CRcoder: An Interactive Web Application and SAS Macro to Support Personalized Clinical Decisions

Gail J McAvay, PhD[1]; Terrence E Murphy, PhD[1,2]; George O Agogo, PhD[1]; Heather Allore, PhD[1,2]

Perm J 2020;24:19.078

E-pub: 12/18/2019

https://doi.org/10.7812/TPP/19.078

ABSTRACT

Introduction: Electronic health care data offer an opportunity to improve clinical decision making through advanced statistical analyses of longitudinal observations.

Objective: To describe a Web application and SAS/STAT macro (SAS Institute Inc, Cary, NC) for computing joint models to estimate the typical and personalized risk of 2 concurrent binary outcomes.

Methods: Features of the Web application design include uploading longitudinal files formatted with constant or time-varying covariates, specification of 2 binary outcomes, specification of a propensity model for treatment, and joint and separate models of the outcomes. In addition we designed an SAS macro for conducting the analysis. Fitting of joint and separate statistical models was implemented using a model specified in the Web application, with subsequent processing by the SAS macro. To illustrate the fitting of models, a sample of older adults with comorbid hypertension and chronic obstructive pulmonary disease from the Medical Expenditure Panel Survey was created to examine the association between polypharmacy (use of ≥ 5 medication classes) and limitations in social activities and mobility.

Results: Relative to separate models, the joint models typically estimated attenuated associations between explanatory variables and the 2 outcomes with smaller standard errors. These joint models yielded estimates of personalized concurrent risk and typical concurrent risk.

Discussion: Clinical decision making based on electronic health data can be improved using joint modeling to generate an individual's probability of concurrent risk.

Conclusion: This user-friendly software performs the advanced statistical analyses needed to estimate typical and personalized concurrent risks.

INTRODUCTION

Clinician decision making is focused on choosing which treatments or interventions are best for an individual patient at a given moment. This process is complicated in that evidence of the effectiveness of treatments is often based on overall treatment effects found in randomized clinical trials. Variability in the effectiveness of treatments at the individual level suggests that leveraging personalized information from individual patients might enhance the decision-making process. An electronic health record (EHR), consisting of diagnostic and treatment data, is collected at the individual level, thereby representing a potentially rich source of information to improve the quality of health care provided to patients.[1,2] However, issues such as selection bias are likely to occur when using an EHR because sicker patients will often have more records in the health care system. Therefore, advanced methods are required that reduce the potential biases inherent to analyses of observational data.[3,4]

Translating the results of such statistical analyses into personalized estimates of risk is a crucial step for providing useful results for clinicians and their patients. It is recognized that many outcomes are correlated (eg, functional disability and hospitalization). This information on correlated outcomes can be captured through joint modeling techniques. The EHR data from individuals collected over multiple time points can be used to develop statistical models and to estimate risk at a personalized level, in addition to risk estimates for groups of individuals with similar characteristics, that is, sharing the same values for a given set of covariates (eg, sex, age, medical conditions).[1]

We describe a Web-based application based on SAS software[5] to apply advanced statistical analyses of EHR data. This software is designed for use by clinicians and quality improvement professionals to study the quality of health care systems and to potentially improve clinical decision making. Practitioners can apply this software to EHRs collected at the patient level to estimate individualized and group risk of multiple correlated outcomes.

For example, the association between a specific medical procedure (the exposure) and the occurrence of an adverse event (outcome 1), and the correlated occurrence of polypharmacy (outcome 2) might be studied to estimate the risk of these 2 outcomes for an individual or a group. In this example, we could estimate the risk of developing the adverse event (outcome 1) associated with the procedure, while also considering the contribution of possible drug interactions reflected by polypharmacy (outcome 2). Another example, for quality control, involves the concurrent associations between an infection control protocol (exposure) and hospital-acquired infection (outcome 1) and admission to an intensive care unit (outcome 2). These types of questions can be studied at the individual or group level. Multiple measures and outcomes, such as presence or absence of recommended care (evidence-based medicine), postoperative infections, hospital readmissions, medication errors, functional ability, and discharge status are other examples of health care outcomes that could be examined at the individual or group level.

