Simple Errors in Interpretation and Publication Can Be Costly

TO THE EDITORS:

The study by Haas et al[1] that compares different risk stratification methods is enlightening, although the authors have overstated their results and a discrepancy between the abstract and their results both contribute to the possibility that their study could be misinterpreted in costly ways.

It is good practice to include CIs on all point estimates, and CIs are at least available for the main results presented in Table 2 (on following page). But the authors fail to interpret those values, and as a result conclude—incorrectly—that the Adjusted Clinical Group (ACG) model is the best. Given the information available in the paper, it is correct to say that the results are consistent with there being no difference between ACG and Minnesota Tiering (MN). In fact, the paper's figures (which lack CIs, unfortunately) and C statistic values suggest that any differences in outcomes for all models might be statistically distinct at times but are practically trivial. It is also questionable whether the C statistic is even appropriate for comparing models with different drivers, and whether it accurately portrays the receiver operating characteristic curve upon which it rests (it can't).[2,3]

The specific error in the abstract states that the C statistic for the ACG model for predicting the highest 10% of cost users is 0.81, but Table 2 shows that the correct value is 0.76 (95% CI, 0.75-0.76). Since this overlaps with the results for the MN model (0.74; 95% CI, 0.74-0.75), the ACG model is not clearly "superior to the others." If the authors meant to refer to readmissions, again the overlap in CIs between ACG and MN show that they are essentially equivalent models in this outcome as well. The term "superior" is a value-laden word that can be easily overinterpreted without a close read of the results; its use here presents both clerical and conceptual errors.

It may well be that the ACG model is superior to these other models, but this paper did not demonstrate that—in fact, this study's results actually support the idea that all of these models are practically equivalent. Academics might call that conclusion a "negative result" and decide not to publish because of the excess difficulty of getting it through review. But as an industry professional, I would consider "negative results" like these—when properly interpreted—"fiscally important results."[4]

AJMC is read by thousands of on-the-ground healthcare industry workers, and time-constrained administrators often take abstracts at face value, having no time to evaluate the results themselves. In fact, this paper made its rounds among our administrators when it was published, and may (or may not) have contributed to our subsequent decision to purchase one of these models. Purchasing risk models is expensive, and internal switching, implementation, and training costs multiply that well beyond the cost of the model. I would ask that editors and reviewers make absolutely sure that authors' conclusions are warranted by their results before publication—improper interpretation can be as damaging and/or costly as clerical errors when subjected to the decision-making hurricane that is today's healthcare industry.

Sincerely,
Dwight Barry, PhD

Address correspondence to: Dwight Barry, PhD, Group Health Cooperative, 320 Westlake Ave N, Ste 100, Seattle, WA 98109. E-mail: barry.d@ghc.org.

REFERENCES

1. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care*. 2013;19(9):725-732.

2. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115(7):928-935.

3. Hand DJ. Measuring classifier performance: a coherent alternative to the area under the ROC curve. *Mach Learn*. 2009;77:103-123.

4. Trouble at the lab. *The Economist*. 2013;409(8858):26-30. ■

RESPONSE:

The letter "Simple Errors in Interpretation and Publication Can Be Costly," a response to the article "Risk-Stratification Methods for Identifying Patients for Care Coordination," is primarily based on 1 line in the abstract. The author of the letter did identify a typographical error in the abstract which included the C statistic for readmissions instead of that for high-cost users (we have submitted an erratam to correct the error). Although we agree that the word "superior" could have been softened, the ACG did perform better than other models in predicting healthcare utilization. For example, as can be seen in **Table 2** (below) of the original paper, for predicting the top 10% high-cost users, the ACG had non-overlapping CIs with all models except for MN Tiering, which is based on the ACG.

We believe the abstract was clear. Unfortunately, as abstracts are limited with regard to space, we would hope the reader would draw their impressions from the entire paper. The conclusions of the paper stated that although the ACG was generally better at predicting utilization, all models had good concordance, suggesting that choosing any model would be more beneficial than none. All models, excluding ACG and MN Tiering, are free and publically available.

