

Analysis of Maryland Poisoning Deaths Using Classification And Regression Tree (CART) Analysis

Carol Pamer, RPh, MS^{1,3}, Tracey Serpi, MS, PhD², Joseph Finkelstein³, MD, PhD, MA

¹Office of Surveillance and Epidemiology, Food and Drug Administration, Silver Spring, MD

²Center for Preventive Health Services, Maryland Department of Health and Mental Hygiene, Baltimore, MD

³Chronic Disease Informatics Group, University of Maryland School of Medicine, Baltimore, MD

Abstract

Our study is a cross-sectional analysis of Maryland poisoning deaths for years 2003 and 2004. We used Classification and Regression Tree (CART) methodology to classify 1,204 Maryland undetermined intent poisoning deaths as either unintentional or suicidal poisonings. The predictive ability of the selected set of variables (i.e., poisoned in the home or workplace, location type where poisoned, place of death, poison type, victim race and age, year of death) was extremely good. Of the 301 test cases, only eight were misclassified by the CART regression tree. Of 1,204 undetermined intent poisoning deaths, CART classified 903 as suicides and 301 as unintentional deaths. The major strength of our study is the use of CART to differentiate with a high degree of accuracy between unintentional and suicidal poisoning deaths among Maryland undetermined intent poisoning deaths.

Introduction

In the U.S. and globally, deaths due to poisoning have been increasing steadily over time, with more recent steep increases.¹ Poisonings may be due to illicit or licit drugs, chemicals, gases, or other substances. The U.S. increases, however, have been largely attributed to unintentional and undetermined intent drug poisonings. The illness burden of these potentially preventable deaths is high. In the U.S. between 1999 and 2005, there were 180,347 deaths attributed to poisoning, 120,596 of which were unintentional deaths, 37,435 suicides, and 21,792 of undetermined intent.² An economic analysis for 4,862 self-inflicted fatal poisonings in the U.S. for year 2000 estimated the medical costs of these poisonings at 17 million

dollars and lost productivity costs at 4.8 billion dollars.³ Poisonings were the leading cause of injury deaths in Maryland in the year 2004 (718 deaths) and occurred more frequently than motor vehicle accident (660), firearm (654), fall-related (430), and suffocation deaths (316).⁴ In the same year, Maryland's age-adjusted undetermined intent death rate ranked 2nd (10.94 per 100,000 population) among the 50 states and the District of Columbia.²

Considerable ongoing public health efforts attempt to reduce the rate of poisonings. An important complement to those programs is an accurate surveillance system. Currently, the U.S. Centers for Disease Control and Prevention (CDC) National Center for Health Statistics (NCHS) produces the National Vital Statistics System (NVSS) as the primary U.S. mortality surveillance system.⁵ The NVSS is based upon official death certificates. One problem in completing death certificates involves assessing the intent behind a decedent's death.⁶ Due to these uncertainties, a number of investigators believe that a proportion of suicides are misclassified as unintentional or undetermined intent injury deaths.⁷ This has important implications for the surveillance of fatal poisoning deaths.

In 2003, the CDC began implementation of a new surveillance system, the National Violent Death Reporting System (NVDRS).⁸ Participating states abstract multiple records and compile case data in electronic format for suicides, homicides, deaths due to legal intervention, unintentional firearm injury deaths, and deaths of undetermined intent. Deaths due to unintentional poisoning are not included. The abstracted records include death certificates, medical examiner and coroner records, law enforcement

records, and crime laboratory records. Ultimately, state-level data are sent to the CDC to create the NVDRS. With more than 300 variables available for study, the NVDRS provides a richer data source for surveillance of these types of injuries versus the NVSS, which is based only on official death certificates. An abbreviated version of NVDRS data can also be freely downloaded via the National Archive of Criminal Justice Data website.⁹ Maryland is one of the original reporting NVDRS states. The Maryland state-level surveillance system is the Maryland Violent Death Reporting System (MVDRS).¹⁰ A number of studies of poisoning deaths have been conducted with usual statistical methods, such as multivariable logistic regression modeling, to determine which characteristics are most prevalent among the different types of poisoning deaths and which of those characteristics may predispose a poisoning death to be classified as undetermined intent. Another statistical methodology which has been used extensively to develop classification schemes and to partition data is CART (Classification and Regression Tree Analysis).

