<div align="center">Original Article</div>

Modeling the Risk of Esophageal Squamous Cell Carcinoma and Squamous Dysplasia in a High Risk Area in Iran

Arash Etemadi MD PhD•[1,2], Christian C. Abnet PhD MPH[2], Asieh Golozar MD MPH[1,2], Reza Malekzadeh MD[1], Sanford M. Dawsey MD[2]

Abstract

Background: Identifying people at higher risk of having squamous dysplasia, the precursor lesion for esophageal squamous cell carcinoma (ESCC), would allow targeted endoscopic screening.

Methods: We used multivariate logistic regression models to predict ESCC and dysplasia as outcomes. The ESCC model was based on data from the Golestan Case-Control Study (total n = 871; cases = 300), and the dysplasia model was based on data from a cohort of subjects from a gastroenterology clinic in Northeast Iran (total n = 724; cases = 26). In each of these analyses, we fit a model including all risk factors known in this region to be associated with ESCC. Individual risks were calculated using the linear combination of estimated regression coefficients and individual-specific values for covariates. We used cross-validation to determine the area under the curve (AUC) and to find the optimal cut points for each of the models.

Results: The model had an area under the curve of 0.77 (95% CI: 0.74–0.80) to predict ESCC with 74% sensitivity and 70.4% specificity for the optimum cut point. The area under the curve was 0.71 (95% CI: 0.64–0.79) for dysplasia diagnosis, and the classification table optimized at 61.5% sensitivity and 69.5% specificity. In this population, the positive and negative predictive values for diagnosis of dysplasia were 6.8% and 97.8%, respectively.

Conclusion: Our models were able to discriminate between ESCC cases and controls in about 77%, and between individuals with and without squamous dysplasia in about 70% of the cases. Using risk factors to predict individual risk of ESCC or squamous dysplasia still has limited application in clinical practice, but such models may be suitable for selecting high risk individuals in research studies, or increasing the pretest probability for other screening strategies.

Keywords: Esophagus, risk modeling, squamous cell carcinoma, squamous dysplasia

Introduction

Esophageal squamous cell carcinoma (ESCC) is still the most prevalent type of esophageal cancer in the world[1] and the first cause of cancer mortality and morbidity in Golestan Province in the Northeast of Iran.[2] Despite recent advances in cancer care, ESCC still has a five-year survival of less than 20% in the U.S.[3] In Golestan Province, it has an even poorer prognosis, with a median survival of 7 months and a five-year survival of only 3.3%.[4] Thus, it is very important to diagnose ESCC in very early stages, when treatments with curative intent are possible. Squamous dysplasia is the precursor lesion for ESCC,[5] and it can be detected by endoscopy with Lugol's iodine staining and biopsy of unstained areas.[6] But endoscopic screening of all adults is not practical. Limiting this screening to a subset of individuals with a higher risk of squamous dysplasia, and ultimately ESCC, would reduce the clinical burden and likely increase the acceptance of this screening method. In the past, different statistical models have been applied to predict the risk of cancer development in individuals, the most well-known of which is the Gail model for breast cancer.[7] Models have also been developed to predict the risk of having precursor lesions for both esophageal squamous cell carcinoma[8] and adenocarcinoma.[9–10] The purpose of this study was to try to develop a statistical model using known risk factors to predict individual risk of squamous dysplasia or ESCC in Northeast Iran.

Materials and Methods

We built and tested two separate models for ESCC and squamous dysplasia, each using a different set of data. We used data from the Golestan Case-Control Study to build the ESCC model. This study was conducted from 2003 to 2007, and included 300 biopsy-proven ESCC cases and 571 age and sex-matched neighborhood controls. Details of this study have been published before.[11] We used data from individuals visiting Atrak Clinic, a gastroenterology research clinic in Gonbad City, in Golestan Province, to build the dysplasia model. Atrak Clinic was set up by the Digestive Disease Research Center of Tehran University of Medical Sciences in 2001, and patients complaining of GI symptoms are referred to it by local physicians. Between 2002 and 2007, 724 individuals with GI symptoms visited this clinic and underwent video endoscopy with Lugol's iodine staining and completed the same questionnaire used in the Golestan Case-Control Study.[11] All unstained lesions were biopsied and sent for histological examination. Overall, 26 individuals with dyplastic lesions were identified in this group.

