

# Predicting Patients at High Risk of Hospital-Acquired Conditions Using the LOGISTIC Procedure

Jennifer Hargrove, Pamela Hipp, Glenn Sidle, SAS Institute Inc.

## ABSTRACT

Hospital-acquired conditions (HACs) are health-care complications that occur during inpatient hospital stays. Given that HACs can potentially be avoided if hospitals follow standard clinical evidence-based guidelines, Medicaid and Medicare reduce reimbursement for these conditions. This paper demonstrates how the LOGISTIC procedure can identify patients at high risk of developing HACs during an inpatient stay. PROC LOGISTIC predicts the probability of an outcome given a set of covariates by using logistic regression models and maximum likelihood estimation. Features in the procedure enable users to select covariates associated with the highest probability of having an event. This paper demonstrates how to use PROC LOGISTIC to identify patients at an elevated risk of HACs based on patients' prior medical history. Steps include: 1) identifying patient cohorts (training and test cohorts); 2) defining the HAC outcome event; 3) selecting covariates associated with developing an HAC; 4) identifying the best fit model in the training cohort; 5) examining model fit in the test cohort; and 6) assigning patients an HAC risk score. Results can be used to flag patients at high risk of developing HACs and to improve a hospital's overall health-care quality. Risk scores applied to electronic medical data can be used to prevent HACs in real time upon patient admission.

## INTRODUCTION

Hospital-acquired conditions (HACs) are medical diagnoses that are not initially present upon hospital admission (Centers for Medicare and Medicaid Services, 2018a). Common types of HACs include adverse drug events, pressure ulcers, falls, and hospital associated infections (for example, catheter-associated urinary tract infections, clostridium difficile infections, and surgical site infections) (AHRQ, 2018). In the United States, an estimated 1 in 25 hospital patients are diagnosed with a hospital acquired infection every year (Centers for Disease Control and Prevention, 2016). HACs are associated with prolonged hospital stays, high medical costs (Zimlichman et al., 2013), and increased risk of patient mortality (Umscheid et al., 2011). Since these conditions are potentially avoidable if evidence-based guidelines for clinical care are followed, the Centers for Medicare and Medicaid Services (CMS) use HACs as a measure of **a hospital's** quality of care (Centers for Medicare and Medicaid Services, 2018a; Umscheid et al., 2011). To stress the importance of preventing HACs, CMS limits hospital reimbursements for HACs and mandates that hospitals report whether a diagnosis is present when patients are first admitted to a hospital (Centers for Medicare and Medicaid Services, 2018a).

Logistic regression models are often used in epidemiologic studies to examine whether an exposure event impacts the likelihood of a binary outcome. Given that HACs are a binary event (the patient did or did not acquire an HAC), logistic regression models can be used to predict whether a patient is at increased risk before the event occurs. Features of the SAS<sup>®</sup> procedure allow users flexibility when entering risk factors that are associated with the likelihood of having both the exposure and the outcome event. The output of the model identifies which risk factors are associated with the highest likelihood that an HAC will occur, given admission to a hospital.

If we can identify patients at high risk of developing HACs, then we can potentially prevent these conditions from occurring when a patient is admitted to a hospital. This paper demonstrates how to use the LOGISTIC procedure to identify risk factors that place patients at elevated risk of developing HACs based on their prior medical history.

## IDENTIFICATION OF MODEL CRITERIA

### POPULATION OF INTEREST

The first step in building a logistic regression model is defining your patient population. The population depends on the question that you are trying to answer. In our example, we are trying to answer the question “Which patients are at elevated risk of developing an HAC?” Therefore, our patient population would include patients who had at least one hospital stay during a prespecified time. In some settings, specific inclusion or exclusion criteria are needed to limit bias in your model.

Inclusion criteria are rules required for a patient to be eligible for your patient population. In our HAC example, inclusion criteria would require that patients had at least one hospital stay during a prespecified time. If patients are admitted multiple times during a specified period, the inclusion criteria might limit the initial logistic regression **model to a patient’s initial hospital stay**. Once the final model is defined and validated, the model could be run **upon a patient’s admittance to assess the risk of developing an HAC** during that stay. Other inclusion criteria might depend on your specific patient population. For example, if you are using insurance claims data, you might require that patients remain continuously enrolled for a minimum time prior to the initial hospital stay. Requiring continuous enrollment will help ensure that you are not missing any prior diagnosis or health events that could impact **a patient’s** risk of developing an HAC.

