**Supplementary Material**

##### **Methods**

**The eight-step approach to developing the MRSA risk score**

#### **Step 1: Bivariate analyses**

Following the Hosmer-Lemeshow guideline, only potential exposure variables with a Chi-square test p-value ≤0.25 were included in the multivariate logistic regression model [[1]](https://paperpile.com/c/onTJdj/fofW). However, potential exposure variables with a p-value greater than 0.25 were retained in the model if they were considered to have high clinical relevance and were easy to measure (**Table 1**). Based on the odds ratio (OR) from the bivariate logistic regression model, the strength of association of each exposure variable was categorised as follows: (i) low for non-significant odds ratios; (ii) intermediate for significant odds ratios between 1.0 and 2.0 (inclusive) or between 0.5 and 1.0 (inclusive); and (iii) high for significant odds ratios above 2.0 or below 0.5. Clinical relevance refers to the degree of relevance of the variable to the colonisation of MRSA in the LTCFs based on **Table 3** from a systematic review published by Rodríguez-Villodres et al. [[2]](https://paperpile.com/c/onTJdj/6TJp).

The “ease of measurement” for variables collected in LTCFs refers to how easily they can be accessed. A measurement is considered high in ease if variables can be collected reliably and accurately with minimal time and effort. For example, sex and age can be found in health records without further investigation. On the other hand, a measurement is considered low in ease if the variable requires significant time and effort to collect or is complex in nature. For instance, education level is often self-reported and prone to recall bias. Intermediate ease of measurement falls between these two extremes, requiring some effort and expertise to collect the data accurately. An example is assessing a resident’s degree of self-care dependence, which requires training but can be done with moderate effort and resources.

#### **Step 2: Backward stepwise elimination for the individual-effects modelling**

Variable selection was performed using a backward elimination technique. This involved including all potential predictor variables in an initial multivariate model and then, one by one, removing the variable with the highest p-value. After each step, the model was re-fitted and the process repeated until only significant variables remained. The threshold for statistical significance was set at a Wald test p-value of 0.05 [[3,4]](https://paperpile.com/c/onTJdj/LC11+YtZR). Akaike’s information criterion (AIC) was employed to select the final model [[4,5]](https://paperpile.com/c/onTJdj/YtZR+wJCF) and the one with the lowest AIC was chosen as the “best” model [[6]](https://paperpile.com/c/onTJdj/5q8i).

The delta AIC, which represents the difference between the “best” model and all other models generated from the data, was also calculated [[7]](https://paperpile.com/c/onTJdj/rRsU). A delta AIC value <2 indicates strong evidence in favour of the candidate model. This means that the candidate model is nearly as good as the best model. If the delta AIC is between 4 and 7, the candidate model has significantly less support. If the delta AIC is 8 or above, there is essentially no support for the candidate model and suggests that the candidate model is unlikely to be the best model.

#### **Step 3: Exploration of interaction terms**

The presence of significant interactions among potential exposure variables was investigated within the final model.

#### **Step 4: Stepwise method to build the mixed-effects model**

To accommodate the hierarchical structure of the data, the individual-effects model was expanded to a mixed-effects model. This approach allows for disentangling the influence of individual variables on the outcome from the impact of facility-level variables [[8]](https://paperpile.com/c/onTJdj/e9af).

To ensure transparency and thorough reporting, the Logical Explanations and Visualisations of Estimates in Linear Mixed Models checklist from Aguinis and colleagues was followed [[9]](https://paperpile.com/c/onTJdj/njGn). This involved reporting each step of the model-building process, including the Wald statistics, standard errors, and variance components [[10]](https://paperpile.com/c/onTJdj/xbWB). The mixed-effects model construction involved four parts: (i) a null model, (ii) a model with random intercept and fixed slope, (iii) a model with random intercept and random slope, and (iv) a model with cross-level interactions. In the final model, adjusted ORs with 95% confidence interval (CI) were presented. For visualising the variation in the mixed-effects model, the *merTools* package in R was used [[38]](https://paperpile.com/c/onTJdj/Tbck5). First, the prediction intervals were explored by generating a distribution of predictions for each resident. Second, the relative magnitude of the parameters in the final model was explored using confidence interval plots. Third, a caterpillar plot for visualising the random-effects terms was created.

