Comparison of methodologies to characterize haemoglobin variability in the US Medicare haemodialysis population

Thomas J. Arneson, David Zaun, Yi Peng, Craig A. Solid, Stephan Dunning and David T. Gilbertson

Chronic Disease Research Group, Minneapolis Medical Research Foundation, Minneapolis, MN, USA

Keywords: dialysis; haemoglobin levels; haemoglobin variability

Introduction

Anaemia is a common complication of renal failure [1]; its appropriate management is an important aspect of dialysis patient care [2]. Due to developments in anaemia management over the past 20 years, average haemoglobin levels have risen steadily and transfusion needs declined [1]. The concern about what constitutes appropriate anaemia management has increased due to results from clinical trials [3–5] and increasing costs of treatment [1]. Considerable month-to-month haemoglobin level change has been demonstrated [1,6–8] and two recent studies, using different methods to characterize haemoglobin variability, came to different conclusions regarding the association between haemoglobin variability and mortality [9,10].

Though many studies have analysed haemoglobin level distributions within dialysis patient populations (inter-patient variability) [8,11–13], and some have used a measure of individual-level haemoglobin variability over time (intra-patient variability) as an outcome [14,15], few have focused on describing the extent and nature of intra-patient variability [6,7,9,10,16]. Yang et al. compared residual standard deviation and absolute haemoglobin change methodologies [10], and Gilbertson et al. compared two methodologies based on fluctuations across thresholds [17]. Whether haemoglobin variability affects outcomes independently of underlying patient morbidity and treatment remains unresolved; additional research addressing this topic is likely. Understanding the different ways to characterize haemoglobin variability will help evaluate results of these studies.

Our study compares four methodologies to describe individual-level haemoglobin variability: standard deviation, residual standard deviation, fluctuation across thresholds and haemoglobin cycling, applying each to a prevalent and an incident cohort of US Medicare haemodialysis patients.

Methods

Data sources and study population

Patients were identified from the Centers for Medicare & Medicaid Services (CMS) end-stage renal disease (ESRD) database. Data were from Medicare Part A and Part B claim files, and haemoglobin levels from erythropoiesis-stimulating agent (ESA) claims. Each claim reports ESA dosing information and a patient haematocrit level (divided by 3 to obtain haemoglobin values). Most ESA claims are billed monthly; average monthly haemoglobin levels were calculated for patients with more than one claim during a month.

Included patients had Medicare as primary payer with Part A and Part B coverage throughout the study, and one or more ESA claim with a valid haematocrit level in each of the 6 study months to allow tracking of haemoglobin values over the entire period. Patients who changed modality or payer status before the end of the 6-month period were excluded.

Prevalent patients were receiving haemodialysis on 31 December 2003, and July–December of 2004 (n = 129 084). Average haemoglobin levels remained constant over the 6-month study interval (Figure 1, panel A). Incident patients initiated haemodialysis as their first type of renal replacement therapy during a 1-year period starting 1 July 2003. For patients not already enrolled in Medicare at dialysis initiation, Medicare claim data are unavailable for dialysis months 1–3. To ensure complete Medicare claim data for months 1–6, we limited the incident cohort to patients aged ≥65 years at initiation (49.2% in 2004 [18]). The incident cohort shows a pattern of sharp average haemoglobin level increase in the early months, then a plateau and modest decline (Figure 1, panel B).
This methodology measures variability in a single, scaled number that is relatively easy to understand conceptually. However, the method cannot discern patterns of variability because it reflects only variability from the mean. It does not distinguish between linear increases or decreases, cyclical patterns of increase and decrease and erratic patterns above and below the mean from month to month. Notably, a steady increase (or decrease) in haemoglobin levels results in a non-zero standard deviation. To register a standard deviation of zero, each of the six monthly haemoglobin levels would need to be identical.

