Case 3: Vance County EMS

Steph Reinke, Olivia Fu, Srika Gopal

October 20, 2025

1 Background and Motivation

Emergency medical service (EMS) response times can drastically impact patient outcomes. Vance County, North Carolina is a growing county that is interested in evaluating their current EMS ambulance placements. The county has three major regions, the North, Central, and the South. The population is mostly concentrated in the city of Henderson, located in the Central region. Currently, there are two EMS stations – one in the South with one ambulance, and one in the Central region with three ambulances. The North region does not have a dedicated station, raising concerns about delayed response times for residents in that area.

To address this, county officials are considering relocating one of the existing ambulances to the North. Two potential station locations are being evaluated: Near North and Far North. By either reassigning an ambulance from the Central or South station, four placement scenarios were proposed:

- Scenario 0 (Current Setup): 1 in South, 3 in Central
- Scenario 1: 0 in South, 3 in Central, 1 in Near North
- Scenario 2: 0 in South, 3 in Central, 1 in Far North
- Scenario 3: 1 in South, 2 in Central, 1 in Near North
- Scenario 4: 1 in South, 2 in Central, 1 in Far North

Research question: Where should the ambulances be stationed to best serve Vance county?

2 Data and Exploratory Analysis

To address this research question, we examined a dataset with real EMS call records from Vance County. Select identifying information, such as dates and addresses, was modified or removed to comply with HIPAA regulations; however, for the purposes of our analysis those changes are not relevant. The dataset includes key operational details for each call, such as location, emergency level, dispatch time, arrival time, etc. For each call, estimated travel times were obtained using the Google Maps API for all existing stations and the two proposed Northern station locations. Google Maps provided four types of travel time estimates for each call–station pair: best-guess, optimistic, pessimistic, and unadjusted. These estimates enabled us to compare response times across the four proposed ambulance placement scenarios.

We started by examining the spatial distribution of EMS calls across Vance County. We mapped call locations across Vance County and colored them by response time (Figure 1). Calls are densely concentrated in the Central region around the city of Henderson and become increasingly sparse with distance, particularly in the far North and South. Response times are shortest in the Central region and tend to be longer in the North and South. The vast majority of calls have response times under 10 minutes, with nearly all under 20 minutes.

Because four types of travel time estimates were available from the Google Maps API, we conducted an accuracy assessment to determine which estimate most closely aligned with actual recorded travel times. As shown in Table 1, although the optimistic estimate had the lowest mean absolute error (MAE) and root mean squared error (RMSE), it exhibited a positive bias, suggesting a tendency to underestimate travel times. We therefore chose to use the best-guess estimate (BG), which had the lowest bias. This can be seen visually in Figure 2, where the distribution of BG residuals is most centered around zero compared to the other estimates.

To simulate EMS responses, we established a dispatch rule in which each call is assigned to the nearest station with an available ambulance. More specifically, we used an availability matrix to identify overlapping calls. If the dispatch time of a call occurred before a previously dispatched ambulance became available, it was classified as a conflict. We then adjusted each station's available ambulances by subtracting those already in use. Among stations with at least one available vehicle, the call was dispatched from the station with the shortest estimated travel time.

With this dispatch rule in place, we generated estimated travel times for all calls under each proposed scenario. Figure 3 summarizes the distribution of travel times across scenarios. Scenario 3 shows a greater proportion of calls reached within 0–6 minutes and fewer calls exceeding 15 minutes relative to the current setup. This preliminary evidence suggests that Scenario 3 may be the optimal choice in travel time performance.

3 Model Rationale, Implementation, and Evaluation

3.1 Rationale & Selection

We selected a two-stage modeling approach: a logistic regression to model whether estimated travel time changes from the baseline scenario, followed by a linear mixed-effects model to quantify how travel time changes when it does differ. This approach is justified both by the data structure and the nature of our dispatch system. In our simulation, many calls have identical estimated travel times across all four proposed scenarios and the baseline (S0). For example, calls located in the central often have the same optimal dispatch station under all scenarios, resulting in no variation. Including these repeated values directly in a continuous-response model would violate normality and constant variance assumptions because of the large number of zero differences.

To address this, our first model is a logistic regression predicting whether estimated travel time differs from baseline (change vs. no change). This allows us to estimate the probability that each scenario leads to a change in dispatch outcome.

Conditional on a change occurring, we then model the magnitude of that change using a linear mixed-effects model. The response variable is the difference in estimated travel time relative to baseline. We include a `Scenario*Region` interaction to assess both overall scenario effects and whether their impacts differ across geographic regions. Because each call contributes multiple observations (one per scenario), there is clear grouping in the data. Therefore, we include a random intercept at the call level to account for within-call correlation.

Exploratory residual analysis revealed heteroskedasticity across regions, so we incorporate region-specific residual variances using a `varIdent` variance structure. Attempts to also model variance by scenario resulted in singular fits, likely due to limited variability or nesting of scenarios within call-level random effects. A competing GLS model without random effects and with `Scenario*Region` variance structure had substantially worse AIC/BIC values (Table 2), which supports that accounting for call-level correlation and regional heterogeneity provides a better fit.

Final model specifications:

$$\begin{aligned} &\text{logit}\left[P(\text{Change}_i)\right] = \beta_0 + \beta_{\text{Scenario}} \\ &Y_{ij} = \beta_0 + \beta_{\text{Scenario}_j} + \beta_{\text{Region}_k} + \beta_{\text{Scenario}_j \times \text{Region}_k} + u_i + \varepsilon_{ijk}, \quad u_i \sim N(0, \sigma_u^2), \quad \varepsilon_{ijk} \sim N(0, \sigma^2_{\text{Region}_k}) \end{aligned}$$

where Y_{ij} is the travel time difference from baseline for call i under scenario j.