We agree with the correspondent that receiver operating characteristic curves should not be the only basis of comparison between models. In fact, the analysis did not rely strictly on C statistics, but also focused on calibration, particularly regarding the accuracy of identifying patients at the high end of each of the outcome distributions as displayed in **Table 3** and **Figure 2** (on following page).

Take-Away Points

Published literature should be more critically evaluated when the chances of safety, outcome, or cost implications are high.

- Abstracts can use overstated results or terms to drive readership and citation rates.
- Since the abstract may be the only part of a paper that administrators read, errors in abstracts can have significant real-world impacts on decision making.
- Errors in interpretation of results can also have significant real-world impacts on decision making that can ripple across the entire spectrum of healthcare management.

Sincerely,
Lindsey R. Haas, MPH
Paul Y. Takahashi, MD
Nilay D. Shah, PhD
Robert J. Stroebel, MD
Matthew E. Bernard, MD
Dawn M. Finnie, MPA
James M. Naessens, ScD

Source of Funding: None.

Author Disclosures: The authors (LRS, PYT, NDS, RJS, MEB, DMF, JMN) report no relationship or financial interest with any entity that would pose a conflict of interest with the subject matter of the article.

Authorship Information: Concept and design (LRS, PYT, NDS, RJS, MEB, DMF, JMN); acquisition of data (LRS, JMN); analysis and interpretation of data (LRS, NDS, RJS, MEB, DMF, JMN); drafting of the manuscript (LRS, PYT, MEB, JMN); critical revision of the manuscript for important intellectual content (LRS, PYT, NDS, RJS, MEB, JMN); statistical analysis (LRS, JMN); obtaining funding (JMN); and administrative, technical, or logistic support (DMF).

Address correspondence to: Lindsey R. Haas, MPH, Health Care Policy and Research, Mayo Clinic, 200 First St Southwest, Rochester, MN 55905. E-mail: haas.lindsey@mayo.edu.

REFERENCE

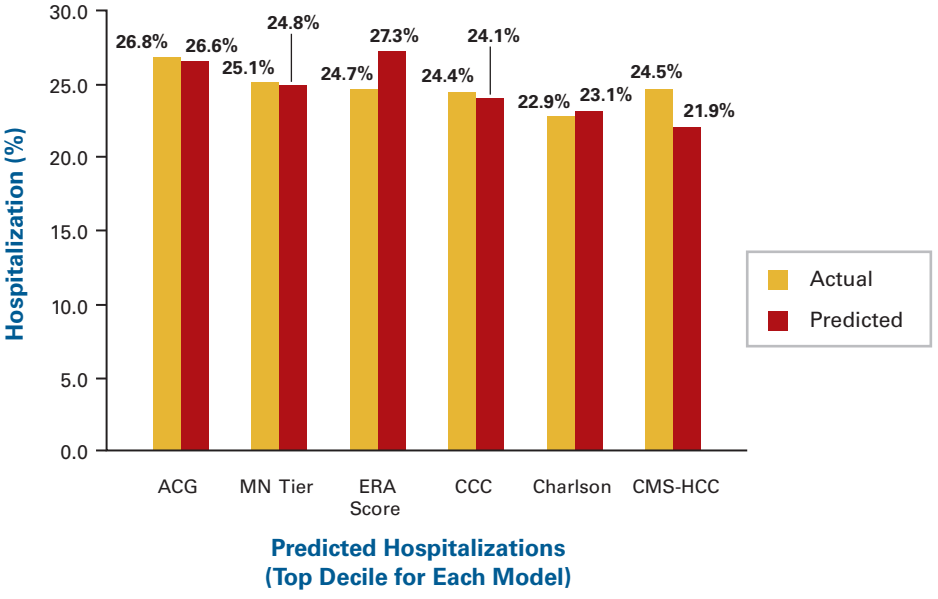
1. Haas LR, Takahashi PY, Shah ND, et al. Risk-stratification methods for identifying patients for care coordination. *Am J Manag Care.* 2013;19:725-732. ■

