KI REPORTS
KIReports.org

Developing Biomarker Panels to Predict Progression of Acute Kidney Injury After Cardiac Surgery

Kathleen F. Kerr[1], Eric R. Morenz[1], Jeremy Roth[1], Heather Thiessen-Philbrook[2], Steven G. Coca[3] and Chirag R. Parikh[2]

[1]Department of Biostatistics, University of Washington, Seattle, Washington, USA; [2]Division of Nephrology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; and [3]Division of Nephrology, Icahn School of Medicine at Mount Sinai, New York, New York, USA

Introduction: Acute kidney injury (AKI) is a frequent complication of cardiac surgery, but only a fraction of cardiac surgery patients that experience postoperative AKI have progression to more severe stages. Biomarkers that can distinguish patients that will experience progression of AKI are potentially useful for clinical care and/or the development of therapies.

Methods: Data come from a prospective cohort study of cardiac surgery patients; the analytic dataset contained data from 354 cardiac surgery patients meeting criteria for AKI following surgery. Candidate predictors were 38 biomarkers of kidney function, insult, or injury measured at the time of AKI diagnosis. The outcome was AKI progression, defined as worsening of AKI Network stage. We investigated combining biomarkers with Bayesian model averaging (BMA) and random forests of classification trees, with and without center transformation. For both approaches, we used resampling-based methods to avoid optimistic bias in our assessment of model performance.

Results: BMA yielded a combination of 3 biomarkers and an optimism-corrected estimated area under the receiver operating characteristic curve (AUC) of 0.75 (95% confidence interval [CI]: 0.68, 0.82). The random forests approach, which nominally uses all biomarkers, had an estimated AUC of 0.74 (95% CI: 0.66, 0.82). A second application of random forests applied to biomarker values after a center-specific transformation had an estimated AUC of 0.80 (95% CI: 0.72, 0.88).

Conclusion: These findings suggest that the application of advanced statistical techniques to combine biomarkers offers only modest improvements over use of single biomarkers alone. This exemplifies a common experience in biomarker research: combinations of modestly performing biomarkers often also have modest performance.

Kidney Int Rep (2019) **4,** 1677–1688; https://doi.org/10.1016/j.ekir.2019.08.017
KEYWORDS: acute kidney injury; Bayesian model averaging; biomarkers; cardiac surgery; prognosis; random forest

AKI is common after cardiac surgery and is associated with adverse outcomes.[1] However, only a fraction of cardiac surgery patients who experience postoperative AKI have progression to more severe stages. Identifying cardiac surgery patients who are at high risk of AKI progression could potentially improve patient care or aid in the development of novel treatment strategies.

Numerous biomarkers of injury, inflammation, repair, and fibrosis have been associated with the development of AKI following cardiac surgery.[2–7] Few studies have investigated biomarkers that can be used to predict AKI progression.[8] We used data from a large multicenter cohort of cardiac surgery patients to seek out prognostic biomarkers of AKI progression. We also investigated whether the prognostic capacity of individual biomarkers could be leveraged to produce a biomarker panel with improved prognostic capacity.

Correspondence: Chirag R. Parikh, Division of Nephrology, Ronald Peterson Professor of Medicine, Johns Hopkins School of Medicine, 1830 E. Monument St., Suite 416, Baltimore, Maryland 21287, USA. E-mail: Chirag.Parikh@jhmi.edu

Received 29 May 2019; revised 5 August 2019; accepted 12 August 2019; published online 30 August 2019

METHODS

Study Population

This is a secondary analysis of the Translational Research Investigating Biomarker Endpoints in AKI

(TRIBE-AKI study). Briefly, this study prospectively enrolled 1219 adults undergoing coronary artery bypass graft (CABG) and/or valve surgery at 6 academic medical centers in North America between July 2007 and December 2009. Study details have been described previously.[3] Patients who developed at least AKI Network stage 1 were included in this analysis.

Sample Collection

Urine and ethylenediamine tetraacetic acid (EDTA) plasma specimens were collected preoperatively and daily for up to 5 postoperative days. The first postoperative samples were collected soon after patients were admitted to the intensive care unit (0–6 hours after surgery). Details regarding sample collection and storage were provided in earlier reports.[3]

Biomarkers

The present investigation considers biomarkers measured closest to the time of AKI diagnosis. For each patient, we defined the biomarker value as the most recent nonmissing postoperative value up to and including the day of that patient's AKI diagnosis. For example, suppose a patient is diagnosed with AKI 2 days after surgery but has missing biomarker values on that day. If the patient had nonmissing biomarker values 1 day after surgery, then those values were used as the biomarker measurements. On the other hand, if the patient had no biomarker measurement at any time between cardiac surgery through the day of AKI progression, then the patient's biomarker value was coded as missing. Supplementary Table S1 summarizes the timing of biomarker measurements used in analysis relative to the first day of clinical AKI. Use of the most-recent-carried-forward measurements reduced the number of patients with missing data for one or more biomarkers from 156 to 84 (of 354).

Supplementary Table S2 lists biomarker assays; procedures are detailed in prior publications.[3,4,6,9–11] Our study included 31 blood plasma biomarkers, 7 urine biomarkers, and the change in serum creatinine level (day of AKI diagnosis vs. pre-surgery). Preoperative serum creatinine level, collected about 3 days prior to surgery (interquartile range: 1–11 days), served as a baseline measurement of kidney function. Pre- and postoperative creatinine values were measured by the same laboratory for each patient at all centers. Supplementary Table S1 provides details of the plasma and urine biomarker measurements.

Outcome

The primary outcome was AKI progression, defined as an increase in AKI Network stage from 1 to 2 or 3.

Statistical Methods

We summarized the association of individual biomarkers and AKI progression using odds ratios comparing patients 1 SD apart in the biomarker. We summarized the prognostic capacity of individual biomarkers using the area under the receiver operating characteristic curve (AUC). AUC is a general measure of how well a combination discriminates between patients who progress and those who do not, and we acknowledge that AUC is an incomplete assessment.[12,13]

We considered 40 candidates (31 plasma biomarkers, 7 urine biomarkers, change in serum creatinine level, and cardiopulmonary bypass surgery time) in developing biomarker combinations. Cardiopulmonary bypass surgery time was included as a binary variable indicating a time >120 minutes. Urine biomarkers were not normalized to urine creatinine level, although urine creatinine level was included as a candidate predictor. All biomarkers were log-transformed, except for serum creatinine level.

Biomarker Combination Methods (BMA and Random Forests)

We used 2 complementary methods to seek biomarker combinations. First, we used methods based on BMA[14] to identify combinations of biomarkers and cardiopulmonary bypass surgery time. Briefly, BMA assigns a prior probability to each candidate predictor that reflects the chance that the predictor is useful for prediction. These prior probabilities induce a prior probability for each biomarker combination. The method uses these prior probabilities and the data to calculate a posterior probability for each candidate predictor. This posterior variable probability reflects the support in the data for the variable as a predictor of the outcome. The BMA methods use a "leaps and bounds" algorithm to strategically search the large space of all possible biomarker combinations to identify the most-promising combinations for further consideration. With 40 candidate predictors, there are >1 trillion possible combinations; the search algorithm makes exploring this space computationally feasible.

In our implementation of BMA, we assigned each candidate predictor a prior probability of one-half, implying that each candidate was *a priori* as likely to be in the model as not. These prior probabilities induce a prior probability for each combination of $(0.5)^{40} = 9.1 \times 10^{-13}$, because there are 40 candidate predictors. In our application of BMA, we selected the combination of all biomarkers with posterior probability exceeding 0.5, which is sometimes called the median probability combination.