3.2 Implementation

For the logistic model, we created a binary indicator of whether estimated travel time changed from baseline and fit a generalized linear model using the `glm()` function in R with a binomial family and scenario as the predictor.

For the linear mixed-effects stage, we retained only rows where travel time differed from baseline and modeled the continuous difference using the `lme()` function in the `nlme` package. We specified a random intercept for each call (`random = ~1 | row_val`) and allowed the residual variance to differ by region using `weights = varIdent(form = ~1 | region)`.

3.3 Evaluation

To evaluate whether the model provided a good fit for the data, we examined multiple residual plots. The residuals are normalized to account for the model's variance structure. In the residuals vs. fitted values plot (Figure 4), the residuals are randomly scattered around zero without clear patterns or clustering, which suggests that the model's mean structure is appropriate and that the variance has been reasonably modeled. The horizontal layering is expected due to the discrete set of fitted values from the scenario–region combinations.

Residuals plotted by scenario (Figure 5) suggest some differences in spread, particularly for S1/S2 relative to S3/S4. While this indicates possible scenario-level heteroskedasticity, the model could not support this additional structure without convergence failures. The residuals by region plot (Figure 6) shows some remaining variance differences across regions, but the variation is improved compared to the model without variance adjustment.

We also examined the Q-Q plots to assess the normality assumption. For the random effects (Figure 7), the points closely align the diagonal line with only minor deviations in the lower tails, suggesting the normality assumption for random effects is reasonable. For the residuals (Figure 8), the Q-Q plot shows deviations in the tails, which may be due to repeated identical travel time estimates for certain scenarios or the discrete nature of the response variable. The central portion aligns closely with the theoretical line, indicating that the normality assumption is approximately satisfied for most observations.

4 Results

From the logistic regression results (Table 3), we estimated the probability that a call's estimated travel time would change relative to the baseline for each scenario. As shown in Table 4, 25.4% of calls exhibited a change under Scenario 1, 23.6% under Scenario 2, 10.8% under Scenario 3, and 9.3% under Scenario 4. This indicates that the majority of calls are unaffected by ambulance redistribution across all scenarios, which is an important consideration when evaluating the overall effectiveness and operational impact of each proposed strategy.

To quantify the magnitude of change in estimated travel time for calls that were affected, we computed estimated marginal means using the `emmeans()` function in R. As shown in Table 5, Scenario 3 has the lowest estimated marginal mean of –41.07 seconds, indicating that it reduces estimated travel time by approximately 41 seconds on average compared to the current setup. However, this effect is not statistically significant at the 0.05 level ($p = 0.1138$). Thus, while Scenario 3 appears to perform best among the four proposals, the overall improvement is not strong enough to conclude a statistically significant benefit at the county-wide level.

When evaluating results by region (Table 6), we observed substantial variation across regions. In the Central and South regions, all scenarios result in positive estimated marginal means, indicating longer travel times compared to the baseline. This is expected, as the scenarios reallocate an ambulance from the Central or South region to the North. In contrast, the North region experiences substantial improvements, with Scenario 3 achieving the largest reduction in estimated travel time (-477.7 seconds, $p = 0$). Scenario 1 yields a similar reduction, while Scenarios 2 and 4 perform less favorably, suggesting that moving an ambulance to the near north site is more effective than moving it farther away. These differences in fitted values across regions and scenarios are illustrated in Figure 9.

Overall, Scenario 3 provides the greatest improvement in the region it is intended to benefit (North), but its benefits at the county level are limited by negative impacts on the Central and South regions and the relatively small proportion of calls that actually experience a change.

5 Limitations and Future Directions

Our analysis has several limitations arising from both modeling assumptions and data constraints. First, the residuals grouped by scenario (Figure 5) show slight differences in spread, suggesting potential scenario-level heteroskedasticity not fully captured by our variance structure. Additionally, although the Q-Q plots (Figures 7 and 8) show that residuals generally follow a normal distribution, tail deviations indicate minor violations of the normality assumption. These deviations may result from the discrete nature of travel time differences and repeated identical values across scenarios.

The model also does not incorporate external factors that influence real-world EMS response times, because these were either unavailable in the dataset or could not be incorporated without strong assumptions. These include traffic conditions, rush-hour patterns, road closures, seasonal effects, and staffing availability. Without these factors, our estimates likely underestimate the true variability in response times.

Further limitations relate to input assumptions and model structure. The estimated travel times are based on a single best-guess value from Google Maps and do not capture real-time fluctuations. Our variance modeling accounted only for regional differences, although other sources, such as call urgency, may also contribute to variability. Finally, the dispatch rule used was intentionally simplified and does not fully reflect the complexity of real EMS operations, which are influenced by concurrent calls, on-the-ground decision making, and staffing constraints.

Future work includes integrating dynamic factors such as road conditions, staffing patterns, traffic, and call urgency (emergency vs. non-emergency) into the dispatch rule to more accurately reflect operational decision-making. We could also refine the estimated travel times by calibrating Google Maps predictions to match observed response times or by incorporating multiple travel time estimates rather than relying on a single best-guess value. These enhancements would produce a more realistic and informative model that evaluates optimal ambulance allocation not only by region, but also on call severity and real-time conditions.

6 Conclusion

Based on our analyses, if Vance County aims to improve response times in the North region, the optimal reallocation is Scenario 3, which moves one ambulance from the Central station to the Near North station. Under this scenario, approximately 10.8% of calls experience a change in estimated travel time relative to the current setup. Among those affected calls, Scenario 3 yields an average reduction of about 41 seconds compared to baseline; however, this overall effect is not statistically significant due to increased travel times in the Central and South regions, from which ambulance resources were removed. When focusing specifically on the North region, Scenario 3 produces a substantial improvement, reducing travel times by an estimated 478 seconds (approximately 8 minutes), which is operationally meaningful in emergency response contexts where outcomes depend critically on timely arrival.

