Technical Notes on the Standardized Mortality Ratio (SMR)

For the Dialysis Facility Reports

July 2015
Technical Notes on the Standardized Mortality Ratio
For the Dialysis Facility Reports

Table of Contents

Table of Contents ........................................................................................................ 2
Introduction .................................................................................................................. 3
Assignment of Patients to Facilities for the SMR Calculation .................................... 3
  General Inclusion Criteria for Dialysis Patients ......................................................... 4
  Identifying Facility Treatment Histories for Each Patient .......................................... 4
  Days at Risk for Each Patient-Record ........................................................................ 4
Model for Calculating Expected Mortality .................................................................... 5
  Missing Data ............................................................................................................. 6
Calculation of Expected Deaths at a Facility ............................................................... 6
  Example .................................................................................................................... 8
  Caveats ..................................................................................................................... 10
References .................................................................................................................... 11
Introduction

The Standardized Mortality Ratio (SMR) in Table 1 of the Dialysis Facility Reports (DFR) is designed to compare the mortality at a facility to the mortality that would be expected given national death rates for patients with similar characteristics. The SMR equals the ratio of the actual number of deaths among patients treated by a facility to the expected number of deaths for the facility. Similarly, the regional SMR values in the DFR are calculated as the ratio of the total number of observed deaths among patients from each region to the number of expected deaths among patients from each region.

Qualitatively, the degree to which the facility’s SMR varies from 1.00 is the degree to which it exceeds (>1.00) or is under (<1.00) the national death rates for patients with the same characteristics as those in the facility. For example, an SMR=1.10 would indicate that the facility’s death rates typically exceed national death rates by 10% (e.g., 22 deaths observed where 20 were expected, according to the facility’s patient mix). Similarly, an SMR=0.95 would indicate that the facility’s death rates are typically 5% below the national death rates (e.g., 19 observed versus 20 expected deaths). An SMR=1.00 would indicate that the facility’s death rates equal the national death rates, on average.

The SMR is adjusted for age, race, ethnicity, sex, diabetes, duration of ESRD, nursing home status, comorbidities at incidence, body mass index (BMI) at incidence, calendar year, and race-specific state population death rates. The SMR indicates whether patients treated in the facility had higher or lower mortality than expected when adjusted for age, race, ethnicity, sex, diabetes, years of ESRD, comorbidities, BMI, year, and population death rates.

Beginning in 2011, the mortality rate for a facility in a particular calendar year is compared to the US mortality rates for dialysis patients in that same year rather than to the average mortality rates over the 4-year period. The advantage of this is that the reference year for a particular estimate will be the same in each DFR and therefore the SMR value will change less between DFRs. In the past, because these statistics were compared to a different reference population in each DFR, the values changed more over time, and could change somewhat even for the same year across reports. The use of a different reference year for each year's estimate will allow you to identify trends over time at your facility beyond the overall US trend over time. In other words, if the SMR for your facility decreases over the time period, this means that mortality at your facility has decreased more over that time period than the overall US average mortality decreased. If mortality at your facility decreased over the four year period at the same rate that overall US mortality decreased over this time period, the SMR for your facility would be the same for each year.

Assignment of Patients to Facilities for the SMR Calculation

This section describes the methods we used to assign patients to a facility in order to calculate the SMR. Because some patients 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 that are described below.
General Inclusion Criteria for Dialysis Patients

A patient’s follow-up in the database can be incomplete during the first 90 days of ESRD therapy. For the purposes of this report, we entered a patient’s follow-up into the tabulations only after that patient had received chronic renal replacement therapy for at least 90 days. Mortality and survival during the first 90 days do not enter into the calculations. This minimum 90-day period assures that most patients are eligible for Medicare insurance — either as their primary or secondary insurer. It also excludes from analysis patients who died during the first 90 days of ESRD, since such patients may have incomplete data.

In order to exclude patients who received only temporary dialysis therapy, 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, deaths and survival during the first 60 days of treatment at a facility do not affect the SMR of that facility.

