OncoPredict
Cancer Trial Completion Prediction

10,000
Total Trials
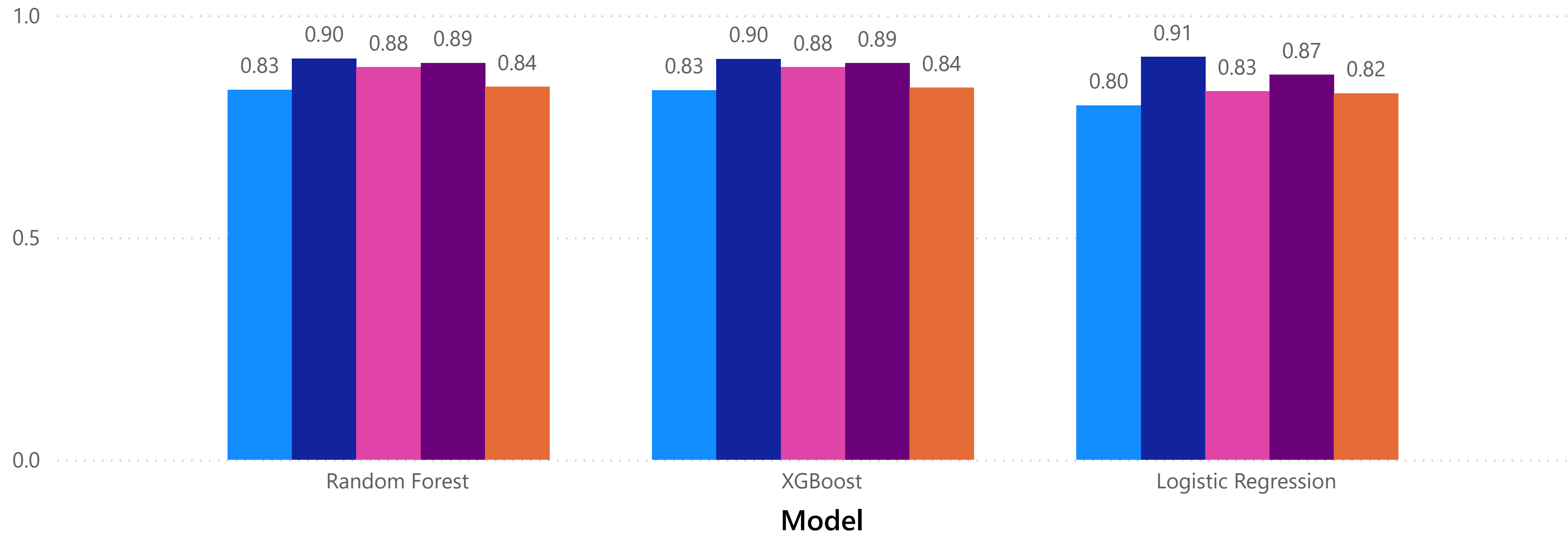
79.1%
Overall Completion Rate

2,000
Test Set Trials

Model Performance Comparison

- Accuracy
- Precision
- Recall
- F1_Score
- ROC_AUC

Random Forest
- 0.83
- 0.90
- 0.88
- 0.89
- 0.84

XGBoost
- 0.83
- 0.90
- 0.88
- 0.89
- 0.84

### Logistic Regression
- 0.80
- 0.91
- 0.83
- 0.87
- 0.82

Model

XGBoost achieves 83.15% accuracy, outperforming baseline by 3.4%. Model correctly identifies 83% of at-risk trials, enabling early intervention

Model Performance Deep Dive

Model	Accuracy	F1_Score	Precision	Recall	ROC_AUC
Random Forest	0.83	0.89	0.90	0.88	0.84
XGBoost	0.83	0.89	0.90	0.88	0.84
Logistic Regression	0.80	0.87	0.91	0.83	0.82