Note that BMA seeks linear combinations of biomarkers. In particular, it does not easily handle interactions among biomarkers. As a complementary approach, we also used the random forests method of aggregating classification trees. In this ensemble approach, we took 1000 bootstrap samples of our data and fit a classification decision tree to each sample using a randomly selected set of 6 biomarkers. In each bootstrap sample, we used Gini impurity as the splitting criterion. Predicted risks of AKI progression from a single tree can be obtained as the frequency of the outcome in the nodes of the tree.[15] To obtain predicted risks for individual patients from a collection of trees, we averaged these predicted risks using only the trees that, due to the random bootstrap sampling, did not include that patient ("out-of-bag" trees). Obtaining predictions by averaging across out-of-bag trees should protect against overfitting and avoid optimistic bias in assessment of performance.

It is difficult to understand the role of individual predictors in a random forest, which is sometimes referred to as a "black box" prediction instrument. As a first-pass measure of the importance of individual predictors, we repeated the entire random forest procedure 40 times, once for each candidate predictor, excluding one candidate predictor each time. We measured the decrease in estimated AUC when an individual biomarker was excluded. We caution against overinterpretation of this measure of variable importance. For example, suppose there are 2 biomarkers that are each highly predictive but are also highly correlated with each other. Excluding one of the biomarkers as a candidate predictor will have little effect on the performance of the random forest, but this will understate the importance of the marker.

Optimism Correction

For both BMA and random forests, we used methods to correct for optimistic bias that arises when the same data are used to both fit a model and then evaluate the model's performance. For BMA, we used a bootstrapping procedure to estimate and correct for optimism in estimates of combination performance due to resubstitution bias and model-selection bias (more details follow).[16] For random forests, we used out-of-bag trees to obtain predicted risks for patients, in order to estimate model performance. That is, for a given observation, we used all trees that did not use that observation to grow the tree to make a prediction for that observation, and averaged those predictions.

Multiple Imputations of Missing Biomarker Data

There was a non-negligible amount of missing biomarker data, which was a particular consideration for our investigation of biomarker combinations. Among the 83 individuals with at least one missing biomarker measurement, 14.5% experienced progression. Among the 271 patients with no missing data, 10.0% experienced AKI progression.

Table 1. Demographics and clinical variables of study population as a whole and stratified on acute kidney injury (AKI) progression status

		AKI progression	
Patient characteristic	**Overall (n = 354)**	**Event (n = 39)**	**Non-event (n = 315)**
Demographics			
Age (yr), mean (SD)	72 (9.7)	72 (9.6)	72 (9.7)
Male sex	255 (72)	26 (67)	229 (73)
White race	333 (94)	36 (92)	297 (94)
Clinical variables			
Preoperative eGFR (ml/min per 1.73 m^2), mean (SD)	63.3 (19.8)	57.1 (23.6)	64.1 (19.0)
Diabetes	161 (46)	21 (54)	140 (44)
Hypertension	297 (84)	35 (90)	262 (83)
Congestive heart failure	120 (34)	18 (46)	102 (32)
Type of surgery			
CABG or valve	267 (75)	30 (77)	237 (75)
CABG and valve	87 (25)	9 (23)	78 (25)
Status of procedure			
Elective	255 (72)	25 (64)	230 (73)
Emergent or urgent	99 (28)	14 (36)	85 (27)
Cardiac catheterization <48 h prior to surgery	16 (4.5)	2 (5.1)	14 (4.5)
Preoperative myocardial infarction	94 (27)	12 (31)	82 (27)
Reoperation	47 (13)	8 (21)	39 (12)
CPB > 120 (min)	172 (64)	23 (59)	149 (48)

CABG, coronary artery bypass grafting, CBP, cardiopulmonary bypass; eGFR, estimated glomerular filtration rate.
Values are n (%), unless otherwise indicated.
There were 2 missing data values for cardiac catheterization, 5 for preoperative myocardial infarction, and 5 for CPB surgery time.

We used a multiple imputation approach to retain the 83 patients with missing biomarker data in our search for biomarker combinations. Next, we provide an overview of our approach for BMA analysis and provide details in the online Supplementary Methods.

First, we created 10 versions of our dataset into which missing biomarker data were imputed. Second, in each imputed dataset, we used BMA to estimate the posterior probability of every candidate predictor. Third, we identified the predictors for which the posterior probability estimates, averaged over the 10 imputed datasets, exceeded 0.5. Fourth, we fit a logistic regression model using the selected predictors in each of the 10 imputed datasets. For each predictor, the final regression coefficient was

Table 2. Summary of biomarkers included in this investigation

		AKI progression	
Biomarker	**Overall (N = 354)**	**Event (n = 39)**	**Non-event (n = 315)**
Postoperative serum creatinine level (mg/dl)	1.3 (1.1, 1.5)	1.4 (1.2, 1.9)	1.2 (1, 1.5)
Change in serum creatinine level (mg/dl)	0.4 (0.3, 0.5)	0.5 (0.4, 0.7)	0.4 (0.3, 0.5)
Urine markers (most recent measurement up to AKI diagnosis)			
Creatinine (mg/dl)	67 (36.1, 112)	41.7 (24, 72.3)	70.7 (38.8, 121)
IL-18 (pg/dl)	49.5 (13.6, 143)	70.4 (14.9, 314)	46.4 (13.8, 125)
NGAL (ng/dl)	27.9 (12.7, 85.2)	62.4 (8.7, 503)	27.8 (13.6, 82.8)
Albumin (mg/dl)	31 (13.5, 68.5)	44.1 (19.6, 79.2)	28.9 (13.1, 62.8)
KIM-1 (ng/dl)	3.1 (1.1, 7.3)	2.4 (1.1, 4.3)	3.2 (1.1, 7.9)
L-FABP (ng/ml)	15.5 (3.4, 67.4)	21.4 (5.6, 371)	15.2 (3.3, 60.9)
Cystatin C (mg/l)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)	0.2 (0.1, 0.3)
Plasma markers (most recent measurement up to AKI diagnosis)			
Pro BNP (pmol/l)	255 (106, 568)	281 (103, 440)	253 (115, 572)
IL-6	60.5 (34.6, 105)	107 (61.8, 219)	57 (32.7, 97.4)
TNI (ug/l)	2.2 (1.1, 4.7)	3.3 (1.9, 5.7)	2 (1, 4.5)
TNTHS (ug/l)	564 (332, 1055)	921 (484, 1482)	530 (330, 946)
CKMB (ug/l)	20.5 (10, 39)	34.4 (20.7, 56.4)	19 (9.6, 36.2)
hFABP	30.8 (19.8, 58.2)	99.4 (43.5, 141)	28.4 (18.7, 51.2)
MCP-1	292 (203, 451)	421 (295, 537)	278 (198, 445)
EGF	0.9 (0.9, 3.1)	0.9 (0.9, 0.9)	0.9 (0.9, 3.5)
VEGF	76.6 (25.1, 137)	38.9 (3.9, 94)	79.3 (31.4, 139)
bFGF	6.1 (3, 11.8)	9.1 (3.9, 15.2)	5.8 (3, 11)
PlGF	32.7 (22, 47.3)	27.9 (15.5, 40.7)	33.2 (22.7, 48)
Tie-2	3688 (3125, 4261)	3810 (3010, 4257)	3679 (3132, 4263)
VEGF R1	255 (122, 660)	704 (261, 1208)	229 (117, 570)
VEGFd	344 (275, 452)	326 (271, 434)	349 (276, 452)
VEGFc	112 (88, 146)	134 (102, 165)	111 (88, 143)
IFNg	7.1 (4, 12.1)	11.1 (8.4, 20.3)	6.7 (3.8, 11.6)
IL-10	6.3 (2.3, 35.1)	26.9 (4.4, 117)	5.6 (2.3, 28.9)
IL-12 p70	1.5 (0.9, 2.7)	2 (1.3, 5.2)	1.4 (0.9, 2.5)
IL-13	4.3 (2.4, 7.4)	6.8 (4.1, 10.4)	4.1 (2.3, 7)
IL-1b	0.4 (0.2, 0.7)	0.7 (0.5, 1.5)	0.4 (0.2, 0.7)
IL-2	0.9 (0.5, 1.8)	1.8 (0.9, 3.4)	0.9 (0.5, 1.6)
IL-4	0.3 (0.1, 0.5)	0.4 (0.3, 0.8)	0.2 (0.1, 0.4)
IL-6	60.5 (34.6, 105)	107 (61.8, 219)	57 (32.7, 97.4)
IL-8	20.2 (11.8, 36.2)	47.4 (19.8, 96.4)	19 (11.2, 32.6)
TNFα	4.1 (2.9, 6.1)	5.9 (4.4, 9.2)	3.9 (2.8, 5.8)
IL-18	468 (374, 574)	472 (367, 571)	467 (378, 574)
KIM-1	323 (260, 411)	321 (269, 387)	323 (257, 412)
TNFR-1 (ng/ml)	6.8 (4.8, 10)	10.8 (7.2, 13.1)	6.5 (4.7, 9.4)
TNFR-2 (ng/ml)	9.1 (6.9, 12.4)	12.2 (8.7, 15.4)	8.8 (6.8, 11.8)
ST-2 (ng/ml)	95.6 (10.1, 187)	43.2 (8.1, 124)	100 (11.5, 190)
Galcetin-3 (ng/ml)	11.5 (8.1, 16.4)	16 (9, 21.1)	11 (8, 15.7)

