Available online at www.isas.org.in/jisas
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 65(1) 2011 99-103

Associating Risk Factors with Farm Injuries using Classification Tree

Shande Chen[1*], Karan P. Singh[1*], Sejong Bae[1] and Alfred A. Bartolucci[2]
[1] *Department of Biostatistics, University of North Texas Health Science Center, USA*
[2] *Department of Biostatistics, University of Alabama at Birmingham, USA*

Received 21 October 2008; Revised 22 January 2011; Accepted 27 January 2011

SUMMARY

Agriculture is an occupation with risks, especially, when farmers use more machines for farming. Therefore, it is important to associate risk factors with agricultural related injuries, which has implications to make policies for preventing injuries. Lyman, *et al*. (1999) and others have applied multiple logistic regression in agricultural research. However, it is challenging to incorporate appropriate interactions in the logistic regression model. In this paper, the approach of classification tree is used to associate risk factors with farm injuries, which can automatically handle possible interactions. Based on a classification tree, it is also possible to develop different prevention programs for various subgroups given by the tree. We used a data set of 1,051 subjects from the states of Alabama and Mississippi, USA, including African-American farm workers and Caucasian and African-American owners (Source: UAB Injury Center). Condition of farm machinery (Excellent/Good vs. Fair/Poor) first splits the tree. For fair/poor condition of farm machinery, large farms (749 or more acres) give high injury rates (0.28), while for farms with fewer than 749 acres, the estimated injury rate is 0.16 without farm safety training and it is reduced to 0.063 with farm safety training. For excellent/good condition of farm machinery, history of farm injury then splits the tree. Factors that further split the tree include "How Often Do I Hurry When Doing Farm Work on Machinery" wearing a seat belt, education, primary commodity, race, alcohol consumed per week and farm safety training. Furthermore, classification trees are developed separately for Caucasians and African-Americans.

Keywords: Machines, Farm injury, Socio-demographic characters, Risk factors, Comorbid medical conditions, Nonparametric approach, S-PLUS, Classification trees.

1. INTRODUCTION

Agriculture is considered a dangerous occupation (National Committee for Injury Prevention and Control 1989; Bell *et al*. 1990; Rossignol and Pineault 1993; Lyman *et al*. 1999), especially, when more and more machines are being used for farming (Gerberich *et al*. 1998). This is certainly the situation in many developing countries. Therefore, it is important to associate risk factors with agriculture-related injuries, which has policy implications for preventing injuries. In addition, such studies for developed countries may provide useful information for developing countries for their current and near future agriculture industry. Several researchers have tried conventional regression methods to address the main question of farm injuries. For example, Lyman *et al*. (1999) used multiple logistic regression and McGwin *et al*. (2000) utilized the Cox regression for such purpose.

However, in general, the association between the predicators and the injury could be very complicated. That is, non-linear relationships and complex interactions are rather difficult to properly incorporate in these conventional regression models. In this paper, we apply a classification tree approach to associate risk factors with farm injuries, which can automatically handle possible non-linearity and interactions. Based on

**Corresponding authors* : Shande Chen and Karan P. Singh
E-mail address :hshande.chen@unthsc.edu and karan.singh@unthsc.edu

a classification tree, it is also possible to develop different prevention programs for various subgroups given by the tree. That may lead to policy implications.

We used a retrospective data set of 1,051 subjects with farm injuries from the states of Alabama and Mississippi, USA, which include injury data on African-American farm workers and Caucasian and African-American owners. Since the data contain a large proportion of African-American farmers, ethnic differences associated with injury and risk factors for separate ethnic populations are also studied.

2. MATERIALS AND METHODS

2.1 Study Data

The data set used in this paper is retrospective described in detail in Lyman *et al*. (1999) and McGwin *et al*. (2000). The data were collected from five rural counties in the state of Alabama and four rural counties in the state of Mississippi, USA. These counties have large African-American populations (50%) so African-American farmers and farm workers are well represented. The study subjects were recruited between January 1994 and June 1996, and were followed until April 1998. The average follow-up duration was 2.5 years. Baseline data were collected from 1,686 active farmers and farm workers, and among them 1384 farmers completed the follow-up data. For this study, 122 females and 16 Caucasian farm workers were excluded due to small numbers for meaningful representation. Thus, a total of 1,246 subjects (685 Caucasian owners, 321 African-American owners and 240 African-American workers) are used for this study.

