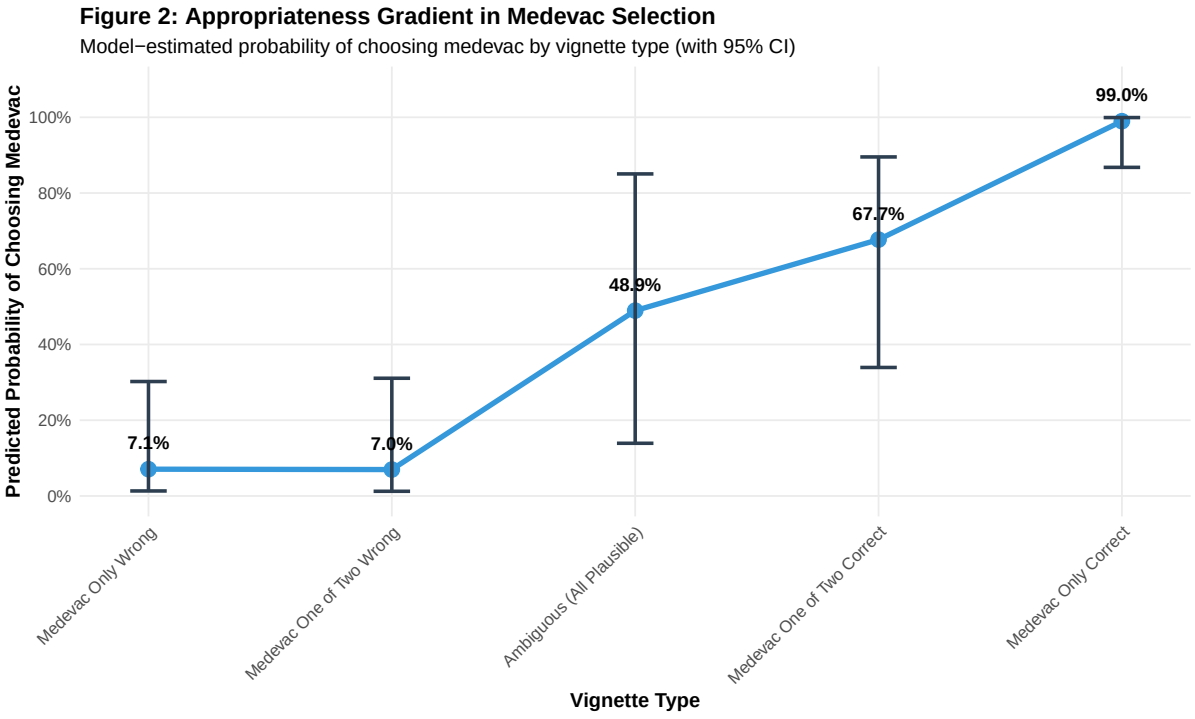
Table 2: Table 2: Descriptive Statistics by Vignette

Vignette Type	Q#	Medevac	Commercial	Remain	Modal	Agree
Medevac Only Wrong						

Commercial or Remain	2	20.0% (n=4)	45.0% (n=9)	35.0% (n=7)	Commercial
Commercial or Remain	5	0.0% (n=0)	75.0% (n=15)	25.0% (n=5)	Commercial
Commercial or Remain	6	25.0% (n=5)	45.0% (n=9)	30.0% (n=6)	Commercial
Commercial or Remain	14	0.0% (n=0)	20.0% (n=4)	80.0% (n=16)	Remain
Medevac One of Two Wrong					
Remain Only	8	5.0% (n=1)	15.0% (n=3)	80.0% (n=16)	Remain
Commercial Only	9	0.0% (n=0)	75.0% (n=15)	25.0% (n=5)	Commercial
Remain Only	11	0.0% (n=0)	25.0% (n=5)	75.0% (n=15)	Remain
Commercial Only	15	65.0% (n=13)	30.0% (n=6)	5.0% (n=1)	Medevac
Ambiguous (All Plausible)					
All Options Reasonable	3	20.0% (n=4)	0.0% (n=0)	80.0% (n=16)	Remain
All Options Reasonable	12	45.0% (n=9)	40.0% (n=8)	15.0% (n=3)	Medevac
All Options Reasonable	20	65.0% (n=13)	25.0% (n=5)	10.0% (n=2)	Medevac
Medevac One of Two Correct					
Medevac or Commercial	7	95.0% (n=19)	5.0% (n=1)	0.0% (n=0)	Medevac
Medevac or Commercial	10	70.0% (n=14)	30.0% (n=6)	0.0% (n=0)	Medevac
Medevac or Remain	16	75.0% (n=15)	0.0% (n=0)	25.0% (n=5)	Medevac
Medevac or Remain	17	50.0% (n=10)	0.0% (n=0)	50.0% (n=10)	Medevac
Medevac or Remain	18	25.0% (n=5)	0.0% (n=0)	75.0% (n=15)	Remain
Medevac Only Correct					
Medevac Only	1	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac
Medevac Only	4	95.0% (n=19)	0.0% (n=0)	5.0% (n=1)	Medevac
Medevac Only	13	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac
Medevac Only	19	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac

