Physician Medevac Propensity Analysis

Comprehensive Analysis of Clinical Appropriateness and Practice Variation

Medevac Interrater Study Team

2025-12-18

Table of contents

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):

- **Always Medevac** (n=4): Medevac is clearly the best answer
- **Possible Medevac** (n=9): Medevac may or may not be appropriate depending on clinical judgment
- **Special Considerations** (n=3): Medevac or Remain are both reasonable; conflict between physiology and logistics
- **Never Medevac** (n=4): Medevac is clearly not appropriate

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 "Never Medevac" to "Always Medevac"
   - 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
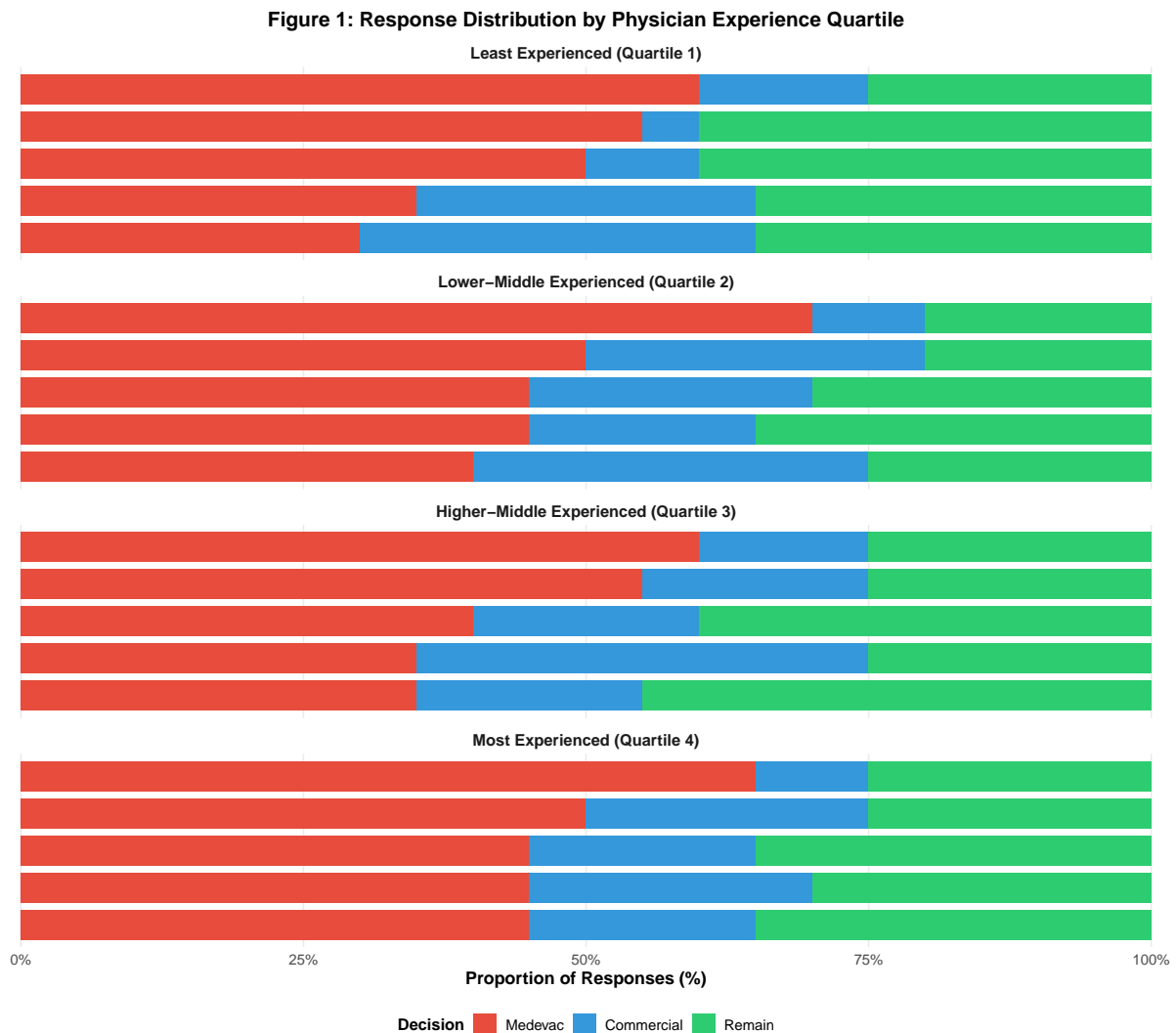   - Suggests additional unmeasured factors influence decision-making