The baseline data include information on demographics (e.g., age, race, education), farm and farming (e.g., commodities, safety training, equipment usage), and behavior (e.g., risk taking, alcohol assumption), number of hours per week for farm-related work and farm activities (e.g., machinery operation) as well as previous farming-related injuries. During the follow up, injury information was collected. As in some previous studies including the studies done by Lyman, *et al*. (1999) and McGwin *et al*. (2000), the agricultural injury is defined as unintentional physical injury or poisoning which occurred during an agricultural activity and required medical attention or resulted in at least one-half day of restricted activities. The primary outcome is the occurrence of the agricultural injury during the follow up.

This outcome is a dichotomous response variable in the analysis. The risk factors or independent variables are race (white or African-American), age category, education level, history of farm injury (yes or no), comorbid medical conditions (yes or no), farm owner or worker, full-time farmer (40 or more hours per week on farming), other non-farm job (yes or no), full-time non-farm job (yes or no), farm safety training (yes or no), farm safety training for chemicals (yes or no), milliliters of alcohol consumed per week, behavior for turning off machinery when removing stuck objects, behavior for wearing seat belt, attitude towards farm safety (very careful or not), attention level towards farm safety (very attentive or not), how often tired when doing farm work (often or not often), how often in an hurry when doing farm work (frequently or not), primary commodity farmed, size of farm in acres, use of tractor on farm (yes or no), number of regularly used farm machines, and condition of farm machinery (excellent, good or fair/poor).

2.2 Statistical Methods

In this paper, we use classification and regression trees (CART) (Breiman *et al*. 1984) to establish a relationship between the farm injury with risk factors. Since the dependent variable is binary, our data analyses are based on the classification tree. The predictors in a tree method can be a mix of continuous and categorical variables. In general, the tree method is considered as a nonparametric approach, and can easily handle possible interactions among the predictors. Due to this nature, the tree methods are often used in data mining and exploratory data analysis. S-PLUS is used for the classification trees. The minimum size to split a node in the analyses is 100, which seems to be adequate for such a large sample. After the tree is fully grown, deviance against number of terminal nodes is plotted. Based on the reduction in the deviance, pruning the tree can be performed based on the cost-complexity measure. We prune a tree to have about 75% reduction in deviance. Classification trees were developed for all the data and separately for Caucasian owners and for African-Americans.

3. RESULTS

In the sample, about 55% subjects are Caucasian owners, 25% are African-American owners, and 20% are African-American farm workers. Lyman *et al*.

(1999) and McGwin *et al*. (2000) provide detailed summaries for socio-demographic characters and other characters/risk factors.

For all the subjects, the classification tree actually used condition of farm machinery, history of farm injury, how often in a hurry when doing farm work, primary commodity farmed, size of farm, race, alcohol consumed per week, farm safety training, education, behavior for wearing seat belt, and farm safety training for chemicals in tree construction. Due to missing information 1,051 observations were used for the tree development. The overall injury rate is 0.105 for these 1,051 subjects. The tree first splits according to the machinery condition. For those with excellent/good machinery condition, the branch then splits according to whether they had a history of farm injury. Further, splitting variables are how often in a hurry when doing farm work, primary commodity farmed, size of farm, race, farm safety training for chemicals, education, behavior for wearing seat belt. For those with fair/poor machinery condition, the branch then splits according to the farm size. For smaller farms, the branch splits according to whether they received farm safety training. The fully grown tree has 15 terminal nodes. A plot for deviance against size is obtained as follows.

Based on the reduction on the deviance, the tree is pruned in order to have 10 terminal nodes as described in Table 1.

For African-Americans, the classification tree actually used condition of farm machinery, whether they are a full-time farmer, education, or number of regularly used farm machines. A total of 444 African-Americans farmers and farm workers were in the tree development. The overall injury rate is 0.115, slightly higher than that for all subjects. Again, the tree first splits according to the machinery condition. The fully grown tree has 7 terminal nodes. Based on the reduction on the deviance, we prune the tree to have 5 terminal nodes as summarized in Table 2.

