

연속주행시간과 졸음사고간 통계적 상관관계 분석

김덕녕¹ · 김수진² · 최재현² · 조종석^{1*}

¹한국도로공사 도로교통연구원, ²명지대학교 교통공학과

The Statistical Correlation Between Continuous Driving Time and Drowsy Accidents

KIM, Ducknyung¹ · KIM, Sujin² · CHOI, Jaeheon² · CHO, Jongseok^{1*}

¹Korea Expressway Corporation Research Institute, Gyeonggi 18489, Korea

²Department of Transportation Engineering, Myungji University, Gyeonggi 17058, Korea

*Corresponding author: ch4815@ex.co.kr

Abstract

During recent 5 years, it was recorded that 20% of total accident frequency and 30% of total number of death have been occurred due to drowsy driving. Drowsy driving accident is result from the loss of driving ability due to driver's accumulated fatigue. Continuous driving time can be measured as a surrogate variable to quantify the level of fatigue. The main purpose of this research is to investigate statistical correlation between the proportion of continuous driving vehicle (more than 2 hours) and the number of drowsy accidents. To carry this out, continuous driving time was measured using GPS route-guidance trajectory data. Also, accident frequency, traffic volume and segment length were collected to estimate safety performance function (SPF) for Jungbunearuk expressway in Korea. Through various types of estimated SPFs, statistical correlation was analyzed based on estimated statistical indices. This research can provide theoretical background for enforcement to regulate commercial vehicle driver's continuous driving time. In addition, throughout the trajectory data expansion, it is expected that strategy for anti-drowsy driving facilities installation can be established based on the suggested methodology.

keywords: continuous driving time, drowsy accident, GPS driving trajectory, over-dispersion, safety performance function

초록

최근 5년간 집계된 고속도로 교통사고자료에 따르면 전체 사고건수의 20%, 사망자수의 30%가 졸음운전으로 발생되고 있다. 졸음사고는 운전자의 피로도가 누적되었을 때 주행능력을 상실한 운전자의 인적 요인으로부터 발생되며, 정성적으로 체감되는 피로도는 운전자의 연속주행시간을 통해 간접적으로 분석될 수 있다. 본 연구는 2시간 이상 연속 주행한 차량의 비율과 해당 링크에서 집계된 졸음사고간의 통계적 상관관계를 규명하는데 주된 목적을 둔다. 이를 위해 네비게이션 주행 궤적 자료를 활용하여 링크별 연속주행차량의 비율을 산출하였으며, 중부내륙 고속도로의 링크별 졸음사고 건수, 교통량, 구간길이를 변수로 하는 안전성능함수를 추정하였다. 본 연구에서는 다양한 형태의 안전성능함수가 추정되었으며, 도출된 통계치의 비교를 통해 연속주행시간과 졸음사고 발생간의 통계적 상관성이 다각적으로 분석되었다. 본 연구 결과는 최근 화물차 및 여객버스 운전자의 연속주행을 단속하는 제도에 학술적 근거를 제공할

J. Korean Soc. Transp.
Vol.35, No.5, pp.423-433, October 2017
<https://doi.org/10.7470/jkst.2017.35.5.423>

pISSN : 1229-1366
eISSN : 2234-4217

Received: 26 June 2017

Revised: 22 August 2017

Accepted: 27 September 2017

Copyright ©
Korean Society of Transportation

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

수 있을 것으로 판단하며, 궤적자료의 양적 질적 확대를 통해 향후 졸음방지 시설물의 설치 위치를 결정하는 전략 수립에 활용 가능할 것으로 기대한다.

주요어: 연속주행시간, 졸음사고, GPS 주행궤적, 과분산, 안전성능함수

서론

졸음운전사고는 운전자의 피로도가 누적되었을 때 주행능력을 상실한 운전자의 인적요인으로부터 발생하게 된다. Highway safety manual (AASHTO, 2010)에 따르면 인적요인으로부터 발생하는 교통사고를 최대 93%로 집계하고 있으며, 국내에서 발생하는 사고자료 역시 유사한 맥락을 보인다. 최근 5년간 발생된 고속도로 사고자료(C급 이상)에 따르면 사고건수의 20%, 사망자수의 30%가 졸음운전으로 발생(사고원인 중 1위)된 것으로 보고되고 있다. 이러한 배경하에 ‘졸음사고 치사율 과속사고의 2.4배’ 등의 문구를 VMS에 표출하여 운전자들의 경각심을 고취시키고 휴식을 유도하는 정책이 시행되고 있다.