Author Affiliations

[1] Department of Internal Medicine, Yale University School of Medicine, New Haven, CT

[2] Biostatistics Department, Yale University School of Public Health, New Haven, CT

Corresponding Author

Heather Allore, PhD (heather.allore@yale.edu)

Keywords: decision making, electronic health records, joint modeling, longitudinal analyses, personalized concurrent risk

In this study, we illustrate the use of the Web-based CRcoder SAS macro to estimate group and individual (personalized) risk for 2 correlated patient-centered outcomes: Limitations in mobility and social activity difficulties based on exposure to 5 or more medication classes (polypharmacy).

METHODS

Implementation

The CRcoder (Concurrent Risk Coder) Web application (http://crcoder.phs.wakehealth.edu) was designed to develop analyses through a user-friendly interface. The information collected is transferred to an SAS[5] program that reflects the specifics of the study design and can be run locally behind the user's firewall. This design was chosen to address Health Insurance Portability and Accountability Act regulations on the confidentiality of health records, by limiting access to approved personnel at the individual site.

The software interface consists of a series of tabs that gather information specified by the user, to design an analysis. (Figure 1 shows an example of a tab.) Steps for transfer of information about the dataset being used for analysis are briefly described in Table 1, data steps 1 to 2. Once this information is transferred, steps 3 to 4 in Table 1 describe the creation of the analytic design. The final step is submission of the SAS program (step 5).

An overview of the analytic procedures and the results are displayed in Table 1. If propensity scoring is selected as a method to balance receipt of treatment (exposure), then the analysis begins with the estimation of a propensity model (Table 1, Analytic procedure). Propensity modeling techniques are particularly important when one is using observational data.[6,7] Their basic function is to minimize bias owing to preexisting differences in treatment groups or other binary characteristics to be studied. A logistic regression model where treatment (or any binary exposure) is the dependent binary variable and covariates are the predictors of receipt of treatment (exposure) is estimated. Inverse probability of treatment weights are then created on the basis of the predicted probabilities from the logistic regression

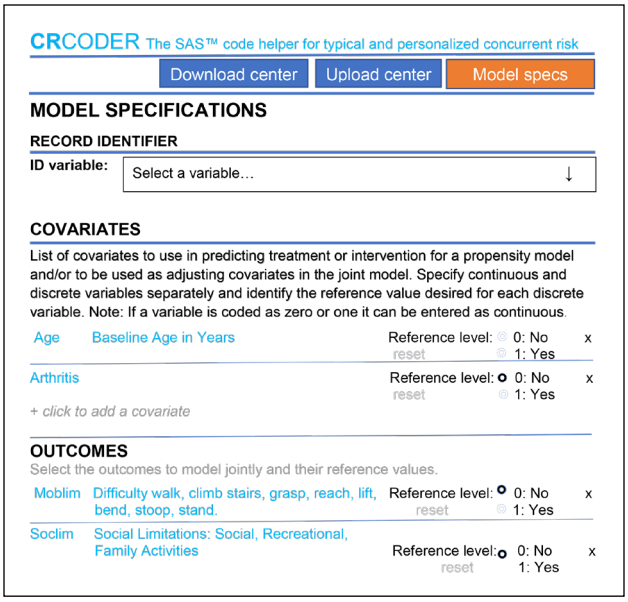

Figure 1. CRcoder model specification tab.
CRcoder = Concurrent Risk Coder; ID = identifier; Moblim = mobility limitation; Soclim = social limitation; specs = specifications.