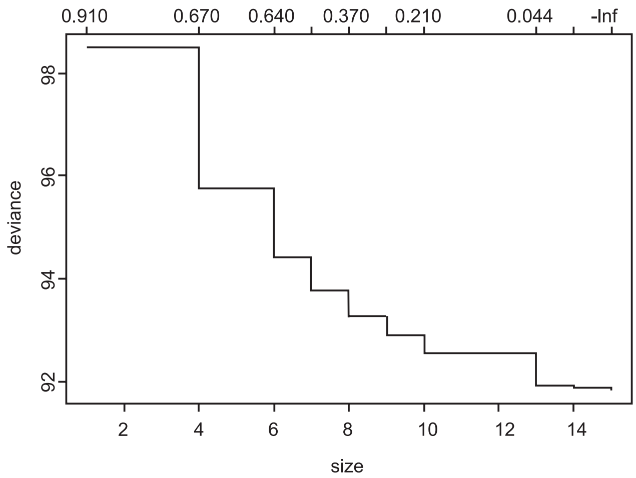

Fig. 1. Deviance against tree size using all the subjects

There were 607 farmers included in the tree for Caucasian owners. The overall injury rate is 0.097. The predictors used for the tree are primary commodity farmed, comorbid medical condition, attention level towards farm safety, number of regularly used farm machines, history of farm injury, education, and whether they had other non-farm jobs. The tree first splits according to the commodity, and the fully grown tree has 9 terminal nodes.

Table 1. Final classification for farm injury using all the subjects

Machine Condition	Excellent/Good							Fair/Poor		
Previous Injury	No				Yes					
Frequently in Hurry	No		Yes							
Commodity*	*a*, *b*, *c*, *d*, *g*	*e*, *f*	*a*, *b*, *c*, *f*	*d*, *e*						
Acres								< 749		≥749
Education (4-yr College or beyond)					No		Yes			
Seat Belt					Wear	Never				
Safety Training								No	Yes	
n	406	68	89	77	55	86	68	88	64	50
Rate	.039	.132	.067	.195	.127	.233	.074	.159	.063	.280

**a* = Aquaculture, *b* = Field Crop, *c* = Forestry, *d* = Livestock, *e* = Mixed, *f* = Other, and *g* = Vegetable.

Table 2. Final classification for farm injury using African-Americans

Machine Condition	Excellent/Good			Fair/Poor	
Commodity*				*a,d,e,f*	*b,c*
Full-Time Farmer		No	Yes		
Education (High School or Beyond)	No	Yes			
N	101	148	85	55	55
Rate	.109	.034	.153	.109	.291

**a* = Aquaculture, *b* = Field Crop, *c* = Forestry, *d* = Livestock, *e* = Mixed, *f* = Other, and *g* = Vegetable.

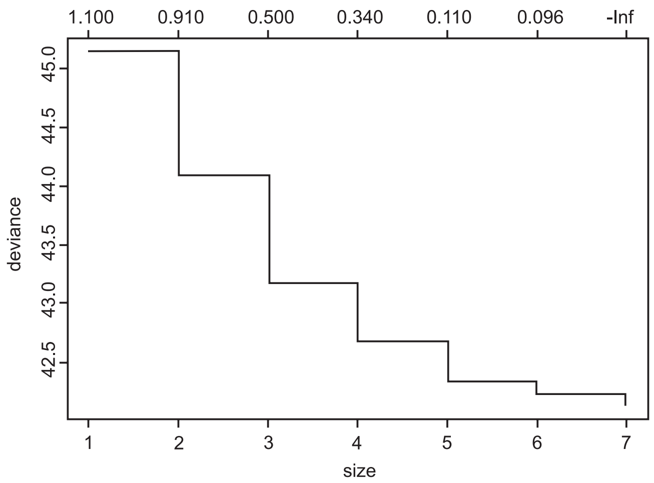

Fig. 2. Deviance vs. Tree size—African-Americans

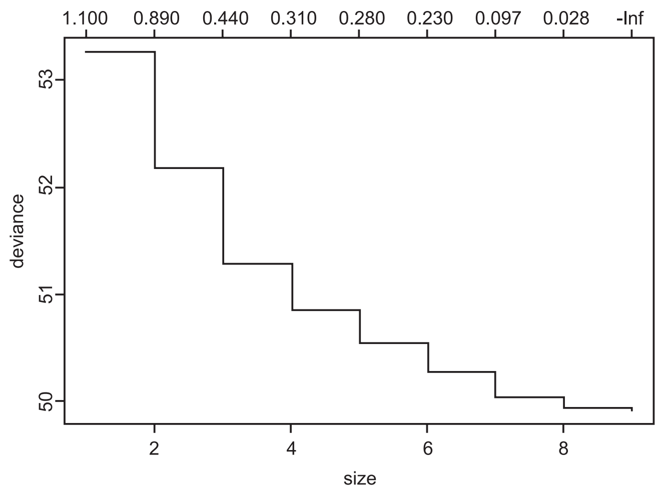

Fig. 3. Deviance against tree size using caucasian owners

Based on the reduction on the deviance, the tree is pruned to have 6 terminal nodes given in Table 3.