Table 2 Notes: - **Gwet's AC1**: Chance-corrected agreement statistic chosen over Cohen's Kappa due to superior performance with class imbalance. AC1 is less affected by prevalence and marginal probability asymmetry, making it more appropriate when decision options have unequal base rates across vignettes - **Median Conf. [IQR]**: Median confidence rating with interquartile range (1-10 scale) - Vignettes grouped by vignette type (matching Table 1 classification)

5 Figure 2: Appropriateness Gradient



---

6 Primary Analysis

6.1 Mixed-Effects Model: Vignette Type + Confidence + Experience

Note: Experience and confidence are moderately correlated (r = 0.11), but both are included as mean-centered continuous predictors to capture their independent effects.

6.2 Table 3: Model Results - All Fixed Effects

Table 3: Table 3: Model Fixed Effects (Vignette Type + Confidence + Experience)

Predictor	Odds Ratio	95% CI	P-value
Vignette Type			
Medevac Only Wrong (reference)	1.00	ref	ref

Medevac One of Two Wrong	0.99	(0.08 - 11.76)	0.991
Ambiguous (All Plausible)	12.60	(1.06 - 149.28)	0.044*
Medevac One of Two Correct	27.56	(3.00 - 253.02)	0.003*
Medevac Only Correct	1260.35	(52.19 - 30435.95)	<0.001*
Decision Characteristics			
Confidence (per 1-point increase)	1.39	(1.19 - 1.63)	<0.001*
Years of Experience (per 1-year increase)	0.98	(0.94 - 1.03)	0.527

* indicates statistically significant ($p < 0.05$)

Interpretation:

- **Confidence**: OR = 1.39 per 1-point increase (p = <0.001)
- **Experience**: OR = 0.98 per 1-year increase (p = 0.527)

6.3 Table 4: Model Results - Variance Components

Table 4: Table 4: Model Variance Components

Component	Estimate
Random Effect Variances	
Physician	0.241
Vignette	2.158
Intraclass Correlations	
ICC (Physician)	4.2%
ICC (Vignette)	37.9%
Median Odds Ratio	
MOR (Physician)	1.60

Key Findings:

- **ICC (Physician)** = 4.2%: Proportion of total variance attributable to physician differences

- **MOR (Median Odds Ratio)** = 1.60: The median increase in odds of choosing medevac when comparing a randomly selected physician with higher propensity to one with lower propensity. An MOR of 1.60 means that if two identical patients saw two randomly selected physicians, the odds of receiving medevac would differ by a factor of 1.60 on average (50% of the time)
- Even after accounting for vignette appropriateness, confidence, and experience, meaningful physician-level differences persist

