Predictive analytics, text mining, and drug-impaired driving in automobile accidents

03 April 2013 | PHILIP S. BORBA, Ph.D.

Drug-impaired driving is an increasingly difficult problem for property-casualty insurers, law enforcement officers, prosecutors, judges, and policy makers. From a nationally representative survey, 16% of weekend nighttime drivers tested positive for illicit drugs or medications.[1] One in eight high school seniors responding to a 2011 survey reported driving after smoking marijuana in the two weeks preceding the survey. One in three drivers who died in fatal crashes and had known drug-test results tested positive for drugs (illicit substances as well as over-the-counter and prescription medications).[2] Even as the total number of drivers killed in motor vehicle crashes declined 21% from 2005 to 2009, the involvement of drugs in fatal crashes increased by 5% over the same time period.[3]

Why the interest (especially on the part of property-casualty insurance stakeholders) in identifying drug- impaired driving for drivers involved in an automobile accident? Let's begin with three reasons.

- First, claim triage. Knowing that a driver (whether the insured driver or the other driver) might have been under the influence of a medication, prescription, drug, or illegal narcotic will help the insurer assign a claim adjuster or other specialist who is able to efficiently determine whether a drug-impaired condition existed at the time of the accident. As described below, determining whether a driver was driving under the influence of a drug (DUID) is much more complicated than determining whether a driver was under the influence of alcohol. Furthermore, the triage assignment could be specific to whether the driver had been taking a medication (such as an over-the-counter product), prescription, drug, or illegal narcotic.

- Second, assignment of liability, and especially subrogation opportunities. Finding that the other driver was DUID may be cause for a subrogation recovery against that driver, or provide enough additional evidence to increase the likelihood or size of the recovery.

- And third, knowing a driver had been involved in an accident while on a medication, prescription, drug, or illegal narcotic may be a reason to non-renew or to renew at a higher rate.

In this article, we demonstrate how information not commonly captured in automobile insurer data systems but available in automobile accident descriptions can improve an insurer's ability to predict accident severity. We extract from accident descriptions information not typically captured in insurer data systems to capture whether one of the drivers in the accident was on a medication, prescription, drug, or an illegal narcotic. We found that the additional information in accident descriptions improved the ability to predict the severity of an accident. Narrative data can feed predictive analytics, improve claim-triage and subrogation recovery opportunities, and power a more intelligent approach to renewals and rate-classification. With DUID representing a measurable (but largely unrecognized) source of increased accident severity, automobile insurers have an opportunity to extract value from text mining and better manage the risk posed by driving under the influence of drugs.

Detecting DUID is difficult

Detecting drivers who are under the influence of a drug is much more complicated than detecting drivers who are driving under the influence of alcohol (or driving while intoxicated, DWI). Alcohol passes through the body in a reasonably predictable manner and has a reasonably consistent impact on a driver's ability to operate a vehicle safely. Field tests can be performed efficiently for DWI with acceptable accuracy. Furthermore, it does not matter whether the blood alcohol content (BAC) was due to the intake of beer, wine, or hard alcohol; the sex, age, or body mass of the individual; or the length of time since consumption. In most states, a BAC of 0.08 grams per deciliter or higher is considered a per se case of DWI.

By contrast, tests for medications and prescriptions are more difficult to perform. It may take days, weeks, or months to obtain results. For the impairment-impact of drugs on an individual's ability to operate a motor vehicle, there is no corollary to a BAC standard. Detecting drug-impaired driving is a complex problem due to the large number of substances with the potential to impair driving and impose the risk of an accident, the variations in the ways that different drugs can impair driving, the lack of basic information concerning the drugs that impair driving, and the differences in the ways that the drugs can affect the body and behavior.[4]

Identifying the presence of drugs in auto accidents

In recent years, there has been an increased effort to train law enforcement to recognize drivers that may be DUID. The typical case is for an officer to perform a standardized field sobriety test for a driver's blood alcohol content. If the BAC is found to exceed the statutory limit, the officer is unlikely to test for drugs, and consequently the incidence of DUID may be understated. If the BAC does not exceed the statutory limit, the officer may seek evidence for a DUID charge. In most states, a Drug Evaluation and Classification (DEC) program has been made available to law enforcement personnel and many officers have been trained to be a Drug Recognition Expert (DRE). Nevertheless, identifying drivers under the influence of a drug is much more complicated than the testing for alcohol with a breathalyzer or urine test. For drug impairment, the tests may require a broader range of specimens (e.g., blood, urine, oral fluid, sweat, hair) and the present technology often requires lab tests that may take days, weeks, or months for findings.