#### **Step 5: Exploration of cross-level interaction terms**

In the mixed-effects model, cross-level interaction terms were included to investigate the moderating effect of the group-level variable on the relationship between the individual-level variable and MRSA colonisation.

#### **Step 6: Model validation**

To evaluate the robustness of the model, internal validation was conducted using bootstrap resampling with 400 repetitions. If the model produces highly variable results when applied to different samples, the model may have poor accuracy and may not be robust.

***Bootstrapping method***

Each of the 400 bootstrap samples was tested on the original sample and the performance was measured using the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. To account for optimism bias resulting from overfitting, the difference in performance between the bootstrap samples and the original sample was calculated and the difference across all samples was averaged [[11,12]](https://paperpile.com/c/onTJdj/lFIG). The bias-corrected AUC was obtained by subtracting the estimated optimism bias from the ROC obtained from the original sample.

##### ***Sensitivity Analysis***

Details are described in the main text.

### **Step 7: Model Performance Evaluation**

The performance of the model was evaluated using discrimination, calibration, and decision curve analysis.

##### ***Discrimination***

A 10-fold cross-validation technique was applied to evaluate the final model. This involved testing the model’s performance on each of the ten randomly shuffled folds of the training set, with one fold held out as the test set.

To address the potential classification imbalance between MRSA carriers and non-MRSA carriers, random over-sampling of MRSA carriers was employed on the training dataset within each fold [[13,14]](https://paperpile.com/c/onTJdj/CJND+ZDfl). Oversampling the MRSA carriers allows provision of a more accurate representation of this group, thereby improving the model’s estimation. The performance of the model was compared with and without oversampling based on its ability to distinguish between positive and negative cases using the AUC [[13,14]](https://paperpile.com/c/onTJdj/bJCqC+ZdJsN). This curve plots the true positive rate against the false positive rate at different classification thresholds. A value of 1 indicates perfect discrimination, while a value of 0.5 suggests no discrimination. A model with a value greater than 0.7 is considered useful, while one between 0.8 and 0.9 indicates good diagnostic accuracy [[15]](https://paperpile.com/c/onTJdj/JA4nI). In addition to AUC, the determination of the optimal threshold for clinical use was also considered. This threshold was determined using the Youden index, which maximises the diagnostic accuracy statistics [[16]](https://paperpile.com/c/onTJdj/JYfSm). Furthermore, the performance of the model was evaluated by calculating the balance between sensitivity and positive predictive value (PPV) using estimates of the area under the precision-recall (AUCPR) curve. This curve plots PPV against the sensitivity at different classification thresholds. A high AUCPR value indicates good performance in both sensitivity and PPV. To assess the overall probabilistic prediction of the model, the Brier score, which ranges from zero to one, was used. The Brier score measures the mean squared difference between the predicted probabilities and the actual outcomes. A lower score indicates better calibration of the predicted probabilities by the model [[17]](https://paperpile.com/c/onTJdj/qfPs). These metrics provide a comprehensive assessment of the model performance and can guide further improvements and optimisations in a clinical setting.

##### ***Calibration***

The model's predictive capability for MRSA and non-MRSA carriers was assessed by employing a calibration plot with loess smoothing [[18]](https://paperpile.com/c/onTJdj/n7m5H). This plot visually illustrates the level of agreement between the predicted probability of MRSA carriage and the actual proportion observed within each decile, following the guidelines outlined in the TRIPOD statement [[19,20]](https://paperpile.com/c/onTJdj/ulSi). In addition to the calibration plots, a histogram with a density curve overlain was created to examine the distribution of the predicted MRSA probabilities.