In summary, reallocating an ambulance from Central to Near North significantly improves response times in the North region but introduces trade-offs in other regions and affects only a limited proportion of calls. Given these considerations, we recommend that Vance County weigh the regional benefits against potential losses in coverage elsewhere, as well as the costs and equity implications of reallocating resources.

7 Appendix

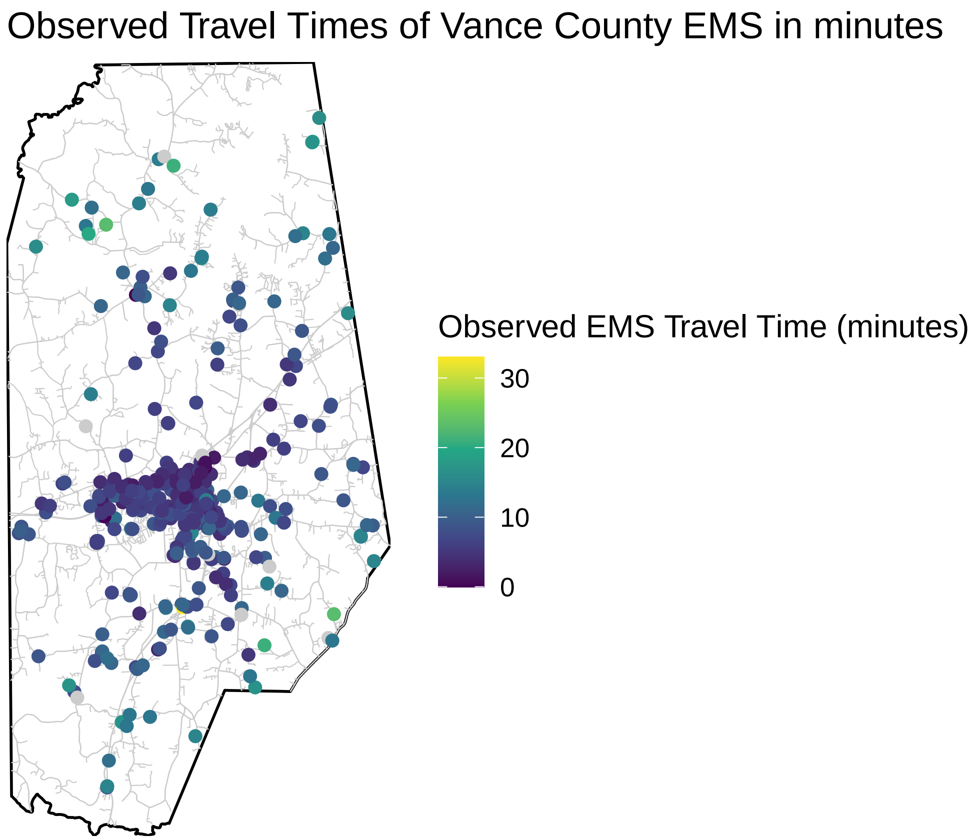

Figure 1: Spatial distribution of EMS calls and observed travel times in Vance County. Calls are densely concentrated in the Central region around Henderson, with longer travel times in the North and South.

Table 1: Comparison of travel time estimator performance based on MAE, RMSE, and bias. The best-guess (BG) estimator is selected due to its minimal bias and balanced error metrics.

Estimator	MAE	RMSE	Bias
est_BG	116.0021	192.2824	-1.737395
est_Op	112.3403	186.4617	16.277311
est_Pe	134.0651	214.2580	-41.674370
est_UA	116.4391	192.1155	-9.140756

Table 2: Model comparison of linear mixed-effects (LMM) and generalized least squares (GLS) models. The LMM with region-specific variance and random intercept provides a significantly better fit.

	Model	df	AIC	BIC	logLik	Test	L.Ratio	p-value
model_lmm	1	16	4320.808	4381.497	-2144.404		NA	NA
model_gls	2	24	4579.422	4670.454	-2265.711	1 vs 2	242.613	0