Table 1. CRcoder Web interface

Data step	Description/output
Steps for designing an analysis	
Download and run programs for creating description files	SAS programs (CONTENTS, FORMATS)
Upload descriptor files created in step 1	Use upload tab of Web application
Specify variables to be used in the analysis, including covariates, treatment/exposure, time, subject identifier, and the 2 outcomes	See Figure 1 for model selection tab for selecting variables
Save the session (step 3) information	Create session file using download tab of Web application
Submit the SAS macro	Compute the model estimates and typical concurrent risk (TCR) and personalized concurrent risk (PCR)
Analytic procedure	
Estimate propensity model using logistic regression.	Dependent variable is treatment/exposure variable. Predictors are covariates specified in data step 3
Output includes estimates, standard errors, odds ratios, confidence intervals, and p values from step 1	Example report is shown in Table S1[a]
Generate propensity scores and inverse probability of treatment weights Assess balance between the 2 treatment/exposure groups	See Table 2 and Figure 2
Estimate joint and separate models using generalized mixed model techniques	SAS GLIMMIX Procedure (SAS Institute Inc, Cary, NC)
Output includes estimates, standard errors, odds ratios, confidence intervals, and p values from step 4	See Table 3 and Tables S2-S5[a]
Create risk estimates (TCR and PCR)	On the basis of model results from step 4
Create individualized report of risks	See Table 4

[a] See Supplemental Material, available at: www.thepermanentejournal.org/files/2019/19-078-SuppMat.pdf

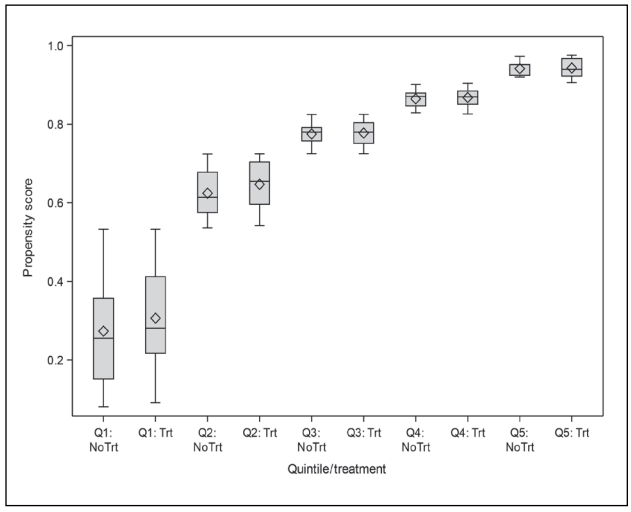

Figure 2. Distribution of propensity scores by quintile and treatment status.
No Trt: < 5 medication classes; Q = quintile; Trt: ≥ 5 medication classes.

model. The final step is assessing whether use of the weights led to a balance of variables between treatment (exposure) groups (Table 2 and Figure 2).[8]

Using the inverse probability of treatment weights, a joint model for 2 longitudinally measured binary outcomes is used to provide risk estimates for each person in the dataset (Table 1, Analytic procedure).[9-11] Specifically, a typical concurrent risk (TCR) estimate describes the risk of experiencing outcomes for groups of individuals with similar characteristics (eg, age 85 years, female, with arthritis). The second measure of risk, the personalized concurrent risk (PCR) describes the personalized risk estimate for an individual, which incorporates the individual's variability on the predictors and outcomes across time. The Supplemental Material (available at: www.thepermanentejournal.org/files/2019/19-078-SuppMat.pdf) gives more detail about the study methods.

Illustrative Example

We illustrate the use of CRcoder with de-identified data from the Medical Expenditure Panel Survey, an observational study based on a national sample.[12] Because this study used existing de-identified data that were publicly available, the institutional review board granted exemption from participant consent (Human Investigation Committee Protocol no. 1510016585 at Yale School of Medicine). We constructed the sample by requiring that older adults (≥ age 65 years) have a diagnosis of both hypertension and chronic obstructive pulmonary disease (N = 536). Our rationale was to select a group of individuals who are at high risk of taking 5 or more medication classes (the exposure), are at risk of both outcomes, and would also be more likely to have other multiple chronic conditions and impairments associated with the 2 index conditions (ie, hypertension and chronic obstructive pulmonary disease). There were 2 years of follow-up data.

