CrossMark

Analysis of model development strategies: predicting ventral hernia recurrence

Julie L. Holihan, MD,[a,*] *Linda T. Li, MD,*[b] *Erik P. Askenasy, MD,*[b]
Jacob A. Greenberg, MD,[c] *Jerrod N. Keith, MD,*[d]
Robert G. Martindale, MD PhD,[e] *J. Scott Roth, MD,*[f]
and Mike K. Liang, MD,[a] *Ventral Hernia Outcomes Collaborative*

[a] Department of Surgery, University of Texas Health Science Center at Houston, Houston, Texas
[b] Department of Surgery, Baylor College of Medicine, Houston, Texas
[c] Department of Surgery, University of Wisconsin, Madison, Wisconsin
[d] Department of Surgery, University of Iowa, Iowa City, Iowa
[e] Department of Surgery, Oregon Health and Science University, Portland, Oregon
[f] Department of Surgery, University of Kentucky, Lexington, Kentucky

ARTICLE INFO

ABSTRACT

Background: There have been many attempts to identify variables associated with ventral hernia recurrence; however, it is unclear which statistical modeling approach results in models with greatest internal and external validity. We aim to assess the predictive accuracy of models developed using five common variable selection strategies to determine variables associated with hernia recurrence.

Methods: Two multicenter ventral hernia databases were used. Database 1 was randomly split into "development" and "internal validation" cohorts. Database 2 was designated "external validation". The dependent variable for model development was hernia recurrence. Five variable selection strategies were used: (1) "clinical"—variables considered clinically relevant, (2) "selective stepwise"—all variables with a P value <0.20 were assessed in a step-backward model, (3) "liberal stepwise"—all variables were included and step-backward regression was performed, (4) "restrictive internal resampling," and (5) "liberal internal resampling." Variables were included with $P < 0.05$ for the Restrictive model and $P < 0.10$ for the Liberal model. A time-to-event analysis using Cox regression was performed using these strategies. The predictive accuracy of the developed models was tested on the internal and external validation cohorts using Harrell's C-statistic where $C > 0.70$ was considered "reasonable".

Results: The recurrence rate was 32.9% ($n = 173/526$; median/range follow-up, 20/1-58 mo) for the development cohort, 36.0% ($n = 95/264$, median/range follow-up 20/1-61 mo) for the internal validation cohort, and 12.7% ($n = 155/1224$, median/range follow-up 9/1-50 mo) for the external validation cohort. Internal validation demonstrated reasonable predictive accuracy (C-statistics = 0.772, 0.760, 0.767, 0.757, 0.763), while on external validation, predictive accuracy dipped precipitously (C-statistic = 0.561, 0.557, 0.562, 0.553, 0.560).

* *Corresponding author.* 6431 Fannin St, MSB 5.254. Tel.: +1 702 321 6559; fax: +1 713 566 4242.
  E-mail addresses: Julie.L.Holihan@uth.tmc.edu, holihanj@gmail.com (J.L. Holihan).

Conclusions: Predictive accuracy was equally adequate on internal validation among models; however, on external validation, all five models failed to demonstrate utility. Future studies should report multiple variable selection techniques and demonstrate predictive accuracy on external data sets for model validation.

Introduction

Hernia recurrence is a frequent problem after ventral hernia repair, with long-term recurrence rates ranging from 37%-73% at 140 mo.[1] There have been many attempts to identify predictors of hernia recurrence[2-4]; however, existing models are all slightly different, leaving it unclear which variables are truly associated with hernia recurrence and which model is most robust. One possible reason for this is that the statistical modeling techniques used to determine these predictors are flawed. There are multiple methods for variable selection, each with its own strengths and weaknesses, yet the external validity and predictive accuracy of these models are seldom reported. Currently, there is limited data to guide clinical researchers on the quality of various statistical modeling techniques.

One of the most frequently used model development strategies in hernia literature is the logistic regression.[5-7] Critics of logistic regression models using automated model selection algorithms, such as backward-stepwise variable elimination, raise the concern that the independent variable sets identified often cannot be reproduced, and spurious variables may be mistakenly identified.[8-10] In addition, predictive models developed using stepwise regression use a full model, where only the number of events in the data set limits the model. Therefore, some full models may be large and unwieldy for the clinician to use in practice. Other variable selection strategies include selecting only variables with a maximum P value requirement (such as $P < 0.20$) on univariate analysis or selecting clinically relevant variables a priori. In this setting, variables not significant on univariate analysis due to confounding or interactions may not be identified and included in model development.

Alternatively, bootstrapping (internal resampling) may be a better strategy, but this has rarely been used in developing models to predict hernia recurrence.[7] A growing body of literature suggests that bootstrap methods reduce bias of over-fitting and yields better estimates of the regression coefficients and 95% confidence intervals.[8,11,12] In direct comparison between bootstrap methods and stepwise regression using Akaike information criterion, the bootstrap strategy was more accurate in predicting cardiovascular surgery outcomes.[13] Despite these advantages, bootstrapping is seldom used in the surgical literature, and further evidence is needed to support this conclusion. A cursory review of Pubmed using the search terms ("bootstrap" OR "re-sampling" OR "re-sampling") AND "surgery" found only 495 studies in the surgical literature that used bootstrap methods. Of these studies, the cardiothoracic ($n = 145$) and urological ($n = 61$) subspecialties represent the majority, whereas a minority were general surgery ($n = 47$).

We hypothesize that bootstrap model selection as compared to other commonly used variable selection strategies will develop a hernia recurrence model with greater external validity and predictive accuracy.

Methods

After institutional review board approval, two multicenter ventral hernia repair databases were assessed. "Database 1" is a prospective database of all consecutive ventral hernia repairs performed at three institutions from 2009-2010. For database 1, all abstractors were trained on a data dictionary before chart review, and 10% of all cases were reviewed by the senior author to ensure quality.[14] Patient demographic, clinical, and intraoperative data were recorded and defined in accordance with the National Surgical Quality Improvement Project (NSQIP).[15] Recurrence was determined by radiographic data, clinical examination at follow-up visit, or reoperation reports. Follow-up was defined as day of surgery until death or last clinical examination.[14] "Database 2" is a retrospective database of all consecutive ventral hernias performed at seven institutions from 2010-2011. Data was extracted by trained reviewers at each institution guided by a standard data dictionary. Quality checks were dependent on the institution but generally consisted of 10% review by the senior author. NSQIP definitions were used for all patient demographic, clinical, and intraoperative data recorded. Hernia recurrence was determined either clinically or radiographically from the medical record. Date of last follow-up was determined by the last documented abdominal examination. For both databases, time-to-event was determined as the initial date that a hernia recurrence was recorded, either from clinical examination or by patient report.

Database 1 was randomly split into 2/3 ("development") and 1/3 ("internal validation") cohorts, whereas database 2 was for "external validation". The primary outcome was hernia recurrence. The development cohort was used to create five statistical models for hernia recurrence, and these were validated on the internal validation and external validation cohorts.

Statistical analysis

Differences between patients with and without a hernia recurrence were compared using Pearson χ^2 or Fisher exact test for categorical variables and two-tailed t test or Mann–Whitney U test for continuous variables. A Kaplan–Meier curve for time to recurrence was created using the development cohort, the internal validation data, and the external validation data to demonstrate if long-term recurrence rates