Exclusion criteria are rules that prevent a patient from being eligible for your patient population. For example, you might want to exclude patients that were in the hospital less than three days to avoid mislabeling outside conditions as hospital acquired. Conditions that present on the first or second day of a hospital stay might be conditions that were acquired prior to admission. In addition, you might want to exclude patients in certain types of facilities, such as nursing facilities or mental health facilities, since these patients might have different underlying reasons for their hospital stays. See Figure 1 for an example of a potential patient population flow chart for an HAC model.

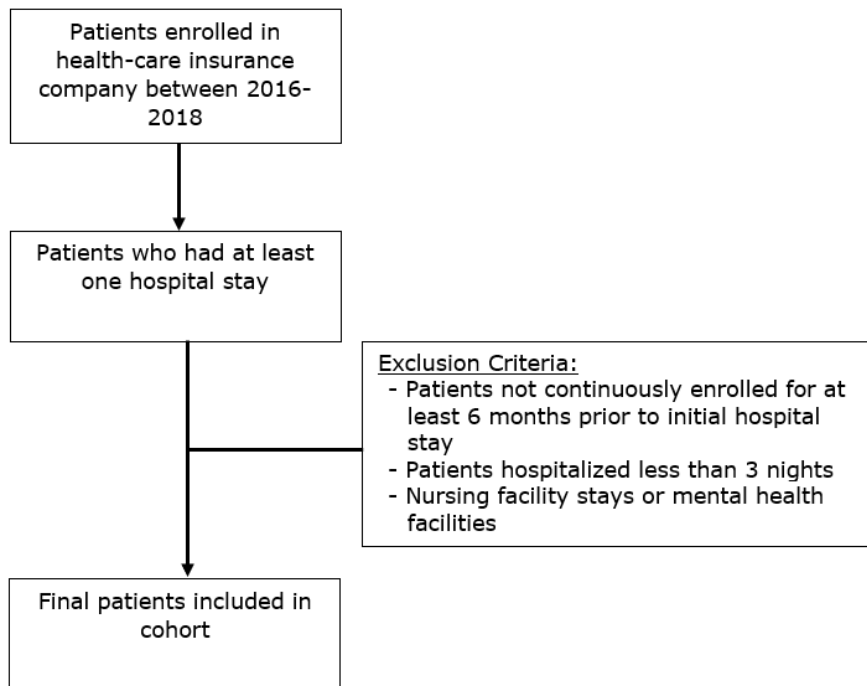


Figure 1. Example Flow Chart for Defining a Patient Population for an HAC Model

## OUTCOME SELECTION

Once you have identified your population, the next step is to identify how you want to define an HAC event. For logistic regression models, the outcome variable must be defined as a binary outcome (HAC versus no HAC). The choice of which HAC to include depends on how frequently the outcome occurs in your population. Logistic regression models can potentially underestimate the probability of an HAC if the event is rare in your population (King and Zeng, 2001; Pavlou et al., 2015). Therefore, you might want to start by building a model that predicts patients at risk of developing the most frequent HAC for your specific population. In health-care settings, International Classification of Disease Clinical Modification (ICD-9-CM and ICD-10-CM) codes are often used to identify health-care outcomes. When selecting your HAC definition, it is important to choose a definition associated with a high predictive value to avoid misclassifying patients and introducing information bias. Information bias can result in underestimated or overestimated effects, depending on the magnitude and whether the bias is the same for patients with and without HACs (Delgado-Rodriguez, 2004). CMS maintains updated HAC definitions on its website and are a good place to start when defining your outcome (Centers for Medicare and Medicaid Services, 2018b).

## COVARIATE SELECTION

The last step in preparing your model is defining your covariates, or individual patient risk factors, associated with having both a hospital stay and acquiring an HAC. When defining covariates, it is important to measure covariates prior to the exposure event. In the field of Epidemiology, the period used to measure covariates is referred to as the *lookback period*. The length of the lookback period depends on the amount of data that you have available and the specific goal of the model. Long lookback periods will help ensure that you are capturing a complete picture of a patient's medical history, but long periods also might limit

your population if you require continuous enrollment as an inclusion criterion. Preen et al. (2006) found that longer lookback periods (longer than one year) were associated with better model fit compared to shorter lookback periods when modeling hospital re-admission events. To examine the impact of the lookback period on your results, you might want to select two different lookback periods and examine the effect on the results depending on the period selected. See Figure 2 for an example of the timeline of events associated with the initial model build.