AKI, acute kidney injury; bFGF, basic fibroblast growth factor; CKMB, creatine kinase-MB; EGF, epidermal growth factor; HFABP, heart-type fatty acid-binding protein; IFNg, human interferon gamma; KIM-1, kidney injury molecule-1; L-FABP, liver fatty acid-binding protein; MCP-1, monocyte chemoattractant protein-1; NGAL, neutrophil gelatinase-associated lipocalin; PlGF, placental growth factor; Pro BNP, pro-B-type natriuretic peptide; ST-2, soluble ST2; Tie-2, tyrosine kinase-2; TNFα, human tumor necrosis factor alpha; TNFR-1, tumor necrosis factor receptor 1; TNFR-2, tumor necrosis factor receptor 2; TNI, troponin I; TNTHS, high-sensitivity troponin T; VEGF, vascular endothelial growth factor; VEGFc, vascular endothelial growth factor-C; VEGFd, vascular endothelial growth factor-D; VEGF-R1, vascular endothelial growth factor receptor-1.
Units are pg/ml unless otherwise indicated. Values are median (interquartile range).
Urine biomarkers and serum creatinine were missing for 4 patients. Plasma biomarker values were missing for 27 patients, except for the following biomarkers, which were missing for 71–73 patients: Pro BNP, TNI, TNTHS, CKMB HFABP, MCP-1.

the average regression coefficient across the 10 imputed datasets. Fifth, we calculated an empirical AUC based on these regression coefficients in each imputed dataset, and averaged these AUC values. In order to correct this estimate of performance for optimistic bias, we took independent bootstrap samples from each imputed dataset, re-running the steps above in each bootstrap sample. Obtaining estimated regression coefficients in a bootstrap dataset, and assessing the combination performance in both the bootstrap dataset and the original dataset, allowed us to estimate and average optimism using a total of 1000 bootstrap datasets. We subtracted this estimated optimism from our nominal estimate of performance to produce an optimism-corrected estimate of performance.

Methods to Handle Study-Center Effect

We aimed to additionally explore the effect of study center on results. Since BMA uses linear combinations of biomarkers, the natural method of center-adjusted performance is to force "center" as a covariate in the linear combinations. However, due to a small number of patients who progressed to AKI in some centers, this approach was not computationally feasible when combined with our resampling approach (described later). For random forests, this integrated approach to center-adjustment is not possible generally. Therefore, to adjust for center with random forests we transformed biomarkers to the percentile scale, using the distribution of the biomarker among those who did not progress to AKI in the same center as the reference distribution. In summary, we applied the random forests approach twice. First, we applied the algorithm to the biomarkers in the same form we used for BMA. In a second application, we applied random forests to the center-specific transformed biomarkers. We note that classification trees, unlike linear model methods such as BMA, are invariant to monotone transformations of predictors. Because the center-specific transformation procedure involves a highly variable nonlinear transformation, we

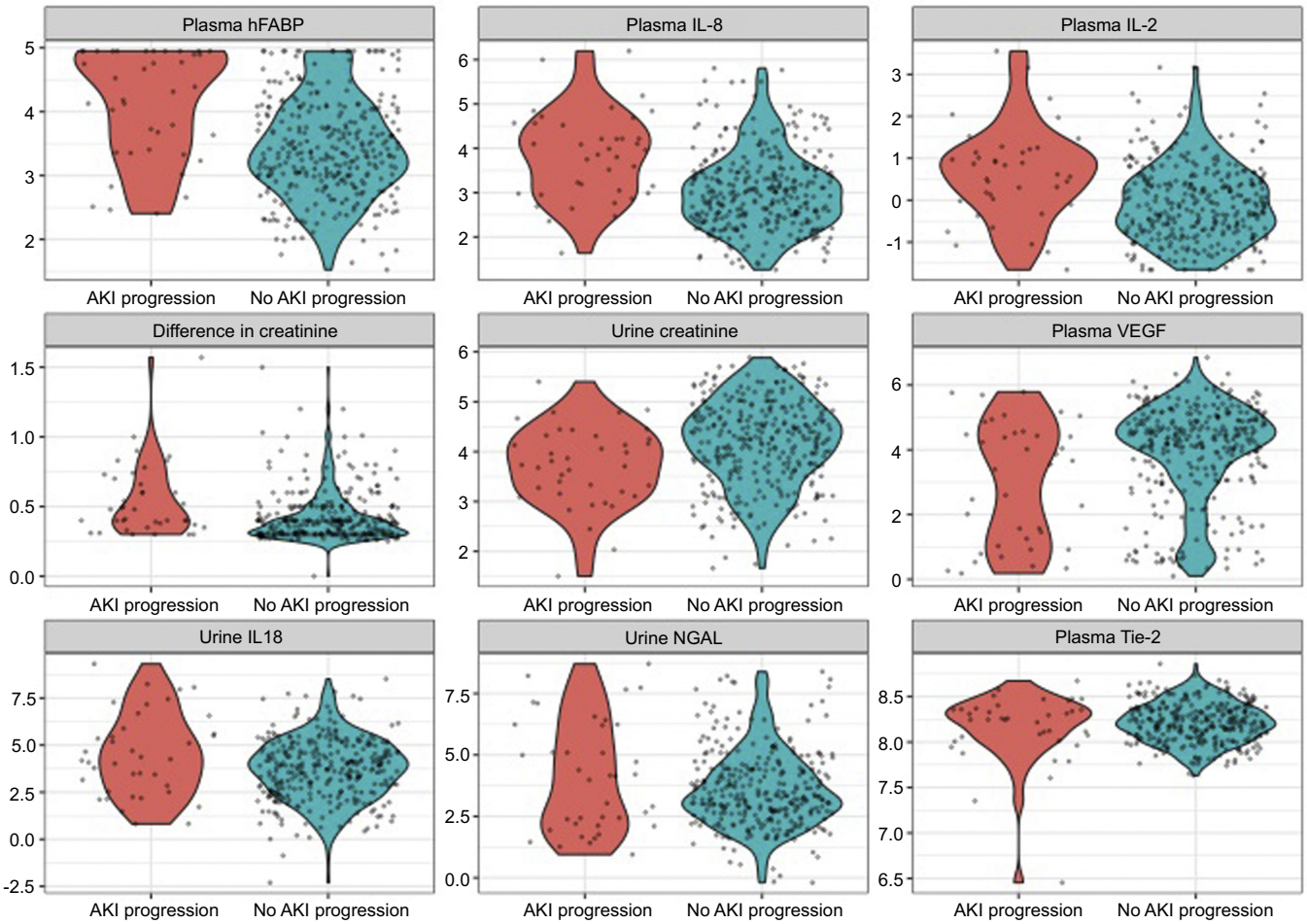

Figure 1. Distributions of select biomarkers in patients who had versus did not have progression of acute kidney injury (AKI). See Supplementary Figure S1 for all biomarkers. hFABP, heart-type fatty acid-binding protein; NGAL, neutrophil gelatinase-associated lipocalin; Tie-2, tyrosine kinase-2; VEGF, vascular endothelial growth factor.

did not apply BMA to the center-specific transformed biomarker values.

