Technical Notes on the Standardized Hospitalization Ratio (SHR)

For the Dialysis Facility Reports

July 2025

Contents

1   Introduction

The Standardized Hospitalization Ratios (SHR) in Table 4 of the Dialysis Facility Reports (DFR) are designed to reflect the number of hospitalization 'events' for the patients at a facility, relative to the number of hospitalization events that would be expected based on overall national rates and the characteristics of the patients at that facility. Numerically, the SHR is calculated as the ratio of two numbers: the numerator ("observed") is the actual number of hospitalization events for the patients in a facility over a specified time period, and the denominator ("expected") is the number of hospitalization events that would have been expected for the same patients if they were in a facility conforming to the national norm. As is evident from the preceding description, the SHR represents the hospitalization analog of the Standardized Mortality Ratio (SMR) used to quantify a facility's mortality experience relative to the national average.

Two types of hospitalization outcome data are used when computing the SHR for the DFR:

  i the number of hospital admissions

  ii the number of days hospitalized

The first provides a summary of hospital admissions within a period, which represents the incidence of hospitalization. The second provides a summary of the total days hospitalized within a period; this gives a picture of the total utilization of hospital resources and relates to the overall prevalence of hospitalization. We use number of hospital admissions for illustrative purposes throughout, though number of days in hospital could also have been accommodated in the same way.

To compute a facility's "expected" count, we utilize a regression model that contains outcome-related factors or covariates, such as age, sex, diabetes, duration of ESRD, nursing home status and body mass index (BMI).

The degree to which the facility's SHR varies from 1.00 measures the performance of the facility in reducing overall hospitalizations. For example, a facility's SHR=1.30 indicates that the facility's covariate-adjusted hospitalization rate exceeds the national hospitalization rate by 30%, e.g. 260 observed hospital admissions versus 200 expected. Similarly, an SHR=0.90 would indicate that the facility's hospitalization rate is 10% below the national hospitalization rates (e.g., 180 hospital admissions observed versus 200 expected). An SHR=1.00 would indicate that the facility's overall hospitalization rate equals the national hospitalization rate.

In the DFR, we also report SHRs for a given region. A region's SHR is calculated as the ratio of the total number of observed hospital admissions among patients from that region, to the covariate-adjusted expected number of hospital admissions for that region's patients. The regional SHRs are provided for comparison purposes, so that each facility's SHR can be compared to the SHR for the region in which it is located.

The document is organized as follows. Section 2 gives required administrative details. In Section 3 and 4, we detail a two-stage modeling approach for computing the expected hospitalizations given patients' characteristics. Section 5 gives a detailed description for computing p-values and confidence intervals, which accounts for possible uncertainty or random variations that may be beyond the control of the facilities. We give some final remarks in Section 6.

2   Assignment of Patients to Facilities for the SHR Calculation

As patients can receive dialysis treatment at more than one facility in a given year, we assign each patient day to a facility (or no facility, in some cases) based on a set of conventions below, which largely align with those for the Standardized Mortality Ratio (SMR). We detail patient inclusion criteria, facility assignment and how to count days at risk, all of which are required for the risk adjustment model.

2.1   General Inclusion Criteria for Dialysis Patients

Because a patient's follow-up in the database can be incomplete during the first 90 days of ESRD therapy, we only include a patient's follow-up into the tabulations after that patient has received chronic renal replacement therapy for at least 90 days. Thus, hospitalizations, mortality and survival during the first 90 days of ESRD do not enter into the calculations. This minimum 90-day period also assures that most patients are eligible for Medicare insurance, either as their primary or secondary insurer. It also excludes from analysis patients who die or recover during the first 90 days of ESRD.

We assigned patients to a facility only after they had been on dialysis there for at least 60 days. This 60 day period is used both for patients who started ESRD for the first time and for those who returned to dialysis after a transplant. That is, hospitalizations during the first 60 days of dialysis at a facility do not affect the SHR of that facility.