##### ***Decision Curve Analysis***

A finding may have statistical significance indicating that it is unlikely to have occurred by chance but it may not hold practical significance for residents, providers, or other relevant parties. In other words, the results may be statistically significant, but they may not be meaningful in real-world settings. On the other hand, a finding can be clinically but not statistically significant due to factors such as small sample sizes or experimental design manipulations [[21]](https://paperpile.com/c/onTJdj/fJFCk).

In order to determine how useful the risk model is in a clinical setting, a decision curve analysis was used. This involved plotting the expected net benefit of the risk model against various threshold probabilities and comparing it to theoretical competing models. These models assumed either that all residents were colonised with MRSA and treated or that no residents were colonised with MRSA and treated [[22]](https://paperpile.com/c/onTJdj/NiYw). The net benefit was calculated by considering the proportion of true and false positives while weighing the relative harm of a false-positive and false-negative result. The threshold probabilities represent the likelihood of MRSA colonisation at which a clinician or LTCF manager would decide to take action, such as further testing or decolonisation. The optimal treatment threshold probability was determined as the probability of MRSA colonisation at which the expected benefit of screening outweighed the potential harm. Clinicians or administrators can choose the threshold probability that best aligns with their expected accuracy while also considering factors such as the costs and disruptions associated with interventions and the prevalence of MRSA colonisation in the LTCFs [[23]](https://paperpile.com/c/onTJdj/NiYw). Depending on their priorities, clinicians might prefer lower thresholds to avoid missing MRSA colonisation or higher thresholds to avoid unnecessary interventions.

#### **Step 8: Risk Model to Score Conversion**

Details are described in the main text.

**References:**

[1] [Hosmer DW Jr, Lemeshow S, Sturdivant RX. Applied Logistic Regression. John Wiley & Sons; 2013.](http://paperpile.com/b/onTJdj/fofW)

[2] [Rodríguez-Villodres Á, Martín-Gandul C, Peñalva G, Guisado-Gil AB, Crespo-Rivas JC, Pachón-Ibáñez ME, et al. Prevalence and Risk Factors for Multidrug-Resistant Organisms Colonization in Long-Term Care Facilities Around the World: A Review. Antibiotics (Basel) 2021;10. https://doi.org/](file://C:\Users\kkokwok\Dropbox\CUHK\phd_supervision\valerie\paper2\MS\version2\Rodríguez-Villodres%20Á,%20Martín-Gandul%20C,%20Peñalva%20G,%20Guisado-Gil%20AB,%20Crespo-Rivas%20JC,%20Pachón-Ibáñez%20ME,%20et%20al.%20Prevalence%20and%20Risk%20Factors%20for%20Multidrug-Resistant%20Organisms%20Colonization%20in%20Long-Term%20Care%20Facilities%20Around%20the%20World:%20A%20Review.%20Antibiotics%20(Basel)%202021;10.%20https:\doi.org\)[10.3390/antibiotics10060680](http://dx.doi.org/10.3390/antibiotics10060680)[.](http://paperpile.com/b/onTJdj/6TJp)

[3] [Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Fam Med Community Health 2020;8:e000262. https://doi.org/](http://paperpile.com/b/onTJdj/LC11)[10.1136/fmch-2019-000262](http://dx.doi.org/10.1136/fmch-2019-000262)[.](http://paperpile.com/b/onTJdj/LC11)

[4] [Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagn Progn Res 2019;3:16. https://doi.org/](http://paperpile.com/b/onTJdj/YtZR)[10.1186/s41512-019-0060-y](http://dx.doi.org/10.1186/s41512-019-0060-y)[.](http://paperpile.com/b/onTJdj/YtZR)

[5] [Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. Wiley Interdiscip Rev Comput Stat 2019;11:e1460. https://doi.org/](http://paperpile.com/b/onTJdj/wJCF)[10.1002/wics.1460](http://dx.doi.org/10.1002/wics.1460)[.](http://paperpile.com/b/onTJdj/wJCF)

[6] [Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control 1974;19:716–23. https://doi.org/](http://paperpile.com/b/onTJdj/5q8i)[10.1109/tac.1974.1100705](http://dx.doi.org/10.1109/tac.1974.1100705)[.](http://paperpile.com/b/onTJdj/5q8i)