졸음운전이란 졸음이라고 하는 운전자의 생체적인 조건(Biometric condition)과 운전이라고 하는 행동조건(Acting condition)이 동시에 주어졌을 때 발생된다(Korea Expressway Corporation Research Institute, 2016). 이 중 생체적인 특성은 운전시간이 지속될수록 졸음강도가 증가되는 것으로 알려져 있으며 졸음강도가 증가되면 주의력이 약화, 감시능력 손상, 차량 통제력이 저하되어 최종적으로 교통사고를 회피할 수 있는 능력이 저하된다(Brown, 1994). 이 때, 운전시간의 지속은 통행 기종점간의 거리, 운전자의 운전습관 그리고 외부환경적 요인(업무 특성 등)에 종속적인 영향을 받는다. 특히 트럭 빛 버스 등 상업차량 운전자의 경우 근로조건 등의 문제로 졸음과 같은 생체적인 조건과 무관하게 운전을 지속해야 하는 상황이 자주 발생하게 된다. 화물자동차 운전자 운전실태조사(Korea Transportation Safety Authority, 2013)에 따르면 화물차 운전자의 일평균 운전시간은 7.9시간이며, 1회 운전 시 평균 2.52시간의 연속운전 시간을 갖는 것으로 조사되었다. 전체 응답자중 80%가 졸음운전을 한적 있다고 응답하였으며, 경험자 중 32%는 주1회 정도 졸음운전을 경험하고 있는 것으로 집계되고 있다.

요컨대, 교통사고 주원인으로 집계되는 졸음운전사고는 운전자의 피로도와 직접적인 연관성을 갖으며, 화물차운전자와 승용차운전자 모두 시간적 부담(정시도착, 시간적 촉박함이 운전 부담이 되는 정도)이 피로도의 가장 큰 요인으로 작용한다(Lee et al., 2008). 이러한 시간적 부담은 운전자의 휴식 대신 연속주행을 강행하게 만드는 주된 요인이 된다. 그러나 ‘연속주행시간-피로도-졸음운전-교통사고’로 연결되는 흐름은 기존문헌에 정리된 바 있으나, 실측 데이터 기반에 통계적 상관성을 규명하는 연구는 미진한 상태이다. Beak et al.(2005)의 연구에서는 부주의 운전이 시작되는 임계치(2시간)를 이용하여 잠재사고비율(potential accident ratio)을 사고예측모형의 설명변수로 제안하였다. 그러나 고속도로 요금소간 OD데이터를 이용하였기 때문에 운전자의 실제 휴식 행태를 고려하는데에는 한계점이 존재한다. Lee et al.(2016)의 연구에서는 네비게이션 자료를 사용하여 운전피로도 지표와 전체사고건수간의 상관관계를 제시하였으나 모형의 선형성(R^2 , Pearson statistic)을 규명하는데 주된 초점을 맞추었다.

본 연구에서는 네비게이션 데이터를 이용하여 i) 휴식여부를 고려한 연속주행시간을 link 단위로 지수화, ii) 해당 지수와 졸음사고로 규명되는 사고발생간의 상관관계를 규명하는 데 주된 목적을 둔다. 본 연구는 연속주행시간이 증대됨에 따른 운전자 개인의 행태적 변화에 초점을 맞추기 보다 링크 기반의 집계 데이터 측면에서 검토를 수행하였다. 즉, 개별차량의 연속주행시간이 링크의 속성을 대표하는 변수(2시간 이상 연속주행 차량 비율)가 되었을 때 해당 링크에서 연간 단위로 발생하는 졸음사고 건수의 중심 경향적 특성을 얼마나 설명할 수 있는지에 초점을 맞추었다. 이를 위해, 개별차량 데이터를 path 개념으로 구축하고 연속주행 지수를 GIS상에 표출하는 내용이 제시되며, 연속주행시간과 졸음사고발생간의 상관관계를 규명하기 위한 모수적 통계기법(parametric statistics)이 제시된다. 또한 연속주행지수에 대한 다양한 통계적 검증을 통해 기존연구와의 차별성을 제시하였으며, 졸음사고를 방지하기 위한 연속주행시간 변수의 활용 방안을 서술하였다. 마지막으로 본연구의 의의 및 향후 연구계획이 결론에 제시된다.