We used multivariate logistic regression models to predict ESCC and squamous dysplasia. The ESCC model was based on the case-control data (total n = 871; cases = 300), and the dysplasia model

Authors' affiliations: [1]Digestive Disease Research Center, Shariati Hospital, Tehran University of Medical Sciences, Tehran, Iran, [2]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, USA.
•**Corresponding author and reprints:** Arash Etemadi MD PhD, 6120 Executive Boulevard, Rockville, MD 20852, USA.
Tel: +1-301-451-0164, Fax: +1-301-496-6829, E-mail: arash.etemadi@nih.gov
Accepted for publication: 16 November 2011

Table 1. The effect of different hypothetical prevalences for dysplasia on positive and negative predictive values of the regression model.

Prevalence of dysplasia (%)	False positive rate (%)	False negative rate (%)	Positive predictive value (%)	Negative predictive value (%)
0.02	96.5	1.1	3.5	98.9
0.04	93.2	2.2	6.8	97.8
0.06	89.9	3.4	10.1	96.6
0.08	86.8	4.6	13.2	95.4
0.1	83.7	5.8	16.3	94.2
0.2	69.5	12.1	30.5	87.9
0.3	57.1	19.1	42.9	80.9
0.4	46.1	26.9	53.9	73.1
The optimal sensitivity (61.9%) and specificity (69.9%) for the dysplasia prediction model was used.				

on the data from Atrak Clinic (total $n = 724$; cases = 26). In each of these we first fit a model including all risk factors known in this region to be associated with ESCC, according to previous reports,[11–15] and then added unintentional weight loss in the past year to the model. The known risk factors included: age, ethnicity, tobacco smoking, opium use, education, marital status, oral health, family history, tea temperature, and water source. We tested each model for goodness of fit using Hosmer-Lemeshow (H-L) chi square. The p value for this test compares the predicted and observed probabilities, and a significant P value shows lack of fit. In general, the base models (without weight loss) had a better goodness of fit (H-L $P = 0.13$ for ESCC and $P = 0.10$ for dysplasia) than the model including weight loss (H-L $P < 0.001$).

Individual risks were estimated using the linear combination of the regression coefficients and individual-specific values for covariates. A receiver operating characteristic (ROC) curve was drawn using the estimates obtained in this way. The area under the ROC curve (concordance index) determines the power of the model to discriminate between each randomly-selected pair of cases and controls, and can be between 0.5 (no discrimination better than chance alone) to 1 (perfect discrimination).[16] Classifying a set of outcomes, using the same observations both to fit the model and to estimate the classification error, will result in biased error-count estimates, and the resulting validation results are overfit.[17] To correct for this bias, we used cross-validation: the whole dataset without a single observation was used to build the model and then based on that model, outcome probability was predicted in that single observation; this was repeated this for all observations. Crossval macro[18] for Stata Statistical Software (Release 11, StataCorp, College Station, TX) was used to draw cross-validated ROC curve and calculate the area under the curve (concordance index) and its 95% confidence interval.

It is also possible to use a one-step approximation to the parameter estimates, as described before,[19] instead of the "leave-one-out" cross-validation. To produce a classification table comparing sensitivities and specificities of different prediction cutpoints for each model, we used this one-step method in Statistical Analyzing System release 9 (SAS Institute, Carey, NC).[19] Estimates are given by:

$$\widehat{\beta}^1_{(j)} = \widehat{\beta} - \frac{w_j(y_j - \widehat{\pi}_j)}{1 - h_{jj}} \widehat{V}(\widehat{\beta}) \begin{pmatrix} 1 \\ x_j \end{pmatrix}$$

Where:
$\widehat{\beta}$ is the maximum likelihood estimate of the parameter vector based on all observations;
$\widehat{\beta}_j$ is the same maximum likelihood estimate without the *jth* observation;

y_j is 1 for an observed event and 0 otherwise;
ω_j is the weight of the *jth* observation;
$\widehat{\pi}_j$ is the predicted event probability based on $\widehat{\beta}$;
h_{jj} is the hat diagonal element assuming single trial with y_j event;
$\widehat{v}(\widehat{\beta})$ is the estimated covariance matrix of $\widehat{\beta}$.

