Using Text Data from Motor-Vehicle Accident Descriptions to Identify Drivers Under the Influence of Legal or Illegal Drugs

presented by:
Philip S. Borba, Ph.D.
Milliman, Inc.
New York, NY

March 11, 2015

Casualty Actuarial Society, Ratemaking & Product Management Seminar, Dallas, TX

Milliman

Casualty Actuarial Society -- Antitrust Notice

The Casualty Actuarial Society is committed to adhering strictly to the letter and spirit of the antitrust laws. Seminars conducted under the auspices of the CAS are designed solely to provide a forum for the expression of various points of view on topics described in the programs or agendas for such meetings.

Under no circumstances shall CAS seminars be used as a means for competing companies or firms to reach any understanding – expressed or implied – that restricts competition or in any way impairs the ability of members to exercise independent business judgment regarding matters affecting competition.

It is the responsibility of all seminar participants to be aware of antitrust regulations, to prevent any written or verbal discussions that appear to violate these laws, and to adhere in every respect to the CAS antitrust compliance policy.

Overview

- Starting Considerations and Definitions
- Reasons to be Interested in Text Data
- Using Text Data, identifying Medications, Prescriptions, and Narcotics in Auto Accidents
- National Motor Vehicle Crash Causation Survey
- Accident Descriptions:
  - 3 examples where medication, prescription, a drug name, or a narcotic is mentioned
  - NMVCCS Accident Descriptions compared to Claim Adjuster Notes
  - Breaking Text Data into Manageable Units – Creating NGrams
- Incidence of Medications, Prescriptions, and Narcotics in Auto Accidents
- Multivariate (Logit) Analyses

Limitations

- Results in this presentation are for demonstration purposes only.
- Data are from public sources and have been reviewed for consistency but have not been audited.
- The analyses and statistical results are intended to demonstrate the principles of text-mining and predictive analytics. Presented methodologies and results may not be appropriate for all applications in the property-casualty insurance industry. Users are strongly advised to review the underlying methodology and data sources when performing a text-mining extraction or predictive analytics.

Starting Considerations

- National Highway Traffic and Safety Administration:
  - From a nationally representative survey, 16% of weekend nighttime drivers tested positive for illicit drugs or medications
  - 1 in 8 high school seniors responding to a 2011 survey reported driving after smoking marijuana within two weeks prior to the survey
  - 1 in 3 deceased drivers with known drug-test results tested positive for drugs (illicit substances as well as OTC and prescription medications)

- The White House: Since 2010, the White House has declared December to be "National Impaired Driving Prevention Month"

- Office of National Drug Control Policy (Exec Office of the President):
  - "Working to Reduce Drugged Driving and Protect Public Health and Safety" April 2012
  - "Working to Get Drugged Drivers Off the Road," November 2010

Milliman

Reasons to be Interested in Meds/Rx, Drugs, and Narcotics

- Studies have linked drugs to increased motor-vehicle accident risk
- In recent years, increase incidence of individuals taking meds/Rx
- Unlike alcohol, difficult to establish an “under the influence” threshold
  - Alcohol: generally, BAC 0.08% (regardless of person or beer, wine, spirits)
  - Meds/Rx: depends on medication and individual
- Difficulty testing for “under the influence”:
  - Alcohol: breathalyzer
  - Meds/Rx: blood or urine (breathalyzer does not work)

Identifying and Measuring Driver Impairment

- Law enforcement officer observes inappropriate behavior

- Law enforcement will begin with a test for alcohol (breath, blood, urine)
  - If positive for alcohol, unlikely to test for drugs
    - Potential for under-reporting of DUID (driving under the influence of drugs)
  - If negative for alcohol, officer may seek evidence for a drug-impaired driving charge
    - Drug Evaluation and Classification (DEC) Program (46 states)
    - Types of tests: blood, urine, oral fluid, sweat, hair.
    - Technology requires lab test, which may take days, weeks, or months

Issues with Identifying DUID and State Laws

- Issues identifying DUID:
  - Which drugs impair driving ability?
  - What drug dosage levels impair driving ability?
  - How frequently do drivers use drugs that impair driving?
  - Which drugs are associated with higher accident rates?

- Different types of state laws:
  - "Incapable": drug renders a driver "incapable" of driving safely.
  - "Under the influence": drug impairs the driver's ability to operate safely or require a driver to be "under the influence or affected by an intoxicating drug.
  - "Per Se": a criminal offense to have a drug or metabolite in one's body/body fluids while operating a motor vehicle (often referred to as "zero tolerance" laws).

  - First two types of statutes:
    - The most prevalent in the United States.
    - The State must prove that "the drug" caused the impaired driving, which is a technically complicated and difficult task.

  - Per Se statutes: favored by many stakeholders (e.g., law enforcement, judges, and prosecutors)
    - Better facilitate the prosecution, conviction, and potential treatment of drugged-driver offenders.
    - As of 2009, Per Se statutes covered roughly 40% of all licensed drivers in the United States.

Milliman

Reasons to be Interested in Text Data

- Able to capture concepts in text data not captured in structured data
  - Many structured data-reporting forms do not capture use of meds, Rx, or narcotics
  - Drivers / occupants may be averse or unaware of reporting meds, Rx, or narcotics
- Claim stratification
  - Able to identify claims with “on medication,” “taking prescription”, etc.
- Univariate and bi-variate analyses
  - What is the incidence of medications in accidents?
  - What types of accidents do medications appear to be an associated (possibly, contributing) factor?
  - Is there a difference by age of driver?
- Multivariate analyses (“predictive analytics”)
  - Does the inclusion of information from text data improve the predictability for target outcomes?