Assessing Variability in Predictive Model Performance

For the model selected by BMA, within each complete dataset arising from multiple imputations, we performed bootstrapping to estimate the AUC variance of the fitted BMA combination. We combined these AUC variances across multiple imputation datasets using the Rubin rule.[17] This gave an estimated AUC variance that we used to construct an AUC confidence interval, which we corrected according to the estimated optimism (see earlier description).

For the random forest prediction models, within each complete dataset arising from multiple imputations, we ran the forest-building algorithm and estimated its AUC using out-of-bag trees for each observation as described. We estimated the variance of these AUCs using DeLong's method. We combined these estimates using the Rubin rule,[17] and then constructed an AUC confidence interval.

Table 3. Area under the ROC curve (AUC) and odds ratios (ORs) for individual biomarkers measured in this study, with 95% confidence intervals

	Original scale		Center-specific transformation	
Log-transformed biomarker	Mean AUC	OR per SD	Mean AUC	OR per SD
Plasma hFABP	0.73 (0.64, 0.82)	2.48 (1.70, 3.61)	0.70 (0.61, 0.79)	2.16 (1.46, 3.18)
Plasma IL-8	0.72 (0.64, 0.81)	2.16 (1.56, 3.00)	0.70 (0.61, 0.78)	2.17 (1.46, 3.22)
Plasma TNF-R1	0.72 (0.63, 0.81)	2.29 (1.57, 3.33)	0.70 (0.61, 0.79)	2.14 (1.44, 3.18)
Plasma VEGF R1	0.70 (0.61, 0.79)	1.90 (1.38, 2.62)	0.69 (0.60, 0.78)	2.10 (1.42, 3.11)
Plasma IL1b	0.70 (0.61, 0.78)	2.06 (1.47, 2.87)	0.68 (0.60, 0.77)	1.99 (1.36, 2.92)
Plasma IFNg	0.69 (0.60, 0.78)	1.89 (1.38, 2.60)	0.67 (0.58, 0.75)	1.88 (1.29, 2.73)
Plasma TNFα	0.69 (0.60, 0.77)	1.72 (1.27, 2.33)	0.66 (0.57, 0.75)	1.79 (1.24, 2.60)
Plasma IL6	0.69 (0.59, 0.78)	2.08 (1.47, 2.95)	0.67 (0.58, 0.76)	1.92 (1.32, 2.79)
Plasma IL-2	0.69 (0.59, 0.78)	1.98 (1.43, 2.75)	0.68 (0.58, 0.78)	1.93 (1.32, 2.81)
Delta serum creatinine	0.68 (0.59, 0.76)	1.59 (1.22, 2.08)	0.64 (0.55, 0.73)	1.77 (1.21, 2.60)
Plasma MCP-1	0.67 (0.59, 0.76)	1.67 (1.21, 2.30)	0.64 (0.56, 0.73)	1.71 (1.19, 2.47)
Plasma IL13	0.67 (0.58, 0.76)	1.62 (1.17, 2.24)	0.65 (0.55, 0.74)	1.72 (1.20, 2.47)
Plasma IL-4	0.67 (0.58, 0.76)	1.83 (1.34, 2.52)	0.64 (0.55, 0.73)	1.66 (1.16, 2.38)
Urine creatinine	0.66 (0.58, 0.74)	0.59 (0.42, 0.81)	0.70 (0.62, 0.78)	0.47 (0.31, 0.69)
Plasma IL10	0.65 (0.56, 0.75)	1.73 (1.24, 2.40)	0.65 (0.56, 0.74)	1.75 (1.21, 2.52)
Plasma TNF-R2	0.65 (0.56, 0.75)	1.72 (1.21, 2.45)	0.63 (0.54, 0.73)	1.62 (1.13, 2.32)
Plasma CKMB	0.65 (0.56, 0.74)	1.67 (1.17, 2.38)	0.64 (0.54, 0.73)	1.65 (1.15, 2.37)
Plasma IL12p70	0.65 (0.55, 0.75)	1.73 (1.26, 2.37)	0.62 (0.52, 0.72)	1.55 (1.09, 2.21)
Plasma galcetin-3	0.64 (0.53, 0.74)	1.59 (1.10, 2.30)	0.62 (0.51, 0.73)	1.54 (1.08, 2.19)
Plasma VEGF	0.63 (0.53, 0.73)	0.61 (0.45, 0.82)	0.61 (0.51, 0.72)	0.66 (0.47, 0.94)
Plasma PlGF	0.62 (0.52, 0.72)	0.63 (0.47, 0.85)	0.63 (0.53, 0.73)	0.62 (0.44, 0.89)
Plasma TNTHS	0.62 (0.52, 0.71)	1.47 (1.05, 2.05)	0.62 (0.52, 0.72)	1.52 (1.07, 2.17)
Plasma bFGF	0.61 (0.52, 0.70)	1.42 (1.04, 1.95)	0.61 (0.52, 0.71)	1.51 (1.06, 2.14)
Plasma VEGFc	0.61 (0.51, 0.70)	1.40 (1.01, 1.93)	0.61 (0.51, 0.71)	1.48 (1.04, 2.10)
Plasma TNI	0.60 (0.51, 0.70)	1.40 (0.99, 1.96)	0.61 (0.52, 0.70)	1.50 (1.06, 2.14)
Urine IL18	0.59 (0.48, 0.70)	1.58 (1.13, 2.21)	0.55 (0.44, 0.66)	1.18 (0.85, 1.65)
Urine KIM1	0.58 (0.50, 0.66)	0.81 (0.59, 1.13)	0.57 (0.48, 0.65)	0.78 (0.56, 1.10)
Urine albumin	0.57 (0.48, 0.67)	1.18 (0.83, 1.66)	0.56 (0.46, 0.66)	1.24 (0.88, 1.74)
Urine LFABP	0.57 (0.47, 0.68)	1.30 (0.92, 1.84)	0.55 (0.43, 0.66)	1.19 (0.85, 1.67)
Plasma ST2	0.57 (0.47, 0.67)	0.79 (0.58, 1.09)	0.54 (0.44, 0.64)	0.88 (0.63, 1.23)
Plasma YKL-40	0.56 (0.47, 0.66)	0.89 (0.65, 1.21)	0.56 (0.47, 0.66)	0.80 (0.57, 1.12)
Plasma VEGFd	0.54 (0.43, 0.64)	0.91 (0.65, 1.27)	0.52 (0.42, 0.62)	0.91 (0.65, 1.28)
Urine NGAL	0.54 (0.42, 0.66)	1.36 (1.00, 1.86)	0.51 (0.39, 0.63)	0.97 (0.69, 1.35)
Plasma KIM1	0.51 (0.42, 0.61)	0.96 (0.69, 1.36)	0.51 (0.42, 0.61)	1.01 (0.72, 1.41)
Plasma PROBNP	0.51 (0.42, 0.60)	1.04 (0.74, 1.46)	0.49 (0.40, 0.58)	1.03 (0.74, 1.43)
Urine CysC	0.51 (0.41, 0.61)	1.06 (0.76, 1.50)	0.53 (0.42, 0.63)	0.90 (0.64, 1.26)
Plasma IL18	0.50 (0.40, 0.60)	0.89 (0.61, 1.29)	0.54 (0.44, 0.64)	0.87 (0.62, 1.21)
Plasma Tie-2	0.48 (0.38, 0.59)	0.77 (0.57, 1.03)	0.47 (0.37, 0.57)	0.90 (0.64, 1.25)
Plasma EGF	0.46 (0.38, 0.53)	0.85 (0.59, 1.24)	0.55 (0.46, 0.63)	1.30 (0.88, 1.90)