In this analysis, we examined the association of polypharmacy (categorized as yes = taking ≥ 5 classes of prescription medications vs no = taking < 5 medication classes) with 2 binary outcomes. The first outcome, mobility limitations, was coded as 1 = at least 1 difficulty vs 0 = no difficulties among the following physical activities: Walking, climbing stairs, grasping objects, reaching overhead, lifting, bending, or stooping. Similarly, the second outcome, social limitations, indicated whether the subject reported any limitations in the following social activities: Participating in social, recreational, or family activities because of an impairment or a physical or mental health problem. The

Table 2. Assessment of balance between high (n = 377) and low (n = 159) polypharmacy treatment groups for baseline covariates[a]

Variable	Label	High group mean	Low group mean	Mean difference	Weighted mean difference	Percent reduction
Propensity score	Propensity score	0.784	0.512	0.272	0.012	95.554
Age	Baseline age, y	73.472	74.384	-0.911	-0.089	90.246
Angina	Angina	0.178	0.101	0.077	0.041	47.146
Arthritis	Arthritis	0.753	0.642	0.112	0.011	89.770
Asthma	Asthma	0.263	0.119	0.143	0.023	83.698
Duration of antihypertensive use						30
No Treatment	Reference	—	—	—	—	—
Don't Know		0.141	0.063	0.078	0.012	84.930
≤ 1-year duration of antihypertensive use		0.098	0.088	0.010	0.002	83.526
2-5-year duration of antihypertensive use		0.228	0.132	0.096	-0.020	78.923
6-10-year duration of antihypertensive use		0.172	0.113	0.059	0.009	84.632
> 10-year duration of antihypertensive use		0.263	0.113	0.149	0.011	92.909
Diabetes	Diabetes	0.395	0.189	0.207	0.038	81.617
Sex	Female sex	0.615	0.553	0.062	0.000	99.973
Special equipment	Used assistive device	0.366	0.258	0.108	0.052	51.890

[a] Polypharmacy was defined as taking ≥ 5 medication classes.

covariates included in the propensity model for polypharmacy included age in years, sex, angina, arthritis, asthma, diabetes, duration of antihypertensive medication use (1 = no treatment; 2 = don't know; 3 = ≤ 1 year; 4 = 2-5 years; 5 = 6-10 years; 6 ≥10 years), and use of any assistive devices (eg, walker, grab bars in the bathtub, or any other special equipment for personal care or everyday activities).

RESULTS

The results generated from the propensity model are displayed in Table 2 (additional model results appear in the Supplemental Material Table S1, available at: www.thepermanentejournal.org/files/2019/19-078-SuppMat.pdf). The means of the propensity score, each covariate by polypharmacy group, and an assessment of balance between the groups obtained by propensity score weighting are shown. The unweighted mean differences for the propensity score and covariates are substantially reduced compared with the weighted differences, with the reduction in differences more than 80% for most variables. The box plot shown in Figure 2 compares the distributions of the propensity scores in the high and low polypharmacy groups, grouped by propensity score quintile. Figure 2 also suggests that the propensity model achieves good balance in the distribution of scores for the 2 groups.

For the 536 older adults in this sample, the prevalence of the 2 outcomes was 23% for social limitations and 54% for mobility limitations. The correlation between the 2 outcomes was 0.42. Results from the joint model and separate models for social and mobility limitations are displayed in condensed form in Table 3. (Supplemental Material Tables S2 to S5, available at: www.thepermanentejournal.org/files/2019/19-078-SuppMat.pdf, provide more detailed model results.) Polypharmacy was marginally associated (p = 0.05) with a greater risk of social limitations in the joint model but not in the separate model, owing largely to the smaller standard errors in the joint model. In contrast, there was no association between polypharmacy and mobility limitations in either the joint or separate model.

Table 4 displays a condensed version of the generated report for 3 subjects in the study. The TCR column is an estimate of average risk for groups of subjects with the same pattern of covariate values. Note that the TCRs for the mobility outcome were much higher than those for the social limitation outcome, reflecting the higher prevalence of mobility limitations in the sample. At the individual level we note that for case 2, the risks were larger than those for case 1, partly because of the presence of arthritis. Case 2 also illustrates the variation in risk across time due to change in the time-varying covariates (eg, asthma developed in year 2).