This built-in delay—and the variety of potential results—may pose a challenge when it comes to accurately tracking DUID instances, since the pertinent information may not be available at the time of the accident or when the police reports are prepared.

The presence of medications, prescriptions, drugs, and illegal narcotics in automobile accidents

We have developed methods to efficiently read, organize, and analyze large volumes of narrative data captured in accident descriptions, adjuster notes, and other reports and documents where narrative information is recorded in an unstructured text format. Within a single narrative report and across reports, the same concept (*such as taking his medications*) can be expressed in numerous ways.

We have developed methods to organize the different-but-similar expressions into a format that can be categorized and then included in statistical analyses. A federal agency database on automobile accidents provided us with an opportunity to showcase these methods. The National Highway Traffic Safety Administration (NHTSA) compiled information on a broad representation of approximately 7,000 passenger automobile accidents. The accidents included single- and multiple-vehicle incidents, drivers of various ages, and accidents occurring in a variety of environmental conditions. The NHTSA compiled a narrative description of each accident, as well as the usual information on number of vehicles, road conditions, weather conditions, and time of day. The narrative described the environmental conditions, vehicular movements, and driver behavior at the time of the accident. On average, the narrative provided approximately 440 (and as many as almost 1,300) words describing the accident.

⊔ Milliman

Processing the accident descriptions

Narrative descriptions for the 7,000 NHTSA accidents were broken into phrases, and similar phrases were grouped together using analytical models. After removing prepositions and uninformative prepositional phrases, the result was a data file with more than 13 million phrases (see Figure 1).

Figure 1: Incidence rates in automobile accidents and percent of accidents with injury for various conditions

Condition	Incidence Among Accidents	Percent With Injury	Injury Frequency Compared to "All Accidents"
All accidents	100%	73%	
Time of day/week			
Night	22%	69%	-
Weekend	22%	73%	+
Environment			
Weather	24%	71%	-
Wet roads	16%	67%	-
Nature of accident			
Multiple vehicles	74%	76%	+
Rear-end	18%	70%	-
Head-on	2%	86%	+
Turned into path	16%	81%	+
Driver condition			
Driver fatigued	13%	76%	+
Alcohol (police report)	6%	82%	+

Next, we used four different themes for identifying the presence of a medication, prescription, drug, or illegal narcotic. First, we identified phrases with a "taking medications" theme. We joined phrases with the word "medications" that indicated a driver may have been taking medications. For example, we joined "on many" and "taking pain" to form "on many medications" and "taking pain medications," respectively.

The second theme followed the same process, replacing "medications" with "prescriptions," which gave us phrases such as "on many prescriptions" and "taking pain prescriptions." These two themes produced approximately 1,100 phrases.[5]

The third theme joined an action and a drug name. The result from these joins was a long list of phrases with "had taken [drug name]," "was on [drug name]," and so on, replacing [drug name] with the names of drugs. For the present analysis, we worked with 3,590 phrases with a drug name. The fourth theme was a list of 52 references to illegal narcotics that we considered red flags when seen on an accident description. This list included "cocaine," "heroin," and "marijuana."

> Accident #1: "The driver of V1 sustained serious injuries during the crash…. The driver of V1 was taking several medications for various health problems…." (380 words)

> Accident #2: [The driver] was transported, treated, and released at a local hospital for a head injury… An associated factor coded to this driver was the use of prescription medications…general health medication with possible side effects… (471 words)

Milliman

In sum, the first two themes were general references to medications and prescriptions, the third theme captured references to drug names, and the fourth theme was a list of illegal narcotics that we considered "red flags" for a driver being under the influence of a drug. For each theme, a binary (0/1) variable was created to capture whether the presence of medications, prescriptions, a drug name, or an illegal narcotic was mentioned in the accident description.

