

Basic Study

# Towards automated calculation of evidence-based clinical scores

Christopher A Aakre, Mikhail A Dziadzko, Vitaly Herasevich

Christopher A Aakre, Division of General Internal Medicine, Department of Internal Medicine, Mayo Clinic, Rochester, MN 55905, United States

Mikhail A Dziadzko, Vitaly Herasevich, Department of Anesthesiology, Mayo Clinic, Rochester, MN 55905, United States

Vitaly Herasevich, Multidisciplinary Epidemiology and Translational Research in Intensive Care, Mayo Clinic, Rochester, MN 55905, United States

**Author contributions:** Aakre CA was responsible for the study design, survey design, statistical analysis, manuscript drafting, revisions, and final paper; Dziadzko MA participated in study and survey design and manuscript review; Herasevich V participated in study and survey design, manuscript drafting, revisions and review.

**Institutional review board statement:** The study was reviewed and approved by the Mayo Clinic Institutional Review Board (IRB #15-009228).

**Conflict-of-interest statement:** The authors do not report any conflicts of interest related to the research contained in this manuscript.

**Data sharing statement:** Technical appendix, statistical code, and dataset are available from the corresponding author at [aakre.christopher@mayo.edu](mailto:aakre.christopher@mayo.edu). Consent was not obtained, but data are anonymized and risk of identification is low.

**Open-Access:** This article is an open-access article which was selected by an in-house editor and fully peer-reviewed by external reviewers. It is distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

**Manuscript source:** Invited manuscript

**Correspondence to:** Christopher A Aakre, MD, Division of

General Internal Medicine, Department of Internal Medicine, Mayo Clinic, 200 First St SW, Rochester, MN 55905, United States. [aakre.christopher@mayo.edu](mailto:aakre.christopher@mayo.edu)  
Telephone: +1-507-5380621  
Fax: +1-507-2845370

Received: August 28, 2016

Peer-review started: August 29, 2016

First decision: November 21, 2016

Revised: November 30, 2016

Accepted: January 16, 2017

Article in press: January 17, 2017

Published online: March 26, 2017

## Abstract

### AIM

To determine clinical scores important for automated calculation in the inpatient setting.

### METHODS

A modified Delphi methodology was used to create consensus of important clinical scores for inpatient practice. A list of 176 externally validated clinical scores were identified from freely available internet-based services frequently used by clinicians. Scores were categorized based on pertinent specialty and a customized survey was created for each clinician specialty group. Clinicians were asked to rank each score based on importance of automated calculation to their clinical practice in three categories - "not important", "nice to have", or "very important". Surveys were solicited *via* specialty-group listserv over a 3-mo interval. Respondents must have been practicing physicians with more than 20% clinical time spent in the inpatient setting. Within each specialty, consensus was established for any clinical score with greater than 70% of responses in a single category and a minimum of 10 responses. Logistic regression was performed to determine predictors of automation importance.

## RESULTS

Seventy-nine divided by one hundred and forty-four (54.9%) surveys were completed and 72/144 (50%) surveys were completed by eligible respondents. Only the critical care and internal medicine specialties surpassed the 10-responder threshold (14 respondents each). For internists, 2/110 (1.8%) of scores were "very important" and 73/110 (66.4%) were "nice to have". For intensivists, no scores were "very important" and 26/76 (34.2%) were "nice to have". Only the number of medical history (OR = 2.34; 95%CI: 1.26-4.67;  $P < 0.05$ ) and vital sign (OR = 1.88; 95%CI: 1.03-3.68;  $P < 0.05$ ) variables for clinical scores used by internists was predictive of desire for automation.

## CONCLUSION

Few clinical scores were deemed "very important" for automated calculation. Future efforts towards score calculator automation should focus on technically feasible "nice to have" scores.

**Key words:** Automation; Clinical prediction rule; Decision support techniques; Clinical decision support

© The Author(s) 2017. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core tip:** We report the results of a modified Delphi survey assessing the importance of automated clinical score calculation to practicing internists and intensivists. Although few scores were identified as "very important" for automation, clinicians indicated automated calculation was desired for many commonly used scores. Further studies of the technical feasibility of automating calculation of these scores can help meet these clinicians' needs.

Aakre CA, Dziadzko MA, Herasevich V. Towards automated calculation of evidence-based clinical scores. *World J Methodol* 2017; 7(1): 16-24 Available from: URL: <http://www.wjgnet.com/2222-0682/full/v7/i1/16.htm> DOI: <http://dx.doi.org/10.5662/wjm.v7.i1.16>

## INTRODUCTION

Clinical scoring models are ubiquitous in medical literature, but relatively few are routinely used in clinical practice<sup>[1]</sup>. In general, models have been created to predict clinical outcomes, to perform risk stratification, to aid in clinical decision making, to assess disease severity, and to assist diagnosis. Clinicians have rejected clinical scoring models for many reasons - they lack external validation, they do not provide clinically useful predictions, they require time-intensive data collection, they involve complex mathematical computations, they use arbitrary categorical cutoffs for clinical predictors, they employ imprecise predictor definitions, they require data elements not routinely collected, or they have poor

accuracy in real practice<sup>[1]</sup>. Even among scores accepted by clinicians in clinical practice guidelines<sup>[2-4]</sup>, these same weaknesses can be barriers to consistent, widespread use.