---

2 Table 1: Sample Characteristics and Study Design

Table 1: Table 1: Sample Characteristics and Study Design

Characteristic	Value
Physicians	
Total Physicians	20
Years in Practice	
Median [IQR]	8 [4 - 13]

Range	1 - 30 years

Practice Location

Regional Hub	7 (35%)
Referral Hub (ANMC)	8 (40%)
Mixed	5 (25%)

Vignettes

Total Vignettes	20
Never Medevac	4
Possible Medevac	9
Special Considerations	3
Always Medevac	4

Responses

Total Responses	400
Appropriate Response to Always Medevac	98.8% (n=79)
Appropriate Response to Never Medevac	88.8% (n=71)

Inter-Rater Agreement by Vignette

Median [IQR]	75.0% [65.0% - 83.8%]
Range	45.0% - 100.0%

Confidence (1-10 scale)

Median [IQR]	8 [6 - 10]
Range	1 - 10

3 Figure 1: Response Distribution by Physician

Figure 1: Response Distribution by Physician Experience Quartile

4 Table 2: Descriptive Statistics by Vignette

Table 2: Table 2: Descriptive Statistics by Vignette

Q#	Medevac	Commercial	Remain	Modal	Agreement %	Median Conf. [IQR]

Never Medevac

14	0.0% (n=0)	20.0% (n=4)	80.0% (n=16)	Remain	80	8 [6-8]
5	0.0% (n=0)	75.0% (n=15)	25.0% (n=5)	Commercial	75	8 [7-10]
2	20.0% (n=4)	45.0% (n=9)	35.0% (n=7)	Commercial	45	8 [6-8]
6	25.0% (n=5)	45.0% (n=9)	30.0% (n=6)	Commercial	45	7 [6-8]

Possible Medevac

7	95.0% (n=19)	5.0% (n=1)	0.0% (n=0)	Medevac	95	10 [10-10]
3	20.0% (n=4)	0.0% (n=0)	80.0% (n=16)	Remain	80	7 [5-8]
8	5.0% (n=1)	15.0% (n=3)	80.0% (n=16)	Remain	80	8 [6-8]
9	0.0% (n=0)	75.0% (n=15)	25.0% (n=5)	Commercial	75	8 [6-9]
11	0.0% (n=0)	25.0% (n=5)	75.0% (n=15)	Remain	75	8 [6-8]
10	70.0% (n=14)	30.0% (n=6)	0.0% (n=0)	Medevac	70	8 [7-9]
15	65.0% (n=13)	30.0% (n=6)	5.0% (n=1)	Medevac	65	8 [6-10]
20	65.0% (n=13)	25.0% (n=5)	10.0% (n=2)	Medevac	65	7 [5-9]
12	45.0% (n=9)	40.0% (n=8)	15.0% (n=3)	Medevac	45	6 [5-8]

Special Considerations

16	75.0% (n=15)	0.0% (n=0)	25.0% (n=5)	Medevac	75	9 [5-10]
18	25.0% (n=5)	0.0% (n=0)	75.0% (n=15)	Remain	75	9 [8-10]
17	50.0% (n=10)	0.0% (n=0)	50.0% (n=10)	Medevac	50	7 [5-8]

Always Medevac

1	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac	100	10 [10-10]
13	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac	100	10 [9-10]
19	100.0% (n=20)	0.0% (n=0)	0.0% (n=0)	Medevac	100	10 [8-10]
4	95.0% (n=19)	0.0% (n=0)	5.0% (n=1)	Medevac	95	10 [9-10]