bFGF, basic fibroblast growth factor; CKMB, creatine kinase-MB; CysC, cystatin C; EGF, epidermal growth factor; hFABP, heart-type fatty acid-binding protein; IFNg, human interferon gamma; KIM-1, kidney injury molecule-1; LFABP, liver fatty acid-binding protein; MCP-1, monocyte chemoattractant protein-1; NGAL, neutrophil gelatinase-associated lipocalin; PlGF, placental growth factor; PROBNP, pro-B-type natriuretic peptide; ST-2, soluble ST2; Tie-2, tyrosine kinase-2; TNFα, human tumor necrosis factor alpha; TNF-R1, tumor necrosis factor receptor 1; TNF-R2, tumor necrosis factor receptor 2; TNI, troponin I; TNTHS, high-sensitivity troponin T; VEGF, vascular endothelial growth factor; VEGFc, vascular endothelial growth factor-C; VEGFd, vascular endothelial growth factor-D; VEGF-R1, vascular endothelial growth factor receptor-1.
ORs are per biomarker SD. We report mean AUC and OR values across the 10 imputed datasets resulting from our multiple imputation procedure.

Software

All analyses were completed using R 3.1.2; the *BMA* package was used for the BMA analyses, the *ranger* package for random forests, the *mice* package for multiple imputations, and the *pROC* package for AUC variances.

RESULTS

Table 1 describes the characteristics of the study population by AKI progression status. There were 354 patients in the full dataset, including 39 patients who progressed to a higher AKI stage. All patients were diagnosed with stage 1 AKI either 1, 2, or 3 days after surgery (33%, 38%, and 29%, respectively). Among the 39 patients with AKI progression, 21 (54%) progressed from stage 1 to stage 2, and 18 (46%) progressed from stage 1 to stage 3. Thirteen of the 18 stage-3 patients received dialysis. Fourteen (4%) patients died in the hospital; 7 each in the subsets of patients with and without AKI progression. Table 2 and Figure 1 summarize the distribution of the biomarkers in those with and without AKI progression (see also Supplementary Figure S1). Table 3 summarizes biomarker AUC and odds ratios (ORs) on the original scale and after the center-specific transformation. A handful of biomarkers show modest predictive capacity in the data, with point estimates for AUC in the range 0.68 to 0.72 and with wide confidence intervals. In terms of individual AUC, the center-specific transformation appeared to lower AUC values.

Applying BMA to develop biomarker combinations, 3 markers had posterior variable probabilities exceeding 0.5. The markers and their mean posterior probabilities across multiple imputation datasets

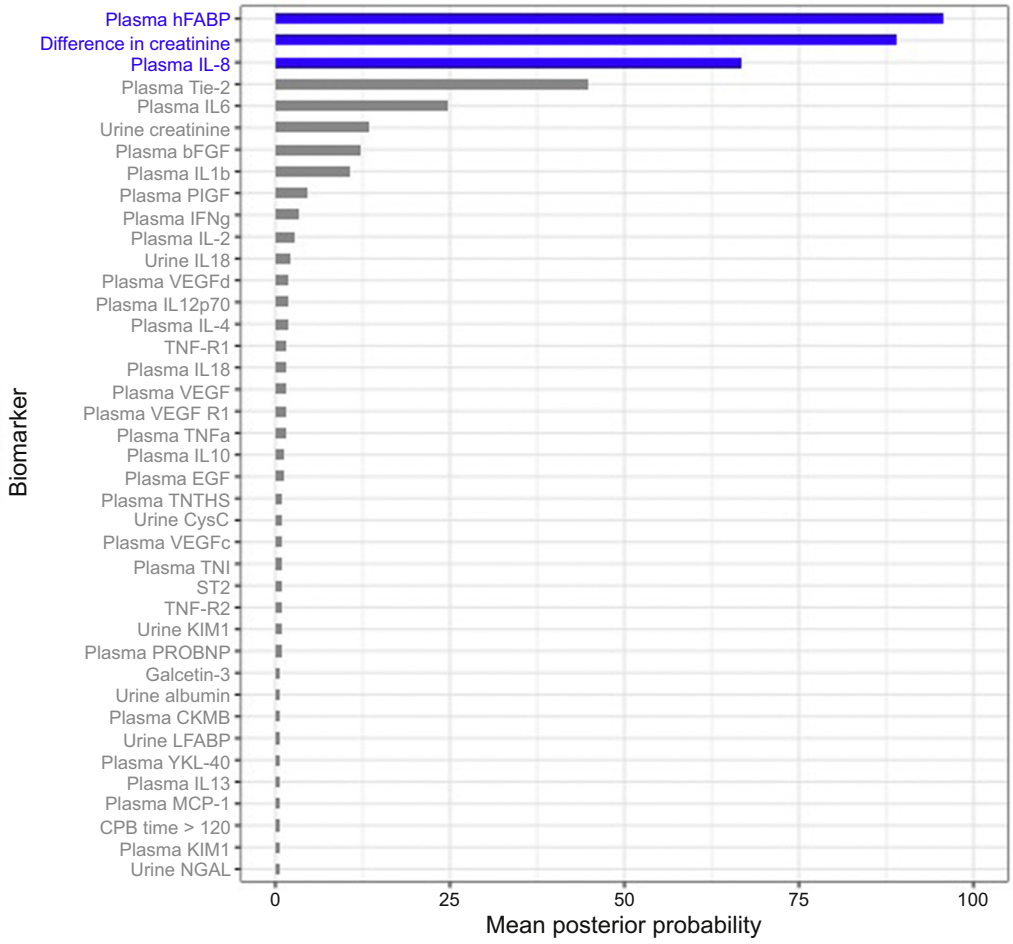

Figure 2. Biomarker posterior probabilities using Bayesian model averaging. The 3 biomarkers in blue have posterior probabilities exceeding 50%; these biomarkers are included in the Bayesian model averaging combination. bFGF, basic fibroblast growth factor; CKMB, creatine kinase-MB; hFABP, heart-type fatty acid-binding protein; IFNg, human interferon gamma; KIM1, kidney injury molecule-1; LFABP, liver fatty acid-binding protein; MCP-1, monocyte chemoattractant protein-1; NGAL, neutrophil gelatinase-associated lipocalin; PIGF, placental growth factor; PROBNP, pro-B-type natriuretic peptide; ST-2, soluble ST2; Tie-2, tyrosine kinase-2; TNFα, human tumor necrosis factor alpha; TNF-R1, tumor necrosis factor receptor 1; TNI, troponin I; TNTHS, high-sensitivity troponin T; VEGF, vascular endothelial growth factor; VEGF R1, vascular endothelial growth factor receptor 1.

were the following: plasma heart-type fatty acid-binding protein (95.8%), change in serum creatinine level (89.0%), and plasma IL-8 (66.8%). Figure 2 displays the posterior probabilities for all biomarkers. The optimism-corrected AUC estimate for the 3-biomarker combination was 0.75 (95% CI: 0.68, 0.82). A random forest of classification trees built using untransformed biomarkers had an AUC of 0.74 (95% CI: 0.66, 0.82). According to our measure of variable importance for random forests, the most important biomarkers were plasma IL-8, heart-type fatty acid-binding protein, Tie-2, and urine neutrophil gelatinase-associated lipocalin (Figure 3). When we built a random forest using the biomarkers after the center-specific transformation, the estimated AUC was 0.80 (95% CI: 0.72, 0.88). The strongest predictors after center transformation were plasma Tie-2, followed by plasma heart-type fatty acid-binding protein, urine creatinine, and tumor necrosis factor receptor 1 (Figure 4).