Definitions

- NHTSA – National Highway Traffic Safety Administration
  - Federal agency established in 1970 to carry out safety programs.

- NMVCCS – National Motor Vehicle Crash Causation Survey
  - Research-designed survey by NHTSA collecting information on accidents between July 3, 2005 and December 31, 2007.
  - On-scene and post-accident data collection.

- Structured data
  - Data reported in numeric or categorical form.
  - Numeric data includes dollar amounts, age, number of vehicles in an accident.
  - Categorical data includes assignment of other types of information to a specific character or number (such as a “rear-end crash” assigned to “22” or “weather-snow” to “2”, in fields for accident type or weather condition).

- Text data
  - Data provided in text form, such as a claim adjustor note, accident description, deposition, or other reports. Books, magazine articles, and research reports or other examples of text data.

THREE PARTS TO THIS PRESENTATION

- Problem
  - Valuable information in text data is not being captured in structured data
  - At time of an accident, some information may be easily coded to structured data
  - After the accident, new information may not be lifted into structured data
  - (Extra attention on this point because the claim analytics objective will affect the extraction of information from the text data)

- Solution
  - Accessing text data can be costly
  - Efficient extraction of information from text data is imperative
  - Assembly process must be flexible to accommodate changing analytical needs
  -

- Analysis
  - Descriptive statistics
  - Predictive analytics: multivariate analyses

- Short-hand references
  - “meds”: medications (eg, over-the-counter medications)
  - “Rx”: prescriptions

THE "PROBLEM"

- Where can information in text data be useful?
  - Causal factors: distracted driving (esp. cell phone/texting)
  - Causal factors: use of meds/Rx, specific drugs, narcotics
  - Participant profiles: use of meds/Rx, specific drugs, narcotics
  - Recovery initiatives: assigning liability (subrogation)
  - Claim abuse: fraud detection

- Overriding objective:
  - Making claim adjustment process more efficient – lower losses and/or reducing LAE

Meds/Rx, Drug Names, Narcotics: Three Different Types of Text Data

- Meds/Rx: same information can be expressed in a variety of ways
  - "on many medications"
  - "taking pain medications"
  - "taking a prescription"
  - "taking his prescriptions"
  - NVMCCS file: over 1,000 unique four-word combinations included "medication" or "prescriptions

- Drug names: large number of infrequently-used names
  - NMVCCS databook: 500+ drug names
  - International Narcotics Control Board: "Yellow List," 50th edition, December 2011
  - PsychCentral.com: 100+ drug names
  - Over-the-counter v. prescription, varying side effects

- Narcotics: law-enforcement implications, changing legal thresholds, state differences
  - Specific names: cocaine, heroin
  - Lesser seriousness: marijuana
  - Heterogeneous references: methadone, opiate

Milliman

National Motor Vehicle Crash Causation Survey

- National Motor Vehicle Crash Causation Survey (NMVCCS)
  - Conducted by the National Highway Traffic Safety Administration (NHTSA)
  - Sample of accidents investigated between July 3, 2005 and December 31, 2007.
  - Primary focus of Survey: Determine the critical pre-accident events and reasons underlying the critical factors.
  - Looked into factors related to drivers, vehicles, roadways, and the environment.
  - Considerable attention to behavioral considerations and factors.

- Data collection process
  - On-site data collection by NMVCCS researchers.
  - Accidents occurring between 6am and midnight.
  - Accident must have resulted in a harmful event.
  - EMS must have been dispatched.
  - Police present when NMVCCS researcher arrived.
  - At least one of the first 3 vehicles involved must be present at the accident scene.
  - Completed police report.

National Motor Vehicle Crash Causation Survey

- Data files
  - 22 files
  - Accident Description, Pre-Crash Assessment (PCA), Occupant
  - Contents are static (not updated)

- Case weights
  - To make the sample representative of all similar types of accidents in the US.
  - Case weights not used in present analyses. Present analyses are from the prospective of an insurer's book of business, rather than a research or policy analysis.

National Motor Vehicle Crash Causation Survey

- Files of special interest to this presentation
  - Structured data
  - Date and time of accident
  - Type of accident (eg, rear end)
  - Police report indicated whether there were injuries
  - Vehicle equipment: presence of a cell phone
  - PCA: whether the driver was engaged in a conversion, weather conditions
  - Drivers: driver fatigue, presence of alcohol

  - Text data
  - Accident Description
    - > One record per accident
    - > 8,000 bytes
    - > Vehicles are identified in various references: V1, Vehicle 1, Vehicle #1, Vehicle One
    - > References not always consistent within the same accident description

Milliman

NMVCCS Sample – Descriptive Statistics

- Table presents:
  - Statistics for:
    - Time of day/week
    - Environment
    - Nature of the accident
    - Driver condition
  - Incidence rates among the accidents
  - Percent of accidents with injury

- Incidence among accidents:
  - Night: 22%
  - Multiple vehicles: 74%
  - Head on: 2%
  - Alcohol: 6%

- Percent with injury:
  - All accidents: 73%
  - Night: 69%
  - Head on: 86%
  - Alcohol: 82%

Condition	Incidence Among Accidents	Percent with Injury	Percent with Injury Compared to "All Accidents"
All accidents (N = 6,949)	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%	+

Milliman

NMVCCS Accident Descriptions

- One record for each accident. Maximum length = 7,800 bytes.

