692 Computer Algorithms Are Superior to Physician Assessment of the Risk for Postoperative Complications When Applied in Interactive Software

Sahil Puri[1]; Dmytro Korenkevych, PhD[2]; Tezcan Ozrazgat-Baslanti, PhD[3]; Ryan Cobb[1]; Petar Momcilovic, PhD[2]; Daisy Zhe Wang, PhD[1]; Azra Bihorac, MD, MS, FASN, FCCM[3]

Departments of [1]Computer & Information Science Engineering, [2]Industrial and Systems Engineering, and [3]Anesthesiology, University of Florida, Gainesville, Florida

Introduction

Postoperative complications (PC) are associated with a two-fold increase in the 30-day mortality and cost and adverse long-term consequences.

We hypothesize that clinical data-driven predictive risk models for PC are more accurate than physicians' risk assessments (RA) and that knowledge exchange between computer algorithms and physicians improves the prediction of physicians.

Methods

Using generalized additive models applied to preoperative clinical data, we have calculated probabilistic risk scores for each PC .

We have developed a prototype of interactive, web-based software that allows the exchange of risk prediction between computer algorithms and physicians in five phases:
1. Present patient information as a case synopsis; 2. Solicit physicians' prediction using risk analog score (RAS); 3. Present a computer's score listing the top features contributing the most to the risk of individual patient; 4. Re-solicit physicians' RA using the same RAS; 5. Ask the expert about her choice of therapy based on RA.

We enrolled eight ICU physicians in the study. Each physician completed numeracy and cognitive reflection tests prior to testing. Using the software, we compared computer algorithms' and physicians' RA for 79 patients for predicting five major postoperative complications:

1. Acute kidney injury defined by Risk, Injury, and Failure; and Loss; and End-stage kidney disease (RIFLE-acute kidney injury [AKI])
2. ICU admission > 48 hours (ICU > 48 hours)
3. Mechanical ventilation > 48 hours (MV > 48 hours)
4. Cardiovascular (CV) complications
5. Severe sepsis

Figure 1: Screenshots of the predictive software.

Table 1. Physicians' characteristics.

Physician Characteristics	Overall (n=8)
Female gender, n (%)	4 (50)
Attending doctor, n (%)	6 (75)
Specialty, n (%)	
Anesthesiology	4 (50)
Medicine	1 (12.5)
Surgery	3 (37.5)
Years since graduation, mean (SD)	13.4 (10.8)
High numeracy score, n (%)	8 (100%)
Cognitive Reflection Test score, n (%) *Measures people's ability to resist their first instinct*	
Low score: impulsive decision making	1 (12.5)
Intermediate score	6 (75)
High score: patient decision making	1 (12.5)

Results

- The AUC of algorithm scores were significantly higher than physicians' first RA for ICU and for severe sepsis (Table 2). All models have good model calibration, with a Hosmer-Lemeshow goodness-of-fit test p-value > 0.05.
- On average, physicians agreed with 8.2 to 8.8 of the top 10 features contributing to the risk of an individual patient selected by computer algorithm.
- The proportion of physicians who changed RA after knowledge exchange with the algorithm was between 56% and 100% (Table 3).
- There was a significant change in physicians' RA for MV > 48 hours, CV complications, and severe sepsis after knowledge exchange (Table 4).
- The AUC of physicians' RA improved by 0.21 for severe sepsis after knowledge exchange with computer algorithm (Table 4).

Table 2. Comparison of physicians' initial RA with the algorithm predictions.

PC	Prevalence of PC, n(%)	Physicians' First RA AUC (95% CI)	Algorithm Prediction AUC (95% CI)	p-value for difference in AUC
RIFLE-AKI	18 (22.8)	0.76 (0.64, 0.88)	0.86 (0.77, 0.95)	0.072
ICU > 48 hours	28 (35.4)	0.72 (0.59, 0.84)	0.84 (0.75, 0.92)	0.047
MV > 48 hours	12 (15.2)	0.79 (0.66, 0.92)	0.90 (0.82, 0.98)	0.157
CV complications	8 (10.1)	0.70 (0.49, 0.9)	0.73 (0.54, 0.91)	0.636
Severe sepsis	4 (5.1)	0.62 (0.44, 0.79)	0.97 (0.93, 1)	<.0001

Abbreviations: AUC, area under the receiver operator characteristic curve; CI, confidence interval.