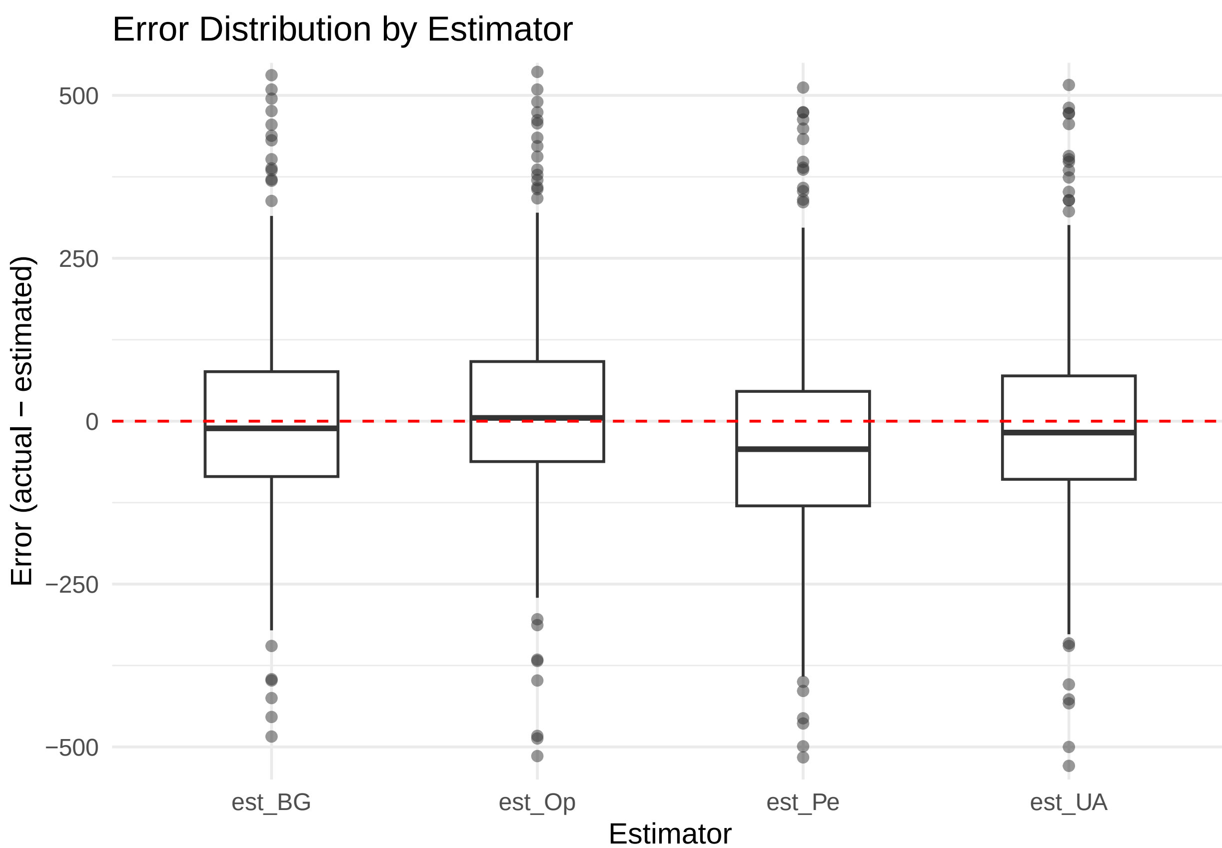

Figure 2: Error distribution of travel time estimators relative to actual response times. The best-guess (BG) estimator shows the lowest bias, supporting its use in subsequent modeling.

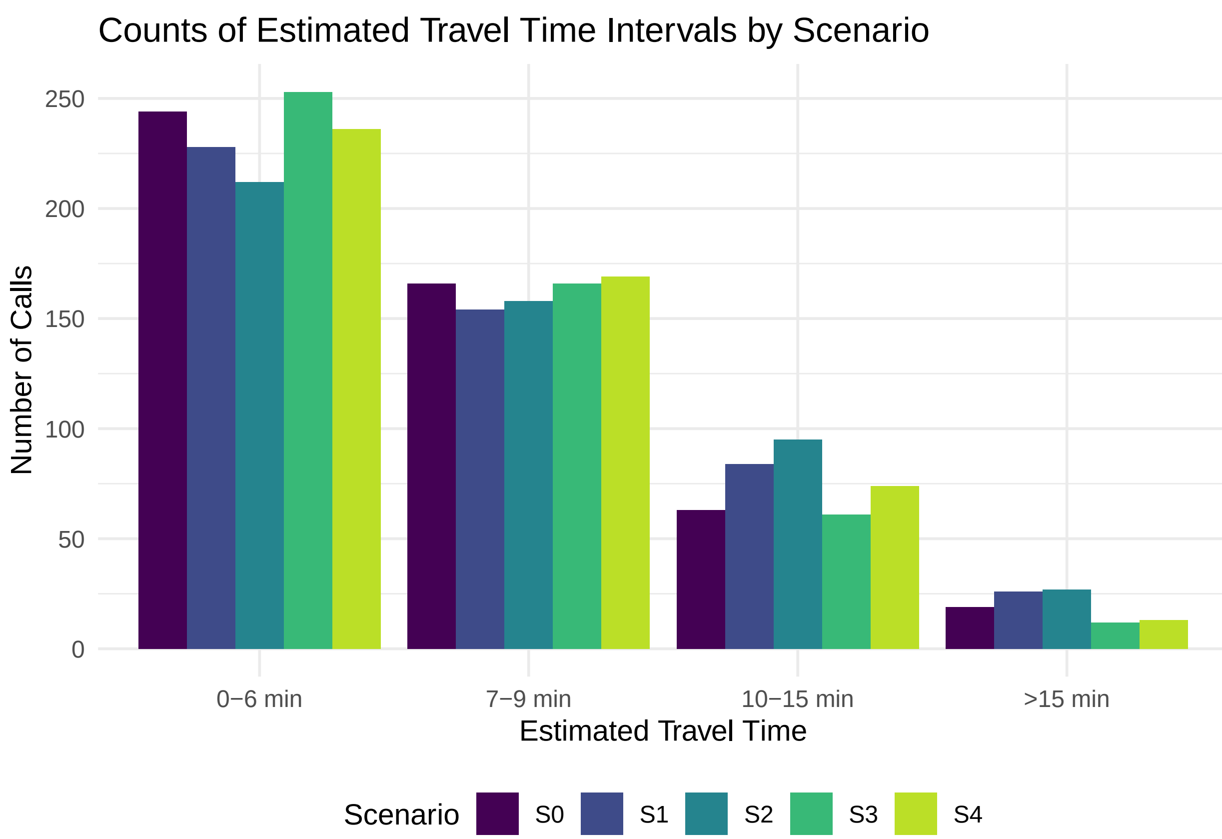

Figure 3: Estimated travel time intervals by scenario. Scenario 3 increases the number of calls reached within 0–6 minutes and reduces long-duration responses (>15 minutes), which suggests a potential improvement over the baseline.

Table 3: Logistic regression results showing the log-odds of travel time differing from the baseline across scenarios.