In the classification table, the optimal cutpoint for the model accuracy was the one with the best combination of sensitivity and specificity. This point also had the same predicted probability of the outcome as the observed prevalence in the sample. We then used the sensitivity and specificity of this optimal point to estimate positive and negative predictive values of the model against different hypothetical values for the prevalence of dysplasia in general population.

Results

ESCC risk prediction

The multivariate model including all known risk factors had a cross-validated area under the curve (AUC) of 0.77 (95% CI: 0.74–0.80) to predict ESCC (Figure 1A). This means that in 77% of randomly selected case-control pairs, the model will predict a higher probability of ESCC for the case compared to the control. The bias-corrected optimal point had a 74.0% sensitivity and a 70.4% specificity. After adding weight loss in the past year to the model (Figure 1B), the AUC increased to 0.87 (95% CI: 0.85–0.89), and the optimal point in the classification table had 80.6% sensitivity and 82.4% specificity. However, the model had a poor goodness-of-fit statistic (H-L $P < 0.001$)

Dysplasia risk

The model with all known risk factors except weight loss (Figure 1C) had a cross-validated area under the curve of 0.71 (95% CI: 0.64–0.79), which showed little change when weight loss was added to the model (Figure 1D). As the figure shows, the ROC curves were not smooth because of the relatively low prevalence of dysplasia (3.6%) and at the optimal cutpoint, a sensitivity of 61.5% and a specificity of 69.5% were observed. Table 1 shows the positive and negative predictive values for this cutpoint with different hypothetical prevalences of dysplasia. As can be seen in the table, with values close to the observed prevalence in this sample, the positive and negative predictive values are 6.8% and 97.8%, respectively.