**Residual standard deviation**

The residual standard deviation method first fits a regression line to haemoglobin values for each patient across the 6-month period, and then calculates the (residual) standard deviation around that line. The regression line is the straight line that produces the least vertical deviation of monthly values from the line; it adjusts for trends of consistently increasing or decreasing haemoglobin levels. The slope of the regression line indicates that trend. If monthly haemoglobin values show perfectly consistent linear increases or decreases, the regression line follows the plot of haemoglobin levels with no vertical deviations, and the residual standard deviation is zero. Least-squares regression fits a line that minimizes the square of the vertical distances from that line, and the residual standard deviation is calculated as the root mean squared error from that regression. The positive slope of the regression lines (Figure 3) indicates general trends of haemoglobin value increase for person A and person B. The larger slope for person B demonstrates a trend of more rapid increase. This method measures variability beyond general trends of increase or decrease. It was first applied to haemoglobin variability by Feldman et al. [19] and was used in a recent analysis of the association between haemoglobin variability and mortality [10].

The average slope of the regression lines for our prevalent cohort is nearly zero, indicating little overall trend for increase or decrease during the 6-month interval (Table 1). The distribution of positive and negative individual slope lines was symmetrical. Conversely, the average slope for the incident cohort is substantially positive. Reflecting general trends of increasing haemoglobin levels during the first 6 months of dialysis, the slope was positive for 75% of the incident cohort. For both cohorts, residual mean standard deviation is less than mean standard deviation, as some variation measured with the standard deviation method is part of a general trend of increase or decrease that is adjusted for in the residual standard deviation methodology.

The residual standard deviation method measures variability in a single, scaled number that adjusts for trends of increase or decrease. This is useful when trends are generally linear, such as during the early months after dialysis initiation or after hospitalization. However, many patients experience non-linear haemoglobin level changes. For example, U-shaped patterns may have a linear slope close to zero. The residual standard deviation method gives little insight into patterns of haemoglobin dynamics beyond general increasing or decreasing trends.

**Definition and application of the four methodologies**

**Standard deviation**

The standard deviation description of haemoglobin variability derives a statistic summarizing monthly variations from the average haemoglobin level during the study period. Monthly values for each patient are used to calculate a mean haemoglobin level for that patient. The distance of each monthly value from the mean is used to calculate the standard deviation for that patient. Individual means and standard deviations are averaged across the entire population. Figure 2 gives a visual example. Person A and person B have the same mean haemoglobin value, but person B has larger vertical deviations from the mean, resulting in a larger standard deviation value.

Mean haemoglobin values were slightly higher in the prevalent than in the incident cohort (11.97 versus 11.91 g/dL (Table 1). Mean standard deviation was higher for the incident than for the prevalent cohort (1.28 versus 0.96).
and incident cohorts, we instead let thresholds be defined by Ebben [6]; it is similar to methodology used in a recent study of the association between haemoglobin variability and mortality [17]. Our initial thresholds were the values developed by combining clinical guidelines and payment policies. Haemoglobin levels were classified as low (<11.0 g/dL), target range (11.0 to <12.5 g/dL) or high (>12.5 g/dL) for each of the 6 months. After initial investigations demonstrated differences in haemoglobin distributions between time intervals and between prevalent and incident cohorts, we instead let thresholds be defined by the distribution in each cohort. We calculated 25th and 75th percentiles of all monthly haemoglobin levels within each cohort to classify low (<25th percentile), mid-range (25th to <75th percentile) and high (≥75th percentile) levels. The percentile cut-offs allow direct comparison across time periods, and adjust for predictably lower haemoglobin level distributions around the time of dialysis initiation.

Applying this methodology to our cohorts reveals considerable fluctuation (Table 2). Fewer than 6% of patients in either cohort remained consistently between the upper and lower haemoglobin level thresholds, and nearly half in each cohort were in the HA category. Between one-fifth and one-fourth of patients in each cohort were in the LAL and LAH categories, with only a small percentage consistently below the lower threshold or above the upper threshold.

The fluctuation-across-thresholds method gives insight into haemoglobin variability through patterns of distribution across fluctuation categories. It does not automatically produce a single, scaled value representing variability. Variability in different populations can be compared by assessing differences in distributions across fluctuation categories. Comparison for each category could offer different insights, but comparing six categories becomes complex; focusing on one or two categories may suffice. To compare general magnitude of variability, the consistently mid-range and HA categories provide measures of staying within a range and varying widely on both sides of the range.

