



# Should (he) stay or should (he) go now?

## Drew Schembre

Swedish Medical Center, Seattle, WA, USA

Correspondence to: Drew Schembre, MD, FASGE, FACG. Swedish Medical Center, Seattle, WA, USA. Email: drew.schembre@swedish.org.

Provenance: This is an invited Editorial commissioned by Dr. Yang Man (Department of Gastroenterology and Hepatology, Songgang People's Hospital, Shenzhen, China).

Comment on: Oakland K, Jairath V, Uberoi R, *et al.* Derivation and validation of a novel risk score for safe discharge after acute lower gastrointestinal bleeding: a modelling study. *Lancet Gastroenterol Hepatol* 2017;2:635-43.

Received: 15 August 2017; Accepted: 29 August 2017; Published: 17 October 2017.

doi: 10.21037/amj.2017.09.02

View this article at: <http://dx.doi.org/10.21037/amj.2017.09.02>

Predicting the future is a tough job but emergency physicians are asked to do it all the time. Judging who will need urgent intervention for GI bleeding and who can be seen more electively represents an almost daily exercise in the ER. Experience and medical knowledge help in this decision making, but comorbidities or uncertain histories diminish even the best clinician's predictive abilities. Risk assessment tools attempt to standardize complex decision making by adding statistical power to what has historically been a subjective gestalt. The high incidence and almost binary natural history of GI bleeding (continue to bleed or stop spontaneously) has made it a popular target for risk stratification tools. In a recent study in *Lancet Gastroenterology and Hepatology*, Oakland *et al.* attempt to provide us with a small crystal ball to predict who can safely be discharged from the ER after presenting with a lower GI bleed (1).

GI bleeding represents one of the most common reasons for emergency room visits and leads to over 500,000 hospital admissions each year in the US (2). Unlike upper GI bleeding (UGIB), which carries a high risk of morbidity and mortality, lower GI bleeds (LGIB) often take a more indolent course, with the majority resolving without urgent intervention. Because of this, triage research has focused on higher risk UGIB, resulting in the creation of predictive tools like the Rockall score, the Glasgow-Blatchford bleeding score (GBBS) and the AIMS65 score (3). LGIB has received less attention. Early tools were clunky and appeared to have limited predictive ability. More recently, Aoki *et al.* created the NOBLADS risk score (NSAIDS, no diarrhea, no abdominal pain, blood pressure <100 mmHg systolic,

antiplatelet medications, albumin <3 g/dL, disease score >2 and syncope) by retrospectively analyzing data from 439 patients admitted for LGIB, confirmed by colonoscopy and validating it among 161 patients admitted to the same institution (4). This score was designed to predict the need for blood transfusion, intervention, length of stay and death. Despite initial enthusiasm for the scoring system, it remains in limited use and has not been widely validated.

Oakland and her colleagues took a similar approach by reviewing a larger group of 2,528 prospectively collected admissions from 143 hospitals for presumed LGIBs in hope of identifying individuals who could safely be discharged from the emergency department. They analyzed 18 predictor variables such as age, sex, blood pressure, hemoglobin level, etc., and selected 7 from which to derive a predictive score. A threshold was identified, below which there was a 95% chance that a patient could be "safely discharged". That is, patients with a score of 8 or lower out of a maximum of 34, would be unlikely to die, require blood transfusion or need endoscopic, radiographic or surgical intervention to control bleeding. This was then validated using a cohort of 288 consecutive patients presenting to the emergency departments of 2 other hospitals with symptoms of LGIB over a four-year period. This scoring system appeared to have greater overall discriminate power than other LGIB scores like BLEED and strate scores as well as NOBLADS. It also provided better predictive ability for safe discharge than the more commonly used UGIB tools like AIMS65, GBBS and the pre-endoscopic Rockall score.

At first glance, the new "Oakland" score sounds like a big step forward for ER triage; wouldn't we all want to predict

the future with 95% certainty? But closer scrutiny raises questions and points to major limitations. First, this scoring system only identified 8% of all patients who could have been safely discharged by their criteria, and 5% of those who meet this relatively low bar, still needed interventions. The 68% of admitted patients in the development cohort ultimately could have been safely discharged from the ER, meaning that the score did not identify 88% of these patients who did not rebleed and could have avoided admission. Further, the percentage of patients with the highest safe discharge score, 8, is not much different than those with a score of 9, 10, 11, 12 or even 13. It is only at 14 and above that the proportion of patients who needed intervention clearly increased. This does not mean that it would necessarily be safe to discharge patients with a score of 12, but it does bring into question the overall strength of the tool as a predictor of any but the most stable appearing patients.