주행계적 데이터를 이용한 연속주행시간 산출

본 연구에서는 연속주행시간을 정량적 변수로 산출하기 위해 차량 내 부착된 네비게이션 주행계적데이터를 활용하였다. 이를 위해, 국가교통DB센터의 교통망 링크 자료(KTDB level-6)가 탑재된 민간 네비게이션 자료를 적용하였으며, 2014년 4개월(3, 4, 10, 11월) 평일(화, 수, 목) 51일 동안 고속도로를 이용한 차량만을 대상으로 한 원시 계적자료가 구축되었다. 본 자료는 하절기 휴가철, 동절기 기상악화, 명절이동의 영향을 배제하기 위해 봄·가을로 계절을 한정하였으며, 연속주행을 강행하는 화물차의 특성을 반영하기 위해 주중통행을 대상으로 자료를 구축하였다.

구축된 해당 원시자료는 주행 경로에 따른 다수의 link 정보를 병렬적인 구조로 제공하고 있어, 다수의 link를 차량의 고유 ID로 매칭하여 이동경로(path)로 변환하는 전처리 과정이 요구된다. 또한 GPS의 교란 등을 원인으로 통행경로가 비합리적인 순차성을 갖는 오류, 차량의 한 위치에서 50분간 정차한 뒤 새로운 통행을 시행했음에도 불구하고 연속적인 하나의 통행으로 기록되는 오류가 포함되어 이를 수정하거나 입력자료에서 제외하는 필터링 과정이 수반된다.

일반도로의 경우 휴식 및 중간 목적지 방문 등으로 연속된 링크간 통행시간의 차이가 발생할 수 있는 반면, 본연구에서 활용한 고속도로 통행 자료의 경우 진출입이 통제되어 휴식을 제외한 통행시간 단절은 발생할 수 없다(단, 정체로 인한 통행시간 차이는 발생 가능). 따라서 고속도로를 주행한 계적 데이터의 경우 연속주행시간의 산출을 위해 휴게소 및 졸음쉼터의 방문 여부를 판단할 수 있는 처리과정이 요구된다. 이를 위해 전국 398개의 정규 휴게소와 170개의 졸음쉼터의 위치정보를 GIS상에 표출하였으며 주행계적정보와의 매칭을 통해 휴게소 방문여부를 판단, 연속주행시간의 연장(expansion) 또는 갱신(reset)이 가능토록 하였다.

i)고속도로 통과 데이터 추출, ii)데이터 매칭, iii)오차보정, iv)휴식여부판단 순으로 전 처리된 자료는 총 14,614,137 통행, 148,322개의 기종점쌍으로 구성된다. 본 연구에서는 Beak et al.(2005) 및 Lee et al.(2016)의 연구와 마찬가지로 피로주행의 여부를 판단할 수 있는 연속주행 시간의 임계치를 2시간으로 설정하였으며, 이를 초과한 차량의 비율(%)을 링크별로 도식화한 결과가 Figure 1에 제시된다. Table 1은 연속주행 2시간 이상의 차량 비율을 내림차순으로 정렬하였을 때 상위 10개 구간(IC-IC 기준)을 나타내며, 1-9위까지가 모두 중부내륙선에 위치하고 있어, 본연구의 공간적 범위로 선정하였다.

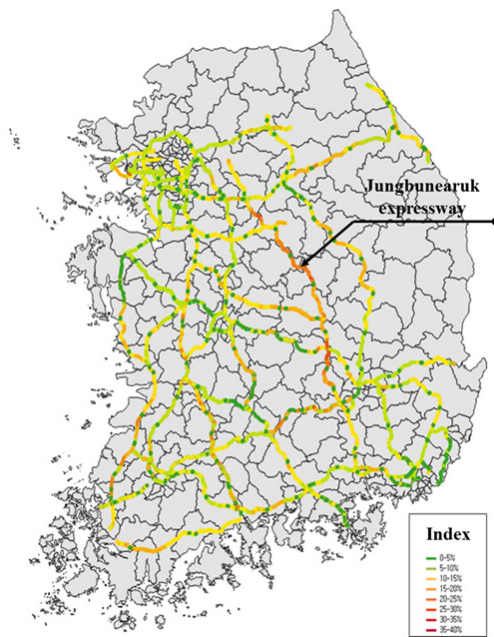


Figure 1. GIS-map for % of continuous driving car (more than 2 hours)