---

6.4 Figure 3: Physician-Level Variation

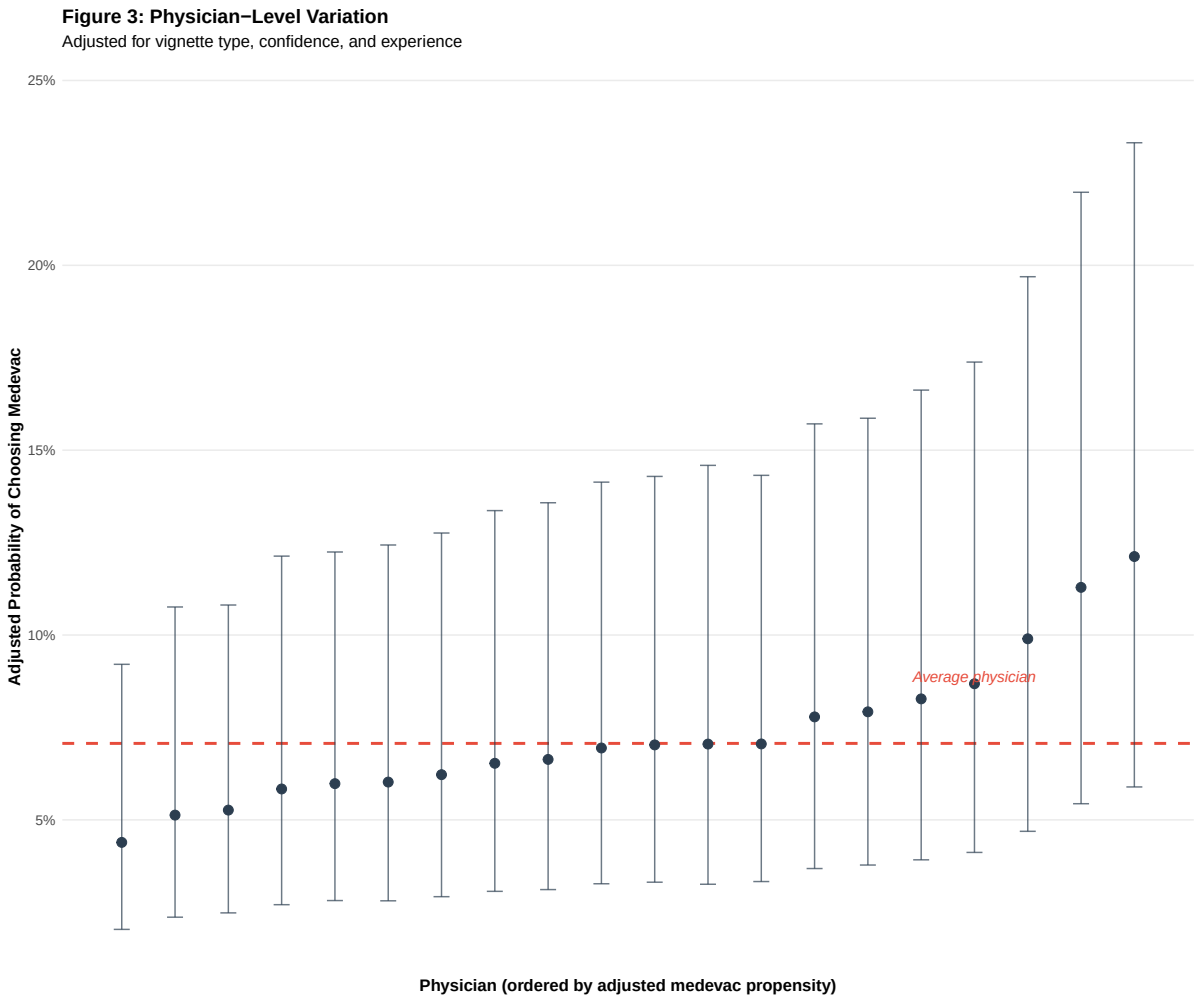


Figure 3 Interpretation: After accounting for vignette appropriateness, confidence, and experience, residual physician variation ranges from 4.4% to 12.1%. The MOR of 1.60 indicates factors beyond measured covariates contribute to practice variation.