The scoring system itself remains fairly complex and not clearly intuitive. Only 7 measures made the cut. Some that would have seemed useful as predictors, like INR, use of NSAIDs, age and many comorbidities, did not appear on complex linear regression analysis to make a difference. NSAID and antiplatelet use were found to be predictive in the NOBLADS study but not in the Oakland study. Similarly, beyond age, there is no surrogate for baseline poor health or comorbidity. Five of the 7 measures, such as pulse, blood pressure and admitting hematocrit come straight from lab and vitals data, but two, presence or absence of blood on digital rectal exam and whether the patient had ever been admitted for a lower GI bleed in the past, require at least some subjective assessment. Although digital rectal examination (DRE) is recommended as part of the initial exam for presumed GI bleeding, it is only performed about half the time in US ERs (5). It is also not clearly stated whether frank blood or melena must be seen or whether heme + stool or smear is adequate for the score. Clearly, DRE represents an important predictor of bleeding and should be performed more widely. But one wonders why, for example, this score is valued at a relatively low level—lone point for a positive test. No blood on DRE might seem like strong evidence against ongoing bleeding. Yet in this scoring system, blood on DRE carries the same weight as being male (1 point for male, 0 for female). Similarly, a history of previous admission for GI bleeding also adds only one point while a relatively normal systolic blood pressure of 129 mmHg buys 3 points and a mildly low hemoglobin (130–159 g/dL), 4 points. But tachycardia up to 109 bpm adds only 2 points.

By these criteria, a normotensive male with no blood on DRE but mild anemia could get admitted. It remains unclear how many permutations of the inputs were done and whether the current scoring system maximizes its predictive ability.

The authors compared their scoring system to other GI bleeding risk scores and found it superior for predicting safe discharge based on relative C statistics for their validation cohort. Interestingly, the C-statistics for NOBLADS scores in this group were worse than for the cohort in which it was developed. In the original NOBLADS study, the derivation cohort consisted of patients admitted to a large tertiary hospital with presumed LGIB who subsequently underwent colonoscopy. This differs from the Oakland group where less than a quarter of the patients underwent any kind of endoscopic evaluation, suggesting that the Oakland group included many more low risk patients who ultimately did not warrant urgent endoscopy. Clearly, availability and utilization of endoscopy varies geographically, however even in the resource constrained UK National Health Service, colonoscopy should be available for patients presumed to be at risk for recurrent bleeding. This raises the question of how many of the Oakland group were very stable at baseline and may not have needed a complex risk tool to determine that they were at low risk for rebleeding. Given the differences between the two scoring systems, further studies should compare their relative abilities in larger and different populations.

Neither this study nor many others have compared the predictive value of the bleeding scores with what remains the most common risk triage tool used in most EDs, physician judgement. A major flaw in all of these derivations remains that the data used in the analysis usually is collected only from patients who ultimately get admitted, excluding potentially large numbers of patients sent home because the physicians' judgement and experience suggested a strong likelihood that those patients could safely receive outpatient care. In one of the few studies of its kind, ER patients with upper GI bleeding were assigned to ICU versus floor beds based on either the GBBS or physicians' best judgement (6). In this study, the clinical decisions by the ER physician outperformed the GBBS in predicting need for ICU care and therapeutic intervention. One wonders whether physician judgement alone might have identified more than 8% of low risk patients with 95% safety.

This raises an interesting question: Would applying physician judgment before using this, or other risk tool, improve the predictive value of the tool itself? Risk scores

are most useful for directing care for patients whose source of bleeding remains unknown. Several sources of lower GI bleeding can be reasonably deduced from history or common ER imaging studies and often follow a predictable course appropriate for predetermined treatment algorithms. For instance, any patient with active bleeding and a history of recent polypectomy is almost certainly bleeding from that site. These patients should always be admitted as these bleeds carry a high risk for significant blood loss and are almost always controllable at colonoscopy. Similarly, a patient with a history of diarrhea followed by bleeding may have an infectious or inflammatory source but probably does not need to be admitted based on risk of blood loss alone. The main group of patients who would benefit from risk stratification remain middle age and older individuals with presumed diverticular. The NOBLADS score, with its exclusions for diarrhea and abdominal pain, seems better designed to identify patients with diverticular bleeding. The ultimate question may be how well the Oakland score—or the NOBLADS or any other risk score—predicts behavior among that subgroup of patients with diverticular bleeding. We know that most diverticular bleeds stop spontaneously, often after an aggressive colon prep. Accurately predicting which of these patients will spontaneously stop bleeding would have great value. But historically, most diverticular bleeds are unpredictable. Therefore, unless we can safely predict which diverticular bleeds will stop on their own, the vast majority of them warrant at least observation. Adding a colonoscopy within a brief admission—a prep and a colonoscopy could easily be performed in a 24 hour period—seems both efficient and good patient care.