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	-1.077	0.104	-10.400	0.000	-1.284	-0.877
ScenarioS2	-0.099	0.148	-0.667	0.505	-0.390	0.192
ScenarioS3	-1.037	0.179	-5.810	0.000	-1.394	-0.693
ScenarioS4	-1.195	0.186	-6.413	0.000	-1.568	-0.836

Table 4: Predicted probabilities of travel time changes from the logistic model. Scenario 1 and Scenario 2 affect approximately one in four calls, while Scenarios 3 and 4 affect fewer than 11 percent.

Scenario	predicted_prob
S1	0.254
S2	0.236
S3	0.108
S4	0.093

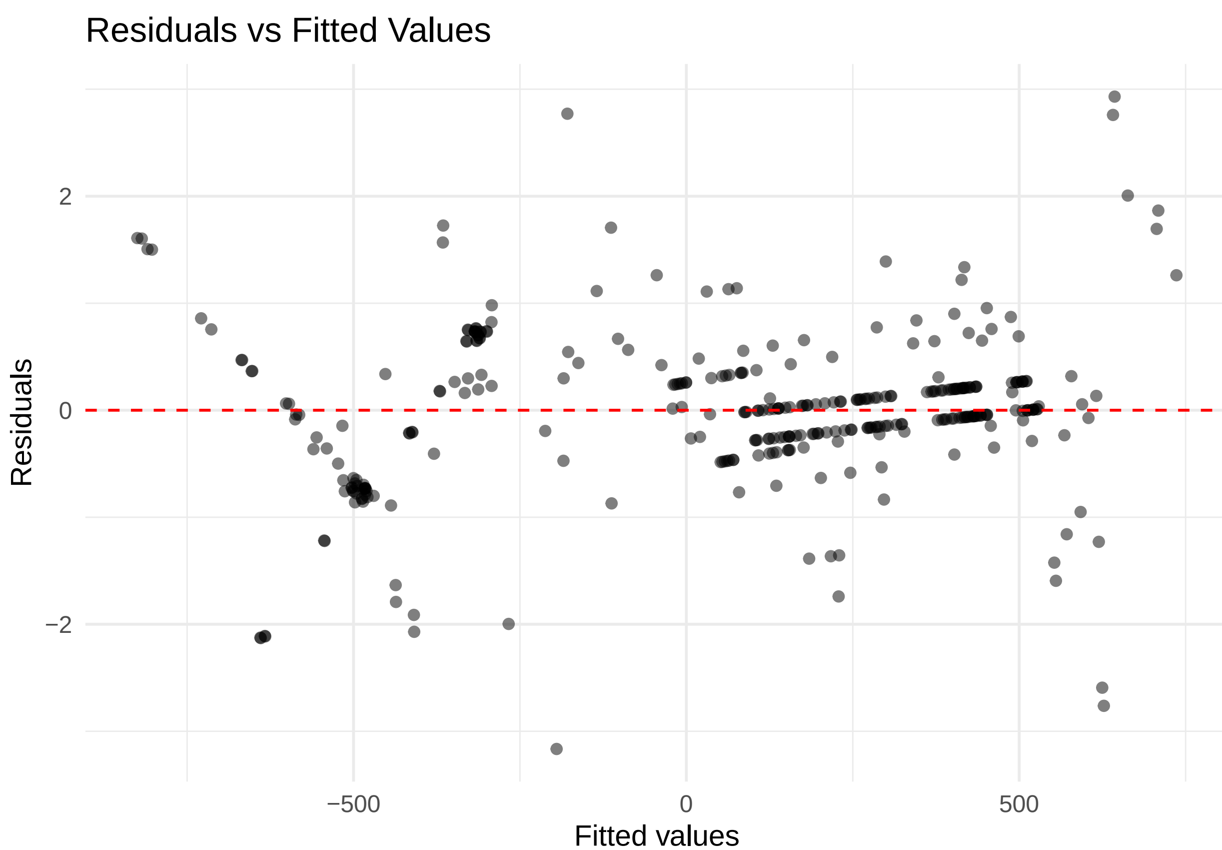

Figure 4: Residuals vs. fitted values showing random scatter around zero, suggesting the model adequately captures the mean structure and that variance has been reasonably model.

