other biologic biosimilar campaigns in Denmark (and some European countries). In contrast, in the US, although the Food and Drug Administration has approved biosimilars for adalimumab, they will not enter the market until 2023 owing to patent disputes with AbbVie, the manufacturer of Humira. Jensen et al highlight the schemes other manufacturers have taken to dissuade biosimilars from formulary inclusion in the US, including use of rebate traps. In 2013, Humira had an annual post-rebate price of $19,000, which had increased to $38,000 by 2018, meaning the delay in market entry of a biosimilar has a big effect not only on increasing spending for Humira, but also increasing spending on the biosimilar, which will be priced using Humira’s price as an anchor.

The rapidity with which the adalimumab biosimilar replaced Humira in Denmark is remarkable, with an almost complete shift immediately after patent expiration. While these kinds of shifts are unlikely to occur so quickly across the entire US, there are some health care systems of comparable size to Denmark (eg, the Veterans Affairs system) and others that are larger (eg, Kaiser Permanente) that have the ability to switch products quickly through use of formularies and a prescriber workforce. For example, Kaiser Permanente has successfully replaced remicade (infliximab) with biosimilars in 80% of patients.2

The lack of competition for biologic agents in the US, including Humira, is a major barrier to more cost-effective care. However, there are some glimmers of hope. There is increasing uptake of other biosimilars, such as infliximab and filgrastim, and the biosimilar-like product insulin glargine, leading to increasing discounts and lower costs.3 A RAND Corporation study projected that fostering biosimilar market adoption for all commonly used biologics could lead to $54 billion in health care savings over the next 10 years.4 With the development of hundreds of new biologics on the horizon and ever-increasing health care spending, we need to take seriously the substantial savings offered by biosimilars and the feasibility, as evidenced by Denmark, of switching to biosimilars quickly once they are available on the market.

Jennifer D. Claytor, MD, MS
Walid Gellad, MD, MPH

Author Affiliations: Department of Internal Medicine, University of California at San Francisco, San Francisco (Claytor); Editorial Fellow, JAMA Internal Medicine (Gellad), Division of General Internal Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (Gellad).


Corresponding Author: Jennifer D. Claytor, MD, MS, 505 Parnassus Ave, M1479, San Francisco, CA 94143 (jennifer.claytor@ucsf.edu).

Conflict of Interest Disclosures: None reported.


Assessing the Agreement of Hospital Performance on 3 National Mortality Ratings for 2 Common Inpatient Conditions

A number of organizations publicly report condition-specific hospital mortality ratings. For those conditions where these ratings overlap, the perception is that the ratings often conflict with each other, making it difficult for health care stakeholders to understand which rating system to trust.1 We sought to understand the level of agreement and disagreement of condition-specific mortality ratings for individual hospitals across 3 national publicly reported rating systems.

Methods | We extracted publicly reported hospital 30-day mortality ratings for 2 common conditions—chronic obstructive pulmonary disease (COPD) and heart failure—from 3 national rating systems, including the US Centers for Medicare & Medicaid Services Hospital Compare, Healthgrades, and US News & World Report Best Hospitals. These rating systems were chosen because they rate hospitals on a national scale and issue ratings for a common set of conditions. The data used for our analysis reflect what was publicly reported in June 2017 and only includes hospitals that received a rating for a particular condition from all 3 rating systems. For purposes of comparing hospital performance across the rating systems, we categorized each rating system’s publicly reported performance categories into 3 groups (good, fair, and poor) based on how consumers would likely interpret the rating. The level of agreement or disagreement in a hospital’s ratings was calculated based on these categorizations. The Johns Hopkins Medicine Institutional Review Board determined that this study did not constitute human subjects research. All analyses were conducted using Excel 2016 (Microsoft).