Score complexity is a frequent barrier to manual calculation, especially given the time constraints of clinical practice. The original APACHE score consisted of 34 physiologic variables; data collection and calculation was time-consuming. Subsequent APACHE scoring models have been simplified to include significantly fewer variables, reducing the risk that needed information was not present<sup>[5-7]</sup>. Other popular scores, such as CHADS<sub>2</sub> and HAS-BLED<sup>[8,9]</sup>, have crafted clever mnemonics and point-based scoring systems for easy use at the point-of-care. Despite these simplifications to support manual calculation, many popular and useful clinical scores have been translated to mobile and internet-based calculators for use at the bedside<sup>[10-12]</sup>. Bringing mobile clinical decision support tools to the point-of-care has demonstrated improvements in clinical decision-making<sup>[13]</sup>, however these tools remain isolated from the clinical data present in the Electronic Health Record (EHR).

In 2009, Congress passed the HITECH act, which aimed to stimulate EHR adoption by hospitals and medical practices. Consequently, as of 2014, 96.9% of hospitals have a certified EHR, and 75.5% have basic EHR capabilities<sup>[14]</sup>. Concurrent with EHR adoption, there has been a renewal of the emphasis on improving quality and safety and practicing evidence-based medicine<sup>[15]</sup>. Integration of useful evidence-based clinical score models into the EHR with automated calculation based on real-time data is a logical step towards continuing to improve patient care.

The goal of this study is to identify the clinical scores recognized by clinicians as important to the scope of their clinical practice. This information will be invaluable for prioritizing further research into methods of score automation and delivery to the right provider for the right patient in the appropriate clinical context.

## MATERIALS AND METHODS

This study was reviewed and approved by the Institutional Review Board at Mayo Clinic in Rochester, MN. This study utilized a modified Delphi methodology to seek a consensus of clinical score calculators important in clinical practice for each represented hospital-based specialty. The Delphi methodology is an iterative process used in studies for the purpose of arriving at a consensus opinion among content experts<sup>[16]</sup>. This approach is often utilized when there is incomplete knowledge about a problem or phenomenon and expert judgment is needed for guidance, such as clinical guideline creation<sup>[17]</sup>. In general, the Delphi methodology consists of a series of rounds where participating content experts are asked to respond to results from the previous round<sup>[16]</sup>. The first round, which serves as a brainstorming session to generate a list of topics for future rounds, can be replaced

**Table 1** Description of modified Delphi methodology

Delphi round 1	Systematic collection of online clinical score calculators	Identified 176 externally validated online clinical score calculators
Delphi round 2	Survey development Survey distribution	Branching survey logic mapped score calculators to applicable specialties Academic and community based clinicians

**Table 2** Survey respondent characteristics

	Completion rate	n of Scores
Anesthesia	2/5 (40%)	49
Cardiology	1/1 (100%)	37
Critical care	14/23 (61%)	75
Dermatology	0/0	1
Emergency medicine	4/6 (67%)	62
Family medicine	2/5 (40%)	107
Gastroenterology	3/3 (100%)	17
Hematology	1/1 (100%)	5
Infectious disease	2/2 (100%)	2
Internal medicine	14/25 (56%)	109
Nephrology	1/1 (100%)	6
Neurology	0/1 (0%)	23
OBGYN	1/1 (100%)	1
Oncology	1/2 (50%)	5
Orthopedics	0/0	3
Pediatric	7/13 (54%)	25
Pulmonology	4/6 (67%)	17
Surgery	2/3 (67%)	66

by a systematic review in many situations<sup>[16]</sup>. The Delphi process used by this study is shown in Table 1.

The list of clinical calculators for the first Delphi round was generated by a prior study performed by our group<sup>[18]</sup>. In brief, 176 externally validated clinical scores were identified in calculator form as internet-based services. While this list of clinical calculators is not all-inclusive, it represents all calculators found on popular medical reference web portals (such as Medscape<sup>[11]</sup> and UpToDate<sup>[19]</sup>) and websites aggregating commonly used clinical calculators<sup>[10-12]</sup>. Each calculator was mapped to clinician pertinent specialties for the purpose of generating a customized survey in the next Delphi round. A survey was created in REDCap<sup>[20]</sup> utilizing branching logic to ensure that each responding clinician would only be presented a subset of clinical scores pertinent to their specialty. Score-specialty assignment was verified by non-study associated clinicians at our institution in each represented specialty.

In the second Delphi round, the survey was distributed to clinicians in academic and community settings throughout the United States *via* specialty group LISTSERV's. Only practicing clinicians with greater than 20% of their clinical time spent in the inpatient setting were eligible to serve as content experts for this Delphi round. Respondents were asked to assess the importance of automatic calculation of each clinical score to their clinical practice. Each survey item could be ranked on a three-point Likert scale - "not needed", "nice to have", or "very important". Consensus for each score was defined by greater than 70% of clinicians

in each specialty rating the score in any category. A target of at least 10 experts from each represented specialty is recommended to attain consensus based on established Delphi methods<sup>[16]</sup>; repeated solicitations were sent to underrepresented specialty groups for 3 mo to maximize participation. Descriptive statistics were obtained for each score, grouped by specialty. Variables for each clinical score were categorized by type of clinical information. Logistic regression was performed to characterize clinical score features predictive of automation importance. Statistical analysis was performed with R version 3.3.1<sup>[21]</sup>.