Methods

Our study is a cross-sectional analysis of Maryland poisoning deaths for years 2003 and 2004. We used Classification and Regression Tree (CART) methodology to classify Maryland undetermined intent poisoning deaths as either unintentional or suicidal poisonings. The CART methodology¹¹ has been used extensively and previously described.¹⁶⁻¹⁸ Briefly, this methodology utilizes recursive partitioning to build hierarchical binary classification trees. The classification trees are automatically developed to forecast target events by considering every possible cutpoint on every independent predictor at every node in the classification tree. An “impurity criteria” is used¹¹ to identify optimal cutpoints which distinguish outcomes at each node. Advantages of CART versus multivariable regression analysis are the ability to utilize large numbers of predictor variables, nonreliance on the underlying distributions for statistical inference, and variables with missing data can still be utilized.¹⁶⁻¹⁷

CART methodology has been successfully utilized across a wide range of medical disciplines. Adams KF Jr et al¹⁹ used CART for prediction of in-hospital mortality, Ambalavanan N et al²⁰ utilized CART to predict death in preterm infants, whereas Bevilacqua M et al²¹ employed the CART approach for prediction of high accident risk situations. CART also has been successfully used for predictive modeling in patients with cardiovascular conditions²²⁻²⁴, cancer²⁵,

work-related disabilities²⁶, upcoming asthma exacerbations²⁷, and drug-induced hyperglycemia in hospitalized patients.²⁸

Two data sources were utilized in our study. Maryland poisoning cases of undetermined intent and the CART learning dataset for suicidal poisonings consisted of the years 2003 and 2004 National Violent Death Reporting System Public Use Datasets (NVDRS-PUD), which have been described previously.¹² The NVDRS-PUD was limited to Maryland residents. Cases in which the weapon was not a poison were excluded. Unintentional poisoning deaths are not included in NVDRS. Because of this, the CART software learning dataset for unintentional poisonings consisted of NCHS Vital Statistics Multiple Cause of Death Data (NCHS-MCOD) files obtained from the National Bureau of Economic Research (NBER) website, which are described elsewhere.^{13,14} Maryland unintentional poisoning deaths with the ICD-10 codes X40 to X49 were retrieved from the NCHS-MCOD files. Ten variables in the NCHS-MCOD file were recoded to match the NVDRS-PUD file. These included: sex, race, age, marital status, type of location where injured, injured at work, level of education, birth place, injured at home, and place of death. The type of drug involved in the unintentional poisoning was derived using the underlying cause of death and 20 additional ICD-10 cause code fields. The type of drug involved was classified as alcohol, street drug, over-the-counter drug, prescription drug, other drug, or carbon monoxide.

There were 1,204 NVDRS-PUD undetermined intent, 172 NVDRS-PUD suicide, and 129 NCHS-MCOD unintentional Maryland poisoning deaths available for study. The NVDRS-PUD data contained 34 variables and NCHS-PUD contained 43 variables, but only 12 variables were common to both data sets. Therefore, the CART analysis variables were limited to twelve variables common to the three manner of death-type files: manner of death (unintentional, suicide, undetermined intent), age (categorized in bands), sex, race (white/black/other), marital status, type of location where injured, injured at work, injured at home, birthplace, place of death, type of poison, and case year. All twelve variables were included in the set of predictors for the CART.

Results

We conducted the CART analysis using eleven variables common to the two dataset types. The cases were divided into groups by the twelfth common variable, manner of death. Table 1 provides a list of

the variables and the importance of the variables in terms of differentiating between unintentional and suicidal poisonings.

