Real-time stats for real-time problems: the development of a risk tool to predict and prevent psychiatric crises in Multnomah County, Oregon

Shannon M. Campbell, MPP; Senior Research & Evaluation Analyst
Multnomah County Mental Health & Addiction Services Division
Presented at the American Evaluation Association annual conference, October 2018
shannon.campbell@multco.us

STEP-BY-STEP GUIDE: HOW TO BUILD A RISK TOOL USING LOGISTIC REGRESSION

- ❑ Define the discrete event you hope to predict.
  Acute care events: behavioral health-driven ER visits, psychiatric emergency services, and inpatient psychiatric hospitalizations, identified via CPT, revenue, and place of service codes in claims data.

- ❑ Select a population with a reasonable chance of this event occurring, sufficient historical data, and similarity to those for whom you would apply this in the future.
  Medicaid members with SPMI (severe and persistent mental illness), with at least one year of Medicaid coverage.

- ❑ Select a past time period where you have data to identify these events. Don't forget to assess whether your data sources are ones you can count on accessing not just today, but also in the future.
  Healthcare claims and call center records (ongoing and updated daily) and enrollment data (ongoing and updated monthly), for the last 2.5 years.

- ❑ Determine a list of potential variables that may predict this event. Talk with subject matter experts; consult the literature. Think not only of potential explanatory variables, but also indicators or common traits of those who have these events occur--even if there's not a theoretically causal relationship.
  Discussed with front-line crisis staff and care coordinators, utilization review, Medicaid management. Examples: high utilizers seem to have many emergency room visits with non-specific pain or respiratory complaints as the primary cause; multiple calls to the crisis line often precede a hospitalization.

- ❑ Use demographic variables as controls to reduce any implicit bias.
  Race, age range, sex, English vs. language other than English as primary language.

- ❑ Run your model. Treat it as you would for any other project. If the model is insignificant, the data violates a number of model assumptions, or so forth, stop. Only proceed further if/when you have a model that holds up under such scrutiny. If you have data where one actor can have multiple events, consider using clustered standard errors to adjust for intragroup correlation.
  Using logit in Stata with vce(cluster id); no model concerns; covariates of interest significant.

- ❑ Check the predictive power of your model. A decent area under the ROC curve is .7 or above; preferably .8 or above. If you do not meet this, you do not have a model that can reliably help you predict events with the necessary level of specificity and should not proceed further.
  Using lroc in Stata after running the logistic model, we had an initial result of 0.85.

❏ Revalidate your results multiple times. Use another sample of different people with similar traits. Test the model against different demographic groups (does it perform as well for different genders, races, ages, etc.?). If results do not vary substantially and area under the curve remains high, proceed.
Examples: logit depvar indvar1 indvar2 … if asian==1, vce(cluster id) or | lroc
if female==1, if english==1, if age30s==1, if asian==1 & female==1 & english==1 & age30s==1, etc.

❏ Use the odds ratios from your original model, sans demographics, as your predictive weights. This will form the basis of your risk score formula. Constant term + (first variable*odds ratio) + (second variable*odds ratio) + (third variable*odds ratio) … + constant term = raw risk score.

❏ Determine the highest possible risk someone could have--possessing every factor that increased risk, and none mitigating it. (You may find it easier to construct all variables to be unidirectional.) You can then scale your range to something more easily understood by others. E.g., if your high score is 25.8, and you want a scale of 0 to 10, divide raw scores by 2.58.