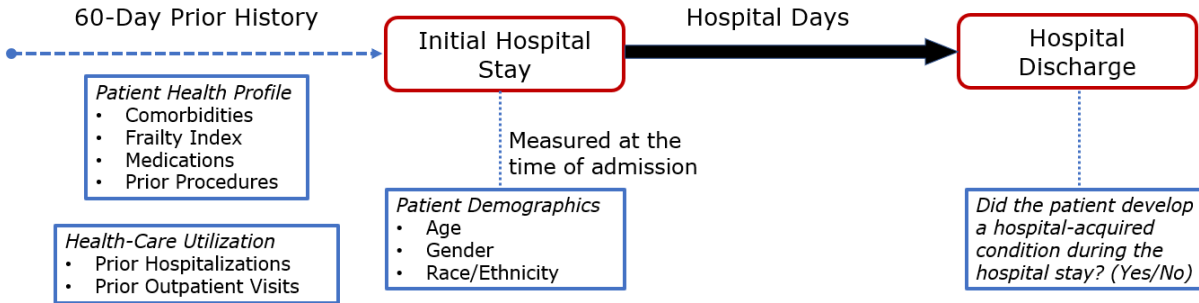


Figure 2. Conceptual Framework of Study Design

Covariates can be defined as continuous, categorical, or as an interaction between two different covariates. In health-care settings, covariates often consist of patient demographics, health-care utilization rates, and comorbidities (based on ICD-CM diagnostic codes or procedure codes). Covariate selection depends on your outcome and the quality of the data. For example, if you want to prevent catheter-associated urinary tract infections, you might want to include history of prior catheterization as one of the covariates in the model. In addition, if you know that a given field is not well captured in your data, you might want to exclude it from your model. See Table 1 for a list of the covariates used in the following example HAC model.

Table 1. Outcome and Covariate Definitions in the HAC Example Model

Outcome	Definition	Type	Value
HAC	Hospital-Acquired Condition	Categorical	0 = No HAC, 1 = HAC
Covariate	Definition	Type	Value
Gender	Patient Gender	Categorical	1 = Male, 0 = Female
AgeCat1	Patient Age Category	Categorical	Patients between the ages of 0–18 (reference)
AgeCat2	Patient Age Category	Categorical	Patients between the ages of 19–35
AgeCat3	Patient Age Category	Categorical	Patients between the ages of 36–64
AgeCat4	Patient Age Category	Categorical	Patients 65 years and older
PriorHosp	Prior Hospitalizations	Categorical	1 = Prior hospitalization in the last 60 days, 0 = No prior hospitalizations in the last 60 days

Comorbid <sup>a</sup>	Number of Comorbidities	Numeric	0–30
PriorOutPt	Prior Outpatient Visits	Categorical	1 = Prior outpatient visit in the last 60 days, 0 = No prior outpatient visit in the last 60 days
FrailtyIndex <sup>b</sup>	Frailty Index Score	Numeric	0–10
<sup>a</sup> Comorbidities are a measure of the number of chronic diseases diagnosed during the lookback period (for example, the Charlson Index) (Valderas et al., 2009).			
<sup>b</sup> <b>Frailty is defined by one’s ability to complete activities of daily living.</b> For an example of a frailty index using health-care claims data, see the article by Faurot et al. (2015).			

## BUILDING THE MODEL

### OVERVIEW OF PROC LOGISTIC AND THE HAC MODEL

PROC LOGISTIC is a SAS procedure that models the probability of a binary outcome given a set of covariates using logistic regression models and maximum likelihood estimation (SAS Institute Inc., 2018). The procedure identifies which covariates are associated with highest probability of the outcome. PROC LOGISTIC offers a wide variety of functions, but the following example shows how you can use the procedure to predict the probability that a patient develops an HAC during a hospital stay. Refer to Table 1 for an explanation of the data elements included in the HAC model. The following demonstrates the equation modeled using PROC LOGISTIC:

$$\text{Logit } P(\text{HAC}=1 | X) = \alpha + \beta_1\text{GENDER} + \beta_2\text{PRIORHOSP} + \beta_3\text{COMORBID} + \beta_4\text{PRIOROUTPT} + \beta_5\text{FRAILTYINDEX} + \sum (\beta_2\text{AGECAT} - \beta_4\text{AGECAT})$$