Table 1. Variable Importance

Variable	Score
Poisoning occurred in the home	100.0
Type of location where poisoned	54.61
Type of poison	43.81
Poisoning occurred at work	39.97
Race	35.97
Place of death	25.12
Age	1.75
Year of death	1.04
Marital status	0.30
Sex	0.00
Place of birth	0.00

Eight of the eleven variables were strong in their ability to differentiate between unintentional and suicidal poisoning deaths. Four of these variables were related to physical location: poisoned in the home or workplace, type of location where poisoned, and place of death. Also important were the type of poison, race and age of victim, and year of death. The three other variables had limited abilities to differentiate between poisoning deaths by intent. These variables were marital status, sex, and place of birth of the victim.

Table 2 provides the results of a test for CART misclassification of the unintentional and suicidal poisoning deaths used to create the algorithm. Of the 301 test cases, only eight were misclassified under the regression tree created by the CART.

Table 2. Performance of the CART Classification Algorithm

Class	Number of Cases	Number Misclassified	%Error
Suicide	172	5	2.91
Unintentional	129	3	2.33

After a classification tree was created using the learning datasets, the algorithm was applied to the Maryland undetermined intent poisoning deaths. Table 3 provides the results of that analysis in which undetermined intent poisoning deaths were classified as either unintentional or suicidal poisoning deaths by the CART algorithm. Of the 1,204 Maryland undetermined intent poisoning deaths, CART classified 903 as suicides and 301 as unintentional deaths.

Table 3. CART classification of the undetermined intent poisoning deaths

Class	Suicide	Unintentional	Total
Undetermined Intent	903	301	1,204

An additional test of the CART classification performance was carried out by utilizing the CART classification results of the undetermined intent poisoning deaths for CART re-training and applying the resulting algorithm to the 1,204 cases from the Maryland dataset. This test was conducted by changing the class of undetermined intent cases to the CART-assigned category (as indicated in the Table 3) of either unintentional or suicidal poisoning. These cases were then analyzed by the CART classification scheme to estimate a classification percent error. Table 4 provides the results of this test.

Table 4. Performance of the CART classification algorithm based on the classified Maryland dataset

CART-assigned Class	Total Cases	Number Misclassified	%Error
Suicide	903	4	0.44
Unintentional	301	1	0.33

Discussion

The predictive ability of the selected set of variables was extremely good. Of the 301 test cases, only eight were misclassified under the regression tree created by the CART. Of the 1,204 Maryland undetermined intent poisoning deaths, CART classified 903 as suicides and 301 as unintentional deaths.

One study has been published in which CART methodology was used to classify undetermined intent poisoning deaths in the state of Utah in 2002.¹⁵ The authors performed the analysis using a data source that was the precursor to NVDRS, the National Violent Injury Statistics System (NVISS). The CART analysis identified several variables which differentiated between unintentional and suicidal poisoning deaths: previous suicidal behavior, drug abuse, physical health problems, depressed mood, and age. Based on the CART classification results, the authors concluded that the official Utah suicidal poisoning death rates could be underestimated by

30%, overall completed suicide rates by 10%, and unintentional poisoning death rates by 61%.

Our study was an investigation of suicidal, unintentional, and undetermined intent poisoning deaths, with a particular emphasis on those occurring in Maryland. We performed a CART analysis to determine which variables are most likely to discriminate between fatal unintentional and suicidal poisonings.

There were limitations to our study. A major issue was that the CART analysis was limited to only those variables common to both the public use NVDRS and NCHS multiple cause of death files. Variables such as day and month of death and county were available in the NCHS file, but not in the NVDRS file. Variables on the circumstances surrounding the death were available in the NVDRS file, but not in the NCHS data. Including variables on the circumstances of the death could improve the analysis using CART. Also, the National Violent Death Reporting System is a new surveillance system, with few published analyses and results. The NVDRS data are secondary data collected from sources that are used for forensic, medico-legal purposes and not for public health research. Also, the NVDRS and NCHS multiple cause of death files are based on death certificates, which are known to vary in terms of accuracy and completeness.