- Three examples in the following slides.
  - Examples are typical of the NMVCCS accident descriptions.
  - Examples are for "….medication," "….. prescription," "….cocaine"
  - Selected to demonstrate different ways each concept may be expressed.

- In claim adjuster notes, much greater variations in expressions (less consistency among adjusters for same insurer, differences in style across insurers)

Milliman

Summary Characteristics of Accident Descriptions

- 6,949 accidents
  - 438 : average number of words in accident descriptions
  - 330 / 514: first and third quartiles for words in accident descriptions
  - 2,436: average number of bytes in accident descriptions
- Similar numbers for cases with weights

	All Cases	With Case Weights
Number of accidents	6,949	5,470
Number of words in accident descriptions		
Average number of words	438	444
Median number of words	411	416
Q1 / Q3 number of words	330 / 514	336 / 520
Maximum number of words	1,294	1,294
Number of bytes in accident descriptions		
Average number of bytes	2,436	2,471
Median number of bytes	2,300	2,324
Q1 / Q3 number of bytes	1,843 / 2,869	1,874 / 2,911
Maximum number of bytes	7,800	7,800

Milliman

Strategies for Extracting Information from Text Data

- Most general: reference to a general term
  - Mention of “medication” or “prescription”
    - “was taking ….”
    - “had taken ….”
  - “Medication” or “prescription” can refer to broad set of OTC, Rx, or other meds
  - Present analysis: approximately 1,100 phrases

- Action associated with a term: action + noun
  - Action associated with a drug name
    - “had taken his [drug name]”
    - “was on [drug name]”
  - With subgrouping, able to control combinations of action+drug
  - Present analysis: 3,590 phrases (10 actions x 395 drug names)

- Most specific: target list of words
  - List of drugs (esp. narcotics) that are red flags
    - Cocaine, heroin, marijuana
  - Present analysis: 52 narcotics

Milliman

Accident Description #1 ("…taking several meds")

Accident #1: V1, a 2002 Dodge Stratus, was traveling westbound on a four-lane, two-way, dry, asphalt roadway with a level grade in daylight conditions. V1 was intending to go straight. V2, a 2004 Honda Accord, was traveling eastbound in the second lane of travel on the same roadway in similar conditions, also intending to go straight. The posted speed limit was 56 kmph (35 mph). The driver of V1 was experiencing low blood sugar and passed out at the wheel, relinquishing control of the car. V1 crossed the double yellow lines and the front of V1 contacted the front of V2. V2 came to final rest on the roadway facing west. V1 came to final rest off the south side of the roadway facing north.

The driver of V1 was a 43-year old diabetic male who reported that he had blacked out due to low blood sugar. Medical records indicated that immediately after the crash, his blood sugar was 32, a dangerously low level. The driver of V1 sustained serious injuries during the crash and was transported to a local trauma facility. The driver of V1 told doctors that he had skipped a meal earlier in the day but had still taken his insulin.

The Critical Pre-crash Event for the driver of V1 was when he traveled over the lane line on the left side of the travel lane. The Critical Reason for the Critical Pre-crash Event was a critical non-performance error due to the diabetic blackout. The driver of V1 was taking several medications for various health problems, including heart problems, high cholesterol, thyroid problems, and diabetes.

The driver of V2 was a 44-year old female who had reported that she had been traveling between 50-64 kmph (31-40 mph) prior to the crash. She had no health related problems and was rested and traveling back to work. She was wearing her prescribed lenses that corrected a myopic (near-sighted) condition. She sustained minor injuries during the crash and was transported to a local trauma facility.

The Critical Pre-crash Event for the driver of V2 was other motor vehicle encroachment, from opposite direction-over left lane line. The Critical Reason for the Critical Pre-crash Event was not coded to the driver of V2 and she was not thought to have contributed to the crash. (380 words, 2,224 bytes)

Accident Description #2 – ("…. prescription")

Accident #2: The crash occurred on a two lane undivided roadway with a posted speed limit of 64 KPH (40 MPH). There was a level curve (radius of curvature 703.125 meters) to the left with a black on yellow warning sign with a suggested speed of 40 KPH (25 MPH). The weather was cloudy, the roadway dry and it was daylight at the time of this weekday afternoon crash.

Vehicle #1, a 2004 Subaru Forester was traveling south on the roadway and negotiating the curve to the left. A non-contact truck approached from the opposite direction .The Driver of Vehicle #1 stated he thought he had had enough room but realized he didn't so he steered right. Vehicle #1 continued off the right side of the road down a small embankment, struck a 65 cm diameter tree with its front and came to rest facing in a southerly direction.

The Subaru Forester (Vehicle #1) was driven by an 82 year old male who was transported, treated and released at a local hospital for a head injury. He stated that he observed the truck approaching, moved to the right, but believed that he was over to far. After that he does not remember any events of the crash. Vehicle #1 was towed due to damage.

The Critical Precrash Event for Vehicle #1 was this vehicle traveling off the edge of the road on the right side. The Critical Reason for the Critical Event was the poor directional control of the driver. An associated factor coded to this driver was the use of prescription medications: (a) gout no disabling side effects, (b) diuretic possible side effects are lethargy, drowsiness, low blood pressure, and (c) general health medication with possible side effects of drowsiness, tiredness, and dizziness. He also takes non-prescription anti-inflammatory drug occasionally. He was wearing prescription glasses that corrected a hyperopic (far-sighted) condition. He was familiar with the roadway and his vehicle.
(471 words, 2,603 bytes)

Accident Description #3 ("…. marijuana")

Accident #3: This is a two vehicle head on type crash that occurred on a two-way, two lane, dry, straight, level, bituminous asphalt roadway. The roadway has one westbound and one eastbound lane separated by a painted centerline stripe. The crash occurred during early morning dawn hours; there no road defects or sight line restrictions noted at the scene, and there were no adverse weather conditions that contributed to this crash. The speed limit on the roadway is posted 105 kmph (65 mph).

