The Journal of Innovation and Social Science Research (JISSR),
eISSN 2591- 6890
Vol. 2 No.1, pp. 14-29, 2025

Comparing Methods for Estimating the Performance of Health Care Providers in Dutch Nursing Homes

Lillian Georgette Nyamizi
Institute of Finance Management (IFM)
Tanzania
lilian.nyamizi@ifm.ac.tz

Abstract

This study evaluates and compares between fixed and mixed effects logistic regression models to estimate the performance of health care providers in Dutch nursing homes, using secondary data from a 2002 cross-sectional survey of 42 nursing homes. The key factors examined include dietary intake, mobility, activity, sensory perception, moisture, ward specifications, and friction-and-shear. These factors are examined in relation to the prevalence of pressure ulcers among elderly residents. The study aimed to assess the statistical power and false positive control of fixed and mixed effect models using the Wald statistic, Likelihood Ratio Test (LRT), and Empirical Bayesian Estimates (EBEs). It also aimed at investigating the influence of shrinkage on inference accuracy. The study results revealed significant variability in coefficients between the two models, with the mixed effects model showing narrower confidence intervals and better control of false positives than the fixed effects model, particularly when using the EBE-based approach. The mixed effects model also identified three nursing homes with lower-quality care that were not detected by the fixed effects model. EBE effectively accounted for the number of patients in each institution thus ensuring accurate performance assessments by shrinking estimates for smaller settings toward the mean. Furthermore, the results suggest that the mixed effects model is a more reliable choice for statistical testing in covariate analysis. However, further research is recommended to refine EBE under heteroscedasticity assumptions and address its limitations. Besides, alternative methods such as Hierarchical Bayes Estimates (HBEs) should be explored for enhanced comparison and performance evaluation accuracy.

Keywords: *Health care providers, Health performance, Pressure ulcers, Dutch nursing homes*

1. Introduction

Under government regulations, a health care provider is defined as a licensed nurse practitioner, doctor of medicine, nurse-midwife, dentist, clinical social worker, or a health facility organization accredited to provide health care services (Thomas and Lobo, 2011). These providers offer various health care benefits that address patient and family needs, including reducing hospitalizations, emergency department visits, and overall care spending. In developed countries, the role of health care providers for the elderly has been a topic of debate for decades. In the 1970s, it was suggested that health care only slightly influenced elderly health outcomes (McKee and Jacobson, 2000; Karanikolos and McKee, 2013). However, emerging evidence indicates that modern health care has a demonstrable positive impact on population health, though methodological limitations affect the precision of these findings.

Accountability and quality improvement in health care rely on appropriate statistical methods to evaluate provider performance. Health accountability encompasses the necessary conditions for delivering high-quality care while ensuring efficient resource use. Health care providers influence key decisions, from surgeon selection to awarding managed care contracts (Delong et al., 1997). In developed countries, the quality of Coronary Artery Bypass Grafting (CABG) surgery is often used as a benchmark for assessing provider performance in health plans or hospitals (Rechel et al., 2013).

Hospitals face increasing challenges due to changes in population demographics, disease patterns, technological advancements, and public expectations. These factors necessitate shifts in care delivery, requiring new configurations of facilities, staff skills, and operational strategies (Rechel et al., 2011). As hospitals focus on short-term care for patients requiring intensive treatment, nursing homes provide care for the elderly who cannot live independently due to social or health issues (Yu et al., 2009). These dynamics underscore the importance of redefining boundaries between hospitals and nursing homes.

For over two decades, the Netherlands has conducted annual audits to assess health care provider performance in various institutions. These audits included monitoring the prevalence of pressure ulcers among elderly patients in hospitals and nursing homes (Meesterberends et al., 2011). Pressure ulcers, caused by prolonged immobility and inadequate blood supply, primarily affect elderly individuals with limited mobility, poor nutrition, or neurological conditions (NICE, 2020). Pressure ulcer prevalence serves as a key indicator of health care quality, though unadjusted rates may not accurately reflect provider performance due to differences in patient susceptibility. Adjusting for case-mix and random variation can help to isolate genuine quality differences (Bours et al., 2002). The Dutch State Inspectorate of Health has emphasized the need for efficient outcome measures to improve nursing home care and reduce pressure ulcer prevalence (Rosen et al., 2006).

International research further highlights the global significance of pressure ulcers. A 2017 study by the European Pressure Ulcers Advisory Board and other organizations, spanning 28 countries, emphasized the need for continued performance evaluation in hospitals, home care, and nursing homes. In Wales, a national audit revealed a decrease in pressure ulcer prevalence from 26.7% in 2008 to 8.9% in 2015 across 66 hospitals (Clark et al., 2017). Similarly, audits in the Netherlands between 2001 and 2008 showed a national decline in pressure ulcer prevalence (Dietz et al., 2008). Institutions often compare their performance against national benchmarks, with adjustments for patient severity to ensure fair evaluations.

This study aims to compare the statistical power of fixed and mixed effect models using the Wald statistic, Likelihood Ratio Test (LRT), and Empirical Bayesian Estimates (EBEs). It also explores the influence of shrinkage on LRT and EBE-based inferences. By evaluating the predictive power of these methods in identifying patients at risk for pressure ulcers, the study seeks to determine the more valid approach for assessing health care provider performance.

2. Methodology

2.1. Data

This study utilized a quantitative approach to collect and analyse secondary data from the national pressure ulcer incidence audit conducted in the Netherlands (Ahmed et al., 2019). A total of 42 nursing homes, comprising 6,281 patients, participated voluntarily in the study. However, 336 patients confirmed to have pressure ulcers were excluded from the analysis, as nursing homes are not accountable for treating pressure ulcers. Consequently, data from 5,945 patients were used for the analysis.