[7] [Frank J. Fabozzi, Sergio M. Focardi, Svetlozar T. Rachev and Bala G. Arshanapalli. Appendix E Model Selection Criterion: AIC and BIC. In: Frank J. Fabozzi, Sergio M. Focardi, Svetlozar T. Rachev, Bala G. Arshanapalli, Markus Hoechstoetter, editor. The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications, John Wiley & Son, Inc.; 2014, p. 399–403.](http://paperpile.com/b/onTJdj/rRsU)

[8] [Blalock HM. Contextual-effects models: Theoretical and methodological issues. Annu Rev Sociol 1984;10:353–72. https://doi.org/](http://paperpile.com/b/onTJdj/e9af)[10.1146/annurev.so.10.080184.002033](http://dx.doi.org/10.1146/annurev.so.10.080184.002033)[.](http://paperpile.com/b/onTJdj/e9af)

[9] [Aguinis H, Gottfredson RK, Culpepper SA. Best-Practice Recommendations for Estimating Cross-Level Interaction Effects Using Multilevel Modeling. Journal of Management 2013;39:1490–528. https://doi.org/](http://paperpile.com/b/onTJdj/njGn)[10.1177/0149206313478188](http://dx.doi.org/10.1177/0149206313478188)[.](http://paperpile.com/b/onTJdj/njGn)

[10] [Monsalves MJ, Bangdiwala AS, Thabane A, Bangdiwala SI. LEVEL (Logical Explanations & Visualizations of Estimates in Linear mixed models): recommendations for reporting multilevel data and analyses. BMC Med Res Methodol 2020;20:3. https://doi.org/](http://paperpile.com/b/onTJdj/xbWB)[10.1186/s12874-019-0876-8](http://dx.doi.org/10.1186/s12874-019-0876-8)[.](http://paperpile.com/b/onTJdj/xbWB)

[11] [Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;102:148–58. https://doi.org/](http://paperpile.com/b/onTJdj/lFIG)[10.1002/bjs.9736](http://dx.doi.org/10.1002/bjs.9736)[.](http://paperpile.com/b/onTJdj/lFIG)

[12] [Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73. https://doi.org/](http://paperpile.com/b/onTJdj/XJY5)[10.7326/M14-0698](http://dx.doi.org/10.7326/M14-0698)[.](http://paperpile.com/b/onTJdj/XJY5)

[13] [Lunardon N, Menardi G, Torelli N. ROSE: a Package for Binary Imbalanced Learning. The R Journal 2014;6:79. https://doi.org/](file:///C:\Users\kkokwok\Dropbox\CUHK\phd_supervision\valerie\paper2\MS\version2\Lunardon%20N,%20Menardi%20G,%20Torelli%20N.%20ROSE:%20a%20Package%20for%20Binary%20Imbalanced%20Learning.%20The%20R%20Journal%202014;6:79.%20https:\doi.org\)[10.32614/rj-2014-008](http://dx.doi.org/10.32614/rj-2014-008)[.](http://paperpile.com/b/onTJdj/CJND)

[14] [Wongvorachan T, He S, Bulut O. A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining. Information 2023;14:54. https://doi.org/](http://paperpile.com/b/onTJdj/ZDfl)[10.3390/info14010054](http://dx.doi.org/10.3390/info14010054)[.](http://paperpile.com/b/onTJdj/ZDfl)

[15] [Mandrekar JN. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. J Thorac Oncol 2010;5:1315–6. https://doi.org/](http://paperpile.com/b/onTJdj/JA4nI)[10.1097/JTO.0b013e3181ec173d](http://dx.doi.org/10.1097/JTO.0b013e3181ec173d)[.](http://paperpile.com/b/onTJdj/JA4nI)

[16] [Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian J Intern Med 2013;4:627–35.](http://paperpile.com/b/onTJdj/JYfSm)