While this method provides more detail about the nature of variation than the first two methods described, much important detail remains hidden within the categories, as demonstrated in Figure 4. In each row, all three graphs would be classified as belonging to the same group. Despite obvious pattern differences, all first-row examples would be considered consistently mid-range, all second-row examples LAL and all third-row examples HA. Enhancing the fluctuation-across-thresholds method by taking mean haemoglobin into account to stratify the categories holds promise. For example, including the subsets of LAL and LAH patients with mean haemoglobin values between the mid-range thresholds with the consistently mid-range group would create a group with monthly haemoglobin values always between the lower and upper thresholds or exceeding them infrequently and shallowly enough to maintain a mean between the thresholds.

Haemoglobin cycling

Work published by Fishbane and Berns [7] inspired this methodology, which defines and measures patterns

| Table 1. Standard deviation and residual standard deviation methods |
|-------------------------|-------------------------|
|                         | Standard deviation method | Residual standard deviation method |
|                         | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| Cohort                  |          |          |          |          |
| Prevalent 2004          | 11.97 (0.80) | 0.96 (0.52) | −0.002 (0.35) | 0.75 (0.44) |
| Incident 2003–2004*     | 11.91 (0.92) | 1.28 (0.56) | 0.274 (0.41) | 0.95 (0.47) |

SD, standard deviation.

*Months 1–6, age ≥65 years.

Fluctuations across thresholds

In this method, patients are classified based on starting haemoglobin values (low, target range, high) and whether and how (increase or decrease) values change over subsequent months. Variability is examined by initial values, direction of change and amplitude of change. The possible combinations of these variables can be consolidated as follows:

- **Consistently low**: haemoglobin below lower threshold in each of the 6 months
- **Consistently target range**: haemoglobin between lower and upper threshold in each of the 6 months
- **Consistently high**: haemoglobin above upper threshold in each of the 6 months
- **Low amplitude low (LAL)**: haemoglobin between lower and upper thresholds and below lower threshold; no values above upper threshold
- **Low amplitude high (LAH)**: haemoglobin between lower and upper thresholds and above upper threshold; no values below lower threshold
- **High amplitude (HA)**: haemoglobin below lower threshold; and above upper threshold may be values between thresholds

This methodology was derived from analyses published by Ebben et al. [6]; it is similar to methodology used in a recent study of the association between haemoglobin variability and mortality [17].
Table 2. Haemoglobin fluctuation across thresholds, fixed cut-offs and 25th and 75th percentiles as cut-offs

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Fixed cut-offs, g/dL</th>
<th>Consistently low/ mid-range</th>
<th>Consistently high</th>
<th>LAL</th>
<th>LAH</th>
<th>HA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>High</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Prevalent 2004</td>
<td>11.0</td>
<td>12.5</td>
<td>1304 (1.01)</td>
<td>7671 (5.94)</td>
<td>3453 (2.68)</td>
<td>19 162 (14.84)</td>
</tr>
<tr>
<td>Incident 2003–2004</td>
<td>11.0</td>
<td>12.5</td>
<td>268 (1.69)</td>
<td>291 (1.83)</td>
<td>183 (1.15)</td>
<td>2170 (13.65)</td>
</tr>
<tr>
<td>Percentiles, g/dL</td>
<td></td>
<td></td>
<td></td>
<td>25th</td>
<td>75th</td>
<td></td>
</tr>
<tr>
<td>Prevalent 2004</td>
<td>11.3</td>
<td>12.7</td>
<td>2261 (1.75)</td>
<td>6544 (5.07)</td>
<td>2070 (1.60)</td>
<td>29 511 (22.86)</td>
</tr>
<tr>
<td>Incident 2003–2004</td>
<td>10.9</td>
<td>12.9</td>
<td>234 (1.47)</td>
<td>899 (5.66)</td>
<td>43 (0.27)</td>
<td>3791 (23.85)</td>
</tr>
</tbody>
</table>

LAL, low amplitude low; LAH, low amplitude high; HA, high amplitude.

Fig. 4. Examples of undetected variability using the fluctuation-across-thresholds methodology.