The data was collected in 2002, a year deliberately chosen due to the high prevalence of pressure ulcers and their substantial economic impact. During this period, the annual cost of pressure ulcer prevention and treatment in Dutch nursing homes increased dramatically, from USD 362 million to approximately USD 3 billion (Dassen et al., 2004). The occurrence of pressure ulcers was measured using the Braden scale, which consists of six subscales that assess pressure ulcer determinants. These include mobility, dietary intake, activity, sensory perception, friction-and-shear, and moisture (Braden and Maklebust, 2005). The subscales friction-and-shear and moisture were rated on a scale from 3 (most favourable) to 1 (least favourable), while all other determinants were rated from 4 (most favourable) to 1 (least favourable).

The data was categorized into three groups. First, variables were adjusted for case-mix differences among nursing homes and were grouped from most to least favourable. These variables included mobility (no limitation, slightly limited, very limited, and completely immobile), dietary intake (excellent, adequate, probably inadequate, and very poor), activity (walks frequently, walks occasionally, chair-fast, and bedfast), sensory perception (no impairment, slightly limited, highly limited, and completely limited), friction-and-shear (no apparent problem, potential problem, and problem), moisture (rarely moist, occasionally moist, very moist, and constantly moist), and ward specification (ward 9, ward 10, and ward 91).

Second, variables were created to capture the remaining differences between nursing homes in the occurrence of pressure ulcers, revealing differences in performance and quality of care. These included 41 effect variables representing the 42 nursing homes. Third, the presence of pressure ulcers was treated as a binary response, categorized as 1 if a patient had pressure ulcers and 0 if they did not.

2.2. Data Exploration

Frequency distribution tables were used to identify key aspects of the data, such as distribution patterns and outliers. A correlation matrix was employed to explore relationships between variables by providing pairwise correlations and the strength of these relationships.

2.3. Fixed Effects Models

Fixed effects models were used to control for time-invariant variables with time-invariant effects thus allowing unobserved variables to correlate with any observed variables. In this study, a logistic regression model was used to analyse the data, with pressure ulcers considered a successful response if a patient had one. A logistic regression model is a generalized linear model with a binomial random component and a logit link function (Gao et al., 2018). It assumes a linear relationship for the logit of the probability:

$logit\ [\pi(x)] = \log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \propto\ + \beta x$. The logistic function defines the relationship between x and $\pi(x) = \frac{\exp(\propto + \beta x)}{1+\exp(\propto + \beta x)}$. The notation $\pi(x)$ reflects the dependency of the success probability on the value of x, where π is the proportion of presence of ulcers response, (x) is the independent variable, and the parameter β determines the rate of increase or decrease of the s-shaped curve. The sign β indicates whether the curve ascends or descends, and the rate of change increases as $|\beta|$ increases. When the model holds with $\beta = 0$, the equation $\pi(x) = \frac{\exp(\propto + \beta x)}{1+\exp(\propto + \beta x)}$ implies a constant value. Then $\pi(x)$ is identical to all x, so the curve becomes a horizontal straight line. The binary response is independent x.

Multiple logistic regression is used for model building by fitting a complex form model: $\log\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$, where all the covariates are included in the model (Agresti, 2018). Parameter β_i refers to the effect x_i on the log odds of success, controlling for other xs. The $\exp(\beta_i)$ is a multiplicative effect on the odds of a unit increase in $x_i$at fixed levels of the other xs.

Substantial risk factors used to control for case mix were identified from the 2002 occurrence survey data. Nursing homes were considered to have an effect if the chance difference was significantly different from zero at the 5% level. Beta coefficients were used to estimate the probability of pressure ulcer occurrence. Nursing homes were classified as being of poor quality if the probability of pressure ulcer occurrence was above average, as performing well if the probability was below average, and as average if the probability was around the average level.

2.3.1. Wald Statistic and Likelihood-Ratio Model Comparison Tests

The Wald statistic and likelihood ratio were used to analyse the performance of the significance test of the hypothesis $H_0: \beta = 0$ about parameters in Generalized Linear Models (GLM). The test statistic is $Z = \frac{\hat{\beta}}{ASE}$, where $\hat{\beta}$ is the estimated parameter and *ASE* is the Asymptotic Standard Error, which has an appropriate standard normal distribution when $\beta = 0$ and Z is the standard normal table in one-sided or two-sided p-values. The two-sided alternative Z^2 has a chi-squared distribution with DF=1, and the p-value is the right-tail chi-squared probability above the observed value (Molenberghs and Verbeke, 2013).

The likelihood test uses the likelihood function through the ratio of two maximizations, the maximum over the possible parameter values that assumes the null hypothesis (l_0) and the maximum over the more extensive set of possible parameter values for the entire model, permitting the null or the alternative hypothesis to be true (l_1). When the null hypothesis is $H_0: \beta_0 = \beta_1 = \beta_2 = \cdots = \beta_k = 0$, l_0 is the likelihood function calculated at the (α, β_i) combination for which actual data would have been most likely; l_0 is the likelihood function calculated at the α value for which the data would have been most likely, when $\beta_i = 0$. Then l_1 is always at least as large as l_0, but l_0 refers to maximizing over a restricted set of the yield parameter values l_1. The likelihood ratio test statistic equals $-2\log\left(\frac{l_0}{l_1}\right) = -2[\log(l_0) - log(l_1)] = -2\,(L_0 - L_1)$ the maximized log-likelihood for the model to the maximized log-likelihood for the simple model without the same parameters.

The Hosmer-Lemeshow test was used to assess the goodness of fit. This test calculates Pearson chi-square statistics based on observed and expected values. To evaluate whether a model excluding specific terms could be assumed, maximized log-likelihood values of nested models were compared. The test statistic follows a chi-square distribution for large samples, with degrees of freedom equal to the difference between the number of parameters in the full model and the reduced model. The significance of each term in the model was assessed using the Wald statistic and the Likelihood Ratio Test, with their corresponding p-values reported.