Identifying Facility Treatment Histories for Each Patient

For each patient, we identified the dialysis provider at each point in time using data from a combination of Medicare-paid dialysis claims, the Medical Evidence Form (Form CMS-2728), and paid dialysis claims. 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 a facility’s analysis upon receiving a transplant. 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 paid 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 entered into analysis after 60 days of continuous therapy at a single facility.

Days at Risk for Each Patient-Record

After patient treatment histories are defined as described above, periods of follow-up time (or patient-records) are created for each patient. A patient-record begins each time the patient is determined to be at a different facility and at the start of each calendar year. The number of days at risk starts over at zero for each patient record so that the number of days at risk for any patient-record is always a number between 0 and 365 (or 366 for leap years). Therefore, a patient who is in one facility for all four years gives rise to four
patient-records and is analyzed the same way as would be four separate patients in that facility for one year each. When patients are treated at the same facility for two or more separate time periods during a year, the days at risk at the facility is the sum of all time spent at the facility for the year so that a given patient can generate only one patient-record per year at a given facility. For example, consider a who patient spends two periods of 100 days assigned to a facility, but is assigned to a different facility for the 165 days between these two 100-day periods. This patient will give rise to one patient-record of 200 days at risk at the first facility, and a separate patient-record of 165 days at risk at the second facility.

The number of days at risk \( t_i \) in each of these patient-records is used to calculate the expected number of deaths for that patient-record as described in the “Calculation of Expected Deaths at a Facility” section below. The SMR for a facility is the ratio of the total number of observed to the total number of expected deaths during all patient-records at the facility.

**Model for Calculating Expected Mortality**

The SMR is based on expected mortality calculated from a Cox model (Cox, 1972; SAS Institute Inc., 2004; Kalbfleisch and Prentice, 2002; Collett, 1994). The model used is fit in two stages. The stage 1 model is a Cox model stratified by facility and adjusted for patient age, race, ethnicity, sex, diabetes, duration of ESRD, nursing home status, patient comorbidities at incidence, calendar year and body mass index (BMI) at incidence. This model allows the baseline survival probabilities to vary between strata (facilities), and assumes that the regression coefficients are the same across all strata. Stratification by facility at this stage avoids biases in estimating regression coefficients that can occur if the covariate distributions vary substantially across centers. The results of this analysis are estimates of the regression coefficients in the Cox model and these provide an estimate of the relative risk for each patient. This is based on a linear predictor that arises from the Cox model, and is then used as an offset in the stage 2 model, which is unstratified and includes an adjustment for the race-specific age-adjusted state population death rates.

The patient characteristics included in the stage 1 model as covariates are age, race, ethnicity, sex, cause of ESRD (diabetes or other), duration of ESRD (<1 year, 1-2 years, 2-3 years, 3+ years at the period start date), nursing home status, comorbidity at incidence, calendar year, BMI at incidence, and interaction terms between race, sex and duration and cause of ESRD. Age as of the period start date is included as a piecewise continuous variable with different coefficients based on whether the patient is 0-13 years old, 14-60 years old, or 61+ years old, and whether the patient is black or not. Ethnicity is included with different coefficients for white and non-white patients. The logarithm of BMI is included as a piecewise continuous log-linear term with different coefficients based on whether the \( \ln \) BMI is greater or less than 3.5. Categorical indicator variables are included as covariates in the stage 1 model to flag records missing values for cause of ESRD, Form CMS-2728, and BMI. These variables have a value of 1 if the patient is missing the corresponding piece of information and a value of 0 otherwise. A categorical indicator variable also flags records with at least one comorbidity at incidence. The stage 2 model includes the age-adjusted population death rates for patients of that race in that
state as a covariate. The example below shows how these coefficients are used to carry out the calculations. In the stage 2 model, there is no stratification and there is a single baseline survival curve, which is estimated along with the estimates of the stage 2 regression parameters. The estimate of the baseline survival curve also arises from the fitting of the Cox model and is analogous the Kaplan-Meier (1958) estimate, except that it is adjusted for variation among patients.

The Microsoft Excel file available with this document indicates the value of the coefficient (β) for each characteristic in the stage 1 and 2 models as well as the associated standard error and a p-value corresponding to a test of the null hypothesis that the corresponding coefficient is 0. The file also includes the baseline survival curve for the stage 2 model. A simple illustrative example is also given at the end of this document.