### DATA PREPARATION

By default, PROC LOGISTIC assumes that you are modeling the probability that an outcome event has not occurred (for example, the probability that HAC = 0). To override this function, you can use the DESCENDING statement to tell SAS to model the probability that the outcome has occurred (for example, the probability that HAC = 1). If one of your covariates is categorical with more than two levels, consider creating dummy variables to separate out the effects of these categorical differences and thus avoid the assumption of an ordinal relationship. The formula for calculating the number of dummy variables to include is *(number of response levels) – 1* (Polissar and Diehr, 1982). For example, if your covariate has four response levels, you would create three dummy variables. The reference response level would not have a dummy variable. Often the category with the largest number of observations is used as the referent level (Polissar and Diehr, 1982). If you want to avoid creating dummy variables, the CLASS and PARAM=REF statements can be used to set the reference criteria for a categorical variable. The choice between using the CLASS statement versus creating dummy variables depends on how your data is structured. If your data is already separated into dummy variables, then you can leave the data as is when entering it into the model. However, if categorical variables are present you can use the CLASS statement to avoid having to create new variables.

Prior to building your model it is important to examine the frequency distribution of the outcome and covariates in your population. If the outcome that you are modeling is rare, you can use the EXACT or FIRTH option in PROC LOGISTIC to help minimize potential bias in

the results (Heinze, G. 2006; SAS Institute Inc., 2018). If patients are missing data for any covariates included in the model, these patients will be excluded from the logistic model. If you believe that the data is not missing at random (for example, there is a difference between patients who have complete data and those who have missing values), failure to account for missing data can result in biased results (Sterne et al., 2009). If you find that data is missing for some of your covariates, multiple imputation strategies exist for computing missing values (Sterne et al., 2009).

Lastly, to validate your final model, it is necessary to divide the data into two distinct samples: training and testing. Separating your cohort into training and testing samples will allow you to test your final model in two distinct unbiased populations. Commonly, a 75:25 ratio is used to split the data, with 75% of the data assigned to the training sample and the other 25% assigned to the testing sample (Giancristofaro and Salmaso, 2007). The testing data is used to validate whether the model identified in the training cohort can accurately predict patients who acquire an HAC in an independent population. Sample code for randomly dividing data into training and testing samples is provided below:

```
DATA temp;
    SET HAC_data;
    N=RANUNI();
RUN;
PROC SORT DATA=temp;
    BY N;
RUN;
DATA training testing;
    SET temp NOBS=NOBS;
    IF _N_<=.75*NOBS THEN OUTPUT training;
    ELSE OUTPUT testing;
RUN;
```

## OUTPUT RESULTS

The output of PROC LOGISTIC can be used to determine if a covariate has a significant association with the outcome event. PROC LOGISTIC provides p-values, odds ratios, and 95% confidence intervals for each covariate included in the model. These three measures can be used to determine whether a given covariate is significantly associated with the outcome. P-values represent the probability that the observed association between a covariate and the outcome is false. A small p-value (typically anything <0.05) indicates a small probability of a false association, so it is reasonable to conclude that the association between covariate and outcome is a true association. However, p-values are subject to random error and fail to provide information related to the precision of the estimate (Poole, 2001). Therefore, it is important to also consider 95% confidence intervals in addition to p-values when examining the results. The width of confidence intervals (ratio between the upper and lower confidence interval) can be used to determine the precision of the association. Smaller confidence intervals indicate a higher level of precision (Poole, 2001). Lastly, odds ratios are useful for describing the magnitude and direction of the association between covariates and the outcome. An odds ratio of 1 indicates that the odds of having an HAC are null (for example, there is no difference in the odds of the event whether a given risk factor is present). An odds ratio greater than 1 denotes higher odds of the outcome, while less than 1 indicates lower odds, all things being equal. The EXPB or ODDSRATIO

options in PROC LOGISTIC can be used to tell SAS to output the odds ratios from the final model. See Table 2 for an example of PROC LOGISTIC output and how you would interpret the odds ratio results.