2.3.2. Residuals for Logit Models

The goodness-of-fit statistic, represented as chi-square, serves as a summary indicator of the overall quality of fit. It provides an additional diagnostic tool to identify any lack of fit in the model. For this study, the Pearson residual was used and it is defined as:

Pearson residual $(e_i) = \frac{observed\ count - fitted\ count}{\sqrt{var(observed\ count)}}$

Each Pearson residual is a component of the chi-square statistic. The Pearson residual (e_i) has an approximately normal distribution when the binomial index is large. If the model holds, the residual has an expected value close to zero but a smaller variance than a standard normal variable. Deviations with absolute values larger than 2 suggest a possible lack of fit (Agresti, 2018). Variables contributing to the occurrence of pressure ulcers were analyzed using a multilevel analysis.

2.4. Mixed Effects Model

Multilevel analysis was performed with patients at level 1 and nursing homes at level 2. The inclusion of case-mix differences and random coefficients was significant in the model analysis. The random coefficient at level 2 represents variations in the occurrence of pressure ulcers across nursing homes, while the random coefficient at level 1 accounts for random variations within nursing homes (Molenberghs, 2011). The random coefficients were adjusted for case-mix, and the level 2 coefficients reflect differences in nursing home performance regarding quality of care.

The mixed effects model is defined as $logit\,\pi(X_i, Z_i) = X_i\beta + Z_i b_i$, $b_i = N(0, D)$ $b_i \ldots \ldots b_N$ are independent, $X_i$and Z_i are $(n_i \times p)$ and $(n_i \times q)$ dimensional matrices of known covariates, respectively. β is a p-dimensional vector containing the fixed effects, b_i is the q-dimensional vector containing random effects, and D is a general $(q \times q)$ covariance matrix. The random effects estimators approximate the Bayes estimators, minimizing Bayes risk under squared error loss. When the effects are normally distributed, the empirical Bayes estimates outperform the fixed effects estimates (Molenberghs and Verbeke, 2013).

2.4.1. Empirical Bayes Estimate (EBE)

The random effects b_i is estimated by using Bayesian techniques. Once the observed values of y_i for Y_i have been collected, the posterior distribution of b_i conditional on $Y_i = y_i$ can also be calculated. The empirical Bayes method is one of the methods used to estimate b_i, where the inference and estimation are based on the posterior distribution. If the density function of $Y_i$conditional of y_i and the prior function of b_i by $f(y_i|b_i)$ and $f(b_i)$ are denoted, we will have: $f(b_i|y_i) = f(b_i|Y_i = y_i)\frac{f(y_i|b_i)f(b_i)}{\int f(y_i|b)f(b_i)db}$. The estimated posterior can estimate b_i, the Bayes estimator for the centre effects. The empirical Bayes estimators are obtained by approximating the posterior mean, which is performed by estimating the model parameters ($\beta$and D) that appear in the multilevel model. The resulting estimates for the random effects are EBE, notated as $\widehat{b_i}$ (Molenberghs and Verbeke, 2010). The empirical Bayes method is characterized as:

a) In Bayesian analysis, the b_i is a prior distribution.
b) The parameters of the prior distribution are estimated based on the observed data.
c) The inference is observed in the estimated posterior distribution.

2.4.2. Cross-validation

Cross-validation was performed in the multiple logistic regression analysis, with 60% of patients from each nursing home used to estimate model parameters and the remaining 40% to test model validity. Cross-validation was also used to evaluate the predictive power of the estimates. The predictive performance of each method was assessed using the Receiver Operating Characteristic (ROC) curve, which plots sensitivity against specificity. In this study, sensitivity refers to the probability of correctly identifying patients with pressure ulcers, while specificity refers to the probability of correctly identifying patients without pressure ulcers, based on the results of the multiple logistic regression analysis.

The ROC curve was generated by plotting observed values against predicted values. The c-statistic, representing the area under the curve, was used to assess the model's discrimination ability, with values above 0.70 considered good. Estimates from the random coefficient method for nursing homes with fewer patients were compared to those from fixed effects models in logistic regression. This comparison ensured appropriate interpretation after adjusting for case-mix using similar risk factors in the random coefficient model (Halfens et al., 2000).

The analysis was conducted using various statistical software tools: SAS for analyzing logistic regression (fixed and mixed effects models), SPSS for determining standard errors in the fixed effects model and conducting ROC analysis, and S-PLUS for graphical displays.

3. Results

3.1. Exploratory Data Analysis

The data consisted of two samples. The first sample, comprising 3,544 of the 5,945 patients, was used for logistic regression analysis. The second sample, containing the remaining 2,401 patients, was used to verify the predictive validity of the model. The frequency distribution of pressure ulcer responses is presented in Table 1. Approximately 12% of the patients in the nursing homes in the Netherlands had pressure ulcers, while 88% did not.

Table 1: Frequency Distribution of Pressure Ulcers Response

Pressure ulcers	Frequency	Percentage (%)	Cumulative frequency
Absent (0)	3111	87.78	3111
Present (1)	433	12.22	3544

The frequency distribution of pressure ulcer responses by each covariate, focusing on patients with pressure ulcers, shows a positive association with sensory perception, activity, dietary intake, friction-and-shear, and mobility. Specifically, as the levels of these covariates improve (from least to most favourable), the proportion of patients with pressure ulcers decreases. The probability of developing pressure ulcers is higher for patients who are completely limited in sensory perception, mobility, activity, have very poor dietary intake, and they experience problems with friction-and-shear. Similar trends are observed for patients with unfavorable ward specifications and moisture levels, where pressure ulcers are more prevalent at the least favorable scores.