Results | Our sample size for comparing the mortality ratings was 3230 hospitals for COPD and 3310 hospitals for heart failure. As outlined in Table 1, most hospitals were rated by Hospital Compare as no different than the US national rate (3088 of 3230 [95.6%] for COPD; 3088 of 3310 [93.3%] for heart failure), with the remaining hospitals rated as being statistically better or worse than the US national rate. Healthgrades rated 2373 hospitals (73.5%) as 3 stars (average) for COPD and 2035 hospitals (61.5%) as 3 stars for heart failure, with the remaining hospitals rated as 1 star (below average) or 5 stars (above average). Best Hospitals assigned a more even distribution of ratings, with the percentage of hospitals in each performance category ranging between 18.4% and 22.9%.

In comparing individual hospital performance on the mortality ratings, we found that only 534 of 3230 hospitals (16.5%) had full agreement in their ratings across the 3 systems for COPD and only 542 of 3310 hospitals (16.4%) had full agreement for heart failure (Table 2). For most hospitals (2609 [80.8%] for COPD; 2607 [78.8%] for heart failure), we found 2 of the 3 ratings systems agreed on the rating, with the third system rating the hospital as 1 category better or worse. For both COPD and heart failure, the 3 rating systems gave discordant results for a nontrivial number of hospitals (87 hospitals [2.7%] and 161 hospitals [4.9%], respectively), where all 3 systems assigned different ratings to the same hospital or 2 of the
ratings agreed at one extreme and the third system assigned the hospital a rating at the opposite extreme.

Discussion | We found for most hospitals the 3 rating systems did not fully agree on the hospital’s mortality performance for a particular condition, but there was generally moderate agreement in the ratings. As similar studies have found, ratings are often sensitive to how the rating is constructed and may not reflect true differences in performance.2–4 While our data set included more than 3000 hospitals for each condition, only certain types of hospitals were included in our analysis. Free-standing pediatric hospitals, federal hospitals, and prospective payment system-exempt cancer hospitals were not included in the analysis. If there are differences in how the rating systems assess the mortality performance of the excluded hospitals, this could result in a level of agreement different from what we found in our particular data set. In addition, our analysis was based on a single snapshot of data from June 2017. If we were to repeat the analysis with a different data snapshot, we could possibly find different results. Our study did not assess the methods used by each rating system, so we cannot draw conclusions about which rating system did the best job of correctly identifying high-performing and low-performing hospitals; we simply compared the agreement and disagreement of the public ratings. As mortality reflects a singular objective construct, one step might be for the mortality rating systems to come together to identify best practices and work to produce a single rating so that stakeholders can center on one truth.

J. Matthew Austin, PhD
Jordan M. Derk, MPH
Allen Kachalia, MD, JD
Peter J. Pronovost, MD, PhD

Table 1. Distribution of Hospital Performance on 3 National 30-Day Mortality Ratings for Chronic Obstructive Pulmonary Disease (COPD) and Heart Failure

<table>
<thead>
<tr>
<th>Publicly reported performance category</th>
<th>Assigned category for assessing rating agreement</th>
<th>Hospitals, No. (%)</th>
<th>COPD (n = 3230)</th>
<th>Heart failure (n = 3310)</th>
</tr>
</thead>
<tbody>
<tr>
<td>US Centers for Medicare &amp; Medicaid Services Hospital Compare</td>
<td>Worse than national rate</td>
<td>Poor</td>
<td>93 (2.9)</td>
<td>77 (2.3)</td>
</tr>
<tr>
<td>No different than national rate</td>
<td>Fair</td>
<td>3088 (95.6)</td>
<td>3088 (93.3)</td>
<td></td>
</tr>
<tr>
<td>Better than national rate</td>
<td>Good</td>
<td>49 (1.5)</td>
<td>145 (4.4)</td>
<td></td>
</tr>
<tr>
<td>Healthgrades</td>
<td>1 star</td>
<td>Poor</td>
<td>558 (17.3)</td>
<td>861 (26.0)</td>
</tr>
<tr>
<td>3 stars</td>
<td>Fair</td>
<td>2373 (73.5)</td>
<td>2035 (61.5)</td>
<td></td>
</tr>
<tr>
<td>5 stars</td>
<td>Good</td>
<td>299 (9.3)</td>
<td>414 (12.5)</td>
<td></td>
</tr>
<tr>
<td>US News &amp; World Report Best Hospitals</td>
<td>Worst</td>
<td>Poor</td>
<td>663 (20.5)</td>
<td>681 (20.6)</td>
</tr>
<tr>
<td>Worse than average</td>
<td>Poor</td>
<td>605 (18.7)</td>
<td>625 (18.9)</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>Fair</td>
<td>597 (18.5)</td>
<td>608 (18.4)</td>
<td></td>
</tr>
<tr>
<td>Better than average</td>
<td>Good</td>
<td>625 (19.3)</td>
<td>659 (19.9)</td>
<td></td>
</tr>
<tr>
<td>Best</td>
<td>Good</td>
<td>740 (22.9)</td>
<td>737 (22.3)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Level of Agreement in Hospital Performance Across 3 National 30-Day Mortality Ratings for Chronic Obstructive Pulmonary Disease (COPD) and Heart Failure