DISCUSSION

Multiple biomarkers were associated with AKI progression among patients diagnosed with AKI following cardiac surgery. However, these associations do not imply high prognostic capacity.[18] Of the 38 biomarkers we examined, the most prognostic individual biomarker had an AUC of only 0.73. The modest prognostic capacity of individual biomarkers motivated our investigation of whether a combination of multiple biomarkers could improve prediction.

Figure 5 summarizes this investigation. The 2 methods we applied—BMA and random forests of classification trees—are complementary approaches in the sense that BMA seeks simple linear combinations of candidate predictors and parsimonious models,

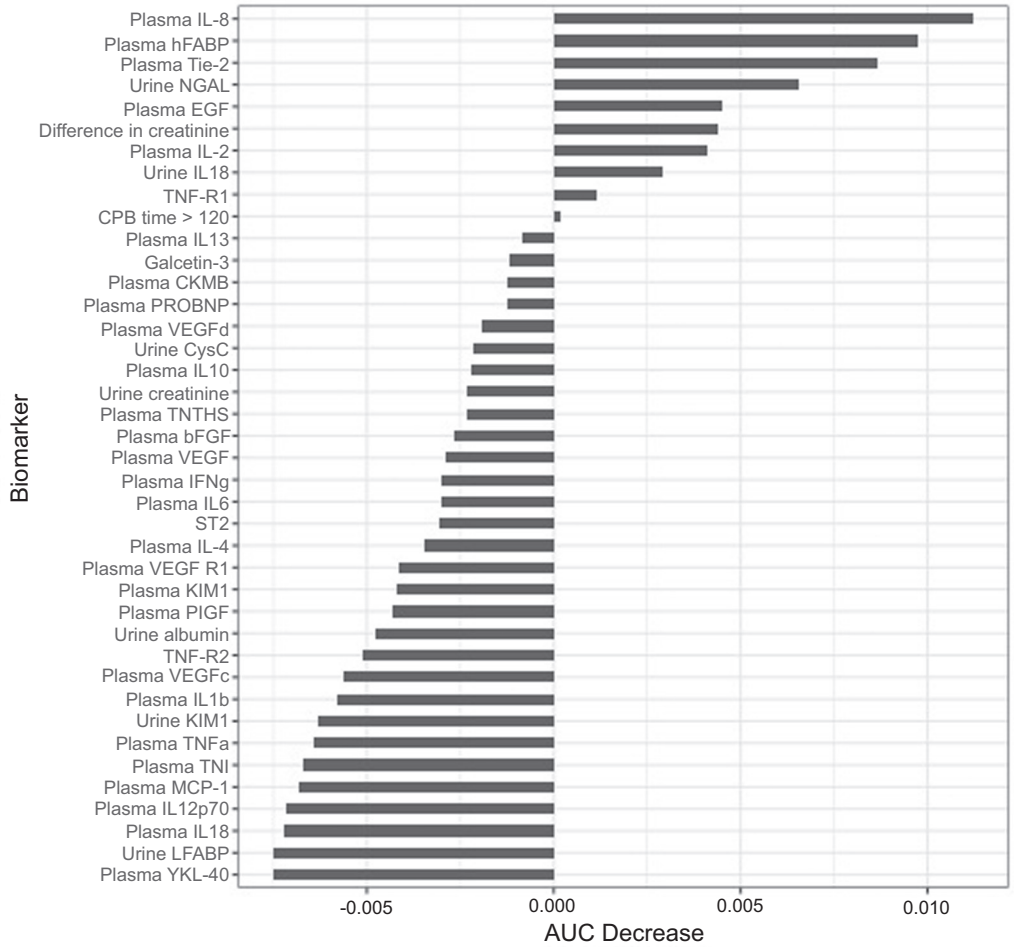

Figure 3. Biomarker importance in the random forest applied to untransformed biomarker data. AUC, area under the receiver operating characteristic curve; bFGF, basic fibroblast growth factor; CKMB, creatine kinase-MB; CPB, cardiopulmonary bypass; CysC, cystatin C; EGF, epidermal growth factor; hFABP, heart-type fatty acid-binding protein; IFNg, human interferon gamma; KIM1, kidney injury molecule-1; LFABP, liver fatty acid-binding protein; MCP-1, monocyte chemoattractant protein-1; NGAL, neutrophil gelatinase-associated lipocalin; PIGF, placental growth factor; PROBNP, pro-B-type natriuretic peptide; ST2, soluble ST2; Tie-2, tyrosine kinase-2; TNF-α, human tumor necrosis factor alpha; TNF-R1, tumor necrosis factor receptor 1; TNF-R2, tumor necrosis factor receptor 2; TNI, troponin I; TNTHS, high-sensitivity troponin T; VEGF, vascular endothelial growth factor; VEGF R1, vascular endothelial growth factor receptor 1; VEGFc, vascular endothelial growth factor-C; VEGFd, vascular endothelial growth factor-D.