Accident #3: The driver was admitted … for a fractured femur and other injuries… a urine sample three hours after the collision tested positive for amphetamines and marijuana…
(584 words)

An injury was reported to have occurred in 73% of the 6,949 accidents in the NHTSA database (see Figure 2).[6] We found a reference to taking or being on a medication in approximately 16% of the accidents, and an injury occurred in 82% of these accidents. Similarly, we found a reference to taking or being on a prescription or a drug in approximately 6.5% of the accidents and an 80% injury occurrence for these subsets of accidents. Finally, we found reference to an illegal narcotic in 2.4% of the accidents and that an injury occurred in 89% of these accidents.

Figure 2: Presence of a medication, prescription, drug, or illegal narcotic: Incidence rates in automobile accidents and percent of accidents with injury

| Condition	Incidence Among Accidents	Percent With Injury	Injury Frequency Compared to "All Accidents"
All accidents	100.0%	73%	
Medication	15.7%	82%	+
Prescription	6.4%	80%	+
Drug	6.6%	80%	+
Illegal narcotic	2.4%	89%	+

Predictive analytics

The variables created from the narrative data were combined with information from the structured data (such as day of the week, time of day, weather conditions at the time of the accident, and the nature of the accident) and used in a multivariate analysis designed to identify the factors associated with whether an injury occurred. The purpose for including the structured data was to take into consideration the information commonly available on auto accidents. If the variables from the narrative data did not improve the statistical results from the predictive analytics then the time and effort used to extract the information may not be worthwhile.

The analytic procedure was a logit analysis where the outcome measure was whether an injury occurred with the accident.[7] The purpose of the analysis was to test whether the inclusion of the information from the accident descriptions improves the ability to predict accident severity. The database did not have information on the economic loss of the accident, and consequently we used whether an injury occurred as a proxy for accident severity; that is, we presumed that accidents with injuries are more expensive and serious than accidents without them.

In the logit analysis, we used information from the structured data to develop variables that would otherwise be included in an accident-severity analysis. The structured-data variables were whether the accident occurred at night, on a weekend, with poor weather, on a wet road surface, involved multiple vehicles, a rear-end or a head-on collision, or turning into the path of another vehicle, and if alcohol was present. We tested for the influence of four variables from the narrative data: reference to the presence of (1) a medication, (2) a prescription, (3) a drug, and (4) an illegal narcotic. For each of the four variables, we found that automobile accidents where the accident description indicated the presence of a medication, prescription, drug name, or illegal narcotic for at least one of the drivers increased the likelihood an injury occurred with the accident. The higher probabilities were statistically significant.

Milliman

For each logit analysis, the chart in Figure 3 presents the probabilities for the reference group and with the inclusion of the narrative-data variable. For example, in the medications analysis, the probability of an injury was 0.57 for an accident that occurred in the daytime, on a weekday, in good weather, on a dry road surface, as a single-vehicle accident that was not a rear-end, head-on, or turning-into-the-path accident, and where alcohol was not present. For the same conditions, the probability of an injury increased to 0.75 if the accident description mentioned that a driver was taking or was on a medication—an increase of 18 percentage points or 30% over the no-medications probability. There was a similar finding for the presence of a prescription, drug, or illegal narcotic. Starting with the reference group's 0.57 probability of an injury and holding all other variables constant, we found that the probability of an injury increased to 0.72 when one of the drivers was on a prescription, 0.72 when a drug name was mentioned, and 0.85 when there was mention of a specific illegal narcotic. Each of these findings was statistically significant at the 95% level of confidence.