The covariate distribution for case-mix differences among nursing homes was validated by ensuring that all values were considered from least to most favorable. The correlation matrix indicates that the covariates are slightly correlated, with the highest correlation value being 0.58, suggesting only minor multicollinearity issues.

3.2. Simple Logistic Regression Analysis

Table 2 presents a simple logistic regression analysis. The effect of each covariate—sensory perception, activity, mobility, dietary intake, friction-and-shear, ward specialty (ward 10 and ward 91), and moisture—on pressure ulcers was significant, as the 95% confidence interval for each did not include 1. However, ward 9 showed no significant effect on pressure ulcers, with a p-value of 0.334 and a 95% confidence interval for the odds ratio that included 1.

Table 2: Simple Logistic Regression Analysis Results

Model	Parameter	Estimate	Standard error	Chi-square	Odds ratio (95% C.I)	p-value
1	Intercept	-2.6177	0.1216	463.5065		< .0001
	Sensory perception	0.3565	0.0577	38.1933	1.428 [1.276; 1.599]	< .0001
2	Intercept	-4.1292	0.1917	463.7396		< .0001
	Activity	0.8129	0.0635	164.0646	2.254 [1.991; 2.553]	< .0001
3	Intercept	-3.7661	0.5169	576.1292		< .0001
	Mobility	0.6857	0.0506	183.4468	1.985 [1.798;2.192]	< .0001
4	Intercept	-2.7561	0.1303	447.6603		< .0001
	Dietary intake	0.4360	0.0631	47.8156	1.547 [1.367; 1.750]	< .0001
5	Intercept	-4.1794	0.1773	555.8971		< .0001
	Friction-and-share	1.0309	0.0708	211.9741	2.804 [2.440; 3.221]	< .0001
6	Intercept	-32232	0.0790	864.4070		< .0001
	Ward 9	0.2048	0.2118	0.9345	1.227 [0.810; 1.859]	0.3337
	Ward 10	0.7499	0.1111	45.5776	2.117 [1.703; 2.632]	< .0001
	Ward 91	0.9277	0.2187	17.9907	2.529 [1.647; 3.882]	< .0001
7	Intercept	-2.4308	0.0964	635.4088		< .0001
	Moist 2	0.7732	0.1321	34.2513	2.167 [1.672; 2.807]	< .0001
	Moist 3	0.6259	0.1443	18.8074	1.870 [1.409; 2.481]	< .0001
	Moist 4	0.6709	0.1621	17.1240	1.956 [1.424; 2.688]	< .0001

3.3. Multiple Logistic Regression

3.3.1. Fixed Effects Model

In the fixed effects model, all covariates were included in the final model, even though some were not statistically significant at the 5% level. All variables were deemed biologically relevant for estimating the performance of health care providers in nursing homes. The parameter estimates and the goodness-of-fit statistics for the final model are presented in Table 3. The logistic model indicated that activity, dietary intake, mobility, friction-and-shear, and ward specialty (Wards 10 and 91) were significantly associated with pressure ulcers (p-value < 0.05). Sensory perception and occasionally moist patients (moist 1) were marginally significant, with p-values of 0.064 and 0.058, respectively. Significant differences in performance were observed among nursing homes 103, 258, 276, 194, 202, 203, and 204 (p-value < 0.05) after controlling for case-mix differences. The beta coefficients for these nursing homes were all positive, indicating that the probability of patients developing pressure ulcers in these facilities was above average. This suggests lower care quality in these nursing homes. The 95% confidence intervals for the exponential of the beta coefficients did not include 1 thus confirming the statistical significance of the identified variables.

3.3.2. Diagnostics

The Hosmer-Lemeshow statistic is a goodness-of-fit test for logistic regression, particularly in risk prediction models. This test evaluates how well the data fits the model by comparing observed event rates to expected event rates across population subgroups. A poor fit is indicated when the significance value is less than 0.05 (p-value < 0.05). In this study, the overall model shows no evidence of a lack of fit, with a p-value of 0.152, as presented in Table 3.

Similarly, Pearson's chi-square statistic, based on the covariates, was used to assess the model's goodness of fit. This statistic indicated a good model fit, with a p-value of 0.93, suggesting insufficient evidence to reject the null hypothesis. Pearson's chi-square goodness-of-fit test provides an additional diagnostic evaluation of any descriptive lack of fit in logistic regression.

Table 3: Parameter Estimate and Goodness of Fit of the Model

Parameter	Estimate (β)	Standard error	Wald statistic	p-value	95% C. I for exp(β)		Hosmer and Lemeshow goodness of fit test		
					Lower	Upper	Value	DF	p-value
Intercept	-5.655	0.376	226.223	0.003			1.973	8	**0.152**
Sensory	-0.142	0.077	3.439	0.064	0.746	1.008			
Activity	0.325	0.097	11.210	**0.001**	1.144	1.674			
Dietary int.	0.195	0.073	7.212	**0.008**	1.054	1.401			
Mobility	0.305	0.084	13.254	**0.000**	1.151	1.599			
Friction	0.600	0.100	36.304	**0.000**	1.499	2.215			
Moist1	0.284	0.150	3.589	0.058	0.990	1.781			
Moist 2	0.044	0.170	0.068	0.795	0.749	1.460			
Moist 3	-0.270	0.190	2.008	0.156	0.526	1.109			
Ward 9	0.328	0.243	1.818	0.178	0.862	2.234			
Ward 10	0.396	0.139	8.165	**0.004**	1.132	1.950			
Ward 91	1.028	0.291	12.456	**0.001**	1.580	4.951			
N Home10	0.069	0.419	0.027	0.869	0.472	2.435			
N Home31	0.255	0.421	0.368	0.554	0.566	2.943			
N Home43	0.488	0.522	0.874	0.350	0.585	4.536			
N Home56	0.043	0.440	0.010	0.922	0.441	2.472			
N Home58	0.312	0.433	0.520	0.471	0.585	3.194			
N Home67	0.627	0.396	2.509	0.113	0.862	4.066			
N Home70	-0.816	0.648	1.585	0.208	0.124	1.575			
N Home73	-0.580	0.586	0.978	0.323	0.178	1.766			
N Home74	-0.028	0.423	0.004	0.947	0.424	2.227			
N Home103	1.496	0.373	16.054	**0.000**	2.147	9.279			
N Home111	0.070	0.416	0.028	0.866	0.474	2.427			
N Home116	-0.100	0.390	0.065	0.798	0.422	1.942			
N Home117	-1.507	1.032	2.132	0.144	0.029	1.676			