Discussion

Our models were able to discriminate between ESCC cases and controls in about 77%, and between individuals with and without
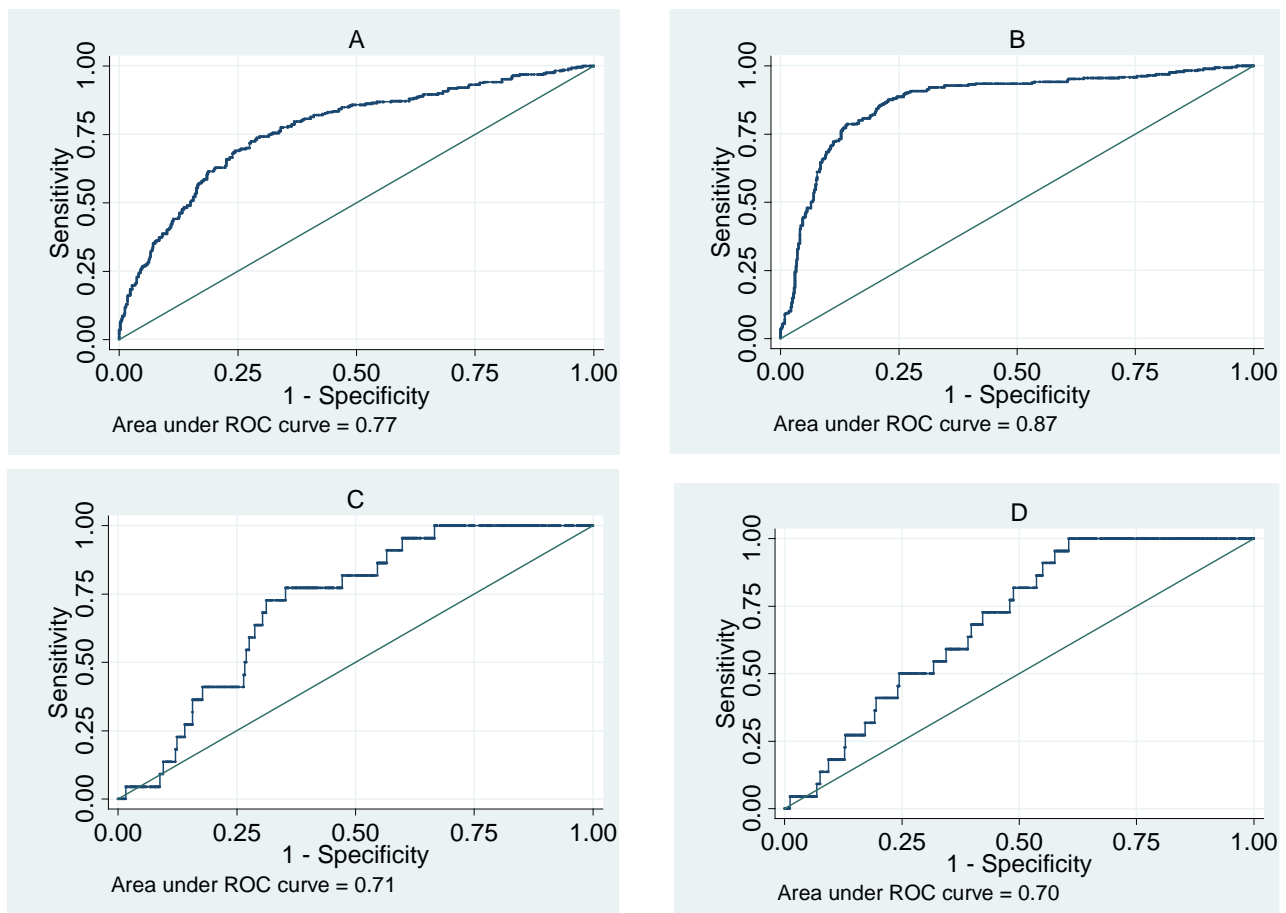
Figure 1. Receiver operating characteristic (ROC) curves comparing the sensitivities and specificities of models to predict esophageal squamous cell carcinoma (ESCC) and squamous dysplasia in Golestan Province. **A)** ESCC risk factors, **B)** ESCC risk factors with weight loss, **C)** dysplasia risk factors, **D)** dysplasia risk factors with weight loss.

squamous dysplasia in about 70% of the cases. The model, however, had only a 6.8% positive predictive value for diagnosing dysplasia, due to the low prevalence of this lesion.

Our results were better than a previous study done in another high risk area in China.[8] In that study, Wei et al. observed a sensitivity of 57% and a specificity of 54% for predicting squamous dysplasia, in spite of the fact that many components of the model were significantly associated with the risk of dysplasia. The area under the curve was 0.58 in their study without using any cross-validation method; the area under the curve was already very low and using any cross-validation would have decreased it even further.[8] We found better sensitivity and specificity, and a higher area under the curve despite using cross-validation (even the lower bound of the confidence interval in our study was higher than 0.58). The cross-validation method tests the model in individuals other than the ones used for model building, and thus gives a more realistic estimate of the discriminating power of the model.

Studies using questionnaire data and symptoms to predict esophageal adenocarcinoma have had similar results. Models using symptoms and individual characteristics for the diagnosis of Barrett's esophagus, a precursor lesion of esophageal adenocarcinoma had an AUC of 0.72[10] and 0.76.[9] One big difference is that Barrett's esophagus is closely related to gastroesophageal reflux disease (GERD), and thus GERD symptoms can be used in the model, while dysplasia is rarely symptomatic.

In cross-validation studies, a model may have a good ability to discriminate between cases and controls, but may not be able to correctly predict the probability of the event.[20] Calibration refers to the agreement between predicted probabilities and the observed proportions.[16] In our data, the base models (without weight loss) showed good calibration according to the Hosmer-Lemeshow statistic. Interestingly, although the ESCC model with weight loss had the highest area under the curve, it had a poor fit. This model has also little application, since weight loss develops during the symptomatic stage of ESCC when endoscopy is strongly indicated and risk screening is no more useful.