## RESULTS

One hundred forty-four surveys were initiated by respondents. Seventy-nine in one hundred and forty-four (54.9%) were completed and 72/144 (50.0%) were completed by eligible respondents based on based on level of experience and percent of practice spent in the inpatient setting. Only two specialties, internal medicine and critical care medicine, surpassed the 10-respondent threshold with 14 complete responses each (Table 2). Among internists, only 2/110 (1.8%) were deemed very important for automation, while 73/110 (66.4%) were "nice to have". Among intensivists, no scores were deemed very important for automation, however 26/76 (34.2%) were "nice to have" if automation was possible. A summary of score ratings for both specialties can be found in Table 3. Suggestions of missing scores included Centor criteria, Ottawa knee/ankle/foot rules, estimated free water deficit, opioid risk assessment tool, Bishop score, and several screening questionnaires. Too few scores were ranked as "very important" for automation by either specialty to perform regression, however logistic regression was performed on a composite outcome of scores deemed "nice to have" + "very important" (Table 4).

## DISCUSSION

This study assesses clinicians' perspectives on the importance of automating specific clinical scores within the EHR for their clinical practice. We chose a modified Delphi methodology because of our previous study's thoroughness in identifying clinical score calculators across multiple specialty domains and to reduce respondent survey burden. The primary advantage of using a modified Delphi methodology in this study is the ability to capture the valuation of multiple scores by clinicians across varying specialties. The primary disadvantage to this methodology is the recruitment of appropriate content

**Table 3 Summary of importance of automation of specified clinical scores ranked by critical care and internal medicine physicians**

Score name	Year of creation	n of variables	Very important	Very important or nice to have
Critical care				
APACHE II	1985	15	9/14 (64.3%)	12/14 (85.7%)
SNAP II	2001	9	7/11 (63.6%)	9/11 (81.8%)
NRDS scoring system	1998	5	7/12 (58.3%)	10/12 (83.3%)
Post-anesthetic recovery score	1970	5	7/12 (58.3%)	9/12 (75%)
Rotterdam score	1997	4	7/12 (58.3%)	8/12 (66.7%)
SNAP	1993	27	7/12 (58.3%)	9/12 (75%)
SNAP-PE	1993	30	7/12 (58.3%)	9/12 (75%)
SNAP-PE II	2001	12	7/12 (58.3%)	9/12 (75%)
Wells criteria for DVT	2006	9	7/12 (58.3%)	9/12 (75%)
Wells criteria for PE	1998	7	7/12 (58.3%)	10/12 (83.3%)
PAWS	2008	7	6/11 (54.5%)	8/11 (72.7%)
CRIB	1993	5	6/12 (50%)	8/12 (66.7%)
CRIB II	2003	5	6/12 (50%)	8/12 (66.7%)
MSSS	2002	7	6/12 (50%)	8/12 (66.7%)
PELOD score	1999	13	3/6 (50%)	4/6 (66.7%)
SAPS II	1993	16	5/10 (50%)	7/10 (70%)
TIMI risk index	2006	3	5/11 (45.5%)	8/11 (72.7%)
TRISS	1987	9	4/9 (44.4%)	6/9 (66.7%)
Children's coma score	1984	3	3/7 (42.9%)	4/7 (57.1%)
PRISM score	1988	16	3/7 (42.9%)	5/7 (71.4%)
CURB-65	2003	5	5/12 (41.7%)	8/12 (66.7%)
SCORETEN scale	2000	6	5/12 (41.7%)	9/12 (75%)
MEWS score	2006	6	4/10 (40%)	6/10 (60%)
Rockall score	2008	11	3/8 (37.5%)	5/8 (62.5%)
TRIOS score	2001	4	3/8 (37.5%)	5/8 (62.5%)
Geneva score for PE	2006	9	4/11 (36.4%)	7/11 (63.6%)
Injury Severity Score	1974	6	4/11 (36.4%)	8/11 (72.7%)
Lung Injury score	1988	5	4/11 (36.4%)	8/11 (72.7%)
MPMII - admission	1993	14	4/11 (36.4%)	6/11 (54.5%)
MPMII - 24-48-72	1993	14	4/11 (36.4%)	6/11 (54.5%)
LODS score	1996	12	3/9 (33.3%)	7/9 (77.8%)
MEDS score	2003	10	3/9 (33.3%)	6/9 (66.7%)
MESS score	1990	5	4/12 (33.3%)	7/12 (58.3%)
Parsonnet Score	1989	14	4/12 (33.3%)	7/12 (58.3%)
Pediatric coma scale	1988	3	2/6 (33.3%)	3/6 (50%)
RAPS	1987	5	3/9 (33.3%)	7/9 (77.8%)
Surgical Apgar score	2007	3	4/12 (33.3%)	8/12 (66.7%)
ASCOT score	1990	8	4/13 (30.8%)	6/13 (46.2%)
MELD score	2001	4	4/13 (30.8%)	12/13 (92.3%)
PIM2	2003	8	2/7 (28.6%)	5/7 (71.4%)
SWIFT score	2008	6	2/7 (28.6%)	4/7 (57.1%)
Clinical Pulmonary Infection Score	1991	8	3/11 (27.3%)	9/11 (81.8%)
MPM-24 h	1988	15	3/11 (27.3%)	6/11 (54.5%)
Child-Pugh Score	1973	5	3/12 (25%)	11/12 (91.7%)
Decaf score	2012	5	2/8 (25%)	4/8 (50%)
ONTARIO score	1995	6	2/8 (25%)	4/8 (50%)
AKICS score	2007	8	3/13 (23.1%)	7/13 (53.8%)
AVPU scale	2004	4	2/9 (22.2%)	6/9 (66.7%)
PERC rule for PE	2001	7	2/9 (22.2%)	6/9 (66.7%)
RIETE score	1988	6	2/9 (22.2%)	6/9 (66.7%)
BISAP score for pancreatitis mortality	2008	5	2/10 (20%)	4/10 (40%)
Bleeding risk score	2007	4	2/10 (20%)	6/10 (60%)
Clinical asthma evaluation score	1972	5	2/10 (20%)	6/10 (60%)
PIRO score	2009	8	2/10 (20%)	7/10 (70%)
ABC score for massive transfusion	2009	4	2/11 (18.2%)	6/11 (54.5%)
ACLS score	1981	4	2/11 (18.2%)	7/11 (63.6%)
MOD score	1995	7	2/11 (18.2%)	8/11 (72.7%)
MPM - admission	1988	10	2/11 (18.2%)	6/11 (54.5%)
sPESI	2010	8	2/11 (18.2%)	7/11 (63.6%)
ABIC score	2008	4	2/12 (16.7%)	5/12 (41.7%)
CRUSADE score	2009	8	2/12 (16.7%)	6/12 (50%)
Pediatric trauma score	1988	6	1/6 (16.7%)	2/6 (33.3%)
LRINEC Score for Necrotizing STI	2004	5	1/8 (12.5%)	4/8 (50%)
Panc 3 score	2007	3	1/8 (12.5%)	3/8 (37.5%)
Pancreatitis outcome score	2007	7	1/8 (12.5%)	3/8 (37.5%)
TASH score	2006	7	1/8 (12.5%)	4/8 (50%)