Table 2. Performance of Risk Stratification Models in Predicting Outcomes

Model Performance (Logistic Regression Models)	Hospitalizations	ED Visits	Readmission	Top 10% Cost
ACG observed rate (lowest decile, highest decile)	(2%, 27%)	(6%, 31%)	(0%, 6%)	(2%, 34%)
ACG area under ROC (C statistic, 95% CI)	0.73 (0.72-0.73)	0.67 (0.66-0.68)	0.81 (0.80-0.83)	0.76 (0.75-0.76)
MN Tiering observed rate (lowest decile, highest decile)	(2%, 25%)	(6%, 30%)	(0%, 5%)	(2%, 33%)
MN Tiering area under ROC (C statistic, 95% CI)	0.71 (0.70-0.72)	0.66 (0.65-0.66)	0.79 (0.78-0.81)	0.74 (0.74-0.75)
ERA observed rate (lowest decile, highest decile)	(2%, 26%)	(9%, 27%)	(1%, 5%)	(2%, 31%)
ERA area under ROC (C statistic, 95% CI)	0.71 (0.70-0.72)	0.61 (0.60-0.61)	0.78 (0.76-0.79)	0.72 (0.71-0.73)
CCC observed rate (lowest decile, highest decile)	(2%, 24%)	(9%, 26%)	(0%, 5%)	(2%, 30%)
CCC area under ROC (C statistic, 95% CI)	0.69 (0.69-0.70)	0.61 (0.60-0.62)	0.77 (0.76-0.79)	0.72 (0.72-0.73)
CCI observed rate (lowest decile, highest decile	(2%, 23%)	(9%, 24%)	(1%, 5%)	(3%, 29%)
CCI area under ROC (C statistic, 95% CI)	0.68 (0.67-0.68)	0.59 (0.59-0.60)	0.75 (0.73-0.77)	0.70 (0.70-0.71)
CMS-HCC observed rate (lowest decile, highest decile)	(2%, 25%)	(10%, 23%)	(1%, 5%)	(3%, 29%)
CMS-HCC area under ROC (C statistic, 95% CI)	0.67 (0.67-0.68)	0.58 (0.58-0.59)	0.74 (0.73-0.77)	0.70 (0.70-0.71)

ACG indicates Adjusted Clinical Group; CCC, Chronic Comorbidity Count; CCI, Charlson Comorbidity Index; CMS-HCC, Centers for Medicare & Medicaid Services Hierarchical Condition Category; ED, emergency department; ERA, Elder Risk Assessment; MN, Minnesota; ROC, receiver operating characteristic.

Figure 2. Top Decile and Percentages of Predicted and Actual Hospitalizations



ACG indicates Adjusted Clinical Group; CCC, Chronic Comorbidity Count; Charlson, Charlson Comorbidity Index; CMS-HCC, Centers for Medicare & Medicaid Services Hierarchical Condition Category; ERA, Elder Risk Assessment; MN, Minnesota.

Table 3. Hybrid Model (Tier 4 or Tier 3 With ERA >10) Compared With Other Models

Model	Number	Mean Age, y	Hybrid, %	Any Hospitalization, %	Hospitalizations per 1000 Patients	Any ED Visit, %	Mean Cost
Hybrid	2347	67	100	41	834	38	$18,501
ACG[a]	2347	69	61	40	791	34	$18,101
MN Tiering[a]	2347	71	76	37	741	36	$16,575
CCC[a]	2347	72	55	35	705	30	$16,399
CMS-HCC[a]	2347	74	46	39	761	30	$17,541
ERA[a]	2347	76	58	38	727	30	$15,204
CCI[a]	2347	75	41	34	637	27	$15,291

ACG indicates Adjusted Clinical Group; CCC, Chronic Comorbidity Count; CCI, Charlson Comorbidity Index; CMS-HCC, Centers for Medicare & Medicaid Services Hierarchical Condition Category; ED, emergency department; ERA, Elder Risk Assessment; MN, Minnesota.
[a]Patients were identified based on the highest estimated probability of hospitalizations.