4. DISCUSSIONS

The approach of the classification tree is used to associate many potential risk factors with farm injury. From the tree for both races, the condition of farm machinery is an important factor related with the farm injury. Without adjusting other factors, the injury rates for fair/poor and excellent/good are 0.158 and 0.092, respectively, which give a risk ratio of 1.71. Among all terminal nodes, large farms (749 or more acres) with fair/poor condition on farm machines were the worst for farm injury with a rate of 28%, which is about 2.7 times the overall rate. For smaller farms with fair/poor condition on farm machines, safety training could greatly reduce the injury rate. With excellent/good condition on farm machinery, those who had a history of farm injury were more likely to have a new farm injury. College education or wearing a seat belt could

Table 3. Final classification for farm injury using caucasian owners

Commodity*	b,c		a,d,e,f,g			
# of Farm Machines	≤ 19	> 19				
Attention Level			Very attentive		Somewhat/Not Attentive	
# of Farm Machines					≤ 13	> 13
Previous Injury			No	Yes		
n 157	50	185	67	89	59	
Rate	.013	.120	.070	.149	.225	.136

**a* = Aquaculture, *b* = Field Crop, *c* = Forestry, *d* = Livestock, *e* = Mixed, *f* = Other, and *g* = Vegetable.

help reduce injury. For those without a history of farm injury, injury more likely occurred to those who were frequently in a hurry when doing farm work. Further, primary commodity farmed may be related to the chance of injury. Mixed commodity seemed more dangerous while livestock or other commodities may also cause a higher injury rate.

For African-Americans, fair/poor condition for farm machinery is still a big risk factor, causing more farm injuries, especially in field crop and forestry farms. With excellent/good condition on farm machines, full-time farmer/farm workers had more of a chance to have an injury. Among these, those with better education had less a chance to be injured. For Caucasian owners, the machinery condition was not used in the actual tree development. Primary commodity farmed was related to the chance of injury; that is to say there were lower injury rates in field crop and forestry. For field crop and forestry farms, injury is more likely to occur if there were a large number (≥ 19) of regularly used farm machines. For the farms with the rest of the commodities, attention level towards farm safety, number of regularly used farm machines and history of farm injury contributed to the injury. However, for those with a lower attention level, a smaller number (< 13) of regularly used farm machines was associated with a higher injury rate.

In general, farm commodity, farm size and number of farm machines could be related to the farm injury rate. Maintaining excellent/good condition on farm machines seems very important to preventing farm injuries. Furthermore, good education, good behaviors (e.g., not hurrying to do farm work, wearing a seat belt) and safety training can reduce farm injury. These are conditions that can be improved by certain interventions, e.g., by the government. Since those with a history of injury are more likely to be injured again, special attention and training may be required.

The analyses in this paper used a retrospective data set. Because of lost to follow-up, there may be some bias introduced. In general, the tree method provides less formal statistical inference, and hence the results in this paper are more exploratory in nature. Nevertheless, our results could provide some useful aids for farm injury prevention. The follow-up times of the subjects are not the same, and certainly may have impact on the injury variable. In our trees, we ignored the follow-up time assuming it is pretty much random. Adjusting the follow-up time in classification trees will be a future topic of study.

ACKNOWLEDGEMENTS

The authors would like to thank Drs. G. McGwin and J. Roseman of the University of Alabama at Birmingham for providing the data used in this paper.

REFERENCES

Bell, C.A., Stout, N.A., Bender, T.R., Conroy C.S., Crouse WE, Myers, J.R. (1990). Fatal occupational injuries in the United States, 1980 through 1985. *JAMA,* **263**, 3047-3050.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth and Brooks/Cole.

Gerberich, S.G., Gibson, R.W., French, L.R., *et al*. (1998). Machinery-related injuries: Regional rural injury study-I (RRIS-I). *Accid. Anal. Prev.,* **30**, 793-804.

Lyman, S., McGwin, G. Jr., Enochs, R., Roseman, J.M. (1999). History of agricultural injury among farmers in Alabama and Mississippi: Prevalence, characteristics, and associated factors. *Am. J. Ind. Med.,* **35**, 499-510.

McGwin, G. Jr., Enochs, R., Roseman, J.M. (2000). Increased risk of agricultural injury among African farm workers from Alabama and Mississippi. *Am. J. Epidemiol.,* **152**, 640-650.

National Committee for Injury Prevention and Control. (1989). *Injury Prevention: Meeting the Challenge*. Oxford University Press, New York.

Rossignol, M., Pineault, M. (1993). Fatal occupational injury rates, 1981 through 1988. *Am. J. Public Health,* **83**, 1563-1566.