POSSUM score	1991	18	1/9 (11.1%)	3/9 (33.3%)
Revised Trauma score	1981	3	1/9 (11.1%)	5/9 (55.6%)
24 h ICU trauma score	1992	4	1/10 (10%)	7/10 (70%)
HIT Expert Probability Score	2010	11	1/11 (9.1%)	6/11 (54.5%)
Bronchiectasis severity index	2014	10	1/12 (8.3%)	4/12 (33.3%)
Oxygenation index	2005	3	1/13 (7.7%)	7/13 (53.8%)
CT severity index	1990	1	0/12 (0%)	6/12 (50%)
Glasgow coma scale	1974	3	0/13 (0%)	10/13 (76.9%)
SOFA	2001	6	0/13 (0%)	8/13 (61.5%)
Internal medicine				
Wells criteria for DVT	2006	9	10/14 (71.4%)	13/14 (92.9%)
Wells criteria for PE	1998	7	10/14 (71.4%)	13/14 (92.9%)
CHA2DS2-VASc	2010	7	9/14 (64.3%)	13/14 (92.9%)
TIMI risk index	2006	3	9/14 (64.3%)	13/14 (92.9%)
TIMI risk score for UA/NSTEMI	2000	7	9/14 (64.3%)	13/14 (92.9%)
TIMI risk score for STEMI	2000	9	9/14 (64.3%)	13/14 (92.9%)
CURB-65	2003	5	8/14 (57.1%)	13/14 (92.9%)
STESS score	2008	4	8/14 (57.1%)	13/14 (92.9%)
Duke criteria for IE	1994	8	6/13 (46.2%)	12/13 (92.3%)
PESI	2006	11	7/12 (58.3%)	11/12 (91.7%)
Revised cardiac risk index for pre-operative risk	1999	6	7/12 (58.3%)	11/12 (91.7%)
SOFA	2001	6	6/12 (50%)	11/12 (91.7%)
ABCD2 score	2006	5	5/12 (41.7%)	11/12 (91.7%)
Charlson Comorbidity index	1987	1	2/12 (16.7%)	11/12 (91.7%)
PERC rule for PE	2001	7	5/11 (45.5%)	10/11 (90.9%)
sPESI	2010	8	4/11 (36.4%)	10/11 (90.9%)
MOD score	1995	7	3/11 (27.3%)	10/11 (90.9%)
MPM - 24 h	1988	15	4/10 (40%)	9/10 (90%)
MPM - admission	1988	10	3/10 (30%)	9/10 (90%)
MEDS score	2003	10	2/10 (20%)	9/10 (90%)
PIRO score	2009	8	1/10 (10%)	9/10 (90%)
SAPS II	1993	16	4/9 (44.4%)	8/9 (88.9%)
SWIFT score	2008	6	2/8 (25%)	7/8 (87.5%)
Panc 3 score	2007	3	1/8 (12.5%)	7/8 (87.5%)
APACHE II	1985	15	9/14 (64.3%)	12/14 (85.7%)
Parsonnett Score	1989	14	8/14 (57.1%)	12/14 (85.7%)
HIT Expert Probability Score	2010	11	6/14 (42.9%)	12/14 (85.7%)
Ranson's criteria	1974	11	6/14 (42.9%)	12/14 (85.7%)
TRIOS score	2001	4	3/7 (42.9%)	6/7 (85.7%)
4Ts Score	2006	5	5/14 (35.7%)	12/14 (85.7%)
Framingham coronary heart disease risk score	1998	7	5/14 (35.7%)	12/14 (85.7%)
30 d PCI readmission risk	2013	10	2/7 (28.6%)	6/7 (85.7%)
Glasgow coma scale	1974	3	9/13 (69.2%)	11/13 (84.6%)
Modified NIH Stroke Scale	2001	9	7/13 (53.9%)	11/13 (84.6%)
King's College Criteria for Acetaminophen Toxicity	1989	6	4/12 (33.3%)	10/12 (83.3%)
Glasgow-Blatchford Bleeding score	2000	9	3/12 (25%)	10/12 (83.3%)
ATRIA bleeding risk score	2011	6	2/12 (16.7%)	10/12 (83.3%)
Glasgow Alcoholic hepatitis score	2005	4	5/11 (45.5%)	9/11 (81.8%)
MEWS score	2006	6	4/11 (36.4%)	9/11 (81.8%)
Hemorr2hages score	2006	11	2/11 (18.2%)	9/11 (81.8%)
Decaf score	2012	5	4/10 (40%)	8/10 (80%)
MPMII - admission	1993	14	4/10 (40%)	8/10 (80%)
MPMII - 24-48-72	1993	14	4/10 (40%)	8/10 (80%)
Malnutrition universal screening tool (MUST)	2004	3	2/10 (20%)	8/10 (80%)
ASTRAL score	2012	6	1/10 (10%)	8/10 (80%)
GRACE ACS	2006	12	1/10 (10%)	8/10 (80%)
CHADS2	2001	5	7/14 (50%)	11/14 (78.6%)
Multidimensional frailty score	2014	9	7/14 (50%)	11/14 (78.6%)
Geneva score for PE	2006	9	3/9 (33.3%)	7/9 (77.8%)
Pittsburg knee rules	1994	3	3/9 (33.3%)	7/9 (77.8%)
Mayo scoring system for assessment of ulcerative colitis activity	2005	4	1/9 (11.1%)	7/9 (77.8%)
4-yr mortality prognostic index	2006	12	1/9 (11.1%)	7/9 (77.8%)
Rockall score	2008	11	1/9 (11.1%)	7/9 (77.8%)
SHARF scoring system	2004	9	1/9 (11.1%)	7/9 (77.8%)
HAS-BLED	2010	12	5/13 (38.5%)	10/13 (76.9%)
ATRIA stroke risk score	2013	7	3/12 (25%)	9/12 (75%)
Euroscore	1999	17	1/8 (12.5%)	6/8 (75%)
Renal risk score	2011	6	1/8 (12.5%)	6/8 (75%)
ROSE risk score	1996	7	1/8 (12.5%)	6/8 (75%)
LRINEC Score for Necrotizing STI	2004	5	3/11 (27.3%)	8/11 (72.7%)