Table 3. Changes in physicians' RA for events and non-events.

PC	Events		Non-events	
	Physician under-estimated risk initially, n (%)	Physician increased score, n (%)	Physician over-estimated risk initially, n (%)	Physician decreased score, n (%)
RIFLE-AKI	9/18 (50)	5/9 (56)	32/61 (52)	19/32 (59)
ICU > 48 hours	16/28 (57)	12 /16 (75)	27/51 (53)	23/27 (85)
MV > 48 hours	6/12 (50)	5 /6 (83)	45/67 (67)	34/45 (76)
CV complications	2/8 (25)	2 /2 (100)	56/71 (79)	41/56 (73)
Severe sepsis	4/4 (100)	4 /4 (100)	55/75 (73)	38/55 (69)

- Net reclassification improvement for physicians' RA after knowledge exchange with algorithm was significant at 14.1 for MV > 48 hours (Table 4).
- For extreme disagreement cases between physician and algorithm score, the computer was correct in most of the cases (Table 5).

Table 4. Change in physicians' risk score and AUC after knowledge exchange with computer algorithm.

PC	Net classification improvement (95% CI)	Change in physicians risk score mean (SD)		Change in physicians AUC (95% CI)
		Overestimated risk initially	Underestimated risk initially	
RIFLE-AKI	8.48 (−0.38, 17.33)	−0.08 (0.13)[a]	0.07 (0.12)	0.004 (−0.06, 0.07)
ICU > 48 hours	9.27 (−12.55, 31.08)	−0.13 (0.14)[a]	0.10 (0.17)[a]	0.03 (−0.05, 0.11)
MV > 48 hours	14.10 (7.01, 21.18)[a]	−0.10 (0.15)[a]	0.23 (0.25)	0.08 (−0.01, 0.17)
CV complications	6.78 (−0.31, 13.87)	−0.08 (0.12)[a]	0.08 (0.1)	0.05 (−0.07, 0.17)
Severe sepsis	4.01 (−0.70, 8.71)	−0.07 (0.13)[a]	0.17 (0.12)	0.21 (0.07, 0.35)[b]

[a] p-value < 0.05 for change in physicians' scores after knowledge exchange was tested using Wilcoxon signed rank sum test.
[b] p-value < 0.05 for change in AUCs after knowledge exchange using the DeLong test.

Table 5. Performance of physicians and computers for cases with extreme disagreement in risk prediction (absolute difference between scores > 0.40).

	Cases with high discrepancy, n (%)	Cases for which computer was better, n (%)
RIFLE-AKI	13	8 (62)
ICU > 48 hours	14	6 (43)
MV > 48 hours	13	9 (69)
CV complications	16	12 (75)[a]
Severe sepsis	12	12 (100)[a]

[a] p-value < 0.05 for z-test of proportion of times computer was significantly better was higher than 50%
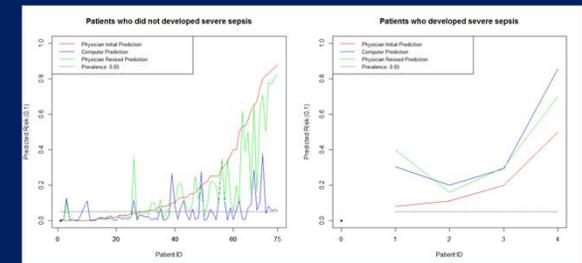

Figure 2. Physicians initial and revised RA compared to the computer algorithms' score.

Conclusion

Clinical data-driven predictive risk models for PC are more accurate than physicians' RA, and the two-way knowledge transfer between algorithms and physicians improves the prediction of physicians.