Parameter	Estimate (β)	Standard error	Wald statistic	p-value	95% C. I for exp(β)		Hosmer and Lemeshow goodness of fit test
N Home118	-1.060	0.772	1.884	0.170	0.076	1.573	
N Home126	-0.389	0.474	0.673	0.412	0.268	1.716	
N Home127	-1.118	1.050	1.132	0.287	0.042	2.563	
N Home130	-0.527	0.538	0.960	0.327	0.206	1.694	
N Home140	0.497	0.457	1.182	0.277	0.671	4.030	
N Home153	0.463	0.411	1.274	0.259	0.711	3.554	
N Home158	0.993	0.411	5.843	**0.016**	1.207	6.034	
N Home165	0.479	0.398	1.447	0.229	0.740	3.523	
N Home171	0.869	0.505	2.959	0.085	0.886	6.422	
N Home172	0.913	0.876	1.086	0.297	0.447	13.981	
N Home174	0.561	0.509	1.216	0.270	0.647	4.747	
N Home176	0.756	0.349	4.690	**0.030**	1.074	4.224	
N Home178	0.714	0.377	3.592	0.058	0.976	4.275	
N Home179	0.072	0.414	0.030	0.863	0.477	2.417	
N Home185	-0.029	0.390	0.005	0.942	0.452	2.008	
N Home189	0.755	1.089	0.480	0.488	0.252	17.980	
N Home191	-0.834	0.602	1.916	0.166	0.133	1.415	
N Home194	0.781	0.353	4.908	**0.027**	1.094	4.362	
N Home196	0.479	0.419	1.306	0.253	0.710	3.669	
N Home197	-4.642	9.917	0.219	0.640	0.000	26.506	
N Home202	1.137	0.402	8.020	**0.005**	1.419	6.851	
N Home203	1.192	0.411	8.408	**0.004**	1.471	7.367	
N Home204	0.879	0.374	5.516	**0.019**	1.157	5.019	
N Home207	0.696	0.578	1.449	0.229	0.656	6.223	
N Home209	-0.821	1.038	0.626	0.429	0.058	3.363	
N Home210	0.703	0.623	1.275	0.259	0.596	6.847	
N Home211	0.190	1.275	0.022	0.882	0.099	6.847	
N Home219	-0.212	0.490	0.188	0.665	0.310	2.111	
N Home220	0.335	0.634	0.278	0.598	0.403	4.846	

3.3.3. Cross-validation

Cross-validation was performed to predict the occurrence of pressure ulcers in the Dutch nursing homes. The second sample (validation data) was used to minimize total misclassification rates and maximize the total number of correct classifications. To achieve this, an appropriate proportion was selected to reduce the misclassification rates for both the first sample (calibration data) and the second sample. A prediction threshold of 0.4 was applied to classify the predicted values for both samples. If the predicted value was greater than or equal to 40%, pressure ulcers were predicted in the patients; if the predicted value was less than 40%, no pressure ulcers were predicted.

The presence of pressure ulcers, as determined by predicted values, was compared with the observed values. Table 4 presents the misclassification rates for the first sample, with a total misclassification rate of 12.6%. The model was validated to predict the occurrence of pressure ulcers in nursing homes.

Table 4: Predicted Pressure Ulcers in Calibration Data (First Sample)

Predicted	**Observed**		**Total number predicted**
	Absent (0)	**Present (1)**	
Absent (0)			
Frequency	3073	407	3480
Percent	86.71	**11.48**	98.19
Row Pct	88.30	11.70	
Col Pct	98.78	94.00	
Present (1)			
Frequency	38	26	64
Percent	**1.07**	0.73	1.80
Row Pct	59.38	40.63	
Col Pct	1.22	6.00	
Total number observed	3111 87.78	433 12.22	**3544** 100.00

Table 5 presents the results of the validated model. The total misclassification rate for the second sample is 12.9%, which does not appear to differ significantly from zero. However, it was necessary to confirm if a significant difference existed. A Wald statistic test was conducted, yielding a value of 0.36 with a p-value of 0.72. Consequently, there is no evidence to reject the null hypothesis, indicating that the misclassification rates for the first and second samples do not differ at the 5% significance level.

Table 5: Predicted Pressure Ulcers in Predictive Validity Data (Second Sample)

Predicted	**Observed**		**Total number predicted**
	Absent (0)	**Present (1)**	
Absent (0)			
Frequency	2067	234	2301
Percent	86.09	**9.75**	95.84
Row Pct	89.83	10.17	
Col Pct	96.50	90.35	
Present (1)			
Frequency	75	25	100
Percent	**3.12**	1.04	4.16
Row Pct	75.00	25.00	
Col Pct	3.50	9.65	
Total number observed	2142 89.21	259 10.79	**2401** 100.00

The Receiver Operating Characteristic (ROC) curve for the fixed effects model was used to evaluate the model's predictive accuracy. The area under the curve (AUC), expressed by the c-statistic, was 0.764, as shown in Figure 1. A value closer to 1 indicates better performance in correctly classifying outcomes. The obtained c-statistic exceeds the standard threshold of 0.70, suggesting that the model discriminates well between individuals with and without pressure ulcers. Therefore, the model is considered to have a good predictive accuracy.