Age-adjusted population death rates (per 100,000) by state and race are obtained from the U.S. Centers for Disease Control National Center for Health Statistics. The 2015 DFR used age-adjusted death rates for 2008-10 from Table 19 of the publication Health, United States, 2014, available at http://www.cdc.gov/nchs/data/hus/hus14.pdf.

Missing Data

Patients with missing data are not excluded from the model. Patients with missing diagnosis are included in the “other” diagnosis group. For the purposes of calculation, BMI missing values are replaced with mean values for patients with similar age, race, sex, and cause of ESRD. When the cause of ESRD is missing, missing values for the BMI are replaced with mean values for patients with similar age and sex. These mean values are included in the attached tables. Patients with missing race are included in the “other” race group strata and classified as non-White in the model. Patients with missing ethnicity are classified as “unknown” ethnicity. No patients were missing age, sex, or date of first ESRD treatment. As mentioned above, indicator variables identifying patients with missing values for cause of ESRD and BMI are also included as covariates in the model.

Calculation of Expected Deaths at a Facility

As described above, each patient typically gives rise to several patient-records. Specifically, a new patient record is defined for each calendar year and each time a patient changes facilities. The i

th

patient record is associated with a risk period t

i

, which specifies the number days that the patient is at risk during that record. Note that each patient record corresponds to a single facility and to a single calendar year.

The Cox model is applied in two stages. Stage 1 yields estimates of the coefficients (β

j

) for the 56 covariates that are measured on individual patients (or patient-records) and included in the model. Using these coefficients, a relative risk or predicted risk is calculated for each patient-record. Stage 2 of the model uses only one covariate, the log of the population death rate for that patient’s race within the state where the patient is being treated. The predicted value for the patient-record from stage 1 is used as an offset in the stage 2 model and the stage 2 analysis is not stratified. The combined predicted
values from stages 1 and 2, and the baseline survival curve from stage 2 of the Cox model are then used to calculate the expected number of deaths for a specific patient-record.

Let $p$ denote the number of patient characteristics in the model and $x_{ij}$ be the specific value of the $j^{th}$ characteristic for the $i^{th}$ patient-record. In stage 1, for patient-record $i$, we denote the measured characteristics or covariates as

$$X_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$$

and use this to define the regression portion of a Cox model in which facilities define the strata. Note that for a categorical characteristic, the $x_{ij}$ value is 1 if the patient falls into the category and 0 otherwise. The output of this model is a set of regression coefficients, $\beta_1, \beta_2, \ldots, \beta_p$ and the corresponding predicted value for the $i^{th}$ patient-record is given by

$$X_i\beta = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}. \quad (1)$$

In stage 2, the only covariate is $x_{i0}$, which specifies the logarithm of the state age-adjusted population death rate corresponding to the race of the patient giving rise to patient-record $i$. The stage 2 model is not stratified, so there is a single baseline survival function assumed. The stage 1 $X_i\beta$ from equation (1) is used as an offset in the analysis. The Stage 2 Cox model gives rise to an estimate of the regression coefficient $\beta_0$ and of the baseline survival function, $S_0(t)$. After stage 2, the linear prediction is

$$A_i = \beta_0 x_{i0} + X_i\beta = \beta_0 x_{i0} + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}$$

Suppose that $t_i$ is the end of follow-up time for patient-record $i$, so that $S_0(t_i)$ is the baseline survival probability at time $t_i$. The survival probability for this patient-record $i$ at time $t_i$ is:

$$S_i(t_i) = [S_0(t_i)]^{exp(A_i)}.$$

The expected number of deaths for this patient-record during follow-up time $t_i$ arises from considerations in the Cox model and can be written as

$$-\ln(S_i(t_i)) = -e^{A_i} \ln [S_0(t_i)].$$

The expected number of deaths at a given facility can now be computed simply by summing these expected values over the totality of patient-records in that facility. Specifically, the expected value is the sum over the $N$ patient-records at the facility giving

$$E = \sum_{i=1}^{N} -\ln[S_i(t_i)] = -\sum_{i=1}^{N} e^{A_i} \ln[S_0(t_i)].$$

Note that, patient-records with 100 days of follow-up, who are otherwise the same, give rise to the same expected mortality even if the 100 day period started at different dates during the year. This approximation is made to simplify the calculations.