[17] [Rufibach K. Use of Brier score to assess binary predictions. J Clin Epidemiol 2010;63:938–9; author reply 939. https://doi.org/](http://paperpile.com/b/onTJdj/qfPs)[10.1016/j.jclinepi.2009.11.009](http://dx.doi.org/10.1016/j.jclinepi.2009.11.009)[.](http://paperpile.com/b/onTJdj/qfPs)

[18] [Kattan MW, Gerds TA. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res 2018;2:7. https://doi.org/](http://paperpile.com/b/onTJdj/n7m5H)[10.1186/s41512-018-0029-2](http://dx.doi.org/10.1186/s41512-018-0029-2)[.](http://paperpile.com/b/onTJdj/n7m5H)

[19] [Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. Br J Surg 2015;102:148–58. https://doi.org/](https://mycuhk-my.sharepoint.com/personal/1155136506_link_cuhk_edu_hk/Documents/Personal/Phd%20studies_Val/Paper%202_risk%20score/Manuscript/Manscript/Manuscript/Manuscript/Collins%20GS,%20Reitsma%20JB,%20Altman%20DG,%20Moons%20KGM.%20Transparent%20Reporting%20of%20a%20multivariable%20prediction%20model%20for%20Individual%20Prognosis%20Or%20Diagnosis%20(TRIPOD):%20the%20TRIPOD%20Statement.%20Br%20J%20Surg%202015%3B102:148–58.%20https:/doi.org)[10.1002/bjs.9736](http://dx.doi.org/10.1002/bjs.9736)[.](http://paperpile.com/b/onTJdj/lFIG)

[20] [Moons KGM, Altman DG, Reitsma JB, Ioannidis JPA, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73. https://doi.org/](http://paperpile.com/b/onTJdj/XJY5)[10.7326/M14-0698](http://dx.doi.org/10.7326/M14-0698)[.](http://paperpile.com/b/onTJdj/XJY5)

[21] [Powell LH, Kaufmann PG, Freedland KE. Clinical Significance. In: Powell LH, Freedland KE, Kaufmann PG, editors. Behavioural Clinical Trials for Chronic Diseases: Scientific Foundations, Cham: Springer International Publishing; 2021, p. 97–124. https://doi.org/](http://paperpile.com/b/onTJdj/fJFCk)[10.1007/978-3-030-39330-4\_5](http://dx.doi.org/10.1007/978-3-030-39330-4_5)[.](http://paperpile.com/b/onTJdj/fJFCk)

[22] [Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565–74. https://doi.org/](https://mycuhk-my.sharepoint.com/personal/1155136506_link_cuhk_edu_hk/Documents/Personal/Phd%20studies_Val/Paper%202_risk%20score/Manuscript/Manscript/Manuscript/Manuscript/Vickers%20AJ,%20Elkin%20EB.%20Decision%20curve%20analysis:%20a%20novel%20method%20for%20evaluating%20prediction%20%20%20models.%20Med%20Decis%20Making%202006%3B26:565–74.%20https:/doi.org)[10.1177/0272989X06295361](http://dx.doi.org/10.1177/0272989X06295361)[.](http://paperpile.com/b/onTJdj/NiYw)

[23] [Guo L, Chen H, Wang G, Lyu Z, Feng X, Wei L, et al. Development of a risk score for colorectal cancer in Chinese males: A prospective cohort study. Cancer Med 2020;9:816–23. https://doi.org/](https://mycuhk-my.sharepoint.com/personal/1155136506_link_cuhk_edu_hk/Documents/Personal/Phd%20studies_Val/Paper%202_risk%20score/Manuscript/Manscript/Manuscript/Manuscript/Guo%20L,%20Chen%20H,%20Wang%20G,%20Lyu%20Z,%20Feng%20X,%20Wei%20L,%20et%20al.%20Development%20of%20a%20risk%20score%20for%20colorectal%20cancer%20in%20Chinese%20males:%20A%20prospective%20cohort%20study.%20Cancer%20Med%202020%3B9:816–23.%20https:/doi.org)[10.1002/cam4.2734](http://dx.doi.org/10.1002/cam4.2734)[.](http://paperpile.com/b/onTJdj/iR9W)