DISCUSSION

We demonstrated the use of the CRcoder for the analysis of EHR longitudinal observational data, for generating TCR and PCR of 2 outcomes, which applies a series of analyses designed

Table 3. Odds ratios for polypharmacy exposure, calculated from joint and separate models[a]

	Social activity limitation		Mobility limitation	
Model	Odds ratio (95% CI)	p value	Odds ratio (95% CI)	p value
Joint model				
High vs low polypharmacy[b]	1.72 (1.00-2.96)	0.052	1.05 (0.67-1.64)	0.846
Separate model				
High vs low polypharmacy[b]	1.64 (0.76-3.56)	0.206	1.22 (0.71-2.09)	0.466

[a] Adjusted for the following covariates: Age, sex, angina, arthritis, asthma, diabetes, duration of antihypertensive use, and use of assistive device.
[b] High: ≥ 5 medication classes; low: < 5 medication classes.
CI = confidence interval.

Table 4. Sample listing of typical and concurrent risks based on a subset of covariates for three individuals

ID[a]	Outcome: Limitation	Survey round	Typical concurrent risk	Personalized concurrent risk	High polypharmacy (≥ 5 medication classes)	Baseline age, y	Angina	Arthritis	Asthma	Diabetes
1	Mobility	1	0.066	0.339	0	66	0	0	0	0
		2	0.066	0.339	0	66	0	0	0	0
	Social	1	0.019	0.124	0	66	0	0	0	0
		2	0.019	0.124	0	66	0	0	0	0
2	Mobility	1	0.421	0.992	1	83	0	1	0	0
		2	0.502	0.994	1	83	0	1	1	0
	Social	1	0.025	0.810	1	83	0	1	0	0
		2	0.037	0.862	1	83	0	1	1	0
3	Mobility	1	0.178	0.992	0	68	0	0	0	0
		2	0.185	0.992	1	68	0	0	0	0
	Social	1	0.013	0.883	0	68	0	0	0	0
		2	0.022	0.928	1	68	0	0	0	0

[a] Subject identifier.

to move in the direction of causal inference. The Web application CRcoder and the SAS macro were designed to provide a tool for clinical decision making that is based on the risk of experiencing 2 correlated binary outcomes. The motivation for designing this application was to broaden the use of advanced techniques for estimating risk to users who may not have experience with the development of SAS programs for weighted joint models.

Potential uses of personalized risk estimates include guiding decisions on individuals' care on the basis of their current risk of future outcomes. Examination of exposure and other factors associated with these risks could be done to suggest changes in the individuals' current treatment regimen (ie, best practices). For example, suggestions to reduce the risk of mobility limitations may include exercise, physical therapy, or medication review, which could be advised on an individual basis.

Alternatively, questions about the associations between implementation of various quality initiatives with risk of outcomes, such as hospital readmission, could be studied by typical or group-level estimates of risk. Screening programs for specific conditions (eg, depression) could examine current vs previous utilization of disease-related services (eg, counseling and medications) at the group level. Medication reconciliation interventions at discharge could be studied for outcomes, such as readmission or admission to a skilled nursing facility.

We envision the CRcoder tool to be best suited when clinicians or quality improvement professionals collaborate with individuals who are knowledgeable on claims data billing procedures or with those who can identify the limitations of different sources of data. The potential problems encountered when one is identifying specific conditions (billing vs clinical use) could be outlined by research or medical informatics personnel. These professionals may not have advanced statistical knowledge but are familiar with extracting the data and conducting statistical analyses. Although the clinician may identify hypotheses, a carefully designed plan, including data items to be used in the propensity modeling to reduce bias and how the data should be extracted, is necessary.