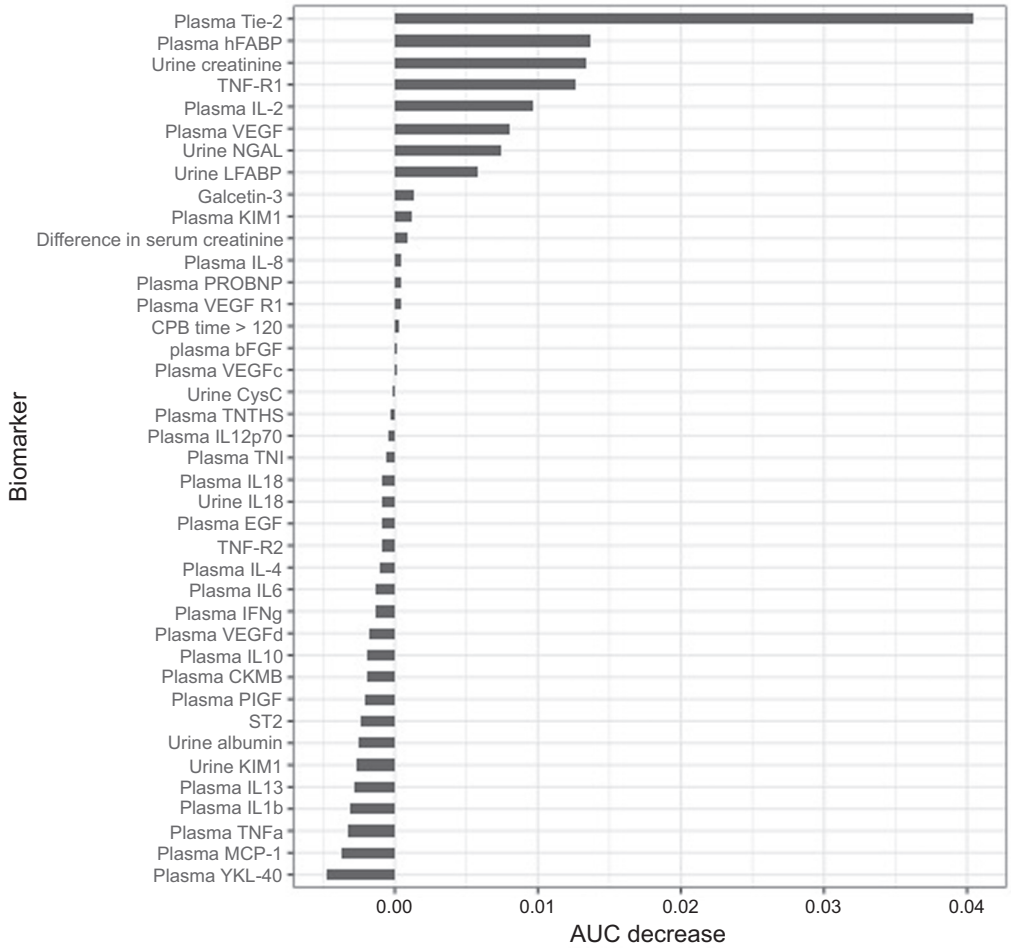

Figure 4. Biomarker importance in the random forest applied to the center-specific transformed marker data. AUC, area under the receiver operating characteristic curve; bFGF, basic fibroblast growth factor; CPB, cardiopulmonary bypass; CKMB, creatine kinase-MB; CysC, cystatin C; EGF, epidermal growth factor; hFABP, heart-type fatty acid-binding protein; IFNg, human interferon gamma; KIM1, kidney injury molecule-1; LFABP, liver fatty acid-binding protein; MCP-1, monocyte chemoattractant protein-1; NGAL, neutrophil gelatinase-associated lipocalin; PIGF, placental growth factor; PROBNP, pro-B-type natriuretic peptide; ST2, soluble ST2; Tie-2, tyrosine kinase-2; TNF-α, human tumor necrosis factor alpha; TNF-R1, tumor necrosis factor receptor 1; TNF-R2, tumor necrosis factor receptor 2; TNI, troponin I; TNTHS, high-sensitivity troponin T; VEGF, vascular endothelial growth factor; VEGFc, vascular endothelial growth factor-C; VEGFd, vascular endothelial growth factor-D; VEGF R1, vascular endothelial growth factor receptor 1.

whereas random forests are well suited to situations in which interactions among candidate predictors are important and the model includes all biomarkers. Although the random forest approach uses all candidate predictors, the bootstrap aggregation component of the algorithm protects against overfitting. Despite these complementary approaches, we did not identify a biomarker combination that was clearly superior to the best-performing individual biomarkers.

A disadvantage of a random forest is that it is a "black box" approach—it is difficult to understand the impact of any given biomarker on predictions. Using our measure of variable importance for random forests, the important biomarkers were roughly similar between the methods.

Although it is counterintuitive, biomarkers that perform best in combination are not necessarily those that have the strongest individual performance.[19,20] It is mathematically possible for a biomarker to have null individual prognostic capacity yet contribute substantially to prediction in combination with other variables. On its own, Tie-2 had almost no predictive capacity in our dataset, yet it had the fourth-highest mean posterior probability in the BMA analysis and appeared as one of the top-performing biomarkers in the random forest analyses. Furthermore, although many share the intuition that combinations of modestly performing biomarkers should be able to achieve substantially better prediction, theoretical results teach us that this intuition may be misguided.[19,20]

The predictive capacity of individual biomarkers in our data tended to be lower after the center-specific transformation. This could suggest that the variability introduced by the transformation hurt performance. An

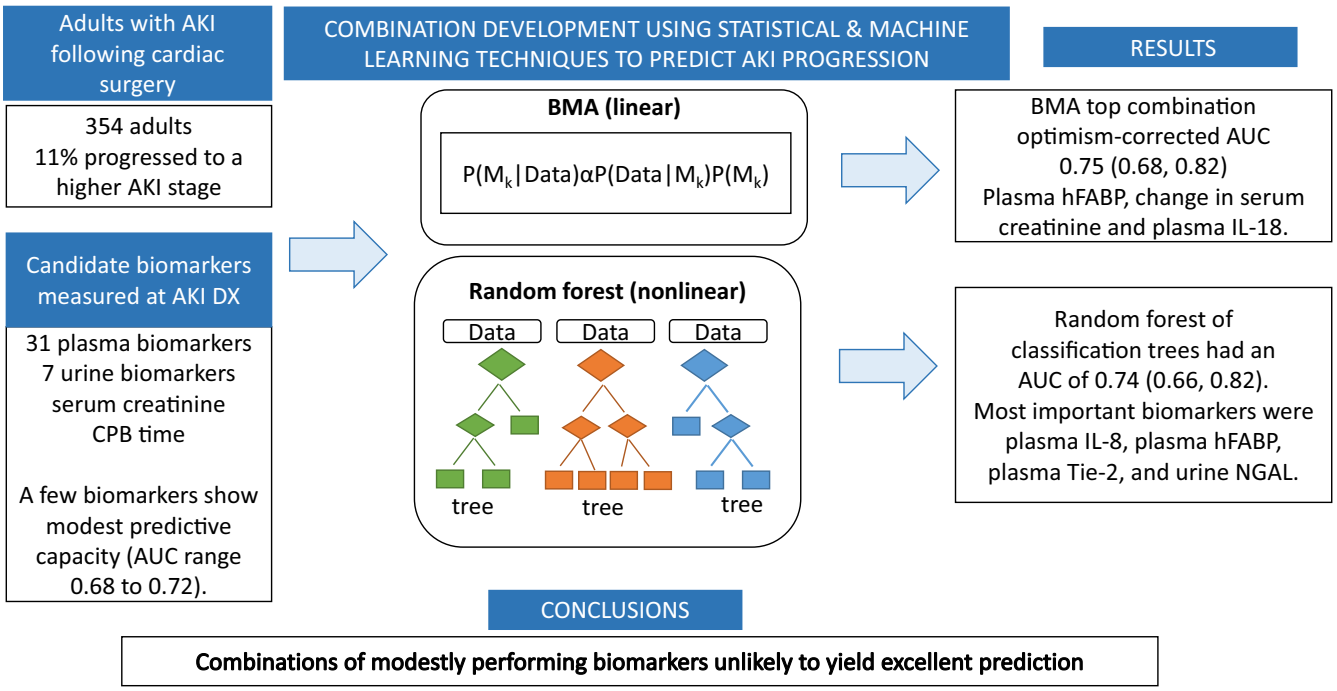

Figure 5. Identifying a panel of biomarkers to predict acute kidney injury (AKI) progression. AUC, area under the receiver operating characteristic curve; BMA, Bayesian Model averaging; CPB, cardiopulmonary bypass; hFABP, heart-type fatty acid-binding protein; NGAL, neutrophil gelatinase-associated lipocalin; Tie-2, tyrosine kinase-2.

alternative explanation is that performance of the untransformed biomarkers was inflated due to confounding effects of center.[21] However, there was not compelling evidence of center effects for these biomarkers that would support the latter hypothesis. Counterintuitively, the most promising result for biomarker combinations was for the center-specific transformed biomarkers. We caution that the center-specific transformation is highly variable, particularly for smaller centers. Therefore, this result should be considered preliminary.

We believe our experience in this endeavor is common across a wide array of fields in which there is interest in using biomarkers to improve risk prediction, prognosis, or other types of forecasting. For example, Wang *et al.*[22] measured 10 biomarkers in 3209 participants in the Framingham Heart study. Adding multimarker scores to conventional risk factors for death or cardiovascular events resulted in only small improvements in predictive capacity. Similar experiences are reported in the literature,[23] and we suspect that many more "negative" results go unpublished.