One of the limitations in our study was the small number of dysplastic cases, which has led to wider confidence intervals for the dysplasia model compared with ESCC models. On the other hand, although the ESCC model had more cases, the prediction of ESCC is not the main purpose of risk screening. The ESCC cases in this series have advanced symptomatic disease, and by the time ESCC has reached this stage, it is usually too late for any intervention.

Unlike high-risk areas in China, where about 30% of the general population have squamous dysplasia,[8] in our sample this rate was only about 4%. Disease prevalence determines post-test probabilities in different populations,[21] so even with the best estimates of accuracy, given this low prevalence, the predictive values will be so low that the use of a risk factor model for individual risk stratification is not advisable. With 62% sensitivity and a 70% specificity,

about 93% of positive cases will be false positives, and the positive predictive value (PPV) for dysplasia will be only around 7%. This implies that, on average, out of 100 endoscopies performed in high-risk individuals (according to the model), only 7 will have dysplasia, while this number is 4 in a randomly-selected sample of the general population. On the other hand, if the prevalence of dysplasia were similar to China, we would have a positive predictive value of 42.9%, which would make it more suitable as an initial selection step for endoscopic screening. The reasons for such a low prevalence of dysplasia, despite the high incidence of ESCC, remains to be determined and may range from differences in ESCC pathogenesis between the two populations (e.g., faster progression from dysplasia to ESCC), to technical differences in the screening methods used.

Lack of reproducibility is a general problem with many risk models.[10] Different populations may have different risk factors for the disease (e.g., the risk factors used in China were different from those used in Iran to build the prediction model). Also, the studies validating these models usually use internal validation (like the cross-validation used in our study), rather than true external validation (which needs testing the model on new data collected from another population).[22] Besides, many clinicians are reluctant to use statistical models for risk assessment and stratification.[21] These models, however, may prove useful in research studies, where a researcher is interested in selecting a group at higher potential risk for developing the precursor lesion and ultimately cancer.

The latest cancer registry data from Golestan show a very high incidence of ESCC, especially in the population above 50 years of age.[23] This high incidence, together with the poor prognosis of ESCC in this region, underlines the importance of finding alternative strategies for early detection. Sponge balloon cytology has recently been shown to be effective for detection of patients with Barrett's esophagus.[24] Similarly, a non-endoscopic esophageal sampling technique coupled to a biomarker is a potential alternative which can be tested against chromoendoscopy for early detection of squamous dysplasia and ESCC. The risk model can be used to increase pretest probability for such screening strategies, by selecting a particularly high risk group.