Even the best scoring system that seeks to divert patients from hospitalization must exist in a system that ensures that these patients can access non-emergent but urgent outpatient care. Unfortunately, in many parts of the world, including many parts of the US, that level of access does not exist. Patients who may be safe to defer work up for a few days may rebleed or develop anemia related complications after a week or more. In this series, 6% of patients bled from malignancies. Many patients with occult cancers generate low risk scores as colon cancers rarely bleed heavily. However, a delay in diagnosis in these patients could be the difference between resectable and unresectable disease.

Bleeding scores really do matter. But the goal of a risk stratification tool should be to identify low risk patients who may not need admission but also to help direct

appropriate levels of care and timely interventions to those who need it most. The Oakland score appears to have some utility identifying the lowest risk patients. It remains to be seen what benefit can be derived from other cut-offs. For example, is there a level at which a patient could be simply observed in house for a short term or above which should be prepped immediately for colonoscopy? Is there a level at which a patient should bypass colonoscopy and go directly for angiographic intervention or surgery? This capability might add greater benefit than simply stratifying admission versus discharge.

Ultimately, even the best bleeding score is only useful if physicians use them. Luckily, most ERs now have access to electronic templates to document most routine presentations. Any scoring tool can be embedded into these EHR templates. No one needs to remember the criteria and the EHR can auto populate most of the objective measures and calculate a score after manual entry of a few simple fields. But the physician still needs to choose the right tool for the right indication and this can be confusing. Despite the claim of superior predictive ability of the Oakland score (and NOBLADS) over the GBBS, the C statistics for the most important variables, rebleeding, need for intervention and transfusion, were nearly identical, 0.74, 0.61 and 0.92 respectively for Oakland versus 0.74, 0.58 and 0.84 for the GBBS. Given this, along with its well demonstrated utility in predicting both low and high risk scenarios for UGIB patients, perhaps it would be simpler to adopt the GBBS for patients presenting with any GI bleeding. This has added appeal because differentiating upper and lower sources of bleeding is not always easy. Minor modifications of GBBS—such as antiplatelet use or albumin level or age—might improve predictive ability within specific subgroups, such as presumed diverticular bleeders, variceal bleeders, etc. With the widespread use of EHRs with imbedded, auto-populating scoring systems, large amounts of data could be collected quickly. Modifications and adoption of newer and better decision making tools should become faster, easier and more acceptable to practitioners. Of course, these tools should not replace physician judgement as the ultimate arbiter of patient care. However, the right tools, used correctly, may help us predict the future a little more clearly.

### Acknowledgements

None.

## Footnote

*Conflicts of Interest:* The author has no conflicts of interest to declare.

## References

1. Oakland K, Jairath V, Uberoi R, et al. Derivation and validation of a novel risk score for safe discharge after acute lower gastrointestinal bleeding: a modelling study. *Lancet Gastroenterol Hepatol* 2017;2:635-43.
2. Zhao Y, Encinosa W. Hospitalizations for gastrointestinal bleeding in 1998 and 2006: Statistical Brief 65. In: *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville: Agency for Healthcare Research and Quality (US), 2006-2008.
3. Simons TG, Travis AC, Saltzman JR. Initial assessment and resuscitation in nonvariceal upper gastrointestinal bleeding. *Gastrointest Endosc Clin N Am* 2015;25:429-42.
4. Aoki T, Nagata N, Shimbo T, et al. Development and validation of a risk scoring system for severe acute lower gastrointestinal bleeding. *Clin Gastroenterol Hepatol* 2016;14:1562-70.e2.
5. Shrestha MP, Borgstrom M, Trowers E. Digital rectal examination reduces hospital admissions, endoscopies, and medical therapy in patients with acute gastrointestinal bleeding. *Am J Med* 2017;130: 819-25.
6. Farooq FT, Lee MH, Das A, et al. Clinical triage decision vs risk scores in predicting the need for endotherapy in upper gastrointestinal bleeding. *Am J Emerg Med* 2012;30:129-34.

doi: 10.21037/amj.2017.09.02

**Cite this article as:** Schembre D. Should (he) stay or should (he) go now? *AME Med J* 2017;2:153.