There are limitations to observational analysis using the EHR that must be considered when one is designing a study and using CRcoder. Selection bias is always a concern because individuals may receive different treatments depending on information captured in the EHR. Moreover, although receipt of a treatment (exposure) may control for confounding by adding covariates as well as weighting with the inverse probability of treatment weights, these methods do not control for unmeasured confounding as randomization may. The importance of propensity methods for addressing some of these potential biases must rely on clinical knowledge of covariates to be used in the model, in contrast to using stepwise procedures for inclusion of variables.

CONCLUSION

We developed a Web-based CRcoder SAS macro to generate empirically based measures of TCR and PCR using longitudinal data. It may be useful for testing treatments or studying quality improvement measures that could affect 2 correlated dichotomous outcomes. ❖

Disclosure Statement

The author(s) have no conflicts of interest to disclose.

Acknowledgments

This work was supported by grants from the US National Institute on Aging of the National Institutes of Health (R01 AG047891, P30AG021342-16S1, R33AG045050, R01AG055681, and U24AG05964).

We acknowledge the important contributions of the participants and staff of the Medical Expenditure Panel Survey, including the US Department of Health and Human Services, the Agency for Healthcare Research and Quality, and the Centers for Disease Control and Prevention. SAS and all other SAS Institute Inc product or service names are registered trademarks or trademarks of SAS Institute Inc, Cary, NC, US.

We especially thank Peter Charpentier, MPH, the Web site developer of CRcoder; and the Pepper OAIC Coordinating Center led by Dalane Kitzman; National Institutes of Health grant #: U24AG05964; and Nicholas Pajewski for installing and maintaining the CRcoder Web site.

Kathleen Louden, ELS, of Louden Health Communications performed a primary copy edit.

How to Cite this Article

McAvay GJ, Murphy TE, Agogo GO, Allore H. CRcoder: An interactive Web application and SAS macro to support personalized clinical decisions. Perm J 2020;24:19.078. DOI: https://doi.org/10.7812/TPP/19.078

References

1. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review. J Am Med Inform Assoc 2017 Jan;24(1):198-208. DOI: https://doi.org/10.1093/jamia/ocw042.
2. Rothman B, Leonard JC, Vigoda MM. Future of electronic health records: Implications for decision support. Mt Sinai J Med 2012 Nov/Dec;79(6):757-68. DOI: https://doi.org/10.1002/msj.21351.
3. Stuart EA, DuGoff E, Abrams M, Salkever D, Steinwachs D. Estimating causal effects in observational studies using electronic health data: Challenges and (some) solutions. EGEMS (Wash DC) 2013;1(3):4. DOI: https://doi.org/10.13063/2327-9214.1038.
4. Stoto MS, Oakes M, Stuart E, Brown R, Zuorovac J, Priest EL. Analytical methods for a learning health system: 3. Analysis of observational studies. EGEMS (Wash DC) 2017 Dec 7;5(1):30. DOI: https://doi.org/10.5334/egems.252.
5. SAS/STAT 14.3 user's guide. Cary, NC: SAS Institute Inc; 2017.
6. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivar Behav Res 2011 May;46(3):399-424. DOI: https://doi.org/10.1080/00273171.2011.568786.
7. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med 2015 Dec 10;34(28):3661-79. DOI: https://doi.org/10.1002/sim.6607.
8. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. Stat Med 2009 Nov 10;28(25):3083-107. DOI: https://doi.org/10.1002/sim.3697.
9. Agogo GO, Murphy TE, McAvay GJ, Allore HG. Joint modeling of concurrent binary outcomes in a longitudinal observational study using inverse probability of treatment weighting for treatment effect estimation. Ann Epidemiol 2019 Jul;5:53-8. DOI: https://doi.org/10.1016/j.annepidem.2019.04.008.
10. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. Stat Sin 2004;14:809-34.
11. Wu L, Liu W, Yi GY, Huang Y. Analysis of longitudinal and survival data: Joint modeling, inference methods and issues. J Prob Stat 2012;640153. DOI: https://doi.org/10.1155/2012/640153.
12. Medical Expenditure Panel Survey [Internet]. Rockville, MD: Agency for Healthcare Research and Quality [cited 2019 Mar 2]. Available from: https://meps.ahrq.gov/mepsweb/.