○ **Logistic Regression**
○ **Random Forest**
○ **XGBoost**

Prediction Breakdown

● False_Negative ● False_Positive ● True_Negative ● True_Positive
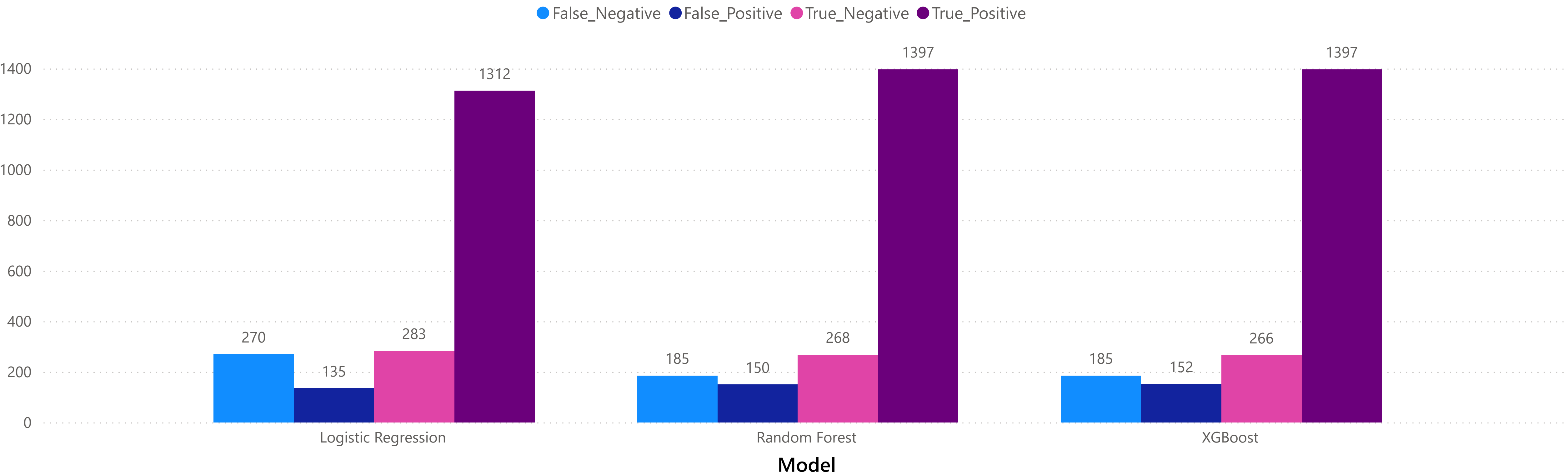
Clinical Trials Dataset Overview

228
Avg Enrollment per Trial

1.3K days
Median Trial Duration

2
Most Common Phase

31.4%
Of Trials Were Large(>500 People)

Distribution of Trials By Current Phase

- 38.55% — ● Phase Unkn...
- 26.8% — ● Phase 2
- 23.15... — ● Phase 1
- 7.7% — ● Phase 3
- 2.6% — ● Phase 4
- ● Other
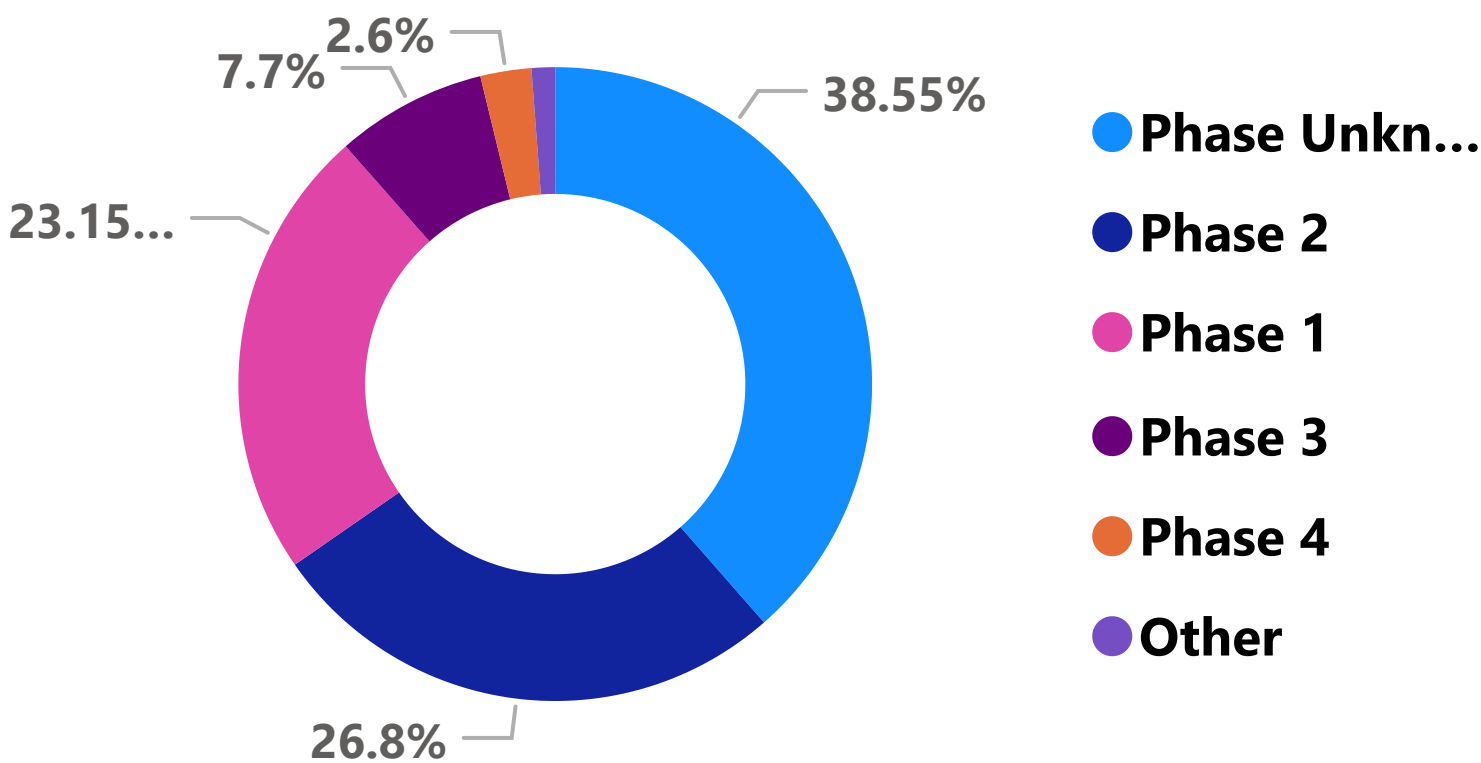
Distribution Of Trials By Number Of People Participating