Table 2. Interpretation of Logistic Regression Results

Covariate	Odds Ratio	95% Confidence Interval	Confidence Interval Ratio <sup>a</sup>	Pr > Chisq <sup>b</sup>	Interpretation of Results
Gender (1 versus 0)	3.014	2.732, 4.885	1.8	0.0023	Males have 3 times the odds of having an HAC event compared to females.
AgeCat1 (1 versus 0)	reference	reference			reference
AgeCat2 (1 versus 0)	1.110	0.893, 5.476	6.13	0.9832	There is no difference in the odds for having an HAC for patients between the ages of 19-35 compared to patients between the ages of 0-18.
AgeCat3 (1 versus 0)	0.985	0.752, 3.001	3.99	0.7632	There is no difference in the odds for having an HAC for patients between the ages of 36-64 compared to patients between the ages of 0-18.
AgeCat4 (1 versus 0)	2.189	1.980, 2.236	1.13	0.0012	Patients over the age of 65 have 2 times the odds of having an HAC event compared to patients between the ages of 0-18.
PriorHosp (ref=0)	6.134	0.012, 1.902	159	0.0678	Patients who have at least one hospital visit in the last 60 days have 6 times the odds of having an HAC compared to patients without prior hospitalizations in the last 60 days.
Comorbid <sup>c</sup> (ref=0)	4.698	3.276, 5.001	1.53	0.0200	Each comorbidity in the last 60 days was associated with a 370% increase in the odds of having an HAC compared to patients who did not have a history of any comorbidities.
PriorOutPt (ref=0)	2.987	1.236, 5.098	4.12	0.7891	Patients who have at least one outpatient visit in the last 60 days have nearly 3

					times the odds of having an HAC compared to patients without prior outpatient visits.
FrailtyIndex (ref=0)	1.083	0.341, 0.659	1.93	0.5612	There is no association between history of being frail and having an HAC event.
<p><sup>a</sup>The confidence interval ratio is not provided in the PROC LOGISTIC output. It was calculated by dividing the upper confidence interval by the lower confidence interval.</p> <p><sup>b</sup> P-value</p> <p><sup>c</sup> Since comorbidities is a continuous variable the odds is associated with a one-unit increase in the number of comorbidities.</p>					

Based on the results in Table 2, the following covariates would be considered as having a statistically significant association with having an HAC: Gender (Male), AgeCat4 (over the age of 65), and Comorbid (history of comorbidity conditions). Prior hospitalizations appeared to be close to the statistically significant cut-off of 0.05; however, the confidence interval ratio tells you that the association is not precise.

**BUILDING THE INITIAL MODEL USING TRAINING DATA**

Now that you have identified the model criteria, it is time to build the model. PROC LOGISTIC provides you the option to manually add covariates, or you can use one of the following automated covariate selection tools: stepwise selection, backward selection, or forward selection. If you choose to use one of the three automated approaches, PROC LOGISTIC provides you the option of specifying the p-value that is used to define significance. The function SLENTRY allows you to specify the p-value for determining whether a covariate gets added to the model, and the function SLSTAY allows you to specify the p-value for determining whether a covariate remains in the model (SAS Institute Inc., 2018). By default, the cut-off criteria for determining significance is 0.05, but depending on your sample size you might want to adjust this level. If your sample size is large, you might want to consider lowering the significance level below 0.05 to reduce the likelihood that the association is due to chance (Sullivan and Feinn, 2012).

In our example, stepwise selection was used to identify significant covariates associated with developing an HAC in the testing model. The stepwise selection process works by adding covariates individually to the model and examining whether the addition significantly improves model fit based on the p-value assigned in the SLENTRY criteria. The difference between stepwise selection and forward selection is that stepwise selection allows fields to be removed once they are added to the model if the effect is no longer significant (SAS Institute Inc., 2018). The result is a final model that only includes covariates significantly associated with the outcome. If you are interested in seeing the steps used to build the final model, the DETAILS function can be used to show the selection process. To verify the fit of the final model, the area under the curve (AUC) value, or C-statistic, can help you determine if the model has a high probability for distinguishing between patients who had an HAC event. Higher C-statistics indicate better discrimination. A C-statistic value of 1 indicates that a model can perfectly predict which patients had an HAC (Giancristofaro and Salmaso, 2007). The AUC plot can be obtained from PROC LOGISTIC using ODS Graphics and the following option: PLOTS(ONLY)=(ROC(ID=OBS)). Hosmer-Lemeshow chi-square values can be used as a measure of bias to examine if model-predicted events match the



observed events (Giancristofaro and Salmaso, 2007). The LACKFIT option tells SAS to generate the Hosmer-Lemeshow test results. Higher p-values indicate that the model is a good fit (for example, expected values are close to the observed values).