In spite of their limitations, a strength of our study is the data sources. Both data sources are based on official law enforcement, forensic investigation, and death records, which are legal documents carefully collected and recorded by professionals. The NVDRS data are based on an overall assessment of multiple records collected by law enforcement and forensic experts, including coroners and medical examiners. These datasets are also readily accessible by researchers at no cost.

The cases in this analysis were reviewed by legal and/or forensic experts prior to their inclusion in the NVDRS and NCHS databases. Although a universal standard (beyond the traditional forensic autopsy) for classifying the intent of poisoning victims does not currently exist, this study could be enhanced by conducting psychological autopsies of the cases using the existing, original case files.²⁹ Undergoing such a process may serve to validate the CART analysis results.

Conclusion

CART methodology allows researchers to classify undetermined intent Maryland poisoning deaths as

either unintentional or suicidal poisonings with a high degree of accuracy.

Acknowledgements

The authors would like to thank Shlyakhov I, PhD for technical assistance with this project. This study was conducted while Ms. Pamer was a student at the University of Maryland, Baltimore. The views expressed are those of the article authors and not of the US FDA.

References

1. Paulozzi LJ, Budnitz DS, Xi YX. Increasing deaths from opioid analgesics in the United States. *Pharmacoepi Drug Safety* 2006; 15:618-27.
2. US Centers for Disease Control and Prevention. National Center for Injury Prevention and Control. WISQARS™. Website: <http://www.cdc.gov/ncipc/wisqars>. Accessed March 13, 2008.
3. Corso PS, Mercy JA, Simon TR, Finkelstein EA. Medical costs and productivity losses due to interpersonal and self-directed violence in the United States. *Am J Prev Med* 2007; 32(6):474-482, 482.e1-e2.
4. Maryland Department of Health and Mental Hygiene Center for Preventive Health Services. Injuries in Maryland: 2004 Statistics on Injury-Related Emergency Department Visits, Hospitalizations, and Deaths. DHMH website: Accessed August 30, 2007.
5. US Centers for Disease Control and Prevention. National Center for Health Statistics. Mortality data from the National Vital Statistics System. Website: <http://www.cdc.gov/nchs/deaths.htm>. Accessed February 11, 2007.
6. Hanzlick R, Goodin J. Mind your manners: part III: individual scenario results and discussion of the National Association of Medical Examiners manner of death questionnaire, 1995. *Am J Forensic Med Pathol* 1997; 18(3):228-45.
7. Phillips DP, Ruth TE. Adequacy of official suicide statistics for scientific research and public policy. *Suicide Life Threat Behav* 1993; 23(4):307-319.
8. Paulozzi LJ, Mercy J, Frazier L, Annest JL. CDC's National Violent Death Reporting System: background and methodology. *Inj Prevent* 2004; 10:47-52.