Bleeding risk score	2007	4	2/11 (18.2%)	8/11 (72.7%)
CT severity index	1990	1	1/11 (9.1%)	8/11 (72.7%)
SCORETEN scale	2000	6	7/14 (50%)	10/14 (71.4%)
REMS	2004	7	2/7 (28.6%)	5/7 (71.4%)
Mayo CABG risk of inpatient death after MI	2007	7	1/7 (14.3%)	5/7 (71.4%)
Mayo PCI risk of inpatient MACE	2007	7	1/7 (14.3%)	5/7 (71.4%)
QMMI score	2001	11	1/7 (14.3%)	5/7 (71.4%)
MELD score	2001	4	0/14 (0%)	10/14 (71.4%)
Nexus criteria for C-spine imaging	1970	5	4/10 (40%)	7/10 (70%)
Birmingham nutritional risk score	1995	7	2/10 (20%)	7/10 (70%)
Canadian CT head rule	2001	9	2/10 (20%)	7/10 (70%)
ACLS score	1981	4	1/10 (10%)	7/10 (70%)
San Francisco syncope rule	2004	5	1/10 (10%)	7/10 (70%)
Mannheim peritonitis index	1993	7	6/13 (46.2%)	9/13 (69.2%)
HADO score	2006	4	3/9 (33.3%)	6/9 (66.7%)
CARE score	2001	3	1/9 (11.1%)	6/9 (66.7%)
ICH score	2001	5	1/9 (11.1%)	6/9 (66.7%)
Adult appendicitis score	2014	8	6/14 (42.9%)	9/14 (64.3%)
IMPACT score	2008	11	6/14 (42.9%)	9/14 (64.3%)
CRUSADE score	2009	8	4/14 (28.6%)	9/14 (64.3%)
PORT/PSI score	1997	20	2/14 (14.3%)	9/14 (64.3%)
CIWA-Ar	1989	10	1/14 (7.1%)	9/14 (64.3%)
LODS score	1996	12	3/8 (37.5%)	5/8 (62.5%)
OESIL risk score	2003	4	2/8 (25%)	5/8 (62.5%)
QRISK2	2010	14	2/8 (25%)	5/8 (62.5%)
Qstroke score	2013	15	2/8 (25%)	5/8 (62.5%)
RIETE score	1988	6	2/8 (25%)	5/8 (62.5%)
EGSYS score	2008	6	1/8 (12.5%)	5/8 (62.5%)
EHMRG	2012	10	1/8 (12.5%)	5/8 (62.5%)
FOUR score	2005	4	1/8 (12.5%)	5/8 (62.5%)
Pancreatitis outcome score	2007	7	1/8 (12.5%)	5/8 (62.5%)
Prostate cancer prevention trial risk calculator	1993	6	6/13 (46.2%)	8/13 (61.5%)
Alvarado score for acute appendicitis	1986	8	5/13 (38.5%)	8/13 (61.5%)
DRAGON score	2012	6	1/10 (10%)	6/10 (60%)
Bronchiectasis severity index	2014	10	3/14 (21.4%)	8/14 (57.1%)
New Orleans head CT rule	2000	8	1/7 (14.3%)	4/7 (57.1%)
POSSUM score	1991	18	1/7 (14.3%)	4/7 (57.1%)
Child-Pugh Score	1973	5	0/14 (0%)	8/14 (57.1%)
Lung Injury score	1988	5	4/9 (44.4%)	5/9 (55.6%)
AVPU scale	2004	4	2/9 (22.2%)	5/9 (55.6%)
Gupta perioperative cardiac risk	2011	5	2/9 (22.2%)	5/9 (55.6%)
HEART score	2008	5	1/9 (11.1%)	5/9 (55.6%)
IgA nephropathy score	2006	8	5/14 (35.7%)	7/14 (50%)
ABIC score	2008	4	4/14 (28.6%)	7/14 (50%)
CAMBS score	1993	4	4/14 (28.6%)	7/14 (50%)
GAP risk assessment score	2012	4	2/8 (25%)	4/8 (50%)
BISAP score for pancreatitis mortality	2008	5	2/10 (20%)	5/10 (50%)
ONTARIO score	1995	6	1/8 (12.5%)	4/8 (50%)
JAMA kidney failure risk equation	2011	7	4/13 (30.8%)	5/13 (38.5%)