Fig. 5. Components of haemoglobin cycling methodology. Each letter represents a monthly haemoglobin value. Decreasing excursion duration, 2 months; increasing excursion amplitude, 1.6 g/dL.

of haemoglobin increase, decrease and cycling. A haemoglobin excursion was defined as a series of decreasing or increasing monthly average haemoglobin values differing by ≥1.5 g/dL, and a haemoglobin cycle as two consecutive excursions in different directions. The duration of an excursion was defined as number of months from its high to low (or low to high) points. Amplitude was defined as the absolute difference between haemoglobin levels at the start and end of the excursion, and velocity as amplitude divided by duration.

In Figure 5, the change from A to B is <0.5 g/dL, not an excursion. The first excursion is the decrease from B to D, and the second the increase from D to F; B–F thus represents a cycle. The value for F (the end of the cycle) is
Table 3. Haemoglobin cycling: occurrence of excursions and cycles

<table>
<thead>
<tr>
<th>Cohort</th>
<th>n</th>
<th>≥1,% (n)</th>
<th>Decreasing,%a</th>
<th>≥1,% (n)</th>
<th>Decrease/increase,%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prevalent 2004</td>
<td>129 084</td>
<td>78.1 (100 811)</td>
<td>50.72</td>
<td>37.8 (47 997)</td>
<td>52.90</td>
</tr>
<tr>
<td>Incident 2003–2004</td>
<td>15 896</td>
<td>92.6 (14 717)</td>
<td>38.13</td>
<td>48.7 (7740)</td>
<td>29.04</td>
</tr>
</tbody>
</table>

aBased on all individual excursions and cycles.

Table 4. Decreasing and increasing excursions: average amplitude, velocity, and duration

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Decreasing excursions</th>
<th>Increasing excursions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Amplitude, g/dL</td>
<td>Velocitya</td>
</tr>
<tr>
<td>Prevalent 2004</td>
<td>2.71</td>
<td>1.51</td>
</tr>
<tr>
<td>Incident 2003–2004</td>
<td>2.71</td>
<td>1.71</td>
</tr>
</tbody>
</table>

aAmplitude divided by duration.
bBetween high and low (or low and high) points.

not as high as the value for B (the beginning), but the cycle definition requires only that F be ≥1.5 g/dL higher than D. The amplitude is measured as the vertical distance (i.e. the total change in haemoglobin) between the beginning and end of an excursion (B–D, D–F). The duration of each excursion in this example is 2 months.

At least three-fourths of patients in both of our cohorts experienced ≥1 excursion, and ≥1 cycle occurred for 38% of the prevalent cohort and 49% of the incident cohort (Table 3). In the prevalent cohort, decrease/increase cycles slightly outnumber increase/decrease cycles. Conversely, in the incident cohort, 71% of cycles are increase/decrease. Increasing and decreasing excursions are about equally common in the prevalent cohort, while increasing excursions predominate in the incident cohort. This preponderance is as anticipated, given the strong trend for increasing haemoglobin levels during the first 6 months of dialysis. Thus, it is also unsurprising that cycles during the first 6 months tend to be increase/decrease cycles. Approximately 35% of the incident cohort experienced an increase/decrease cycle, a frequency that likely reflects the relatively high incidence of acute events, such as infections, that decrease haemoglobin levels after the initial rise during the first months of dialysis.

Characterizing increasing and decreasing excursions separately (Table 4) showed, for both cohorts, decreasing excursions to have higher velocity than increasing excursions, despite lower amplitude, due to shorter duration. This effect is particularly pronounced in the incident cohort.

The strength of the haemoglobin cycling method is its ability to discriminate different types of variation. Number of excursions and cycles, with amplitude and velocity, seems promising for analysing changes over time or differences among groups of prevalent or incident dialysis patients. A particular strength is the ability to distinguish between increasing and decreasing excursions and between decreasing/increasing and increasing/decreasing cycles, as each gives different insights into haemoglobin dynamics. The method is limited by lack of information on haemoglobin values; a cycle from low to mid-range to low could have the same metrics as a cycle from high to higher to high. This limitation could be addressed by adding haemoglobin level requirements to the pattern definition (e.g. decrease/increase cycles starting at ≥11.0 g/dL).