Figure 3: The probability of an injury was higher when there was the presence of a medication, prescription, drug, or illegal narcotic for one or more drivers

■ Reference group ■ With presence of a medication, prescription, drug, or illegal narcotic for one or more drivers
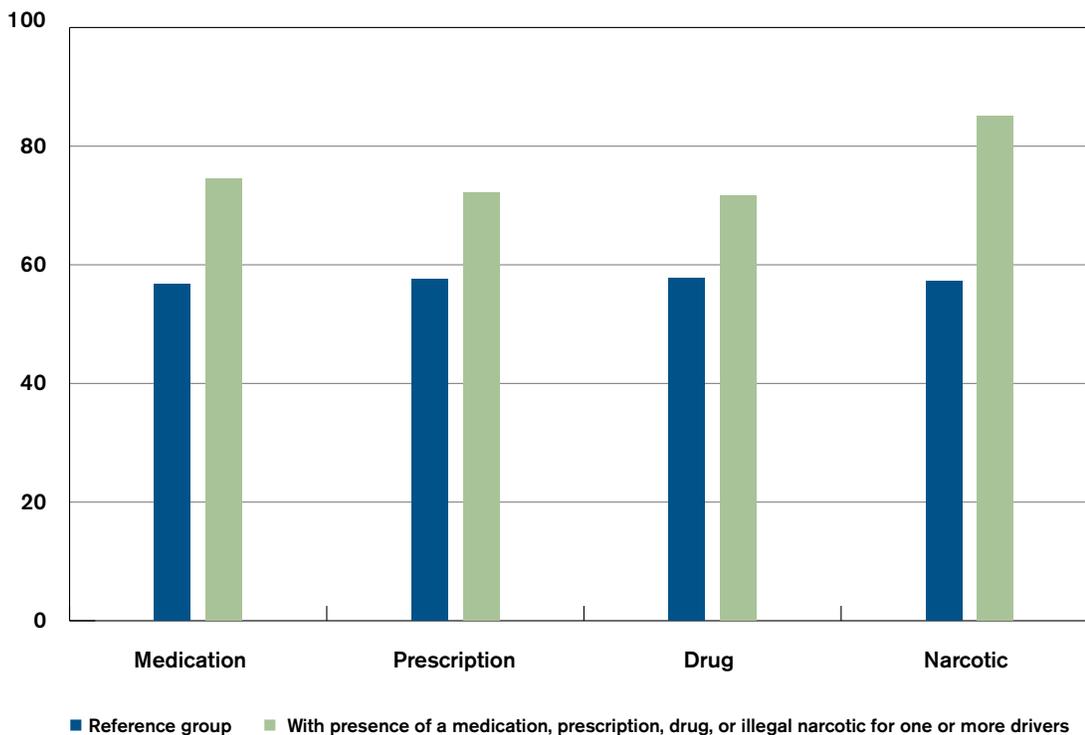
Concluding comment

This analysis demonstrates that narrative data can be used to help insurers improve the results for predictive analytics. Information captured in claim adjuster notes and other text data can be used for improved claim-triage assignments, quicker identification of subrogation recovery opportunities, and to gather information on policyholders for renewal and rate-classifying purposes. As DUID becomes a more prevalent problem, the ability to identify these cases and respond accordingly will become an increasingly important aspect of an automobile insurer's ability to efficiently administer and control claim costs.

1   Richard Compton and Amy Berning. Results of the 2007 National Roadside Survey of Alcohol and Drug Use by Drivers. Washington, DC: National Highway Traffic Safety Administration, Report No. DOT HS 811 175.

2   Office of National Drug Control Policy (October 2011). Drug Testing and Drug-Involved Driving of Fatally Injured Drivers in the United States: 2005-2009.

Milliman

3   Office of National Drug Control Policy (November 2010). Working to Get Drugged Drivers Off the Road. Fact Sheet.

4   National Highway Traffic Safety Administration. Drug-Impaired Driving: Understanding the Problem and Ways to Reduce It: A Report to Congress. Report No. DOT HS 811 268, page 2.

5   For the present analyses, "aspirin," "birth control," and "vitamins" were excluded from the references for medications and prescriptions.

6   Although the NHTSA database provided case weights for the accidents, we did not apply the weights in our analyses. We take the perspective that the analyses are from the perspective of a property-casualty insurer, which is not likely to apply any weights when performing predictive analytics. The application of the case weights would have been more appropriate if the analysis was intended to extrapolate to all accidents or for public policy conclusions.

7   Given that the database was limited to accidents, an analysis on the impact of drugs on accident frequency was beyond the scope of the analysis.

Philip S. Borba is a principal and senior consultant in Milliman's New York office. Contact him at phil.borba@milliman.com**.**

Milliman