experts for each Delphi round<sup>[16]</sup>. Because this study focused on the automated calculation of scores used in inpatient clinical practice, we limited analysis to board-certified clinicians practicing more than 20% of their time in the inpatient setting. This requirement allowed use to gather diverse viewpoints of practicing clinicians in various practice settings.

Clinical scores can play important roles in the clinical decision-making algorithms used daily by clinicians. Mobile and internet-based clinical calculators have made these daily clinical score calculations easier; however the use of these standalone technologies does not reduce the time and effort required for manual data retrieval and entry. Automated retrieval of variables required for score calculation within the EHR eliminates the need for these potentially workflow disrupting standalone smartphone or

web applications<sup>[22]</sup>. Additionally, automated calculation of clinical scores provides a mechanism to improve care standardization, to facilitate adherence to evidence-based practice and clinical guidelines, and to save time<sup>[1]</sup>. However, just as clinicians have rejected many clinical scores for routine usage, our study found that clinicians did not appraise most clinical scores as “very important” for automation.

The clinical score variables examined in this study spanned several broad categories - demographic information, laboratory values, medical history elements, clinical examination findings, clinical judgments, and even other clinical scores. Some categories, such as laboratory values or medical history elements, may require more time-intensive data retrieval compared to others. We predicted that commonly used scores with cognitively

**Table 4 Predictors of desirability of score automation based on number of each variable type in each score**

Automation: Very important/nice to have	OR (95%CI)
Critical care	
<i>n</i> of variables	0.68 (0.23, 1.59)
Clinical history	1.36 (0.36, 4.93)
Vital sign	1.40 (0.53, 4.6)
Medication	4.89 (0.10, 237.52)
Clinical judgment	2.33 (0.76, 9.80)
Examination	0.99 (0.36, 3.14)
Laboratory value	1.48 (0.61, 4.41)
Charted variable (non-vital)	2.26 (0.70, 8.93)
Demographic value	0.20 (0.03, 1.00)
Another score	2.07 (0.39, 12.13)
Internal medicine	
<i>n</i> of variables	0.64 (0.39, 1.04)
Clinical history	2.34 <sup>a</sup> (1.26, 4.67)
Vital sign	1.88 <sup>a</sup> (1.03, 3.68)
Medication	2.89 (0.37, 63.17)
Clinical judgment	1.41 (0.75, 2.74)
Examination	1.56 (0.88, 2.87)
Laboratory value	1.51 (0.90, 2.62)
Charted variable (non-vital)	2.54 (0.85, 8.70)
Demographic value	0.90 (0.41, 1.97)
Another score	0.89 (0.30, 2.17)

<sup>a</sup>*P* < 0.05.

demanding information extraction would be more desirable for automation. However, our regression model did not explicitly include variables representing time-required for data collection or data entry for any score - the key efficiencies gained through automated calculation. Instead, we used the number of variables in the score and variable categorization as surrogates to account for these cognitively demanding tasks. No association between the number of clinical variables and desirability of automation was found for the internal medicine or critical care specialties. Only two scores met the threshold for being "very important" for automation by internists - Wells criteria for DVT<sup>[23]</sup> (10/13, 71.4%) and PE<sup>[24]</sup> (10/13, 71.4%). Although many more scores were deemed "nice to have" by both specialties, regression analysis only identified the number of medical history variables (OR = 2.34; 95%CI: 1.26-4.67; *P* < 0.05) and vital sign variables (OR = 1.88; 95%CI: 1.03-3.68; *P* < 0.05) as predictive of desirability of automation among internists. The time and cognitive workload of performing manual chart review for unknown aspects of the medical history may explain this finding; several tools have been created to meet this clinical need<sup>[25,26]</sup>.