6.5 Figure 4: Appropriateness Calibration by Physician

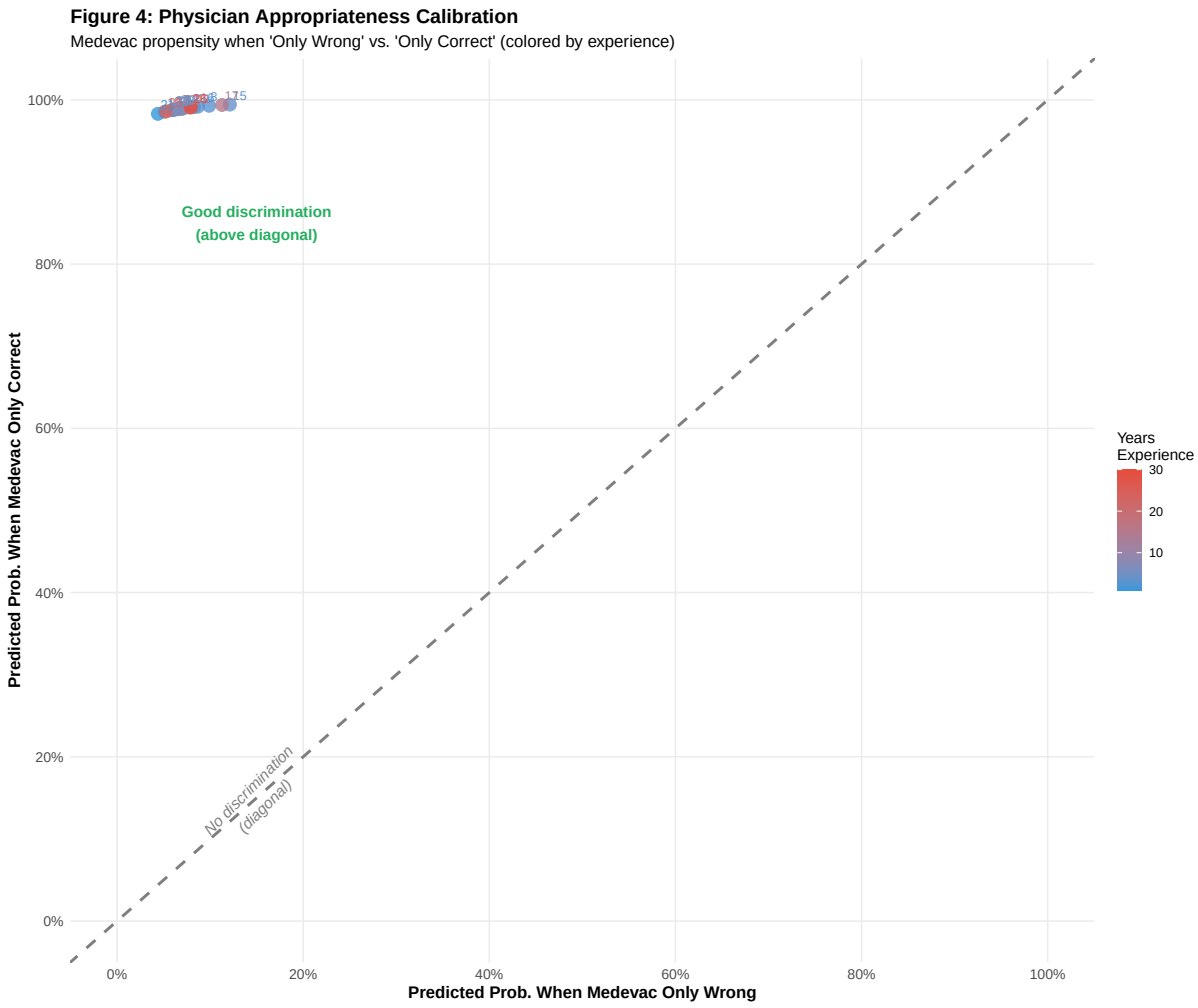


Figure 4 Interpretation: Most physicians are above the diagonal (appropriate discrimination). Distance from diagonal reflects calibration strength. Color shows years of experience - examining whether more experienced physicians show stronger calibration.

---

7 Key Findings

1. Strong appropriateness gradient exists (Table 3, Figure 2) - Odds of choosing medevac increase systematically with appropriateness - Strong evidence of appropriate clinical reasoning

2. Confidence and experience effects - **Confidence**: OR = 1.39 per 1-point increase - **Experience**: OR = 0.98 per year

3. Substantial residual variation persists - ICC (Physician) = 4.2% after accounting for appropriateness, confidence, and experience - MOR = 1.60 - Meaningful physician differences persist beyond measured covariates

The calibration plot (Figure 4) provides an actionable tool for feedback: physicians can see whether they appropriately discriminate between different appropriateness scenarios.

7.1 Methodological Strengths

- **Mixed-effects framework** accounts for clustering (physicians + vignettes)
- **Explicit appropriateness modeling** moves beyond raw agreement
- **Variance decomposition** quantifies sources of variation
- **Multiple metrics** (ICC, MOR) for clinical interpretability

7.2 Implications

This analysis suggests that while physicians generally respond appropriately to clinical scenarios, substantial **calibrated practice variation** remains. Understanding this variation is key to:

1. **Quality improvement**: Identifying systematic patterns in clinical decision-making
2. **Education**: Targeting feedback to appropriateness calibration
3. **Policy**: Accounting for legitimate variation vs. unwarranted variation