2.2   Identifying Facility Treatment Histories for Each Patient

For each patient, we identify the dialysis provider at each point in time. Starting with day 91 after onset of ESRD, we attribute patients to facilities according to the following rules. A patient is attributed to a facility once the patient has been treated there for 60 days. When a patient transfers from one facility to another, the patient continues to be attributed to the original facility for 60 days and then is attributed to the destination facility. In particular, a patient is attributed to their current facility on day 91 of ESRD if that facility had treated him or her for at least 60 days. If on day 91, the facility had treated a patient for fewer than 60 days, we wait until the patient reaches day 60 of treatment at that facility before attributing the patient to the facility. When a patient is not treated in a single facility for a span of 60 days (for instance, if there were two switches within 60 days of each other), we do not attribute that patient to any facility. Patients were removed from facilities three days prior to transplant in order to exclude the transplant hospitalization. Patients who withdrew from dialysis or recovered renal function remained assigned to their treatment facility for 60 days after withdrawal or recovery.

If a period of one year passed with neither dialysis claims nor SIMS information to indicate that a patient was receiving dialysis treatment, we considered the patient lost to follow-up and did not continue that patient in the analysis. When dialysis claims or other evidence of dialysis reappeared, the patient was assigned to a particular facility after 60 days of continuous therapy at the facility.

2.3   Days at Risk for Medicare Dialysis Patients

After patient treatment histories are defined, as described above, periods of follow-up time (or patient-records) are created for each patient. In order to adjust for duration of ESRD appropriately, we define 6 time intervals with cut points at 6 months, 1 year, 2 years, 3 years and 5 years. A new time period begins each time the patient is determined to be at a different facility, at the start of each calendar year or when crossing any of the above duration of ESRD cut points. In the SHR Admissions model only, a new time period also begins at the start of a new COVID-19 diagnosis category, which is described in Section 3.2.

Because we can only identify hospitalizations for the SHR Admissions and SHR Days calculations if they appear in Medicare inpatient claims, we only include patients during time periods in which we are reasonably sure that all of the patient's hospitalizations would be included in Medicare billing records. Therefore, we require that patients reach a certain level of Medicare-paid dialysis bills, have Medicare inpatient claims during the period, or are enrolled in Medicare Advantage during the period. Specifically, a patient-month within a given dialysis patient-period is included in the SHR Admissions or SHR Days calculations when at least one of the following is true:

> (1)   The patient had \$1,200+ of Medicare-paid dialysis claims or at least one Medicare inpatient claim (hospital or SNF) during that month or one of the two prior months.

> (2)   The Medicare Enrollment Database (EDB) indicates the patient was enrolled in Medicare Advantage during the month.

The intention of the criteria is to assure completeness of the relevant hospitalization information for all patients included in the analysis.

The number of days at risk (denoted tik in each of these patient-ESRD-year-facility time periods is used to calculate the expected number of hospitalization events for the patient during that period as described in the "Calculation of Expected Hospitalization at a Facility" section below. The SHR for a facility is the ratio of the total number of observed to the total number of expected hospitalization events during all time periods at the facility.

3   Models for Calculating Expected Hospitalization

The denominators of the SHR-Admissions and SHR-Days stem from the proportional rates model (Lawless and Nadeau, 1995; Lin et al., 2000; Kalbfleisch and Prentice, 2002). This is the recurrent event analog of the well-known proportional hazards or Cox model (Cox, 1972; Kalbfleisch and Prentice, 2002). To accommodate large-scale data, we adopt a model with piecewise constant baseline rates (Cook and Lawless, 2007) and the computational methodology developed in Liu et al. (2010). Model coefficients and baseline survival curves are available for download here:

https://dialysisdata.org/sites/default/files/content/Methodology/SHRModelInfo.xls.