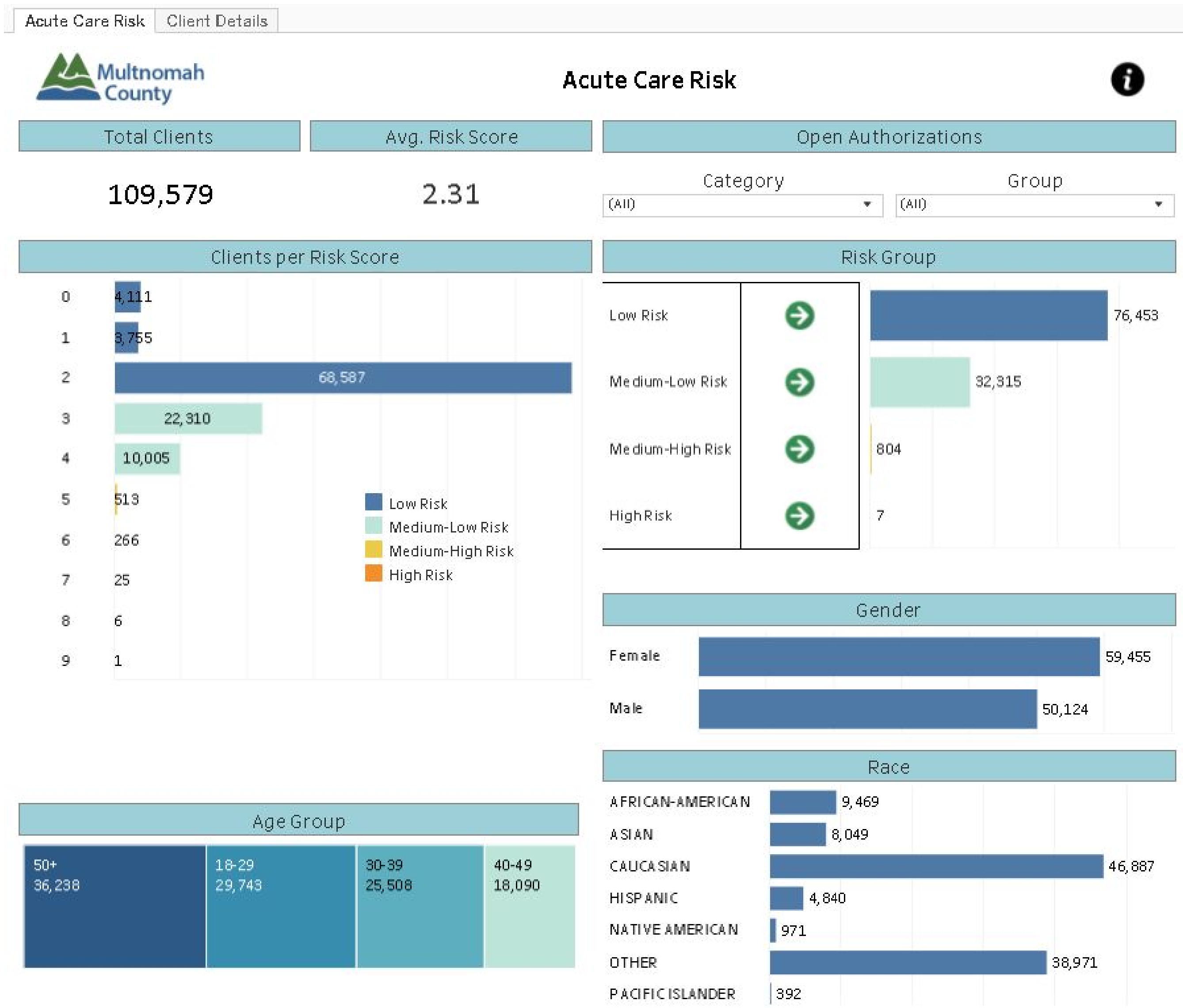
	Client response		Odds ratio
No recent mental health outpatient history (last 120 days)?	1 or 0 (yes/no)	*	4.518399
Multiple SPMI diagnoses (last 12 months)?	1 or 0	*	4.334528
Substance use history (last 12 months)?	1 or 0	*	2.928598
Week with 2+ crisis line calls (last 3 weeks)?	1 or 0	*	2.892232
SSI for disability (any time)?	1 or 0	*	1.696737
History of housing instability (any time)?	1 or 0	*	1.687269
Primary respiratory complaint at healthcare visit (last year)?	1 or 0	*	1.606196
Primary pain complaint at healthcare visit (last year)?	1 or 0	*	1.546471
Constant term	1	*	0.0372867
Subtotal, scaled to range of 0 to 10	=sum of above/2.124772		

❏ Depending upon your intended use, you will likely want to develop a way to automate this process, gathering all the relevant information into a single dataset, dependent on how your data is stored.
All our data was already stored in SQL tables across several databases; we created a stored procedure to bring it all together, apply the formula developed via our analysis, and schedule it to re-compute daily. We also tightened parameters on how far back to look for each variable, given that our original model, a Cox multiple-failure survival analysis, operates on (and controls for) different time periods of observation happening at different times. Our parameters now had to be standardized as relative to a single day.