Table 1. Top 10 ranking for % of continuous driving car

Line	Milepost	% of continuous driving car (more than 2 hours)
Jungbunearuk	102.7-113.9	29.3%
Jungbunearuk	115.5-82.1	26.9%
Jungbunearuk	133.1-123.5	26.7%
Jungbunearuk	187.4-166.6	26.2%
Jungbunearuk	118.4-114.0	25.8%
Jungbunearuk	206.3-209.2	25.7%
Jungbunearuk	76.5-74.1	25.6%
Jungbunearuk	245.8-251.1	25.6%
Jungbunearuk	222.1-225.0	25.4%
Jungbunearuk	143.1-143.6	25.3%
88 Olynpic	139.6-138.4	25.3%

안전성능함수 추정을 통한 통계적 상관관계 분석

1. 안전성능함수의 구축

안전성능함수(safety performance function, SPF)는 다양한 설명변수의 공변량을 통해 사고건수, 사고율 및 사고 심각도를 예측하는 함수로 정의된다. 안전성능함수는 독립변수와 종속변수간의 통계적 상관성을 규명하기 위한 추론통계 기법의 일환으로 활용되며, Equation 1과 같이 사고수정계수(crash modification factor, CMF), 보정계수(calibration factor, C)와 결합하여 해당 구간의 사고건수를 예측($N_{predicted}$)하거나 안전개선사업의 편익을 추정하는 목적으로 사용되기도 한다.

$$N_{predicted} = N_{SPF} \times (CMF_1 \times CMF_2 \times \dots \times CMF_y) \times C \quad (1)$$

다양한 독립변수를 사용한 모형의 경우 사고발생에 영향을 미치는 주요 변수를 규명하기 위해 도로의 기하구조, 운영조건, 제한속도, 차종별비율, 교통안전시설 설치 여부 등이 모형 내 설명변수로 설정된다. Kang et al.(2002)의 연구에서는 고속도로 곡선구간의 사고건수 예측모형 개발을 위해 곡선반경, 편경사, 포장상태 등을 입력변수로 활용하였으며, Yoon(2007)는 고속도로 트럼펫형 IC 연결로의 사고예측모형을 위해 곡선반경, 본선 종단경사, 곡률차이 등을 모형에 반영하였다. Mitra and Washington(2007)에서는 교차로 사고건수 예측모형을 위해, 중앙분리대, 길어깨 폭, 좌/우회전 전용차로 설치여부, 조명시설 설치여부 등을 검토하였으며, Brimley et al.(2012)는 지방부 도로의 사고예측모형 추정을 위해 추월차로 설치 여부, 노면요철포장 설치여부, 중차량 비율 등을 고려하였다. 이와 같이 공변량의 통제를 통해 사고발생 요인을 검증하는 선행 연구는 ‘복합형 안전성능함수(Inclusive SPF)’의 추정과 회귀모수의 통계적 검증에 초점을 맞추고 있다. 해당 모형은 단일모형내에서 예측 정확성을 극대화할 수 있는 장점은 있지만, 향후 시행 가능한 개선대책의 독립적인 효과를 산정하는데 한계점이 존재한다.

교통량, 구간길이 등 사고노출변수(crash exposure variable)로만 구성된 안전성능함수(simple SPF)의 경우 기하구조 및 운영조건 등과 관련된 세부 사항은 별도의 보정계수의 형태로 제시된다. 여기서 교통량 및 구간길이 변수는 이상적 조건하에 사고노출변수에 따른 사고건수의 평균치를 대변하게 된다. Kim et al.(2013)은 클러스터링 기법을 활용하여 3가지 등급으로 세분화된 고속도로 안전성능함수를 개발하였다. Korea Expressway Corporation Research Institute(2014)은 국내 고속도로 본선, 요금소, 연결로 접속부, 터널 및 교량 각각에 대한 안전성능함수를 추정하였다. 해당 방법론은 각 사고수정계수의 독립성을 보장하고 있어 개선사업의 단독적인 효과를 산정하기 용이하고, 지역 간 전이성을 확보할 수 있어 AASHTO(2010)에서 권고하는 방법이다.