The modeling process has two stages. For SHR-Admissions, the model is run separately for each calendar year reported in the DFR, while for SHR-Days, the model is run once for all four calendar years reported in the DFR. At **stage I**, a stratified model is fitted to the national data with piecewise-constant baseline rates and stratification by facility. Specifically, the model is of the following form

$$Pr(\text{hospital admission on day } t \text{ given covariates } X) = r_{0k}(t)\exp(\boldsymbol{\beta}^1 \mathbf{X}_{ik}), \qquad (1)$$

where $\mathbf{X}_{ik}$ is the vector of covariates for the (i, k)th patient and $\boldsymbol{\beta}$ is the vector of regression coefficients. The baseline rate function $r_{0k}(t)$ is assumed specific to the kth facility, which is assumed to be a step function with break points at 6 months, 1 year, 2 years, 3 years and 5 years since the onset of dialysis. This model allows the baseline hospitalization rates to vary between strata (facilities), but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. The stratification on facilities is important in this phase to avoid bias due to possible confounding between covariates and facility effects. The patient characteristics $\mathbf{X}_{ik}$ included in the

SHR-Admissions and SHR-Days stage I models are age centered at 65 years old, squared age centered at 65 years old, sex (male or female), diabetes as cause of ESRD, duration of ESRD (91 days-6 months, 6 months-1 year, 1-2 years, 2-3 years, 3-5 years, or 5+ years as of the period start date), nursing home status, BMI at incidence, comorbidities at incidence, two-way interaction terms between age, age squared, sex and duration and cause of ESRD, Medicare Advantage coverage, and 91 prevalent comorbidities identified through Medicare inpatient claims. Comorbidities at incidence are included as separate indicators for each comorbidity. In addition to these listed covariates, the SHR-Admissions model also includes an adjustment for COVID-19 diagnosis, which is described in more detail in Section 3.2. Since the SHR-Days model is run for all four years at once, calendar year is also included as a covariate. Nursing home status is included as two categorical terms: one to indicate a long-term stay ($\geq$90 days), and one to indicate a short-term stay (1-89 days) in the previous 365 days from the first day of time period at risk. BMI is included as a categorical term indicating underweight, normal, overweight, or obese according to the World Health Organization criteria (WHO, 2020). Missing BMI is imputed as obese. Medicare Advantage status is included as a continuous term that indicates the proportion of time at risk with Medicare Advantage coverage. Categorical indicator variables are included as covariates to flag records missing values for cause of ESRD, and comorbidities at incidence. These variables have a value of 1 if the patient is missing the corresponding piece of information and a value of 0 otherwise. Another categorical indicator variable included as a covariate in the stage 1 model is a flag to note where a patient does not have any of the select comorbidities at incidence. This variable has a value of 1 if the patient has no comorbidities recorded as present and a value of 0 otherwise. The comorbidities at incidence included in the model are: atherosclerotic heart disease, other cardiac disease, diabetes, congestive heart failure, inability to ambulate, chronic obstructive pulmonary disease, inability to transfer, malignant neoplasm, cancer, peripheral vascular disease, cerebrovascular disease, tobacco use (current smoker), alcohol dependence, and drug dependence.

At **stage II**, the relative risk estimates from the first stage are used to create offsets and an unstratified model is fitted to obtain estimates of an overall baseline rate function. That is, we estimate a common baseline rate of admissions, $r_0(t)$, across all facilities by considering the model

$$Pr(\text{hospital admission on day } t \text{ given covariates } X) = r_0(t)R_{ik}, \tag{2}$$

where $R_{ik} = \exp(\boldsymbol{\beta'}\mathbf{X}_{ik})$ is the estimated relative risk for patient i in facility k estimated from the stage I. In our computation, we assume the baseline to be a step function with s unknown parameters, $\alpha_1, \ldots, \alpha_s$, to estimate. These estimates are used to compute the expected number of admissions given a patient's characteristics. Specifically, let t_{iks} represent the number of days that patient i from facility k is under observation in the sth time interval with estimated rate α_s. The corresponding expected number of hospital admissions in the sth interval for this patient is calculated as

$$E_{iks} = \alpha_s t_{iks} R_{ik}. \tag{3}$$

It should be noted that $\mathbf{t}_{iks}$ and hence $\mathbf{E}_{iks}$ can be 0 if patient i from facility k is never at risk during the sth time interval. Summing the $\mathbf{E}_{iks}$ over all s intervals and all N_k patients in a given facility, k, gives

$$E = \sum_{i=1}^{N} \sum_{s=1}^{s} E_{iks} = \sum_{i=1}^{N} \sum_{s=1}^{s} \alpha_s \, t_{iks} \, R_{ik}, \tag{4}$$

which is the expected number of hospital admissions during follow-up at that facility.