References

1. Eslick GD. Epidemiology of esophageal cancer. *Gastroenterol Clin North Am.* 2009; **38:** 17 – 25.
2. Kamangar F, Malekzadeh R, Dawsey SM, Saidi F. Esophageal cancer in northeastern Iran: a review. *Arch Iran Med.* 2007; **10:** 70 – 82.
3. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin.* 2010; **60:** 277 – 300.
4. Aghcheli K, Marjani HA, Nasrollahzadeh D, Islami F, Shakeri R, Sotoudeh M, et al. Prognostic factors for esophageal squamous cell carcinoma-a population-based study in golestan province, iran, a high incidence area. *PLoS One.* 2011; **6:** e22152.
5. Wang GQ, Abnet CC, Shen Q, Lewin KJ, Sun XD, Roth MJ, et al. Histological precursors of oesophageal squamous cell carcinoma: results from a 13 year prospective follow up study in a high risk population. *Gut.* 2005; **54:** 187 – 192.
6. Dawsey SM, Fleischer DE, Wang GQ, Zhou B, Kidwell JA, Lu N, et al. Mucosal iodine staining improves endoscopic visualization of squamous dysplasia and squamous cell carcinoma of the esophagus in Linxian, China. *Cancer.* 1998; **83:** 220 – 231.
7. Gail MH, Brinton LA, Byar DP, Corle DK, Green SB, Schairer C, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989; **81:** 1879 – 1886.
8. Wei WQ, Abnet CC, Lu N, Roth MJ, Wang GQ, Dye BA, et al. Risk factors for oesophageal squamous dysplasia in adult inhabitants of a high risk region of China. *Gut.* 2005; **54:** 759 – 763.
9. Locke GR, Zinsmeister AR, Talley NJ. Can symptoms predict endoscopic findings in GERD? *Gastrointest Endosc.* 2003; **58:** 661 – 670.
10. Gerson LB, Edson R, Lavori PW, Triadafilopoulos G. Use of a simple symptom questionnaire to predict Barrett's esophagus in patients with symptoms of gastroesophageal reflux. *Am J Gastroenterol.* 2001; **96:** 2005 – 2012.
11. Nasrollahzadeh D, Kamangar F, Aghcheli K, Sotoudeh M, Islami F, Abnet CC, et al. Opium, tobacco, and alcohol use in relation to oesophageal squamous cell carcinoma in a high-risk area of Iran. *Brit J Cancer.* 2008; **98:** 1857 – 1863.
12. Abnet CC, Kamangar F, Islami F, Nasrollahzadeh D, Brennan P, Aghcheli K, et al. Tooth loss and lack of regular oral hygiene are associated with higher risk of esophageal squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2008; **17:** 3062 – 3068.
13. Islami F, Boffetta P, Ren JS, Pedoeim L, Khatib D, Kamangar F. High-temperature beverages and foods and esophageal cancer risk--a systematic review. *Int J Cancer.* 2009; **125:** 491 – 524.
14. Islami F, Kamangar F, Nasrollahzadeh D, Aghcheli K, Sotoudeh M, Abedi-Ardekani B, et al. Socio-economic status and oesophageal cancer: results from a population-based case-control study in a high-risk area. *Int J Epidemiol.* 2009; **38:** 978 – 988.
15. Akbari MR, Malekzadeh R, Nasrollahzadeh D, Amanian D, Sun P, Islami F, et al. Familial risks of esophageal cancer among the Turkmen population of the Caspian littoral of Iran. *Int J Cancer.* 2006; **119:** 1047 – 1051.
16. Cook NR. Statistical evaluation of prognostic versus diagnostic models: beyond the ROC curve. *Clin Chem.* 2008; **54:** 17 – 23.
17. Vickers AJ. Prediction models in cancer care. *CA Cancer J Clin.* 2011 Jun 23. doi: 10.3322/caac.20118. [Epub ahead of print]
18. Thompson R. Crossval: a macro for Stata. 1999 [updated February 12, 2000; cited 2011 Sept 1, 2011]; Available from: URL: http://www.biostat.jhsph.edu/~ejohnson/regression/crossval.ado.
19. SAS/STAT^R 9.2 user's guide, Second Edition. Inc. SI, editor: Cary NC : SAS Institute, ; 2009.
20. Vickers AJ, Cronin AM. Traditional statistical methods for evaluating prediction models are uninformative as to clinical value: towards a decision analytic framework. *Semin Oncol.* 2010; **37:** 31 – 38.
21. Moons KG, Harrell FE. Sensitivity and specificity should be de-emphasized in diagnostic accuracy studies. *Acad Radiol.* 2003; **10:** 670 – 672.
22. Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *BMJ.* 2009; **338:** b605.
23. Roshandel G, Sadjadi A, Aarabi M, Khademi H, Keshtkar AA, Sedaghat SM, et al. Cancer incidence in Golestan: report of an ongoing population-based cancer registry in Iran, 2004–2008. *Arch Iran Med.* 2011 [In Press].
24. Kadri S, Lao-Sirieix P, Fitzgerald RC. Developing a nonendoscopic screening test for Barrett's esophagus. *Biomark Med.* 2011; **5:** 397 – 404.