Table 2 Notes: - **Median Conf. [IQR]**: Median confidence rating with interquartile range (1-10 scale) - Vignettes grouped by vignette type (matching Table 1 classification)
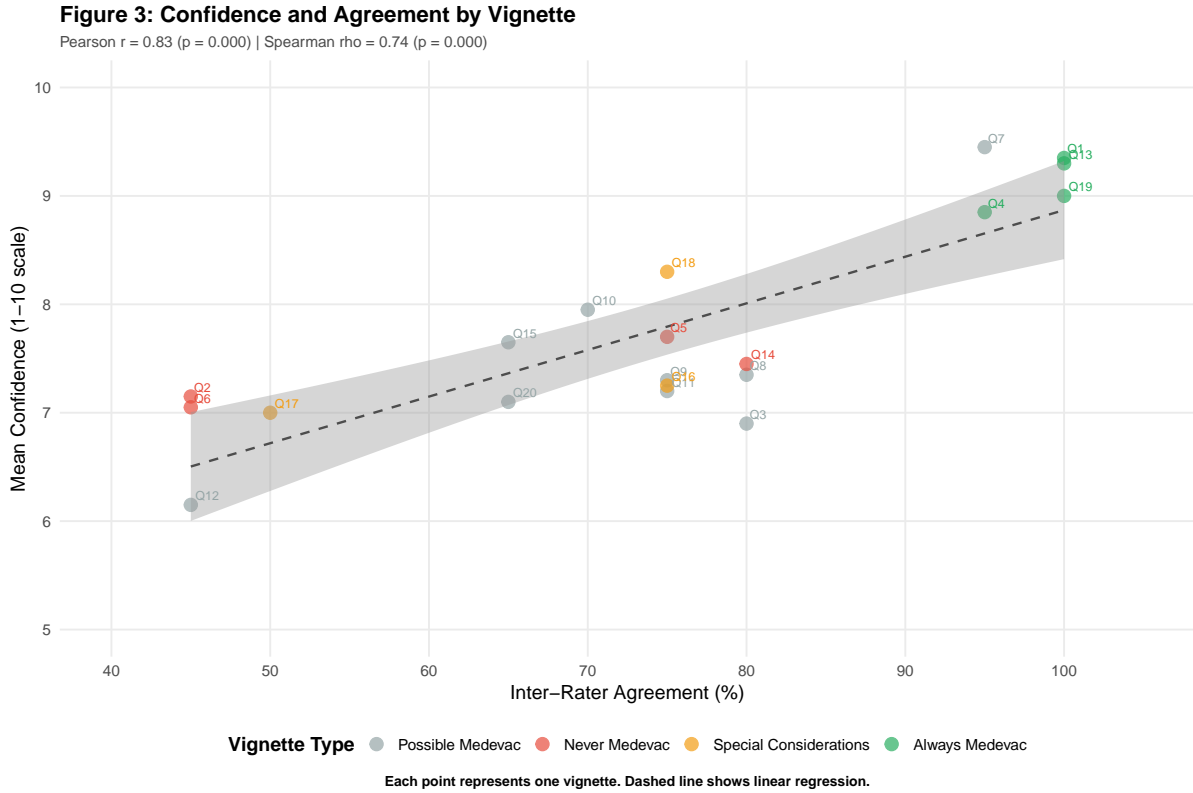
4.1 Figure 2: Decision Distribution by Vignette Type

Figure 2: Decision Distribution by Vignette Type

Key Patterns: - **% Medevac** (green): Strong gradient from Never (0%) to Always (100%) - **% Commercial** (purple): Higher for intermediate scenarios - **% Remain** (red): Higher when medevac is not indicated - Error bars show 95% confidence intervals

4.2 Figure 3: Confidence and Agreement by Vignette

Figure 3: Confidence and Agreement by Vignette

Pearson r = 0.83 (p = 0.000) | Spearman rho = 0.74 (p = 0.000)

Vignette Type ● Possible Medevac ● Never Medevac ● Special Considerations ● Always Medevac

Each point represents one vignette. Dashed line shows linear regression.