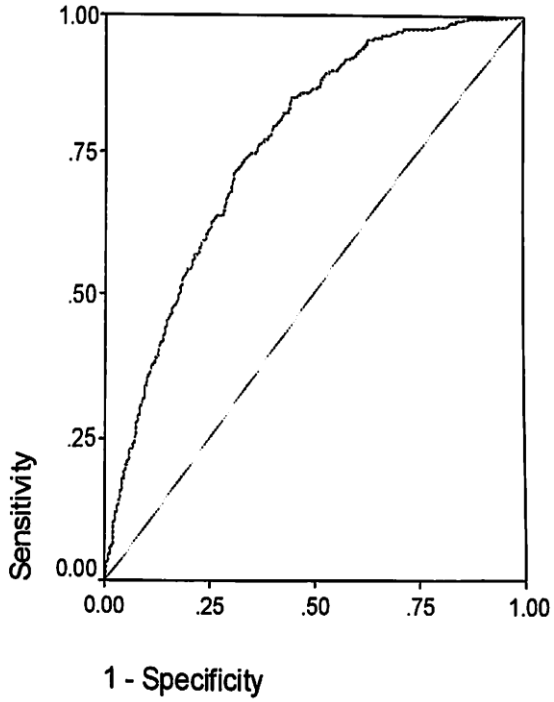

Figure 1: Receiver operating characteristic curve for fixed effects model

3.3.4. Mixed Effects Model

The results of the Empirical Bayes Estimates (EBE) are presented in Table 6. The fixed parameters for sensory perception, activity, dietary intake, mobility, friction-and-shear, Ward 10, and Ward 91 were found to have significant effects on pressure ulcers, with p-values of 0.043, 0.002, 0.009, 0.001, 0.001, and 0.001, respectively. Notably, in the fixed effects model, sensory perception was only marginally significant (p-value = 0.064). The 95% confidence intervals for the parameters in the EBE confirm the significance of these variables, as the intervals do not include zero. The beta coefficient for sensory perception is negative, indicating that impaired sensory perception has a negative effect on the presence of pressure ulcers.

A nursing home's performance was estimated as the mean of the posterior distribution of the corresponding level 2 random coefficient. As shown in Table 6, the variance of the level 2 random coefficient is significantly different from zero (p-value = 0.015), suggesting significant differences in performance between nursing homes after controlling for case-mix differences.

Table 6: Factors related to pressure ulcers in Empirical Bayes Analysis

Parameter	Estimate	Standard error	t-value	p-value	95% C. I	
Fixed						
Intercept	-5.4002	0.2722	-19.84	< .0001	-5.9494	-4.8509
Sensory	-0.1559	0.0749	-2.08	0.0434	-0.3070	-0.0048
Activity	0.3171	0.0950	3.34	0.0018	0.1255	0.5088
Mobility	0.2993	0.0823	3.64	0.0007	0.1332	0.4654
Dietary int.	0.1959	0.0712	2.75	0.0087	0.0523	0.3395
Friction	0.6112	0.0977	6.25	< .0001	0.4139	0.8084
Ward9	0.3749	0.2360	1.59	0.1197	-0.1014	0.8512

Parameter	Estimate	Standard error	t-value	p-value	95% C. I	
Ward10	0.4548	0.1297	3.51	0.0011	0.1931	0.7165
Ward91	0.9569	0.2664	3.59	0.0009	0.4193	1.4945
Moist1	0.3193	0.1471	2.17	0.0356	0.02251	0.6160
Moist2	0.04733	0.1650	0.29	0.7757	-0.2857	0.3803
Moist3	-0.2657	0.1860	-1.43	0.1606	-0.6411	0.1097
Random (level 2variance) Intercept	0.2173	0.08571	2.54	0.0150	0.04434	0.3903

Table 7 shows the EBE of the random effect coefficients. It was evidenced that the 95% confidence interval is smaller for the random coefficients' EBE than the estimates of the fixed coefficients. The shrinkage effect toward the mean on EBE random coefficients implies that the variability of the estimates for the fixed coefficients is larger than that of the random coefficients. The EBE of the random effect coefficients found 3 nursing homes, including 103, 202 and 203 to have significant random coefficients. The 3 nursing homes diverged significantly from the average nursing home in terms of the presence of pressure ulcers even after controlling for case-mix differences. The beta coefficients for these nursing homes are positive, indicating that the patients' probability of pressure ulcers is above average. These results showed evidence that these nursing homes have less patient care quality. In addition, the nursing homes 158, 176 and 204 are at the marginal significance, with positive beta coefficients. In the fixed effects model, these specific nursing homes were significant at a 5% level which indicates that these particular nursing homes have less patient care quality.