Vehicle one (V1), a dark green (teal) 1996 Ford Taurus 4 door, was traveling eastbound when it drifted into the westbound lane and collided front to front with vehicle two (V2). The impact caused V1 to enter a counterclockwise rotation. V1 came to rest in the eastbound lane facing generally northwest. The non-restrained V1 driver, a 25-year-old female, was transported from the scene and admitted to a metropolitan trauma center for a fractured femur and other injuries. The second row left passenger, a 3-year-old female who was in a child safety seat but not restrained, was transported from the scene for injuries and released after treatment. The second row right passenger, a 3-year-old female who was in a child safety seat but not restrained, was transported from the scene for injuries and later released. Hospital personnel took a urine sample from the driver three hours after the collision that tested positive for amphetamines and marijuana. Police at the scene located substances inside the vehicle that was field-tested positive for Methamphetamine and Marijuana. The V1 driver told police that she does not remember what happened but does remember turning onto this road from the Interstate to pull over and sleep. V1 was equipped with first row frontal airbags that deployed as result of this crash. V1 was loaded with numerous clothing and household items in the trunk and inside of the vehicle. V1 was towed due to damage.

V2, a maroon and tan 1993 Dodge Ram 250 pickup with a white fiberglass camper shell, was traveling westbound in the westbound lane. The V2 driver told poice that he saw V1 coming from the opposite direction and drifting into his travel lane and slowed and steered to the right onto the westbound shoulder to avoid the crash, but the 2 vehicles collided front to front. The impact caused V2 to rotate slightly counterclockwise and depart the roadway to the right where it came to rest in the dirt area to the north of the roadway facing generally west/southwest. The V2 driver, a restrained 79-year old male, refused treatment at the scene by EMS but was later taken to an area hospital by friends and admitted for a ruptured spleen and multiple lacerations to his head and arms. According to the medical report the V2 driver has a history of a number of degenerative physical ailments. V2 was towed due to damage.

The critical pre-crash event for V1 was coded: this vehicle traveling, over the lane line on the left side of travel lane. The critical reason was coded to V1 as a driver related factor: sleeping, that is actually, asleep. The V1 driver told police that she did not remember what happened but she does remember exiting onto the roadway where the crash occurred to sleep. There was no evidence of braking by V1 prior to the crash.

The critical pre-crash event for V2 was coded: other vehicle encroachment; from opposite direction, over the left lane line. The critical reason was not coded to V2. (584 words, 3,474 bytes)

Milliman

NMVCCS Accident Descriptions

- Notable differences in the three examples.

- References to “vehicle”:
  - V1, V2 (#1, #3)
  - Vehicle #1, Vehicle #2
  - Other accident descriptions: insert “#” before the number (eg., V#1), spell numeric (eg., Vehicle One)
  - Reference not always consistent within the same accident description. (Significant problem with claim adjuster notes.)

- References to medications, Rx, and drugs with common “under the influence” implications:
  - was taking several medications (#1)
  - use of prescription medications (#2)
  - diuretic side effects (#2)
  - health medication with possible side effects (#2)
  - takes prescription anti-inflammatory drug (#2)
  - tested positive for amphetamines (#3)
  - mention of “red flags” (#3)
  - With claim adjuster notes, some meds/Rx may not be contributing factors to the accident.

NMVCCS Accident Descriptions compared to Claim Adjuster Notes

- NMVCCS accident descriptions are “cleaner” than the typical claim adjuster notes.

- Distinctions with Claim Adjuster notes:
  - Typically span more than one record.
  - Include considerable amount of ancillary information (eg, phone numbers, addresses).
  - Provide claim activity, often with dates (open, closed).
  - Provide insurer-liability information (eg., subrogation).

- Compared to the NMVCCS data, many of these points provide for a much wider scope of information.

- Insurer text data can also include text data beyond claim adjuster notes (eg, medical case manager notes, underwriting notes, depositions, statements).