Key Findings:

- **Correlation**: Significant positive correlation (r = 0.83, p = 0.000)
- **Interpretation**: Physicians are more confident on vignettes with higher agreement, suggesting appropriate calibration.

---

5 Primary Analysis

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

Note: Confidence is measured as within-person deviation (deviation from each physician's average confidence). Between-person confidence differences are intentionally excluded from fixed effects, allowing the physician random effect to capture unmeasured risk tolerance and other stable physician characteristics. This approach provides the best model fit (lowest AIC) while preserving interpretable physician-level variation.

Why Generalized Linear Mixed Models (GLMM)?

We use a mixed-effects logistic regression model to account for the nested structure of our data: multiple responses from each physician across multiple vignettes. Standard logistic regression would violate the assumption of independent observations, as responses from the same physician and to the same vignette are correlated. The GLMM approach includes random intercepts for both physicians and vignettes, along with fixed effects for vignette type, case-specific confidence, experience, and practice location. This allows us to: (1) obtain valid standard errors and p-values that account for clustering, (2) separate within-physician effects (being more confident than usual on a specific case) from stable between-physician differences (risk tolerance), and (3) quantify how much variation exists at both the physician and vignette level beyond what's explained by measured predictors.

5.2 Table 3: Model Results - All Fixed Effects

Table 3: Table 3: Likelihood to Medevac

Predictor	Odds Ratio	95% CI
Vignette Type		
Possible/Ambiguous (reference)	1.00	ref
Never Medevac	0.07	(0.00 - 1.05)
Always Medevac	455.53	(13.40 - 15481.30)
Decision Characteristics		
Confidence (per 1-point increase compared to own average)	1.36	(1.13 - 1.64)
Physician Characteristics		
Years of Experience (per 1-year increase)	0.99	(0.94 - 1.04)
Rural Experience (vs Referral Hub only)	0.37	(0.16 - 0.85)

* indicates statistically significant ($p < 0.05$) Model Quality Statistics: N = 400 observations from 20 physici

Model Quality Interpretation:

The model quality statistics indicate excellent fit and validity:

- **AIC = 311.4** and **BIC = 343.4**: These information criteria balance model fit with complexity. Lower values indicate better models; these values suggest strong predictive performance while avoiding overfitting.

- **Marginal R² = 0.570**: Fixed effects (vignette type, confidence, experience, rural practice) explain 57.0% of variance in medevac decisions. This quantifies how much our measured predictors account for variation in physician choices.

- **Conditional R² = 0.816**: The full model (fixed + random effects) explains 81.6% of variance. The difference between conditional and marginal R² (24.6%) represents variation due to unmeasured physician and vignette characteristics captured by random effects.

- **Sample**: The model leverages all 400 decisions from 20 physicians across 20 vignettes, providing robust estimates while properly accounting for clustering.

Interpretation:

Note: Model includes all 20 vignettes collapsed into 3 categories for parsimony: Possible/Ambiguous (n=12 vignettes: Possible Medevac + Special Considerations, used as reference), Never Medevac (n=4), and Always Medevac (n=4). Total: 400 responses from 20 physicians. The Possible/Ambiguous reference category represents scenarios where medevac is defensible but not definitively indicated.

Significant Predictors (p < 0.05):

- **Vignette Type - Always Medevac**: OR = 455.53 (p = <0.001) - Compared to Possible/Ambiguous scenarios, physicians are 456× more likely to choose medevac for "Always Medevac" scenarios, demonstrating appropriate clinical reasoning
- **Case-Specific Confidence**: OR = 1.36 per 1-point increase (p = <0.001) - Being more confident than usual on a specific case significantly increases medevac likelihood
- **Rural Experience**: OR = 0.37 (p = 0.019) - Physicians with rural/regional practice experience are significantly less likely to choose medevac compared to those with referral hub experience only (63% reduction), suggesting rural practice experience is associated with more conservative disposition decisions