Table 7: Empirical Bayesian Estimates for the Random Coefficients

Nursing homes	Effect	Estimate ($\boldsymbol{\beta}$)	Standard Error	Probit	95% CI	
					Lower	Upper
10	μ	-0.1356	0.2704	0.6186	-0.6813	0.4101
31	μ	0.0009	0.2751	0.9976	-0.5542	0.5559
43	μ	0.1295	0.3379	0.7034	-0.5523	0.8113
56	μ	-0.1451	0.2829	0.6107	-0.7161	0.4259
58	μ	0.0237	0.2843	0.9339	-0.5501	0.5975
67	μ	0.2609	0.2643	0.3292	-0.2725	0.7944
70	μ	-0.4586	0.3452	0.1913	-1.1553	0.2381
73	μ	-0.4027	0.3340	0.2346	-1.0766	0.2712
74	μ	-0.1902	0.2715	0.4875	-0.7380	0.3577
103	μ	0.9419	0.2692	**0.0011**	0.3986	1.4852
111	μ	-0.1347	0.2683	0.6184	-0.6763	0.4069
116	μ	-0.2674	0.2467	0.2846	-0.7653	0.2305
117	μ	-0.4906	0.3833	0.2076	-1.2641	0.2829
118	μ	-0.4729	0.3673	0.2050	-1.2143	0.2684
126	μ	-0.3786	0.2974	0.2101	-0.9788	0.2217
127	μ	-0.3268	0.3968	0.4148	-1.1275	0.4739
130	μ	-0.4151	0.3187	0.1999	-1.0583	0.2281

Nursing homes	Effect	Estimate (β)	Standard Error	Probit	95% CI	
					Lower	Upper
140	μ	0.1662	0.3057	0.5894	-0.4506	0.7831
153	μ	0.1461	0.2735	0.5960	-0.4058	0.6979
158	μ	0.5144	0.2904	0.0838	-0.0717	1.1006
165	μ	0.1532	0.2637	0.5644	-0.3789	0.6853
171	μ	0.3286	0.3453	0.3467	-0.3682	1.0254
172	μ	0.1340	0.4228	0.7529	-0.7192	0.9873
174	μ	0.1553	0.3296	0.6399	-0.5099	0.8205
176	μ	0.4405	0.2210	0.0527	-0.0054	0.8865
178	μ	0.3457	0.2508	0.1753	-0.1603	0.8518
179	μ	-0.1335	0.2666	0.6192	-0.6715	0.4046
185	μ	-0.2154	0.2484	0.3908	-0.7167	0.2859
189	μ	0.0720	0.4393	0.8705	-0.8145	0.9586
191	μ	-0.5035	0.3394	0.1454	-1.1885	0.1815
194	μ	0.4163	0.3380	0.0750	-0.0439	0.8764
196	μ	0.1555	0.2801	0.5817	-0.4097	0.7208
197	μ	-0.2017	0.4346	0.6450	-1.0788	0.6753
202	μ	0.5998	0.2834	**0.0403**	0.0278	1.1718
203	μ	0.6308	0.2881	**0.0341**	0.0495	1.2121
204	μ	0.4620	0.2517	0.0735	-0.0460	0.9699
207	μ	0.2016	0.3649	0.5837	-0.5349	0.9380
209	μ	-0.2550	0.3965	0.5237	-1.0552	0.5452
210	μ	0.1727	0.3786	0.6507	-0.5913	0.9367
211	μ	0.0021	0.4379	0.9962	-0.8815	0.8857
219	μ	-0.2750	0.3049	0.3722	-0.8903	0.3403
220	μ	0.0155	0.3675	0.9667	-0.7261	0.7570
226	μ	-0.4393	0.4116	0.2919	-1.2701	0.3914

4. Discussion, Conclusion and Recommendation

This paper compares two methods for estimating the performance of health care providers in Dutch nursing homes. The indicator measure was examined within a framework of fixed effects models. The data structure for the mixed effects model was empirical, with patients nested within nursing homes.

The data exploration revealed that approximately 12% of patients were estimated to have pressure ulcers. The frequency distribution of patients with pressure ulcers showed a positive relationship with levels of sensory perception, activity, dietary intake, friction-and-shear, mobility, and ward specialty. Specifically, as these variables were measured from the most to the least favorable, there was an increase in the proportion of affected patients. The probability of developing pressure ulcers was higher for patients in the least favorable ranks compared to those in the most favorable ranks.

Simple logistic regression analysis was used to test the significance of covariates on pressure ulcers. It was observed that, apart from Ward 9, which had no significant effect on pressure ulcers (p-value = 0.334), all other covariates were statistically significant (p-value < 0.05). However, all covariates were included in the analyzed model because they were biologically relevant for estimating the performance of health care providers in nursing homes.

Multiple logistic regression analyses were performed to correct for case-mix differences among nursing homes. The results showed that activity, mobility, dietary intake, friction-and-shear, Ward 10, and Ward 91 significantly affected pressure ulcer prevalence. Similarly, six nursing homes, which captured the remaining differences in the prevalence of setting-acquired pressure ulcers, significantly affected the response. The beta coefficients for these nursing homes indicated that the probability of patients developing pressure ulcers was above average. These nursing homes were associated with lower care quality, potentially due to the large number of patients admitted. Additionally, it was found that smaller institutions performed worse than larger ones. This may be attributable to statistical reasons, as smaller nursing homes often yielded larger confidence intervals due to fewer patients.

The overall goodness of fit for the model, as shown by the Hosmer-Lemeshow statistic, displayed no evidence of lack of fit (p-value = 0.152), indicating that the selected model fits the data well. Similarly, Pearson's chi-square statistic, based on the independent variables, provided a p-value of 0.927, further supporting the model's goodness of fit.

Cross-validation was conducted to predict the presence of pressure ulcers in nursing homes and to minimize the total misclassification rates for both the calibration and validation data. The model was deemed appropriate for case-mix adjustment, as the Wald statistic yielded a p-value of 0.72. This indicated no evidence to reject the null hypothesis, confirming that the misclassification rates for the calibration and validation datasets were not significantly different at the 5% significance level. Additionally, the ROC curve was used to assess the model's predictive accuracy and it demonstrated a good discrimination between individuals with and without pressure ulcers.

The performance of nursing homes was also evaluated using the mixed effects model and Empirical Bayes Estimates (EBE). The level 2 random coefficient variance significantly affected pressure ulcer prevalence (p-value = 0.015), suggesting meaningful differences in performance between nursing homes after controlling for case-mix differences. The mixed effects model provided a clear identification of specific nursing homes with lower care quality. The EBE results revealed that three nursing homes were significant for the random coefficients, with positive beta coefficients.