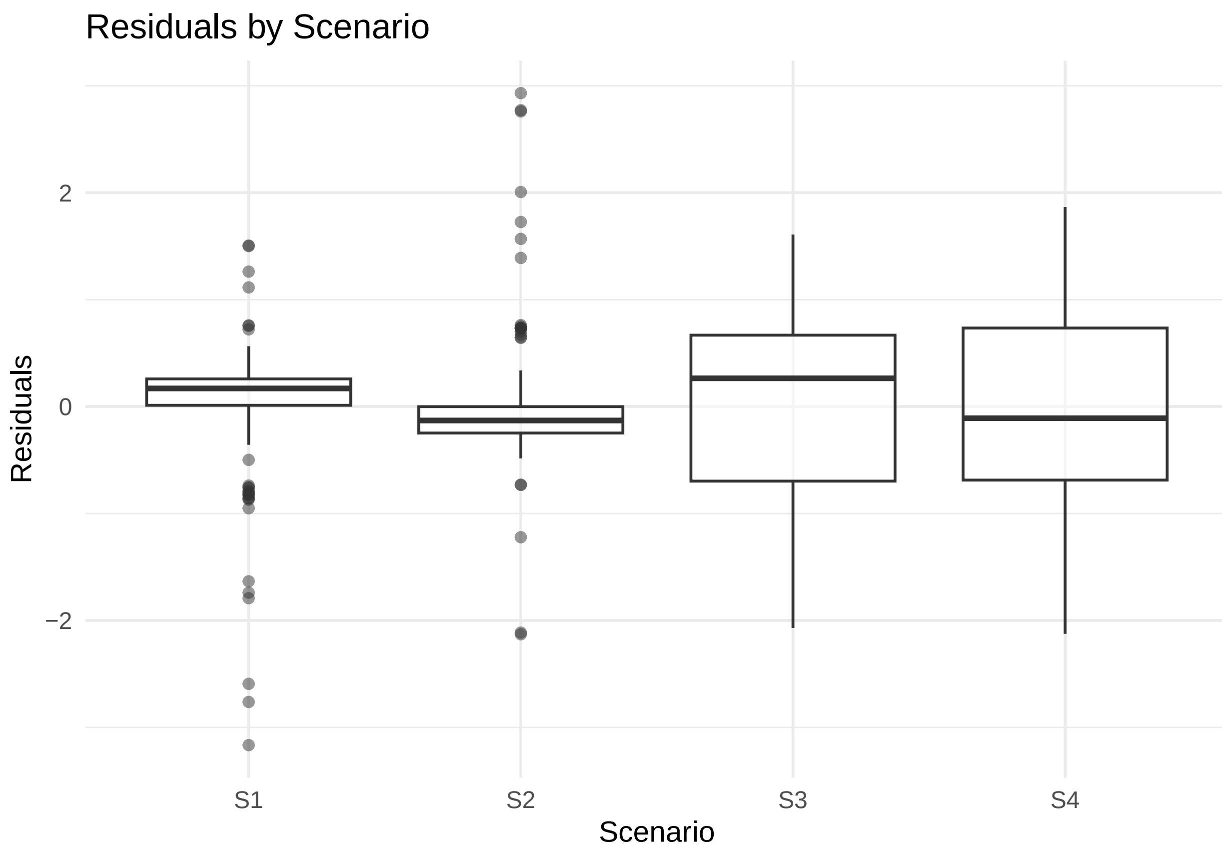

Figure 5: Boxplot of residuals by scenario. Differences in spread suggest possible scenario-level heteroskedasticity that could not be fully modeled due to data constraints.

Table 5: Estimated marginal mean travel time differences by scenario from the LMM. Scenario 3 yields the largest improvement relative to baseline, though not statistically significant at the county-wide level.

Scenario	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
S1	-21.09875	23.21680	140	-66.99961	24.802115	-0.9087706	0.3650330
S2	64.64958	24.13432	140	16.93472	112.364433	2.6787398	0.0082750
S3	-41.06652	25.80494	140	-92.08427	9.951237	-1.5914207	0.1137704
S4	110.67488	26.90801	140	57.47630	163.873455	4.1130835	0.0000663

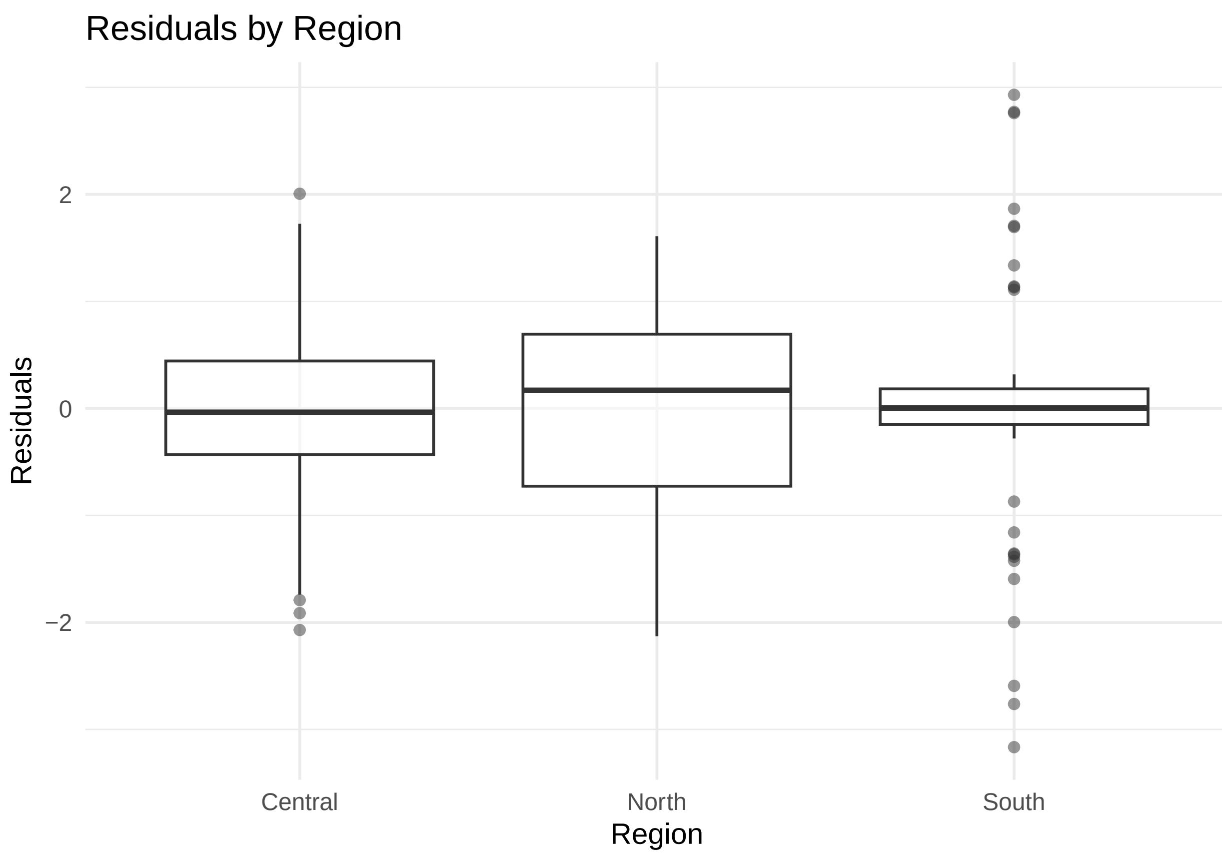

Figure 6: Boxplot of residuals by region. Regional variation is reduced compared to models without variance structure but some heteroskedasticity remains, supporting the use of region-specific variance modeling.