9. National Archive of Criminal Justice Data. National Violent Death Reporting System Series. Website: <http://www.icpsr.umich.edu/cocoon/NACJD/SERIES/00217.xml>. Accessed May 4 2007.
10. U.S. Centers for Disease Control and Prevention: National Violent Death Reporting System website. <http://www.cdc.gov/ncipc/profiles/nvdrs/facts.htm>. Accessed July 17, 2007.
11. Salford Systems, Inc. An Overview of the CART® Methodology White Paper. Website: <http://www.salfordsystems.com/421.php>. Accessed August 31, 2007.
12. National Archive of Criminal Justice Data. National Violent Death Reporting System Series. Website: <http://www.icpsr.umich.edu/cocoon/NACJD/SERIES/00217.xml>. Accessed May 4 2007.
13. National Bureau of Economic Research (NBER). Mortality Data – Vital Statistics NCHS's Multiple Cause of Death Data, 1959-2004. Website: <http://www.nber.org/data/multicause.html>. Accessed May 23 2007.
14. U.S. Centers for Disease Control and Prevention: National Center for Health Statistics Mortality Data, Multiple Cause-of-Death Public-Use Data Files. Website: http://www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm. Accessed May 4 2007.
15. Donaldson AE, Larsen GY, Fullerton-Gleason L, Olson LM. Classifying undetermined poisoning deaths. *Injury Prevention* 2006; 12:338-343.
16. Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
17. Terrin N, Schmid CH, Griffith JL, D'Agostino RB, Selker HP. External validity of predictive models: a comparison of logistic regression, classification trees, and neural networks. *J Clin Epidemiol*. 2003;56(8):721-9.
18. Chipman HA, George EI, McCulloch RE. Bayesian CART Model Search. *Journal of the American Statistical Association* 1998;93(443):935-960.
19. Adams KF Jr, Uddin N, Patterson JH. Clinical predictors of in-hospital mortality in acutely decompensated heart failure-piecing together the outcome puzzle. *Congest Heart Fail*. 2008;14(3):127-34.
20. Ambalavanan N, Van Meurs KP, Perritt R, Carlo WA, Ehrenkranz RA, Stevenson DK, Lemons JA, Poole WK, Higgins RD; NICHD Neonatal Research Network, Bethesda, MD. Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure. *J Perinatol*. 2008;28(6):420-6.
21. Bevilacqua M, Ciarapica FE, Giacchetta G. Industrial and occupational ergonomics in the petrochemical process industry: A regression trees approach. *Accid Anal Prev*. 2008;40(4):1468-79.
22. Girman CJ, Dekker JM, Rhodes T, Nijpels G, Stehouwer CD, Bouter LM, Heine RJ. An exploratory analysis of criteria for the metabolic syndrome and its prediction of long-term cardiovascular outcomes: the Hoorn study. *Am J Epidemiol*. 2005;162(5):438-47.
23. Bukkapatnam S, Komanduri R, Yang H, Rao P, Lih WC, Malshe M, Raff LM, Benjamin B, Rockley M. Classification of atrial fibrillation episodes from sparse electrocardiogram data. *J Electrocardiol*. 2008;41(4):292-9.
24. Möckel M, Danne O, Müller R, Vollert JO, Müller C, Lueders C, Störk T, Frei U, Koenig W, Dietz R, Jaffe AS. Development of an optimized multimarker strategy for early risk assessment of patients with acute coronary syndromes. *Clin Chim Acta*. 2008;393(2):103-9.
25. Kohrt HE, Olshen RA, Bermas HR, Goodson WH, Wood DJ, Henry S, Rouse RV, Bailey L, Philben VJ, Dirbas FM, Dunn JJ, Johnson DL, Wapnir IL, Carlson RW, Stockdale FE, Hansen NM, Jeffrey SS; Bay Area SLN Study. New models and online calculator for predicting non-sentinel lymph node status in sentinel lymph node positive breast cancer patients. *BMC Cancer*. 2008 Mar 4;8:66.
26. Baker NA, Sussman NB, Redfern MS. Discriminating between individuals with and without musculoskeletal disorders of the upper extremity by means of items related to computer keyboard use. *J Occup Rehabil*. 2008;18(2):157-65.
27. Duvvuri SK, Finkelstein J. Computerized Decision Support for Asthma Management. *Value in Health* 2006;9(3):A101-102.
28. Mohr JF 3rd, Peymann PJ, Troxell E, Lodise TP, Ostrosky-Zeichner L. Risk factors for hyperglycemia in hospitalized adults receiving gatifloxacin: A retrospective, nested case-controlled analysis. *Clin Ther*. 2008;30(1):152-7.
29. Scott CL, Swartz E, Warburton K. The psychological autopsy: solving the mysteries of death. *Psych Clin North Am* 2006; 29(3):805 – 822.