<table>
<thead>
<tr>
<th>Level of agreement based on assigned categories</th>
<th>Hospitals, No. (%)</th>
<th>COPD (n = 3230)</th>
<th>Heart failure (n = 3310)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All 3 ratings agree</td>
<td>534 (16.5)</td>
<td>542 (16.4)</td>
<td></td>
</tr>
<tr>
<td>2 Ratings agree and 1 rating differs by 1 category</td>
<td>2609 (80.8)</td>
<td>2607 (78.8)</td>
<td></td>
</tr>
<tr>
<td>2 Ratings agree at 1 extreme and 1 rating is opposite extreme or all 3 ratings disagree</td>
<td>87 (2.7)</td>
<td>161 (4.9)</td>
<td></td>
</tr>
</tbody>
</table>

Availability of Statistical Code From Studies Using Medicare Data in General Medical Journals

Limited access to statistical code (ie, computer programming instructions used to perform analyses from research data) following publication of an article may be a barrier to open science, methodologic rigor, and the reproducibility of research.1,2

Supplemental content

Unlike clinical research data that may raise privacy concerns, sharing statistical code should be straightforward.3 We assessed the availability of statistical code from research articles published in leading general medical journals, focusing on studies using Medicare data.4

Methods | We searched for all studies that cited use of national Medicare data sets (Part A and/or B) published in 6 general medical journals from January 2007 to December 2017. We calculated the percentage of articles that cited studies that were publicly available as code or code in a repository.

Results | We identified 21202 (95.6%) articles that cited studies using Medicare data. Of these, 8343 (39.4%) articles included additional details, such as code or code in a repository. Among these, 3445 (41.3%) articles cited publicly available code.

Discussion | Although published studies often cite the use of Medicare data, they rarely provide access to the code used to perform analyses. This is a barrier to open science and scientific reproducibility. The availability of code from published studies is currently limited, and future studies should strive to make code publicly available to enhance transparency and reproducibility.

Author Affiliations: Johns Hopkins Armstrong Institute for Patient Safety and Quality, Baltimore, Maryland (Austin, Derk, Kachalia); University Hospitals, Cleveland, Ohio (Pronovost).

Accepted for Publication: February 2, 2020.

Corresponding Author: J. Matthew Austin, PhD, Johns Hopkins Armstrong Institute for Patient Safety and Quality, 750 E Pratt St, 15th Floor, Baltimore, MD 21202 (jmaustin@jhmi.edu).


Author Contributions: Dr Austin had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Austin, Derk, Pronovost.

Acquisition, analysis, or interpretation of data: Austin, Kachalia, Pronovost.

Drafting of the manuscript: Austin, Pronovost.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Pronovost.

Administrative, technical, or material support: Derk, Kachalia, Pronovost.

Study supervision: Austin.

Conflict of Interest Disclosures: Dr Austin has received grants from the Leapfrog Group outside the submitted work. No other disclosures were reported.