Participation Size	Trials
0-100	6.7K
100-300	1.8K
300-500	0.5K
500-1,000	0.4K
1,000-2,000	0.2K
2,000+	0.2K
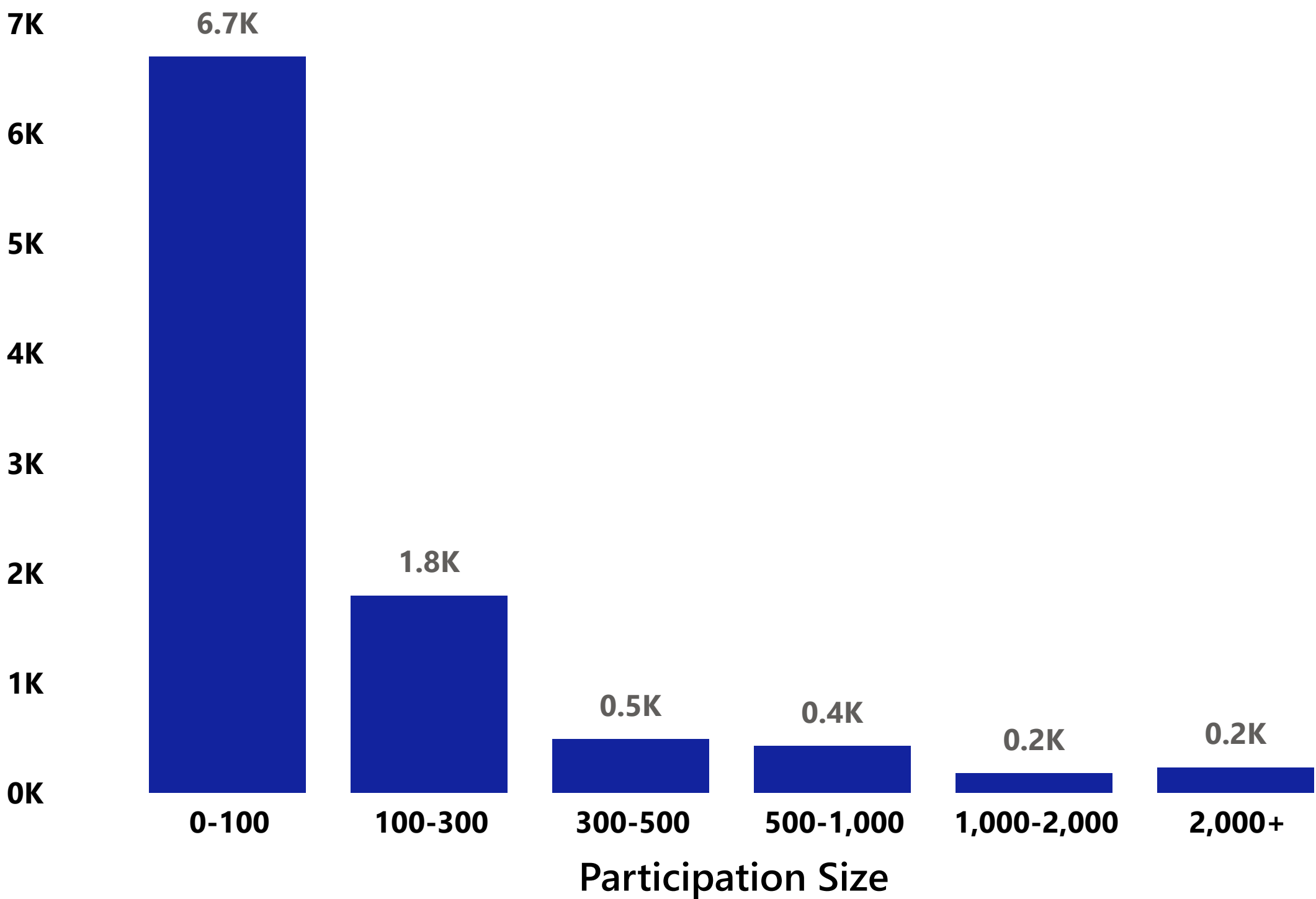
67% of trials enroll <100 participants. Small trials show 2.3x higher termination rates, making enrollment the strongest predictor of trial success.

Focus recruitment efforts early to reach 500+ participants.