When comparing the fixed and mixed effects models, the mixed effects approach, particularly using the EBE method, appears superior. The results showed that the 95% confidence intervals for the EBE random coefficients were smaller than those in the fixed effects model. Additionally, fewer nursing homes were identified as having lower care quality in the mixed effects model compared to the fixed effects model. The EBE method accounts for the number of patients in each institution thus shrinking the estimates for smaller settings toward the overall mean to better reflect the true underlying performance of the setting. Thus, the estimates obtained from the mixed effects model are superior to those from the fixed effects logistic regression model.

Further investigations into the quality of the EBE method under the heteroscedasticity assumption for level 2 random coefficients could be explored. Additionally, future analyses could employ alternative techniques, such as hierarchical Bayes estimates, for improved comparisons.

References

Agrestic, A. (2018). *An Introduction to Categorical Data Analysis* (3rde.d). Wiley Series in Probability and Statistics.

Ahmed, S., Wasim, S., Irfan, S., Gogoi, S., Srivastava, A., & Farheen, Z. (2019). 'Quantitative Vs Qualitative Research'. *Journal of Evidence Based Medicine and Healthcare, 6*(43), 2828-2832.

Braden, B., & Maklebust, J. (2005). 'Preventing Pressure Ulcers with Braden Scale: An Update on this easy-to-use tool that assesses a patient's risk'. *AJN American journal of Nursing,105*(6), 70-72.

Bours G, Halfens, RJ., Abu-Said,H., & Grol RT, .(2002). *Development of a model for Case-mix adjustment of Pressure Ulcer rates*. Medical Care.

Clark, M., Semple,M., Ivins,N., Mahoney, K., & Harding,K. (2017). 'National Audit of Pressure Ulcers and Incontinence-associated Deramtitis in Hospitals across Wales: A Cross-Sectional study'. *BMJ Open,7*(8), 1-8.

Dassen, T., Tannen, A., Bours, G., & Halfen, R. (2004). 'Acomparison of pressure ulcer prevalence: Concerted data collection in the Netherlands and German'. *International Journal of Nursing Studies*, *41*(6), 607-612.

Delong, M.D., Delong, E., Peterson, E.D., Muhlbaier, L.H., Hackett, S., & Mark, D.B. (1997). *Comparing Risk-Adjustment Methods for Provider Profiling*. National Library of Medicine.

Dietz, E., Halfens, R., & Mertens, E. (2008). 'Pressure Ulcer risk screening in hospitals and nursing homes with a general nursing assessment tool: Evaluation of the care dependency scale'. *Journal of Evaluation in Clinical Practice*, *14*(6),1018-1025.

Gao, L., Yang, L., Li, X., Chen, J., Du, J., Bai, X., Yang,X. (2018). 'The Use of Logistic Regression Model to develop a Risk Assessment of Intraoperatively Acquired Pressure Ulcer. *Journal of Clinical Nursing*, *27*(15-16), 2984-2992.

Halfens, R.J.G., Achterberg, T. &Bal, R.M. (2000). 'Validity and reliability of the Braden scale and the influence of other risk factors: A multi-centre prospective'. *International Journal Nursing study, 37*, 313-319.

Karanikolos, M., & McKee, M. (2013). 'Health and health systems in the Commonwealth of Independent States'. *Lancet, 381*(9872), 1145–1155.

McKee, M. & Jacobson, B. (2000). 'The birth of the hospital in the Byzantine Empire'. *Journal of Public health in Europe, Lancet,* 356, pp. 665-770.

Meesterberends, E., Halfens, R., Heinze, C., Lohrmann,C., & Schols, MGA. (2011). 'Pressure ulcer incidence in Dutch and German nursing homes: design of a prospective multicentre cohort study'. *BMC Nursing, 10*(8). Retrieved 7 July 2021 from http://www.biomedcentral.com/1472-55/10/8.

Molenberghs, G. (2011). *Linear Mixed Models*. International Encyclopidia of statistical Science.

Molenberghs, G. &Verkebe, G. (2013). *Generalized Linear Mixed Models*. Springer.

Molenberghs, G. &Verkebe, G. (2010). *Linear Mixed Models for Longitudinal Data*. Spinger.

NICE (National Institute for health and Care Excellence) 2020. *Pressure Ulcers: Prevention and Management*. National Institute for health and Care Excellence clinical guideline.

Rechel B., Kennedy, C., McKee, M., &Bernd,R. (2011). 'The Soviet legacy in diagnosis and treatment: implications for population health'. *Journal of Public Health Policy*, *32*(2),

293–304.
Rechel, B., Robert, B., Richardson, E., Shishkin, S., Shkolnikov, V., Leon, D., Bobak, M,
Rosen, A., Wu, J., Chang, B., Berlowitz, D., Rakovski, C., Ash, A. & Moskowitz, M.(2013). ‘Risk-adjustment for measuring Healthcare Outcomes: Application in VA Long term Care’. *American Journal of Medical Quality*, *16*(4),115-117.
Rosen, J., Mittal,V., Degenhottz,H., Castle,N., Mulsant, B., Nace,D. & Rubin,F. (2006). ‘Pressure Ulcer Prevention in Black and White Nursing Home Residents: A QI Intitative of Enhanced Ability, Incentives and Management Feedback’. *Advances in Skin and Wound Care, 19*(5),262-269.
Thomas, K. & Lobo, B. (2011). *Care Planning in End-of-Life Care*. Oxford: Oxford University Press.
Yu, D. SF., Lee, D. TF., & Kwong, A. NL. (2009).‘Quality of life of older people in residential care home: a literature review’. *Journal of Nursing and Healthcare of Chronic Illness-Wiley online library*, *1*(2),116-125.