A plethora of methods are available for building predictive models. It is possible that another method, perhaps one that has not yet been invented, would yield a combination of our biomarkers that would accurately forecast AKI progression. However, there are reasons to suspect that other methods would not yield better results.[24] Lim *et al.*[25] compared 33 algorithms for building predictive models, including classical statistical algorithms and "machine learning" algorithms. The investigators had 32 datasets in which to compare the predictive performance of the models produced by the algorithms. Classical logistic regression was the second-best algorithm in terms of accuracy. An important point to note is that results were similar among the best-performing algorithms. A recent survey of the literature found no advantage to machine-learning methods over logistic regression methods.[26]

This study had several strengths, including the number of biomarkers measured; the use of biomarkers that are biologically meaningful and measured by high-quality immunoassays; and the use of statistical methods to assess performance while avoiding optimistic bias. Limitations of this study include a modest number of patients progressing in AKI, with the majority of patients with AKI progression (23 of 39) coming from the largest center. The other 5 centers had 2–5 patients progressing in AKI. Missing biomarker data further reduced power, although this reduction was mitigated through application of multiple imputations. Seven patients without AKI progression died in the hospital, raising a potential concern for misclassification, since progression might have been observed had these patients survived. However, these patients comprise only 2% of patients without AKI progression, so the potential impact of hypothetical misclassification is minor. Finally, the outcome of AKI progression may not serve as the ideal outcome to test multiple biomarker combinations due to the fact that the

dimensions that determine "AKI progression" (change in serum creatinine level over a given unit of time) may not be congruent with the biological expression of the biomarkers that we measured.[8]

The application of advanced statistical techniques to combine the prognostic information in a panel of biomarkers produced a small or negligible improvement over individual biomarkers. These findings are instructive for investigators seeking to develop biomarker panels in many fields. In light of empirical experience and theoretical results, investigators should have realistic expectations for biomarker combinations when the performance of individual biomarkers is insufficient for clinical application: combinations of modestly performing biomarkers are unlikely to yield excellent prediction.

DISCLOSURE

SGC and CRP receive financial compensation as consultants and advisory board members for Renalytix AI, Inc., and own equity in Renalytix. SGC has received consulting fees from Goldfinch Bio, CHF Solutions, Quark Biopharma, Janssen Pharmaceuticals, and Takeda Pharmaceuticals in the past 3 years. SGC was on the advisory board for pulseData and received consulting fees and equity in return. All the other authors declared no competing interests.

ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health (NIH) grant R01HL085757 (to CRP) to fund the Translational Research Investigating Biomarker Endpoints in AKI (TRIBE-AKI) Consortium to study novel biomarkers of acute kidney injury in cardiac surgery. CRP is also supported by NIH grant K24DK090203. SGC is supported by the following grants: R01DK115562, U01DK10692, R01HL85757, R01DK112258, and U01OH011326. SGC and CRP were also members of the NIH-sponsored ASsess, Serial Evaluation, and Subsequent Sequelae in Acute Kidney Injury (ASSESS-AKI) Consortium (U01DK082185). The biomarker tests were provided (in kind) by Beckman Coulter (Brea, CA), Randox Laboratories (Crumlin, UK), and Roche Diagnostics (Basel, Switzerland).

INFORMED CONSENT

All participants provided written informed consent, and each institution's research ethics board approved the study.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Supplementary Methods: Classification Trees and Random Forests.

Supplementary Methods: Bayesian Model Averaging (BMA).

Table S1. Timing of biomarker measurements relative to AKI.

Table S2. Biomarker measurement details.

Figure S1. Biomarker distributions.

REFERENCES

1. Thiele RH, Isbell JM, Rosner MH. AKI associated with cardiac surgery. *Clin J Am Soc Nephrol*. 2015;10:500–514.
2. Ho J, Tangri N, Komenda P, et al. Urinary, plasma, and serum biomarkers' utility for predicting acute kidney injury associated with cardiac surgery in adults: a meta-analysis. *Am J Kidney Dis*. 2015;66:993–1005.
3. Parikh CR, Coca SG, Thiessen-Philbrook H, et al. Postoperative biomarkers predict acute kidney injury and poor outcomes after adult cardiac surgery. *J Am Soc Nephrol*. 2011;22:1748–1757.
4. Parikh CR, Thiessen-Philbrook H, Garg AX, et al. Performance of kidney injury molecule-1 and liver fatty acid-binding protein and combined biomarkers of AKI after cardiac surgery. *Clin J Am Soc Nephrol*. 2013;8:1079–1088.
5. de Fontnouvelle CA, Greenberg JH, Thiessen-Philbrook HR, et al. Interleukin-8 and tumor necrosis factor predict acute kidney injury after pediatric cardiac surgery. *Ann Thorac Surg*. 2017;104:2072–2079.
6. Belley-Côté EP, Parikh CR, Shortt CR, et al. Association of cardiac biomarkers with acute kidney injury after cardiac surgery: a multicenter cohort study. *J Thorac Cardiovasc Surg*. 2016;152:245–251.e244.
7. Gocze I, Koch M, Renner P, et al. Urinary biomarkers TIMP-2 and IGFBP7 early predict acute kidney injury after major surgery. *PLoS One*. 2015;10:e0120863.
8. Koyner JL, Garg AX, Coca SG, et al. Biomarkers predict progression of acute kidney injury after cardiac surgery. *J Am Soc Nephrol*. 2012;23:905–914.
9. Schaub JA, Garg AX, Coca SG, et al. Perioperative heart-type fatty acid binding protein is associated with acute kidney injury after cardiac surgery. *Kidney Int*. 2015;88:576–583.
10. Patel UD, Garg AX, Krumholz HM, et al. Preoperative serum brain natriuretic peptide and risk of acute kidney injury after cardiac surgery. *Circulation*. 2012;125:1347–1355.
11. Koyner JL, Garg AX, Shlipak MG, et al. Urinary cystatin C and acute kidney injury after cardiac surgery. *Am J Kidney Dis*. 2013;61:730–738.
12. Kerr KF, Meisner A, Thiessen-Philbrook H, et al. Developing risk prediction models for kidney injury and assessing incremental value for novel biomarkers. *Clin J Am Soc Nephrol*. 2014;9:1488–1496.
13. Steyerberg EW, Pencina MJ, Lingsma HF, et al. Assessing the incremental value of diagnostic and prognostic markers: a review and illustration. *Eur J Clin Invest*. 2012;42:216–228.
14. Hoeting JA, Madigan D, Raftery AE, et al. Bayesian model averaging: a tutorial. *Stat Sci*. 1999;14:382–417.
15. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York, New York: Springer; 2001.

16. Harrell FEJ. *Regression Modeling Strategies*. Corrected ed. New York, New York: Springer; 2001.
17. Rubin DOB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: Wiley; 1987.
18. Pepe MS, Janes H, Longton G, et al. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol*. 2004;159:882–890.
19. Bansal A, Pepe MS. When does combining markers improve classification performance and what are implications for practice? *Stat Med*. 2013;32:1877–1892.
20. Kerr KF, Bansal A, Pepe MS. Further insight into the incremental value of new markers: the interpretation of performance measures and the importance of clinical context. *Am J Epidemiol*. 2012;176:482–487.
21. Meisner A, Parikh CR, Kerr KF. Biomarker combinations for diagnosis and prognosis in multicenter studies: principles and methods. *Stat Meth Med Res*. 2019;28:969–985.
22. Wang TJ, Gona P, Larson MG, et al. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N Engl J Med*. 2006;355:2631–2639.
23. Kattan MW. Judging new markers by their ability to improve predictive accuracy. *J Natl Cancer Inst*. 2003;95:634–635.
24. Hand DJ. Classifier technology and the illusion of progress. *Stat Sci*. 2006;21:1–14.
25. Lim T-S, Loh W-Y, Shih Y-S. A comparison of prediction accuracy, complexity, and training time for thirty-three old and new classification algorithms. *Machine Learning*. 2000;40:203–228.
26. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.