교통사고는 발생(1) 또는 미발생(0)으로 구분되는 Bernoulli trial의 결과로 정의할 수 있으며, 사고 발생/미발생 확률이 고정적이라는 가정 하에 이항분포의 형태로 수식화 할 수 있다. 여기서, 사고발생확률(p)가 매우 작고, 통과교통량 즉, 시행횟수(N)이 매우 클 때 이항분포는 포아송 분포로 근사화 될 수 있다(Olkin et al., 1980). 안전성능함수의 종속변수인 사고건수는 비음정수(non-negative integer)이기 때문에 최소제곱법으로 추정되는 일반회귀모형(ordinary least squared regression) 대신 최대우도법(maximum likelihood estimator)로 추정하는 가산적 회귀모형(count data regression)을 적용한다(washington et al., 2003). 선행연구 결과의 경험적 판단에 비추었을 때, 교통사고 데이터는 그 분산의 크기가 평균을 크기를 초과하는 과분산성(over-dispersion)을 내포하고 있다. 따라서 평균과 분산이 동일하다는 포아송 분포의 제약조건을 완화하기 위해 대부분의 안전성능함수는 Equation 2와 같이 음이항 회귀모형(negative binomial/Poisson-gamma regression)을 통해 추정되는 것이 일반적이다.

$$\lambda_i = \exp(\beta_0 + \sum_k \beta_k X_i) \times \exp(\varepsilon_i) \quad (2)$$

여기서, λ_i 는 구간 i 의 평균적인 사고건수를 의미하며, β_k 는 설명변수 X_i 에 대한 회귀계수를 나타낸다. 또한 과분산성을 반영하기 위한 오차항 $\exp(\varepsilon_i)$ 은 평균 1, 분산 α (과분산 계수)를 갖는 감마분포를 따른다. 음이항 회귀모형 추정 시 도출된 과분산 계수(over-dispersion parameter, α)의 통계적 검정을 통해 사고 건수의 과분산성을 확인할 수 있다. 과분산 계수는 i) 사고건수의 과분산성 여부 및 정도를 판단, ii) McFadden's Pseudo R^2 와 함께 모형의 통계적 적합성(goodness-of-fit) 제시, iii) Empirical Bayes 적용을 위한 가중치(ω) 제공의 역할을 수행한다. 그러나 음이항 회귀모형은 하나의 값으로 추정된 α 값을 통해 과분산성의 경향이 전 구간에 걸쳐 동일하다는 가정을 내포한다. 그러나 Kim(2015)은 고속도로 사고자료의 과분산성 경향(분산/평균)이 해당구간의 AADT 값과 양의 상관관계가 있음을 제시하고 있으며, 과분산 계수의 다양성(varying dispersion)을 반영하기 위한 방법으로 일반화된 음이항 회귀모형(Generalized/heterogeneous/modified negative binomial regression)을 제시하고 있다. 일반화된 음이항 회귀모형은 Equation 3과 같이 과분산 계수(α 또는 $1/\phi$)를 입력변수(Z_i)와 회귀계수(γ_k)의 조합으로 추정한다.

$$\alpha = \exp(\gamma_0 + \sum_k \gamma_k Z_i) \text{ or } \phi = 1/\exp(\gamma_0 + \sum_k \gamma_k Z_i) \tag{3}$$

본 연구에서는 누적주행시간이 줄음관련 교통사고에 미치는 통계적 상관성을 규명하기 위한 방법으로 안전성능함수를 사용한다. KTDB 교통망 링크자료의 구분단위별로 집계된 줄음 및 주시태만 사고건수를 종속변수로 설정하고, 해당 구간의 연평균 일교통량(대/일), Segment 링크 길이(m), 2시간 이상 연속주행 차량 비율(%)이 독립변수로 적용된다. 본 연구의 공간적 범위인 중부내륙 고속도로는 양방향 기준 649개(상행 328개, 하행 321개)의 구간으로 구분되며 각 구간의 시종점 이정에 따라 관련 자료가 매칭/구축된다. 여기서 2시간 이상 연속주행 차량 비율은 2014년 4개월간 집계된 데이터를 반영하였으며, 교통량 자료 역시 해당연도의 교통량 통계 자료를 준용하였다. 본 연구에서는 자료수집의 한계로 4개월간 집계된 연속주행차량 비율이 해당 링크의 속성을 대표하는 것으로 가정하였으며, 2014년 사고건수(건/년), 2012년부터 2014년까지 누계한 사고건수(건/3년)을 종속변수로 사용하여 각각의 모형의 설명력을 비교하였다. Table 2는 종속변수 및 독립변수의 기술통계량을 나타낸다.