The time-benefit gained from reduced workflow disruption may be more apparent in scores pertaining to common clinical scenarios, such as sepsis. During the survey period, the SOFA score was integrated into the operational definition of sepsis<sup>[17]</sup>, likely affecting the valuation of automated calculation by some specialties. The prospective benefit of automated calculation of this and similar scores is readily apparent; one study comparing automated and manual calculation of the SOFA score<sup>[27]</sup> found an average time-savings of about 5 min per

calculation attained by automation<sup>[28]</sup>. Extrapolated to a unit of 12 patients, up to one hour of work could be saved daily through automated calculation of this single score. More complex scores may have even greater time-savings.

This study has several limitations. First, the survey items may not represent all pertinent clinical scores in all specialties surveyed. We did consult with local experts in each specialty to review the completeness of the list of clinical scores. Additionally, respondents were solicited for additional scores to be considered. Many of the suggestions represented either diagnostic criteria (Centor criteria or Ottawa foot/ankle/knee rules) or diagnostic questionnaires (PHQ-9, CAGE, AUDIT) - all are useful clinical tools but not amenable to automated score calculation.

Second, the responding experts may not represent the viewpoints of all clinicians in each field. We sought a heterogeneous group of clinicians within each specialty, representing both academic and community hospital settings nationwide. However, only 6 internists and 6 intensivists that completed our survey volunteered their hospital's name; all were academic health centers. This potential response bias would favor clinical scores used primarily in academic settings, a concern that has been raised for certain scores<sup>[29]</sup>. Additionally, survey response rate was low despite multiple solicitations targeting lesser represented specialties, a likely reflection of physician survey fatigue.

Third, consensus was not reached for most clinical scores for either specialty. Since both specialties had a large number of pertinent clinical scores, it would be expected that consensus could not be reached for many scores. When exploring the programmability of specific clinical scores, researchers may be more inclined to investigate methods for automated calculation of "nice to have" scores that are highly programmable to meet the needs of these clinicians. Further investigation is needed to assess the overall programmability of each clinical score calculator within modern electronic medical record systems utilizing commonly available clinical data and information retrieval techniques.

In conclusion, Internal medicine and critical care physicians assessed evidence-based clinical scores on the importance of automated calculation to their clinical practice. Very few clinical scores were deemed "very important" to automate, while many were considered "nice to have". In order to prioritize automating calculation of some of these "nice to have" clinical scores, further research is needed to evaluate the feasibility of programming each score in the electronic medical record.

## ACKNOWLEDGMENTS

This publication was made possible by CTSA Grant Number UL1 TR000135 from the National Center for Advancing Translational Sciences (NCATS), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NIH.

## COMMENTS

### Background

Numerous clinical scores have been created, but it is not known which scores may be important for automated calculation within the electronic medical record.

### Research frontiers

Automated calculation of important scores can reduce physician's cognitive workload and facilitate practice guideline adherence.

### Innovations and breakthroughs

This study is a comprehensive assessment of importance of automating calculation of clinical scores in the inpatient setting.

### Applications

In this study, clinicians identified specific clinical scores as desirable for automated calculation. This information can guide future research on techniques to automate these scores to meet clinician's needs.

### Peer-review

The authors investigated scoring systems of evidence for clinical application. The aim was clear and results were useful.