Strategies for Extracting Information from Text Data

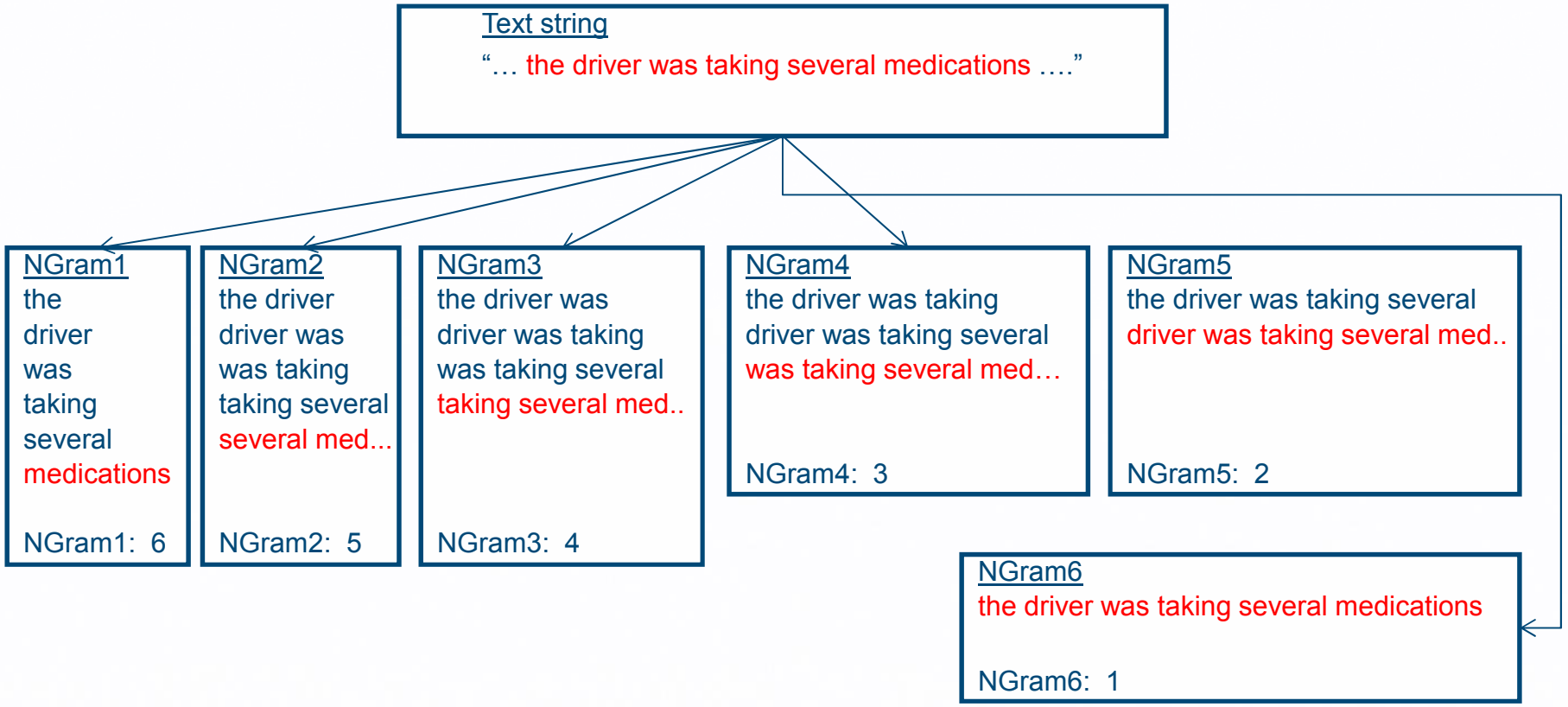
- Most general: reference to a general term
  - Mention of “medication” or “prescription”
    - “was taking ….”
    - “had taken ….”
  - “Medication” or “prescription” can refer to broad set of OTC, Rx, or other meds
  - Present analysis: approximately 1,100 phrases

- Action associated with a term: action + noun
  - Action associated with a drug name
    - “had taken his [drug name]”
    - “was on [drug name]”
  - With subgrouping, able to control combinations of action+drug
  - Present analysis: 3,590 phrases (10 actions x 395 drug names)

- Most specific: target list of words
  - List of drugs (esp. narcotics) that are red flags
    - Cocaine, heroin, marijuana
  - Present analysis: 52 narcotics

Milliman

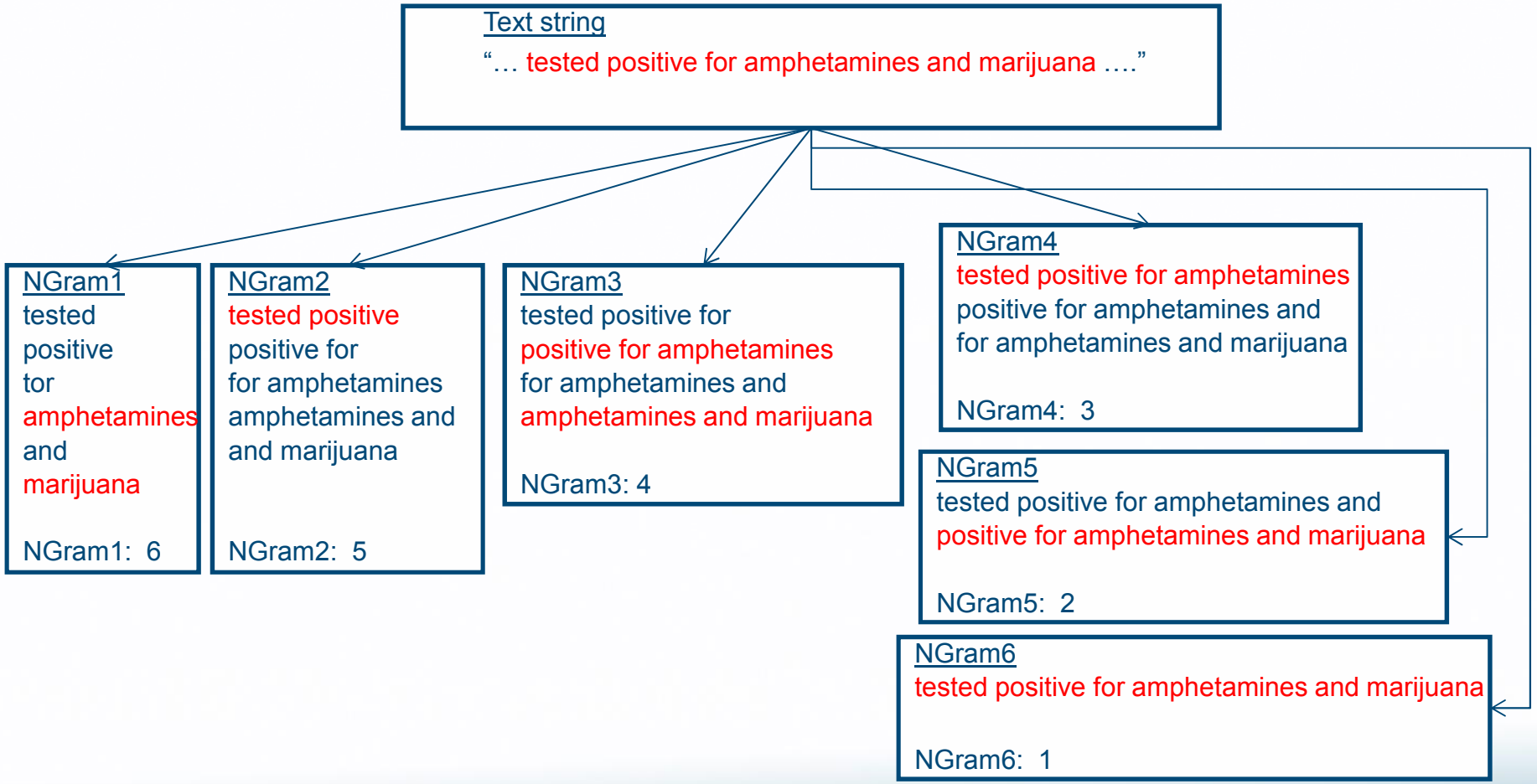
Breaking Text Data into Manageable Units – Creating "NGrams"