Table 2. Descriptive statistic for input data

Model	Dependent variable ('12-'14)		Independent variable ('14)	
	Number of crashes due to drowsy and careless	AADT (vehicle/day)	Segment length (m)	% of continuous driving car (more than 2 hours)
Average	0.52	14,753	1,117	17.8
Std. Dev.	1.23	8,929	1,366	6.4
Max	12.00	44,307	8,200	32.5
Min	0.00	3,153	100	4.2

본 연구에서는 연속주행시간이 줄음사고에 미치는 통계적 상관성을 규명하기 위해 Figure 2에서 제시하는 비교 분석 시나리오를 적용하였다. 기본모형(base model)과 대안모형(alternative model)1은 모두 음이항 회귀모형의 형태로 추정되며, 2시간 이상 연속 주행한 차량의 비율의 반영 여부를 통해 해당 변수의 통계적 의미를 판단하는데 초점을 맞춘다. 대안모형2는 일반화된 음이항 회귀모형의 형태로 안전성능함수가 추정되어 대안 모형 1과 비교된다. 여기서 과분산 함수는 2시간 이상 연속 주행한 차량의 비율을 통해 추정되며, 과분산 계수를 결정하는 요소로서 정/부의 상관성이 검토된다. 대안모형 3의 경우, 독립변수의 구성은 대안모형 2와 동일하나, Hauer(2001) 및 Geedipally et al.(2009)의 연구에 따라 과분산계수를 구간길이의 함수로 정의하는 가장 일반적인 형태를 나타낸다. 더불어 종속변수의 집계범위(건/년 vs. 건/3년)에 따른 통계적 유의성 검토를 위해 비교모형(comparison models)이 각각 제시된다.