Table 6: Estimated marginal mean travel time differences stratified by region. Scenario 3 significantly improves response times in the North region, while increasing travel times in the Central and South.

Scenario	region	emmean	SE	df	lower.CL	upper.CL	t.ratio	p.value
S1	Central	93.5476	39.01992	142	16.41259	170.6826	2.397432	0.0178108
S2	Central	164.4503	40.05862	142	85.26197	243.6386	4.105241	0.0000678
S3	Central	120.9151	41.24058	142	39.39030	202.4400	2.931946	0.0039277
S4	Central	237.5040	41.83505	142	154.80406	320.2040	5.677155	0.0000001
S1	North	-462.4400	51.70835	140	-564.67019	-360.2098	-8.943236	0.0000000
S2	North	-292.2757	54.59975	140	-400.22233	-184.3291	-5.353060	0.0000003
S3	North	-477.7068	52.73204	140	-581.96083	-373.4527	-9.059137	0.0000000
S4	North	-292.8761	56.06538	140	-403.72035	-182.0318	-5.223831	0.0000006
S1	South	305.5962	25.59043	140	255.00251	356.1898	11.941814	0.0000000
S2	South	321.7741	25.61963	140	271.12275	372.4255	12.559671	0.0000000
S3	South	233.5921	38.87932	140	156.72556	310.4586	6.008131	0.0000000
S4	South	387.3967	40.28485	140	307.75135	467.0420	9.616435	0.0000000

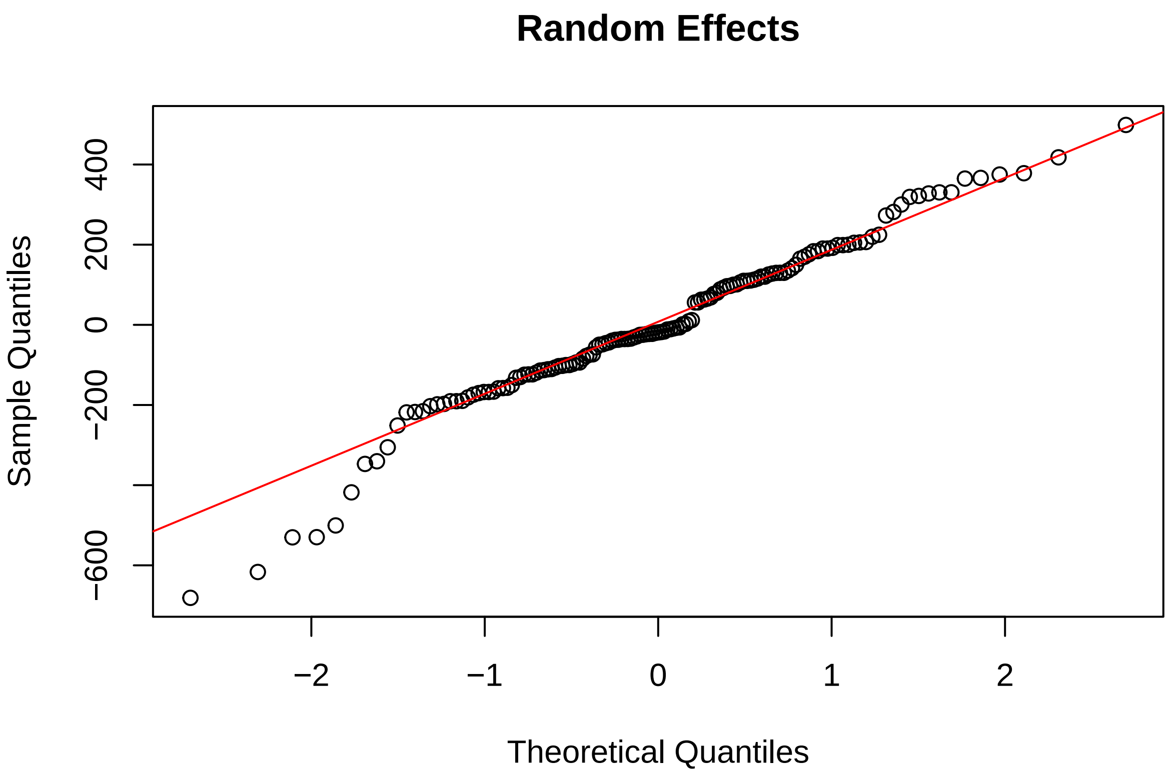

Figure 7: Q-Q plot of random effects showing close alignment with the theoretical normal distribution, indicating that the call-level random intercept is appropriately modele.