- 1 six-word phrase produced 21 NGrams.

Milliman

Breaking Text Data into Manageable Units – Creating "NGrams"



Milliman

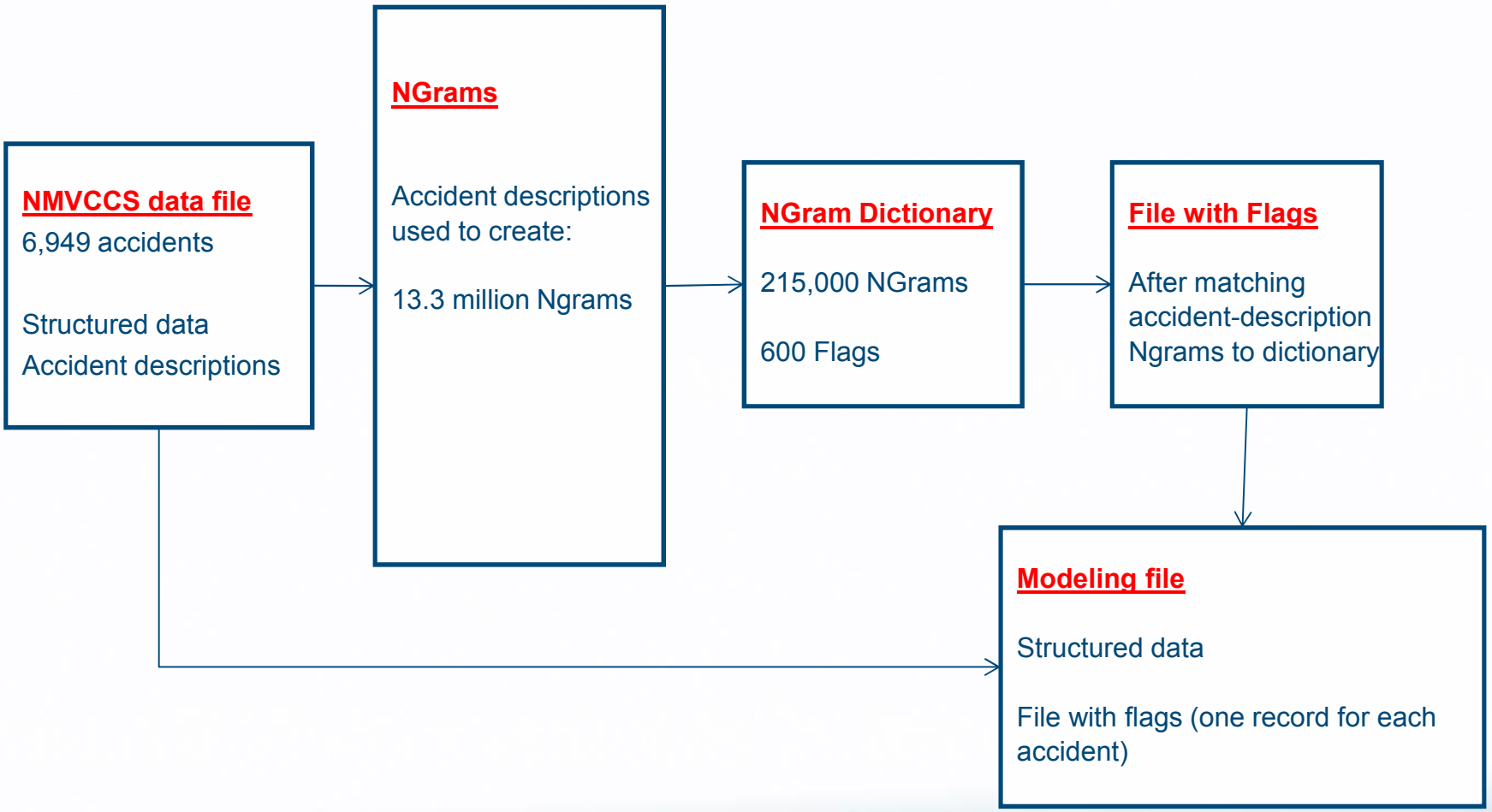
NGrams Created from NMVCCS Accident Descriptions

- Each accident description was parsed into NGram1-NGram6.
- Process removes certain NGram1-NGram3 not expected to be needed in any claim segmentation or analytics.
- For each accident description, unique NGrams are retained. (Repeats can produce misleading emphasis on a particular NGram. Same concept can be expressed with different words.)