## OPERATIONALIZING THE FINAL MODEL

Once you identify the final model in the training data, the last step is to validate the **model's** out of sample accuracy in the testing cohort. The SCORE statement retains the input data and the predicted probabilities associated with the training model and applies them to the testing model. The OUTPUT statement saves the fitted model of the training data, so it can be applied to the testing data. To assess the accuracy of the trained model on the test data, examine how well the trained model correctly predicts whether someone had an HAC event using the output from the SCORE statement. If the model was a good fit, the model will have a high probability of correctly assisinging a patient a HAC status.

The following is example code for building and validating the model using stepwise selection:

```
ODS GRAPHICS ON;
PROC LOGISTIC DATA=Training PLOTS(ONLY)=(ROC(ID=OBS)) DESCENDING;
MODEL HAC = Gender Agecat2 Agecat3 Agecat4 Priorhosp Comorbid Prioroutpt
          Frailtyindex /
SELECTION=STEPWISE
DETAILS
LACKFIT;
OUTPUT OUT=TrainedModel PREDPROBS=INDIVIDUAL;
SCORE DATA=Testing OUT=TestModel;
RUN;
ODS GRAPHICS OFF;
```

## ASSIGNING RISK SCORES

Once the final model has been validated in the testing cohort, the intercept and beta coefficients from this final model can be used to calculate patient risk scores for developing an HAC. **The formula for calculating a patient's risk score** is described below (Pavlou et al., 2015):

**Patient risk score**=  $\alpha$  (model intercept) +  $\beta_1$ \*(patient factor) +  $\beta_2$ \*(patient factor) +  $\beta_3$ \*(patient factor) + etc.

$$\text{Risk of having the outcome} = \frac{\exp(\text{patient risk score})}{(1+\exp(\text{patient risk score}))} * 100$$

Table 3: Example Results from Final Model

Field	Beta Coefficient
Model Intercept	-3.215
Gender	1.356
AgeCat4	0.875

Comorbid	0.762
----------	-------

For example, based on the beta coefficients from the example HAC model in Table 3, a patient's risk score would be defined as follows:  $(-3.215) + 1.356 * (\text{patient gender, 1 or 0}) + 0.875 * (\text{AgeCat4, 1 or 0}) + 0.762 * (\text{number of comorbidities})$ . Since we used females as the reference in the model, female patients would be assigned a value of 0, and male patients would be assigned a value of 1. Similarly, patients over the age of 65 would be assigned a value of 1, and since comorbidities was a continuous variable, values for comorbidity could range 0-30.

Patients estimated to have a high probability of developing HAC events will have values closer to 100%, and patients with values closer to zero have a small probability of developing an HAC.

## CONCLUSION

Preventing HACs can help improve patient quality of care, reduce hospital costs, and **ultimately increase patients' recovery time during hospital stays. Once the initial model is developed and validated, the HAC risk score can be considered as a tool for supplementing clinical decisions in hospital settings. Embedding the risk score algorithm into hospital admitting systems can provide information at the time of admission on the patient's risk for developing a HAC during the hospital stay, providing data-driven early warning of potentially preventable events. Hospital clinicians and administrators can use the HAC risk scores to facilitate and justify the application of evidence-based prevention strategies. A phased approach – in which pilot programs of the HAC risk score are conducted before full roll-out - is crucial to determining operational utility, feasibility of deployment, and clinician acceptance.**

Risk scores, such as the HAC model, require integrated health data systems to have pre-admission patient data readily available to hospitals. As health data integration continues to expand, it is important to be prepared with valid and reliable analytic approaches to avoid inappropriate or misinformed data mining. The approach described here enables determination of a population-specific HAC risk model to maximize validity for your population of interest and includes a testing strategy to verify and quantify reliability.

## REFERENCES

AHRQ (Agency for Healthcare Research and Quality). 2018. "AHRQ National Scorecard on Hospital-Acquired Conditions Updated Baseline Rates and Preliminary Results 2014-2016." Accessed February 15, 2019. Available at

[https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/natlhacratereport-rebaselining2014-2016\\_0.pdf](https://www.ahrq.gov/sites/default/files/wysiwyg/professionals/quality-patient-safety/pfp/natlhacratereport-rebaselining2014-2016_0.pdf).