❏ Revalidate once again. Create a “real-life” use case and test it--e.g., take data from today, calculate risk scores, and check whether or not events occurred afterward and how well the score did in predicting it. Recreate your original statistical model, using the score rather than individual covariates as a predictor.
All scores as of 5/15/2018, and tracking any acute care event afterward for the next few months. We had an lroc result of 0.77, moving from a small specific population of members with SPMI to the entire adult Medicaid membership--not as strong, but still good.

❏ Develop a simple way for end users to access and use this information--for example, a database, a Tableau dashboard, or a daily report. Ensure users understand what this tool can/cannot do. Address how this tool will or will not replace human judgment; address any ethical concerns relevant to your line of work.
We made it clear that this tool wasn’t intended to replace staff judgment, but only intended to serve as an extra data point and an amalgamation of information that many staff already intuitively knew and understood. We also made it clear that this tool should only be utilized to proactively offer services and still honor clients’ autonomy, not to deny services in any way or force services on clients.

Dashboard: allows users to see population-level information, as well as click through to specific client groups via demographic, risk group, current service authorizations, and so forth.



Client details page: includes contact information, current service authorizations, and risk scores.


Acute Care Risk | Client Details

Multnomah County

Client Details

Policy Id	Full Name	Score	Phone	Gender	Race	Age Group	Authorization Type	Referral Id	Vendor Nm	Elig Eff Date
ABC123	LEIA ORGANA SOLO	5	123-456-7890	Female	WHITE	50+	GLOBAL - Level B Adult Global - Primary A...	RESIST999	Lifeworks Northwest	1/1/2014
DEF456	TYRION LANNISTER	7	234-567-8901	Male	OTHER	30-39	Outpatient SUD 1/1/17 - Medication Assi...	WESTER999	CODA CD	2/15/2014
GHI789	HERMIONE GRANGER	5	345-678-9012	Female	AFRICAN-AMERI...	18-29	Null	Null	Null	3/31/2015
JKL012	ARAGORN S O ARATHO...	6	456-789-0123	Male	WHITE	50+	Null	Null	Null	4/1/2015
MNO345	KVOTHE ARLIDENSON	7	Null	Male	OTHER	18-29	Null	Null	Null	5/15/2016
PQR678	SHURI	5	567-890-1234	Female	AFRICAN-AMERI...	18-29	Null	Null	Null	6/30/2016
STU901	RA S AL GHUL	7	678-901-2345	Male	ASIAN	40-49	Exceptional Needs - Supported Employm...	GOTHAM999	Cascadia Behavioral Health	7/1/2017

OTHER RISK TOOLS FOR REFERENCE

- **Allegheny County, Pennsylvania: child welfare**
  https://www.alleghenycountyanalytics.us/wp-content/uploads/2018/02/DevelopingPredictiveRiskModels-package_011618.pdf
- **State of Oregon: criminal justice**
  https://risktool.ocjc.state.or.us/psc/cc/
  Special thanks to Kelly Officer, of the Oregon Criminal Justice Commission, for sharing the CJC methodology with us for this project.

APPRECIATIONS

Many thanks to the following MHASD staff for their contributions:

- **Devarshi Bajpai**
  Medicaid Plan senior manager
- **Heath Barber, Lauren Lopez, Jacob Mestman, Shiva Sangireddy, and Sivakrishna Yedlapelli**
  Decision Support Unit
- **Sarah Adelhart, Rochelle Pegel, and David Sant**
  Utilization Management
- **Jessica Jacobsen and Rachel Phariss**
  Adult Care Coordination
- **Leticia Sainz**
  Call center supervisor