	All Cases
Number of accidents	6,949
Size of NGram	
NGram1	607,260
NGram2	1,998,412
NGram3	2,578,495
NGram4	2,689,556
NGram5	2,725,082
NGram6	2,737,144
Total	13,335,949

Milliman

From Accident Descriptions to Modeling File



Milliman

Incidence of Medication, Rx, Drug, and Narcotics in Accidents and Injuries

- Reference to “medication” was found in 16% of accident descriptions.
- Among accidents with “medication” reference, injury for 82% (compared to 73% for all crashes.
- Injury incidence higher among accidents with reference to a med, Rx, drug name, or narcotic.

Condition	Incidence Among Accidents	Percent with Injury	Percent with Injury Compared to "All Accidents"
All accidents	100.0%	73%	
Medication	15.7%	82%	+
Prescription	6.4%	80%	+
Drug name	6.5%	80%	+
Narcotic	2.4%	89%	+

Milliman

Multivariate (Logit) Analyses

- Outcome measure
  - Injury may have occurred (police report)
  - Are accidents where one of the drivers has been taking meds, Rx, a drug, or a narcotic more likely to result in an injury?

Multivariate (Logit) Analyses

- **Explanatory variables**
  - Time if day/week
    - Night: accident occurred before 7am or after 6pm.
    - Weekend: accident occurred on a Saturday or Sunday
  - Environment
    - Weather: on or more adverse conditions (eg., snow, rain, ice)
    - Wet roads
  - Nature of the accident
    - Multiple vehicles
    - Rear end
    - Head on
    - Turned into path
  - Driver Conditions
    - Driver fatigue: at least one driver in the accident was reported to be fatigued
    - Alcohol: police report recorded presence of alcohol with the driver

Multivariate (Logit) Analyses

- **Explanatory (additional right-hand) variables**
  - Four 0/1 indicators:
    - Medications: mention of driver taking or on “medication”
    - Prescription: mention of driver taking or on “prescription”
    - Drugs: action + drug name (“taking [drug name]”)
    - Narcotics: single-word “red flag” (or per se) references

Logit Regressions: Injury May Have Occurred

- Outcome measure: Injury may have occurred (police report)
  - Are accidents where a driver was taking or on a med, Rx, drug, or narcotic more likely to result in an injury?

- Principal finding:
  - taking or on a med, Rx, drug, or narcotic increases the likelihood of an injury
  - coefficient for each of the four measures statistically significant at the 5% level.

Variable	Medication		Prescription		Drug Name		Narcotic	
Intercept	0.5220	*	0.5726	*	0.5811	*	0.5679	*
Night	-0.2527	*	-0.2672	*	-0.2657	*	-0.2789	*
Weekend	0.0584		0.0509		0.0490		0.0459	
Weather	0.0394		0.0615		0.0569		0.0599	
Wet road surface	-0.2179	*	-0.2341	*	-0.2336	*	-0.2322	*
Multiple vehicles	0.5403	*	0.5395	*	0.5359	*	0.5569	*
Rear end	-0.3009	*	-0.2978	*	-0.3059	*	-0.2942	*
Head on	0.6660	*	0.6675	*	0.6663	*	0.6352	*
Turned into path	0.2957	*	0.3011	*	0.2989	*	0.3156	*
Driver fatigue	0.2262	*	0.2643	*	0.2588	*	0.2452	*
Alcohol	0.7155	*	0.7192	*	0.7076	*	0.6655	*
Medications	0.5488	*	----		----		----	
Prescription	----		0.3771	*	----		----	
Drugs	----		----		0.3439	*	----	
Narcotics	----		----		----		1.1729	*

Milliman

Logit Regressions: Injury May Have Occurred

- Logit coefficients are transformed into odds ratios

- Increase in odds of an injury and statistically significant: presence of meds, Rx, a drug, or narcotic

Variable	Medication		Prescription		Drug Name		Narcotic	
Night	0.777	*	0.766	*	0.767	*	0.757	*
Weekend	1.060		1.052		1.050		1.047	
Weather	1.040		1.063		1.059		1.062	
Wet road surface	0.804	*	0.791	*	0.792	*	0.793	*
Multiple vehicles	1.716	*	1.715	*	1.709	*	1.745	*
Rear end	0.740	*	0.742	*	0.736	*	0.745	*
Head on	1.946	*	1.949	*	1.947	*	1.887	*
Turned into path	1.344	*	1.351	*	1.348	*	1.371	*
Driver fatigue	1.254	*	1.303	*	1.295	*	1.278	*
Alcohol	2.045	*	2.053	*	2.029	*	1.945	*
Medications	1.731	*	----		----	*	----	
Prescription	----		1.458	*	----		----	
Drugs	----		----		1.410	*	----	
Narcotics	----		----		----		3.231	*