Let $O$ be the total number of deaths observed at the facility during the total four year follow up period. As stated above, the SMR is the ratio of the total number of deaths observed to the expected number so that
SMR = O/E.

**Example**

As an example, we calculate the one-year SMR for a hypothetical facility in Florida that treated 5 patients in 2014. Table 1 describes the patients and their history of treatment.

| Table 1. Description of example patients at example facility for 2014 SMR calculation |
|---------------------------------------------|-------------|
| Patient | Start of ESRD | Dates treated at this facility | Characteristics |
| 1       | 4/15/2011     | 4/15/2011 to 12/31/2014         | Male, white, non-Hispanic, 51, diabetic, diabetes as cause of ESRD, alcohol dependence at incidence, BMI = 32.2 |
| 2       | 3/1/2014      | 3/1/2014 to 12/31/2014          | Female, black, non-Hispanic, 70, BMI=23.0 |
| 3       | 11/1/2013     | 11/1/2013 to 11/10/2014 death   | Female, black, non-Hispanic, 78, BMI = 22.1 |
| 4       | 8/15/2006     | 7/15/2014 to 7/31/2014 transfer | Female, Asian, 66, diabetic, BMI = 18.7 |

First we determine which patients are assigned to the facility and for how many days each assignment lasts. Patient 1 started treatment at our example facility in April of 2011 and is assigned to the facility for all of 2014. Patient 2 started renal replacement therapy on 3/1/2014 and is assigned to our facility after 90 days. Patient 3, similarly, started RRT on 11/1/2013 is assigned to our facility after 90 days until her death on 11/10/2013. Patient 4 was only in our facility 16 days, so was never assigned to our facility. Patient 5 started treatment at our facility in 2008 but was treated at another facility from 4/1/2014 through 8/1/2014, so has two treatment periods at the facility which are combined. Table 2 summarizes the assignment periods.

<table>
<thead>
<tr>
<th>Table 2. Patient assignment periods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

For each patient period, we calculate the stage 1 $X_i \beta$ using the comorbidity index weights table, the mean values for imputation of comorbidity index and BMI table, and the
coefficients table in the Excel file. Table 3 shows these details for the example. Note the calculations can be affected by rounding. We show only four decimal places for ease of display.

**Table 3. Stage 1 Calculations for each patient period**

<table>
<thead>
<tr>
<th>Patient</th>
<th>$X_i \beta$</th>
<th>Stage 1 $X_i \beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(age)(-0.0811) + (age-14)(0.1111) + (bmi)(-0.5392) + (vin23)(-0.1553) + (diab)(0.2029) + (vin23*diab)(-0.0694) + (diab_comor)(0.0939) + (alcohol)(0.1269) + (comorb_one)(0.2064)</td>
<td>-1.486</td>
</tr>
<tr>
<td>2</td>
<td>(age)(-0.0811) + (age-14)(0.1111) + (age-60)(0.0016) + (bmi)(-0.5392) + (black)(-0.3187) + (female)(-0.0662) + (vin01)(-0.0562) + (black<em>female)(0.1040) + (vin01</em>black)(0.0245) + (vin01<em>female)(0.0387) + (black</em>age)(0.0149) + (black<em>age-14)(-0.0206) + (black</em>age-60)(0.0051)</td>
<td>-1.462</td>
</tr>
<tr>
<td>3</td>
<td>(age)(-0.0811) + (age-14)(0.1111) + (age-60)(0.0016) + (bmi)(-0.5392) + (black)(-0.3187) + (vin12)(-0.2049) + (female)(-0.0662) + (black<em>female)(0.1040) + (black</em>age)(0.0149) + (black<em>age-14)(-0.0206) + (black</em>age-60)(0.0051)</td>
<td>-1.192</td>
</tr>
<tr>
<td>5</td>
<td>(age)(-0.0811) + (age-14)(0.1111) + (age-60)(0.0016) + (bmi)(-0.5392) + (missc)(0.4481) + (unknown ethnicity)(-0.6553) + (unknown*nonwhite)(-0.1436) + (copd)(0.2007) + (comorb_one)(0.2064)</td>
<td>-0.775</td>
</tr>
</tbody>
</table>