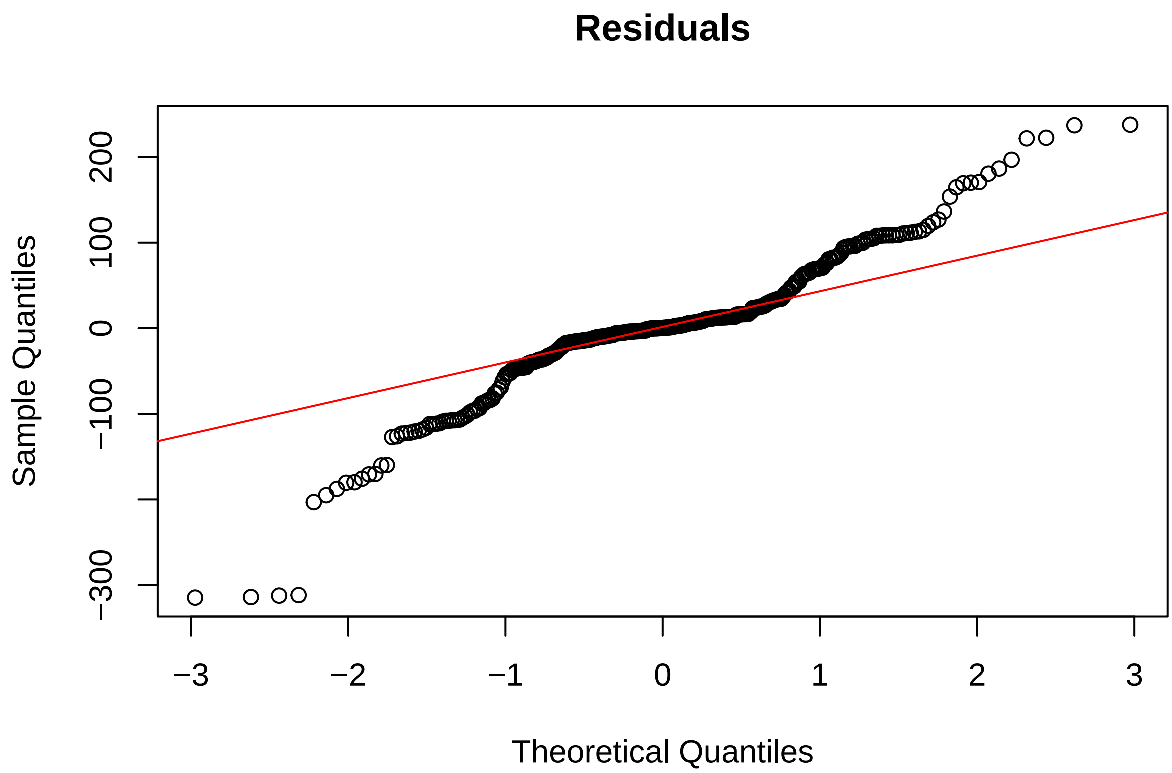

Figure 8: Q-Q plot of normalized residuals for the linear mixed-effects model. Residuals roughly follow the theoretical normal line, with tail deviations indicating minor departures from normality.

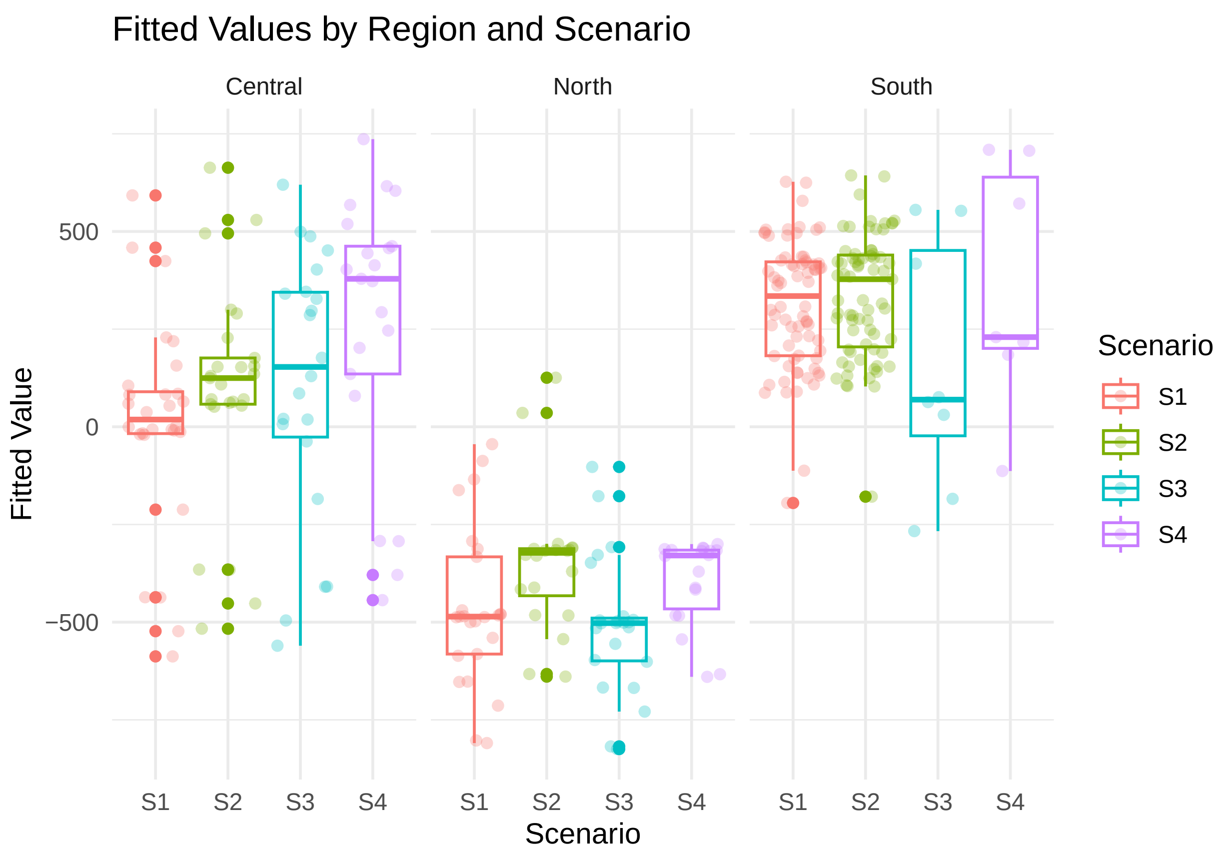

Figure 9: Fitted values by region and scenario. Scenario 3 produces the largest improvement in the North region while increasing fitted travel times in the Central and South, illustrating trade-offs across regions.