Milliman

Logit Regressions: Injury May Have Occurred

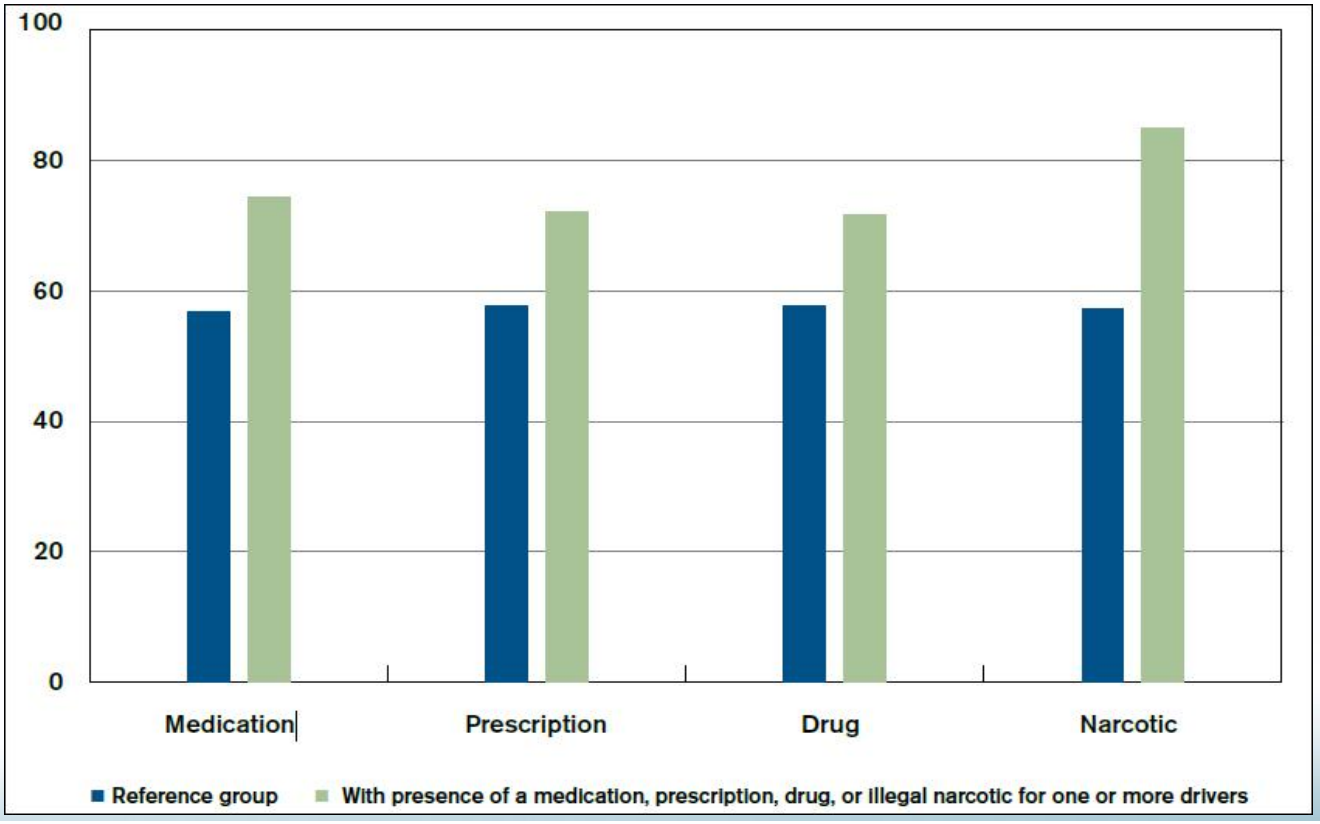
- Odds ratios transformed to probabilities.
- "Reference group": 0 values for control variables in the logistic regression
- Probabilities for presence of meds, Rx, a drug, or narcotic are substantially higher than for the reference group

Probability of an Injury	Medication	Prescription	Drug Name	Narcotic
Reference group	0.567	0.576	0.578	0.572
Daytime				
Weekday				
Good weather				
Dry road surface				
Single vehicle				
Not rear end				
Not head on				
Not turning into path				
Driver not fatigued				
Alcohol not present				
Medications	0.745			
Prescription		0.721		
Drugs			0.716	
Narcotics				0.851

Milliman

Logit Regressions: Injury May Have Occurred

- Odds ratios transformed to probabilities.
- “Reference group”: 0 values for control variables in the logistic regression
- Probabilities for presence of meds, Rx, a drug, or narcotic are higher than for the reference group
- Variables from text data increased the probability that an injury occurred in the accident



Milliman

Logit Regressions -- Summary

- Outcome measure
  - Injury may have occurred
- Control variables (from structured data)
  - Time of day/week
  - Environmental
  - Nature of accident
  - Driver condition
- Preliminary Findings
  - Indication of taking meds, Rx, a drug, or a narcotic increases the chances that an injury occurred with the accident

Summary: Three Parts to the Presentation

- Problem
  - Valuable information in text data is not being captured in structured data
  - At accident, some information may not be easily coded to structured data
  - After the accident, new information may not be lifted into structured data

- Solution
  - Accessing text data can be costly
  - Efficient extraction of information from text data is imperative
  - Assembly process flexible to accommodate changing analytical needs

- Analysis
  - Descriptive statistics
  - Predictive analytics: multivariate analyses

Summary

- Reasons to be Interested in Text Data
- Identifying DUID and State Laws
- National Motor Vehicle Crash Causation Survey
- Accident Descriptions: 3 examples where cell phone use mentioned
- Flags for Med, Rx, Drug Name, and Narcotics Created from Text Data
- Med, Rx, Drug Name, Narcotics: descriptive statistics
- Multivariate (Logit) Analyses