Discussion

We applied four common methods for describing haemoglobin variability to a prevalent and an incident haemodialysis patient cohort. The mean residual standard deviations observed in our 2004 prevalent cohort (0.73 g/dL) was somewhat higher than 0.60 g/dL reported for patients treated at a large dialysis organization in 1996 [10]. Using 11.0 and 12.5 g/dL as cut points in the fluctuation-across-thresholds methodology, results for our prevalent cohort are generally comparable to published results for this methodology in a 2003 prevalent dialysis cohort [6]. Fishbane and Berns [7] used an approach similar to our haemoglobin-cycling methodology in a small 1998–2003 prevalent cohort from one dialysis centre. Direct comparability is limited by their use of a 12-month observation period (versus 6 months), haemoglobin measurements every 2 weeks (versus monthly) and exclusion of patients hospitalized for ≥10 days. However, results appear generally comparable, with similar measures of mean excursion amplitude and duration. They reported ≥1 cycle within 12 months for 90% of their cohort. We found that 38% of our prevalent cohort had ≥1 cycle within 6 months.

Study strengths and limitations

A strength of this study is use of the comprehensive CMS ESRD registry dataset, which includes ~90% of dialysis patients. Haemodialysis patients not included in this dataset are typically covered by employer group health plans and
tend to be somewhat younger and healthier than those in the registry [18]. Whether application of the methodologies in this study to the non-Medicare haemodialysis population would produce similar results is unknown. A similar caveat regarding generalizability applies to patients who have undergone kidney transplant, peritoneal dialysis patients, and chronic kidney disease patients not in need of renal replacement therapy.

An important study limitation derives from the haemoglobin level data sources, ESA claims. Because we required a haemoglobin reading for each of the 6 months analysed for each cohort, patients were required to have claims for ESA administration in each of the 6 months. Among 2004 prevalent dialysis patients who otherwise met inclusion criteria, 27% were excluded because they lacked a haemoglobin reading in ≥1 of the 6 months studied. Similarly, 45% of patients from the 2003–2004 incident cohort were excluded. Excluded patients were likely not receiving ESAs each month because their haemoglobin levels were satisfactory without them. Though one might expect variability to be lower for these patients than for patients regularly receiving ESAs, this is speculation. Results reported here cannot be generalized beyond the population regularly receiving ESAs.

Another limitation arises from data availability around the time of dialysis initiation. Patients not on Medicare at the time of dialysis initiation become Medicare eligible at the beginning of Month 3, possibly delaying data for up to 90 days. Including only patients aged ≥65 years at initiation in the incident cohort ensured data from the first day of dialysis for most patients. This age restriction reduces the size of the incident cohort considerably, but because the ESRD registry is so large, the cohorts are still large. Caution must be used in generalizing the findings for persons aged ≥65 years to the entire Medicare dialysis population. We included Medicare patients of all ages in the prevalent cohort to describe the entire population, and analysed results stratified by age (<65, ≥65 years). We found little difference between these subgroups, but the younger group consistently showed slightly more variability.

Conclusion

The four methodologies used in this study are tools for analysing and describing haemoglobin variability in dialysis patients. Haemoglobin dynamics may possibly be characterized by other methods, but the methods described have been used in recent studies. One cannot be judged better than another, though each has distinctive strengths and limitations. Depending on the study population and the questions of interest, one or more of these methods, used in combination, could offer an advantage.

Acknowledgements. This study was funded by a research contract from Roche Laboratories, Nutley, NJ, USA. The contract provides for the authors to have final determination of manuscript content. The authors wish to thank Chronic Disease Research Group colleagues Shane Nygaard, BA, for manuscript preparation and Nan Booth, MSW, MPH, for manuscript editing.

Conflict of interest statement. None declared.

References

2. KDOQI. Clinical practice guidelines and clinical practice recommen-
6. Ebben JP, Gilbertson DT, Foley RN et al. Hemoglobin level variabil-
ity: associations with comorbidity, intercurrent events, and hospital-
11. Berns JS, Elzein H, Lynn RJ et al. Hemoglobin variability in epoetin-
treated hemodialysis patients. Kidney Int 2003; 64: 1514–1521
15. Salowicz W, Locatelli F, Ryckelyuck JP et al. Once-monthly sub-
16. West RM, Harris K, Gilthorpe MS et al. Functional data analysis ap-
17. Gilbertson DT, Ebben JP, Foley RN et al. Hemoglobin level variabil-