Non-Significant Predictors (p   0.05):

- **Vignette Type - Never Medevac**: OR = 0.07 (p = 0.055) - Compared to Possible/Ambiguous scenarios, physicians are 93% less likely to choose medevac for "Never Medevac" scenarios. The effect is marginally non-significant (p = 0.055), likely due to the very low baseline rate of inappropriate medevac choices
- **Years of Experience**: OR = 0.99 per year (p = 0.663) - Years in practice not significantly associated with medevac utilization after accounting for confidence, vignette type, and practice location

---

5.3 Table 4: Model Results - Variance Components

Table 4: Table 4: Model Variance Components

Component	Estimate
Random Effect Variances	
Physician (log-odds scale)	0.288
Vignette (log-odds scale)	4.101
Intraclass Correlations	
ICC (Physician)	3.8%
ICC (Vignette)	53.4%
Median Odds Ratios	
MOR (Physician)	1.67
MOR (Vignette)	6.90

Physician random effect captures unmeasured risk tolerance and other stable physician characteristics after a
MOR reported as point estimates; 95% confidence intervals not computed due to computational complexity o

Variance Components Summary:

- **Physician ICC**: 3.8% - Modest between-physician variation
- **Vignette ICC**: 53.4% - Substantial between-scenario variation
- **Physician MOR**: 1.67 - Small individual differences
- **Vignette MOR**: 6.90 - Large scenario-to-scenario variation

---

5.4 Model Evaluation: Predictive Performance

5.4.1 Reference Only: Model Predictive Performance

Table 5: Reference Only: Model Predictive Performance

Metric	Value	Interpretation
1. Discrimination		
ROC AUC (c-statistic, 95% CI)	0.837 (0.798 - 0.875)	Quantifies how well predicted risk ranks cases vs
2. Calibration		
Calibration-in-the-large (intercept)	0.364	Systematic over/under-prediction (0 = perfect)