Let **O** be the observed total number of hospital admissions at this facility. The SHR for hospital admissions is the ratio of the observed total admissions to this expected value, or

$$\mathbf{SHR} = \mathbf{O/E}.$$

3.1 Missing Data

Patients with missing data are not excluded from the models. For the purposes of calculation, missing values

for BMI at incidence are is imputed as obese. Missing values for cause of ESRD are replaced with the other/unknown category. No patients were missing age, sex, or date of first ESRD treatment. Indicator variables identifying patients with missing values for cause of ESRD, and comorbidities at incidence are also included as covariates in all three SHR models.

3.2 Adjusting for the COVID-19 Effect in SHR Admissions Model

The COVID-19 pandemic continues to have a profound impact on the US healthcare system including ESRD providers and the high-risk dialysis population since 2020. To assist dialysis surveyors and other stakeholders in investigating the impact of COVID-19 on SHR-Admissions, we have leveraged Medicare claims and EQRS data sources to identify COVID-19 medical events among our ESRD patient population. For the purposes of COVID-19 adjustment in the SHR-Admissions model, patients with a COVID-19 event on February 20, 2020 or later based on inpatient Medicare claims are identified as COVID-19 patients. An inpatient claim record confirms a COVID-19 diagnosis if any COVID-19 diagnosis codes (ICD-10-CM: U071, B9729, J1282) are included as primary or secondary diagnoses. Secondary diagnoses include 2nd through 25th ordered diagnoses.

Using the COVID-19 patient identification explained above, we further defined four distinct stages of the COVID-19 diagnosis: COVID1, COVID2, COVID3 and No COVID. Patients with no evidence of COVID-19 events reported in inpatient claims remain assigned to the No COVID group. For patients with a COVID-19 event, the clock starts at the discharge date of the first COVID-19 inpatient Medicare claim, which is the Index COVID-19 Hospitalization (ICovH), and is then tracked for the following 6 months. The clock is not reset in the case of multiple COVID-19 hospitalizations. The period following the first ICovH is then categorized into three mutually exclusive stages, which are included as co-variables adjusted in SHR-Admissions models for 2020 and beyond:

- COVID1: Patients' time at risk in the first month (days 1-30) after the ICovH discharge date
- COVID2: Patients' time at risk in the second month (days 31-60) after the ICovH discharge date
- COVID3: Patients' time at risk in the 3rd– 6th month (days 61-180) after the ICovH discharge date.

After it has been 6 months since the ICovH discharge, a patient is assigned to the No COVID group, which is used as a reference group in the SHR-Admissions models. Once patients have been reassigned to the No COVID group after advancing through COVID1, COVID2, and COVID3, they remain in this stage for their subsequent time at risk

4   Calculation of SHR P-Values and Confidence Intervals

To overcome the possible over-dispersion of the data, we compute the p-value for our estimates using the empirical null distribution, an approach that possesses more robustness (Efron (2004); Kalbfleisch and Wolfe (2013)). Our algorithm consists of the following concrete steps. First, we fit an over-dispersed Poisson model for the number of hospital admissions

$$\log(E[\mathbf{n}_{ik}]) = \log(\mathbf{E}_{ik}) + \boldsymbol{\theta}_k, \tag{5}$$

where $\mathbf{n}_{ik}$ is the observed number of event for patient i in facility k, $\mathbf{E}_{ik}$ is the expected number of events for patient i in facility k and $\boldsymbol{\theta}_k$ is the facility-specific intercept. Here, i ranges over the number of patients N who are treated in the kth facility. The natural log of the SHR for the kth facility is then given by the corresponding estimate of $\boldsymbol{\theta}_k$. The standard error of $\boldsymbol{\theta}_k$ is obtained from the robust estimate of variance arising from the overdispersed Poisson model.