* bmi = natural logarithm of body mass index  
missc = missing cause of ESRD, 0 for no, 1 for yes  
copd=Incident Comorbidity: Chronic obstructive pulmonary disease  
Next we use the **stage 1 $X_i \beta$** as an offset in step 2 of the model, which includes only the race-specific state population death rate as a covariate.
### Table 4. Stage 2 Calculations for each patient period

<table>
<thead>
<tr>
<th>Patient</th>
<th>( A_i = \beta_0 x_{i0} + X_i \beta )</th>
<th>( A_i )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(popdrate)(.3848) + Stage 1 ( X_i \beta )</td>
<td>-1.515</td>
</tr>
<tr>
<td>2</td>
<td>(popdrate)(.3848) + Stage 1 ( X_i \beta )</td>
<td>-1.503</td>
</tr>
<tr>
<td>3</td>
<td>(popdrate)(.3848) + Stage 1 ( X_i \beta )</td>
<td>-1.233</td>
</tr>
<tr>
<td>5</td>
<td>(popdrate)(.3848) + Stage 1 ( X_i \beta )</td>
<td>-0.801</td>
</tr>
</tbody>
</table>

* popdrate = \( \log \) of the race-specific state population death rate

We also use the Excel file to find the baseline survival probability \( S_0(t_i) \), by finding the corresponding survival value given the number of days at risk in the patient period. Table 5 shows these details for the example. Again, note the baseline survival probabilities are shown to four decimal places in this example.

### Table 5. Baseline survival values

<table>
<thead>
<tr>
<th>Patient</th>
<th>Days (( t_i ))</th>
<th>( S_0(t_i) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>365</td>
<td>0.5897</td>
</tr>
<tr>
<td>2</td>
<td>216</td>
<td>0.7242</td>
</tr>
<tr>
<td>3</td>
<td>285</td>
<td>0.6608</td>
</tr>
<tr>
<td>5</td>
<td>243</td>
<td>0.6990</td>
</tr>
</tbody>
</table>

Finally, we calculate \(-e^{A_i} \ln[S_0(t_i)]\), the expected number of deaths for each of these patients.

### Table 6. Calculate expected deaths for each patient

<table>
<thead>
<tr>
<th>Patient</th>
<th>( A_i )</th>
<th>(-e^{A_i})</th>
<th>( \ln[S_0(t_i)])</th>
<th>Expected deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.515</td>
<td>-0.220</td>
<td>-0.5281</td>
<td>0.1160</td>
</tr>
<tr>
<td>2</td>
<td>-1.503</td>
<td>-0.223</td>
<td>-0.3226</td>
<td>0.0718</td>
</tr>
<tr>
<td>3</td>
<td>-1.233</td>
<td>-0.292</td>
<td>-0.4143</td>
<td>0.1208</td>
</tr>
<tr>
<td>5</td>
<td>-0.801</td>
<td>-0.449</td>
<td>-0.3581</td>
<td>0.1606</td>
</tr>
</tbody>
</table>

The total expected number of deaths in this facility for 2014 is the sum of the expected number of deaths for all the patient periods in that facility, or in this case. Because there was one death in the facility during 2014, the SMR is 1/0.4693, or 2.13.

### Caveats

Calculation of the SMR using this method may differ from the SMR published in the DFR for several reasons. For example, the DFR includes deaths within 60 days after transfer out of a facility, but this information may not be available to other researchers. Other differences in the calculation of days at risk will affect expected mortality and may be associated with events such as transfer, transplant, withdrawal from dialysis, hospitalization, and loss to follow-up. Differences in the coding of patient characteristics may also cause other researcher’s calculations to differ from those published in the DFR.
References