Calibration Slope	0.737	Whether predictions are too extreme or too shru

3. Overall Predictive Accuracy

Brier Score	0.1627	Mean squared error of predicted probabilities (lo

ROC AUC: Values >0.9 = excellent, 0.8-0.9 = good, 0.7-0.8 = acceptable, <0.7 = poor
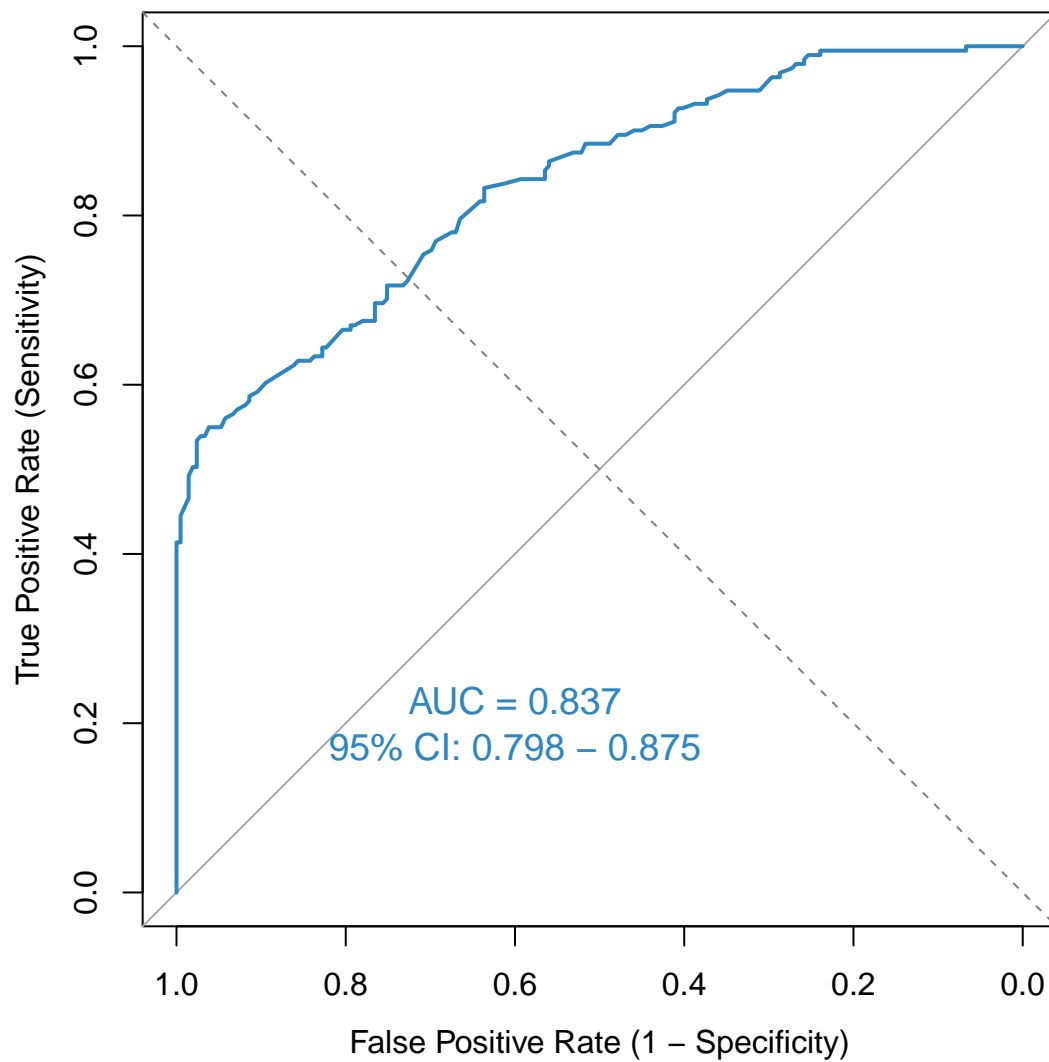
Calibration intercept: Close to 0 indicates no systematic bias

Calibration slope: Close to 1 indicates well-calibrated spread of predictions

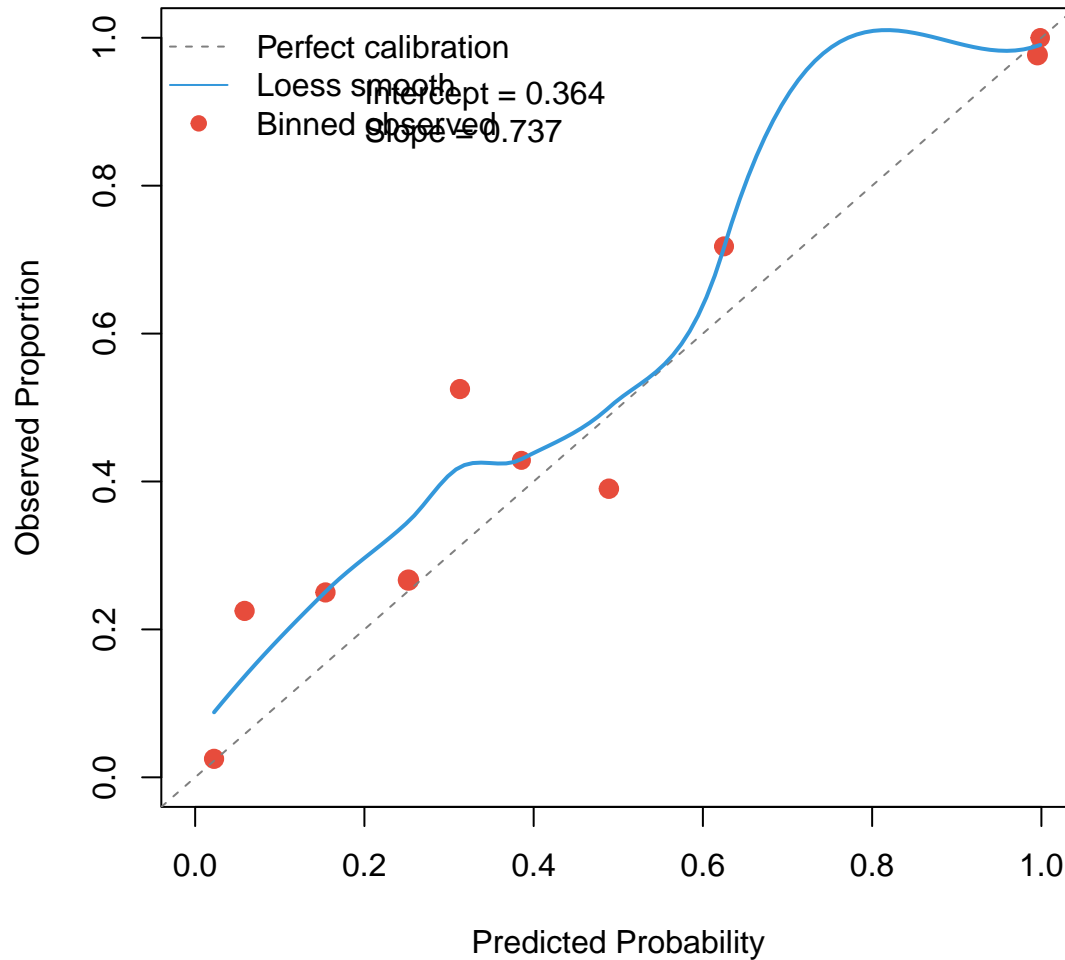Brier score: Range 0-1, with 0 = perfect predictions