Centers for Disease Control and Prevention. 2016. "CDC Winnable Battles Final Report." Accessed February 15, 2019. Available at

<https://www.cdc.gov/winnablebattles/report/docs/winnable-battles-final-report.pdf>.

Centers for Medicare and Medicaid Service. 2018a. "Frequently Asked Questions: Deficit Reduction Act (DRA) Hospital-Acquired Condition (HAC) Reporting." Accessed February 15, 2019. Available at <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/Downloads/FAQ-DRA-HAC-PSI.pdf>.

Centers for Medicare and Medicaid Services. 2018b. "ICD-10 HAC List." Accessed December 12, 2018. Available at [https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/icd10\\_hacs.html](https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/icd10_hacs.html).

- Delgado-Rodriguez, M., and J. Llorca. 2004. "Bias." *Journal of Epidemiology & Community Health* 58(8): 635-641.
- Faurot, K. R., M. Jonsson Funk, V. Pate, M. A. Brookhart, A. Patrick, L. C. Hanson, ... and T. Stürmer. 2015. "Using Claims Data to Predict Dependency in Activities of Daily Living as a Proxy for Frailty." *Pharmacoepidemiology and Drug Safety* 24(1): 59-66.
- Giancristofaro, R. A., and L. **Salmaso**. 2007. "Model Performance Analysis and Model Validation in Logistic Regression." *Statistica* 63(2): 375-396.
- Heinze, G. (2006). "A comparative investigation of methods for logistic regression with separated or nearly separated data." *Statistics in medicine* 25(24), 4216-4226.
- King, G., and L. Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9(2): 137-163.
- Pavlou, M., G. Ambler, S. R. Seaman, O. Guttman, P. Elliott, M. King, and R. Z. Omar. 2015. "How to Develop a More Accurate Risk Prediction Model When There Are Few Events." *BMJ* 351: h3868.
- Polissar, L., and P. Diehr. 1982. "Regression Analysis in Health Services Research: The Use of Dummy Variables." *Medical Care*, 959-966.
- Poole, C.** 2001. "Low P-Values or Narrow Confidence Intervals: Which Are More Durable?" *Epidemiology* 12(3): 291-294.
- Preen, D. B., J. H. CD'Arcy, K. Spilsbury, J. B. Semmens, and K. J. **Brameld**. 2006. "Length of Comorbidity Lookback Period Affected Regression Model Performance of Administrative Health Data." *Journal of Clinical Epidemiology* 59(9): 940-946.
- SAS Institute Inc. 2018. "The Logistic Procedure." In *SAS/STAT® 9.3 User's Guide*. Cary, NC: SAS Institute Inc. Available at [https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug\\_logistic\\_sect036.htm](https://support.sas.com/documentation/cdl/en/statug/63962/HTML/default/viewer.htm#statug_logistic_sect036.htm) (accessed December 27, 2018).
- Sterne, J. A., I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, ... and J. R. **Carpenter**. 2009. "Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls." *BMJ* 338: b2393.
- Sullivan, G. M., and R. **Feinn, R.** 2012. "Using Effect Size—or Why the P Value Is Not Enough." *Journal of Graduate Medical Education* 4(3): 279-282.
- Umscheid, C., M. Mitchell, J. Doshi, R. Agarwal, K. Williams, and P. Brennan. 2011. "Estimating the Proportion of Healthcare-Associated Infections That Are Reasonably Preventable and the Related Mortality and Costs." *Infection Control & Hospital Epidemiology* 32(2): 101-114. Doi: 10.1086/657912.
- Valderas, J. M., B. Starfield, B. Sibbald, C. Salisbury, and M. Roland. 2009. "Defining Comorbidity: Implications for Understanding Health and Health Services." *Annals of Family Medicine* 7(4): 357-363.
- Zimlichman, E., D. Henderson, O. Tamir, C. Franz, P. Song, C. K. Yamin, ... and D. W. **Bates**. 2013. "Health Care-Associated Infections: A Meta-Analysis of Costs and Financial Impact on the US Health Care System." *JAMA Internal Medicine* 173(22): 2039-2046.

## RECOMMENDED READING

- *Base SAS® Procedures Guide*
- *Logistic Regression Using SAS, 2nd Edition (Paul D. Allison)*

- *Logistic Regression: A Self-Learning Text (David G. Kleinbaum and Mitchel Klein)*

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Jennifer Hargrove

[Jennifer.Hargrove@sas.com](mailto:Jennifer.Hargrove@sas.com)

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.