## REFERENCES

- Wyatt JC, Altman DG. Commentary: Prognostic models: clinically useful or quickly forgotten? *BMJ* 1995; **311**: 1539-1541 [DOI: 10.1136/bmj.311.7019.1539]
- Konstantinides SV, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galie N, Gibbs JS, Huisman MV, Humbert M, Kucher N, Lang I, Lankeit M, Lekakis J, Maack C, Mayer E, Meneveau N, Perrier A, Pruszczyk P, Rasmussen LH, Schindler TH, Svitil P, Vonk Noordegraaf A, Zamorano JL, Zompatori M. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J* 2014; **35**: 3033-3069, 3069a-3069k [PMID: 25173341 DOI: 10.1093/eurheartj/ehu283]
- January CT, Wann LS, Alpert JS, Calkins H, Cigarroa JE, Cleveland JC, Conti JB, Ellnor PT, Ezekowitz MD, Field ME, Murray KT, Sacco RL, Stevenson WG, Tchou PJ, Tracy CM, Yancy CW. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2014; **64**: e1-76 [PMID: 24685669 DOI: 10.1016/j.jacc.2014.03.022]
- Kahn SR, Lim W, Dunn AS, Cushman M, Dentali F, Akl EA, Cook DJ, Balekian AA, Klein RC, Le H, Schulman S, Murad MH. Prevention of VTE in nonsurgical patients: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012; **141**: e195S-e226S [PMID: 22315261 DOI: 10.1378/chest.11-2296]
- Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; **9**: 591-597 [PMID: 7261642 DOI: 10.1097/00003246-198108000-00008]
- Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985; **13**: 818-829 [PMID: 3928249 DOI: 10.1097/00003246-198510000-00009]
- Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. *Chest* 1991; **100**: 1619-1636 [PMID: 1959406 DOI: 10.1378/chest.100.6.1619]
- Gage BF, Waterman AD, Shannon W, Boehler M, Rich MW, Radford MJ. Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *JAMA* 2001; **285**: 2864-2870 [PMID: 11401607 DOI: 10.1001/jama.285.22.2864]
- Pisters R, Lane DA, Nieuwlaar R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. *Chest* 2010; **138**: 1093-1100 [PMID: 20299623 DOI: 10.1378/chest.10-0134]
- QxMD.com [Internet]. [accessed 2016 Apr 21]. Available from: URL: <http://www.qxmd.com>
- Topol EJ. Medscape.com [Internet]. [accessed 2016 Apr 21]. Available from: URL: <http://www.medscape.com>
- Walker G. MDCalc.com [Internet]. [accessed 2016 Apr 21]. Available from: URL: <http://www.mdcalc.com>
- Fleischmann R, Duhm J, Hupperts H, Brandt SA. Tablet computers with mobile electronic medical records enhance clinical routine and promote bedside time: a controlled prospective crossover study. *J Neurol* 2015; **262**: 532-540 [PMID: 25476692 DOI: 10.1007/s00415-014-7581-7]
- Charles D, Gabriel M, Searcy T. Adoption of Electronic Health Record Systems among U.S. Non-Federal Acute Care Hospitals: 2008-2014 [Internet]. Office of the National Coordinator for Health Information Technology; 2015. Available from: URL: <https://www.healthit.gov/sites/default/files/data-brief/2014HospitalAdoptionDataBrief.pdf>
- Kohn LT, Corrigan JM, Donaldson MS, editors. To Err Is Human: Building a Safer Health System. Washington (DC): National Academy Press; 2000
- Skulmoski FT, Krahn J, Hartman GJ. The Delphi Method for Graduate Research. *JITE* 2007; **6**: 1-21
- Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, Bellomo R, Bernard GR, Chiche JD, Coopersmith CM, Hotchkiss RS, Levy MM, Marshall JC, Martin GS, Opal SM, Rubenfeld GD, van der Poll T, Vincent JL, Angus DC. The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA* 2016; **315**: 801-810 [PMID: 26903338 DOI: 10.1001/jama.2016.0287]
- Dziadzko MA, Gajic O, Pickering BW, Herasevich V. Clinical calculators in hospital medicine: Availability, classification, and needs. *Comput Methods Programs Biomed* 2016; **133**: 1-6 [PMID: 27393794 DOI: 10.1016/j.cmpb.2016.05.006]
- Post TW, editor. Up To Date [Internet]; 2016. Available from: URL: <http://www.uptodate.com>
- Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**: 377-381 [PMID: 18929686 DOI: 10.1016/j.jbi.2008.08.010]
- R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2016. Available from: URL: <https://www.r-project.org/>
- Levine LE, Waite BM, Bowman LL. Mobile media use, multitasking and distractibility. *Int J Cyber Behav Psychol Learn* 2012; **2**: 15-29 [DOI: 10.4018/ijcbpl.2012070102]
- Wells PS, Anderson DR, Rodger M, Forgie M, Kearon C, Dreyer J, Kovacs G, Mitchell M, Lewandowski B, Kovacs MJ. Evaluation of D-dimer in the diagnosis of suspected deep-vein thrombosis. *N Engl J Med* 2003; **349**: 1227-1235 [PMID: 14507948 DOI: 10.1056/NEJMoa023153]
- Wells PS, Anderson DR, Rodger M, Stiell I, Dreyer JF, Barnes D, Forgie M, Kovacs G, Ward J, Kovacs MJ. Excluding pulmonary embolism at the bedside without diagnostic imaging: management of patients with suspected pulmonary embolism presenting to the emergency department by using a simple clinical model and d-dimer. *Ann Intern Med* 2001; **135**: 98-107 [PMID: 11453709 DOI: 10.7326/0003-4819-135-2-200107170-00010]
- Nease DE, Green LA. ClinfoTracker: a generalizable prompting tool for primary care. *J Am Board Fam Pract* 2003; **16**: 115-123 [PMID: 12665177 DOI: 10.3122/jabfm.16.2.115]
- Hirsch JS, Tanenbaum JS, Lipsky Gorman S, Liu C, Schmitz E, Hashorva D, Ervits A, Vawdrey D, Sturm M, Elhadad N.



- HARVEST, a longitudinal patient record summarizer. *J Am Med Inform Assoc* 2015; **22**: 263-274 [PMID: 25352564 DOI: 10.1136/amiajnl-2014-002945]
- 27 **Vincent JL**, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, Suter PM, Thijs LG. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996; **22**: 707-710 [PMID: 8844239 DOI: 10.1007/BF01709751]
- 28 **Thomas M**, Bourdeaux C, Evans Z, Bryant D, Greenwood R, Gould T. Validation of a computerised system to calculate the sequential organ failure assessment score. *Intensive Care Med* 2011; **37**: 557 [PMID: 21152902 DOI: 10.1007/s00134-010-2083-2]
- 29 **Simpson SQ**. New Sepsis Criteria: A Change We Should Not Make. *Chest* 2016; **149**: 1117-1118 [DOI: 10.1016/j.chest.2016.02.653]

**P- Reviewer:** Doglietto F, Tomizawa M **S- Editor:** Kong JX  
**L- Editor:** A **E- Editor:** Lu YJ





Published by **Baishideng Publishing Group Inc**

8226 Regency Drive, Pleasanton, CA 94588, USA

Telephone: +1-925-223-8242

Fax: +1-925-223-8243

E-mail: [bpgoffice@wjgnet.com](mailto:bpgoffice@wjgnet.com)

Help Desk: <http://www.wjgnet.com/esps/helpdesk.aspx>

<http://www.wjgnet.com>