Supplemental Figure 1: ROC Curve – Model Discrimination

AUC = 0.837
95% CI: 0.798 – 0.875

5.4.3 Supplemental Figure 2: Calibration Plot

Supplemental Figure 2: Calibration Plot

Performance Summary:

- **Discrimination (ROC AUC = 0.837, 95% CI: 0.798-0.875)**: The model demonstrates good ability to distinguish between medevac and non-medevac decisions.

- **Calibration (Intercept = 0.364, Slope = 0.737)**: Moderate calibration - some deviation between predicted and observed probabilities.

- **Overall Predictive Accuracy (Brier Score = 0.1627)**: Good probability accuracy.

6 Key Findings from GLMM

1. Clinical scenario dominates medevac decisions - Physicians are $456\times$ more likely to choose medevac for "Always" vs "Possible/Ambiguous" scenarios (OR = 456, $p < 0.001$) - 93% reduction in medevac likelihood for "Never" scenarios (OR = 0.07, p = 0.055, marginally NS) - Vignettes account for 53% of decision variance - clinical appropriateness is the primary driver

2. Rural practice experience strongly predicts conservative utilization - Physicians with rural/regional experience are 63% less likely to choose medevac (OR = 0.37, p = 0.019) - Effect persists after controlling for case-specific confidence, years of experience, and vignette type - Suggests practice setting shapes disposition decision-making patterns

3. Case-specific confidence influences decisions independently - Being more confident than usual on a specific case increases medevac likelihood 36% per confidence point (OR = 1.36, p = 0.001) - Represents within-person variation - not general confidence level - Confidence acts as a mediator between clinical assessment and disposition choice

4. Modest physician variation remains after accounting for measured factors - Physician ICC = 3.8%, MOR = 1.67 - Small but present individual differences suggest unmeasured factors (e.g., risk tolerance) - Years of experience not independently predictive (OR = 0.99, p = 0.66)

6.1 Methodological Strengths

- **Mixed-effects framework** accounts for clustering at physician and vignette levels
- **Within-person confidence measurement** separates case-specific confidence from stable physician characteristics
- **Explicit appropriateness modeling** moves beyond raw agreement metrics
- **Variance decomposition** quantifies sources of variation (physician vs vignette)
- **Model parsimony** excludes between-person confidence to avoid collinearity with physician random effect

6.2 Implications

This analysis demonstrates that physicians respond appropriately to clinical scenarios with modest individual variation. Medevac decision-making is driven by:

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