Second, we obtain a z-score for each facility by dividing the natural log of its SHR by the standard error from the general linear model described above. These z-scores are then grouped into quartiles based on the number of patient years at risk for Medicare patients in each facility. Finally, using robust estimates of location and scale based on the normal curve fitted to the center of the z-scores for the SHR, we derive the mean and variance of a normal empirical null distribution for each quartile. This empirical null distribution is then used to calculate the p-value for a facility's SHR.

4.1   Example

The uncertainty or confidence intervals are obtained by applying the following steps:

- From the general linear model we obtain the natural log of the SHR (ln SHR) as well as its standard error, (SE). From the empirical null, we obtain a mean (μ) and a standard deviation (σ). The 95% uncertainty interval for the 'true' log standardized hospitalization ratio for this facility is

$$\log(\text{SHR}) - \mu \times \text{SE} \pm 1.96 \times \sigma \times \text{SE}. \tag{6}$$

  Note that 1.96 is the critical point from the standard normal distribution for a 95% interval.

- Exponentiating the endpoints of this interval gives the uncertainty interval for the true SHR.

For example, consider a hypothetical facility whose SHR is 0.927 for which ln SHR = -0.076 with corresponding standard error, SE = 0.118. This facility falls in a quartile where the empirical null has μ = -0.143 and σ = 1.479. The corresponding uncertainty interval for the log SHR is

$$-0.076 - (-0.143) \times 0.118 \pm 1.96 \times 0.118 \times 1.479 = (-0.401, \ 0.283).$$

The 95% interval for the true SHR is then 0.67 to 1.33.

4.2   Flagging Rules

For reporting purposes, we identify outlier facilities from amongst those with at least 5 patient-years at risk during the time period. If the 95% interval lies entirely above the value of 1.00 (i.e. both endpoints exceed 1.00), the facility is said to have outcomes that are "worse than expected". On the other hand, if the 95% interval lies entirely below the value 1.00, the facility is said to be "better than expected". If the interval contains the value 1.00, the facility is said to have outcomes that are "as expected".

5   Final Remarks

This document details the computation of SHR, designed to measure the performance of facilities in reducing hospitalizations. Our proposed methods are based on several statistical publications developed by the investigators of UM-KECC and have been tested in various settings. In general, the validity of any standardized measures largely depend on the validity of risk-adjustment models. As proper choices of risk adjusters are vital in the process, our practical principles lie in scientific relevance and caution of sequel. That is, we choose risk adjusters that are scientifically relevant to the outcome, while avoiding choosing those that may be affected by the quality of care.

References

Cook, R. and Lawless, J. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York. See page 65-67.

Cox, D. (1972). Regression models and life tables (with discussion). *J. Royal statistical Society, Series B,* 34:187–220.

Efron, B. (2004). Large scale simultaneous hypothesis testing: the choice of null hypothesis. *J. Amer. Statist. Assoc.,* 99:96–104.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York.

Kalbfleisch, J. and Wolfe, R. (2013). On monitoring outcomes of medical providers. *Statistics in the Biosciences,* 5:286–302.

Lawless, J. and Nadeau, C. (1995). Some simple and robust methods for the analysis of recurrent events. *Technometrics,* 37:355–364.

Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society Series B,* 62:771–730.

Liu, D., Schaubel, D., and Kalbfleisch, J. (2010). Computationally efficient marginal models for clustered recurrent event data, University of Michigan, *Department of Biostatistics Technical Reports*.

World Health Organization (WHO). (2020). Body Mass Index- BMI. http://www.euro.who.int/en/health-topics/disease-prevention/nutrition/a-healthy-lifestyle/body-mass-index-bmi . Accessed May 13, 2020.

Ding, X., He, Z. and Kalbfleisch, J. (2021). Time dependent COVID-19 effects on Hospitalizations of Dialysis Patients - draft