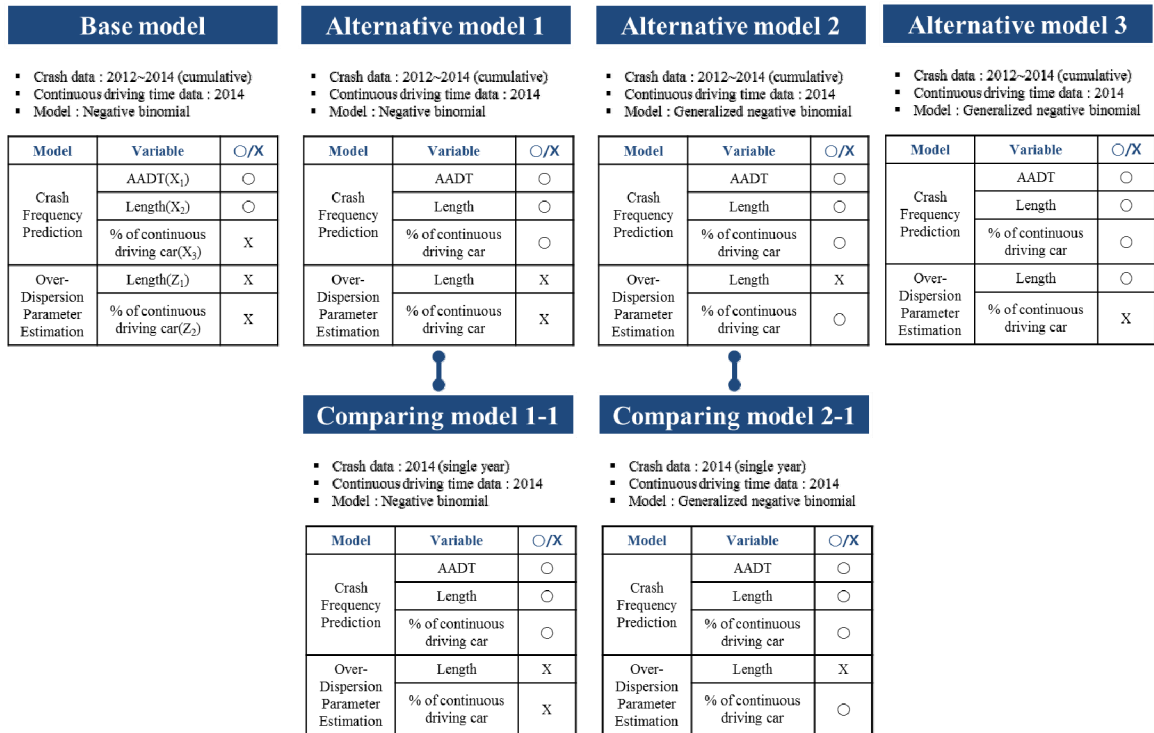


Figure 2. Model estimation scenario

기존연구에서 제시된 안전성능함수는 통계적 적합성을 향상시키는 목적으로 log 변환된 독립변수를 적용하거나 사고건수를 구간길이로 나누어 Km당 사고율로 전환하는 방법을 사용하였으나, 본 연구에서는 연속주행차량 비율의 통계적 의미를 규명하는데 초점을 맞추고 있어 기본적인 음이항 회귀 모형식을 적용하였다. Table 3는 기본 모형, Table 4, 5, 6은 대안모형 1, 2, 3의 추정 결과를 각각 나타내며 추정과정의 편이를 위해 통계패키지 STATA ver. 11.0을 활용하였다.

Table 3. Parameter estimation result for base model (negative binomial regression without variable X_3)

Variable		Coefficient	Std. err.	t-statistic	Significant level
N_{SPF}	β_0	-2.765	0.200	-13.81	99%
	β_1	7.6×10^{-5}	9.39×10^{-6}	8.10	99%
	β_2	3.34×10^{-4}	4.41×10^{-5}	7.57	99%
	α	1.140	0.189	6.03	99%

Table 4. Parameter estimation result for alternative model 1 (negative binomial regression with variable X_3)

Variable		Coefficient	Std. err.	t-statistic	Significant level
N_{SPF}	β_0	-3.705	0.326	-11.36	99%
	β_1	7.47×10^{-5}	9.33×10^{-6}	8.01	99%
	β_2	3.40×10^{-4}	4.26×10^{-5}	7.97	99%
	β_3	4.974	1.262	3.94	99%
	α	1.031	0.174	5.93	99%

Table 5. Parameter estimation result for alternative model 2 (generalized negative binomial regression with variable Z_2)

Variable		Coefficient	Std. err.	t-statistic	Significant level
N_{SPF}	β_0	-3.668	0.321	-11.44	99%
	β_1	6.98×10^{-5}	9.04×10^{-6}	7.72	99%
	β_2	3.98×10^{-4}	4.27×10^{-5}	9.33	99%
	β_3	4.801	1.289	3.73	99%
Function of α	γ_0	1.722	0.465	3.70	99%
	γ_2	-11.341	2.966	-3.82	99%

Table 6. Parameter estimation result for alternative model 3 (generalized negative binomial regression with variable Z_1)

Variable		Coefficient	Std. err.	t-statistic	Significant level
N_{SPF}	β_0	-3.778	0.305	-12.38	99%
	β_1	6.86×10^{-5}	8.41×10^{-6}	4.52	99%
	β_2	4.33×10^{-4}	4.20×10^{-5}	10.30	99%
	β_3	5.153	1.140	4.52	99%
Function of α	γ_0	-1.008	0.320	-3.15	99%
	γ_1	8.81×10^{-5}	2.60×10^{-5}	3.39	99%

모형 추정결과 2시간 이상 연속 주행한 차량 비율에 대한 계수 값이 양수로 도출되었으며, 연속 주행 차량이 많이 혼입된 링크일수록, 졸음 및 주시태만의 사고가 높게 발생하는 것을 확인할 수 있다. 요컨대, 졸음운전에 미치는 다양한 원인 중 연속주행 시간은 졸음과 주시태만으로부터 발생하는 사고의 확률을 가중시키는 역할을 한다. 이는 집계된 데이터의 중심경향적인 결과로서 향후 운전자 개별 데이터를 분석하기 위한 단초를 제공하는 역할을 한다.

2. 통계적 검증

추정된 계수 값은 모두 신뢰수준 99% 이상에서 통계적 유의성을 보이고 있으며, 교통량 및 구간길이의 계수 값이 모두 양수로 도출되었다. 이는 사고 노출도가 높아질수록 사고건수가 높게 기록됨을 의미하며 선행 연구 결과와 동일한 맥락을 취한다. 2시간 이상 연속 주행한 차량 비율에 대한 계수 값 역시 통계적 타당성이 확보되었으며, 사고건수의 과분산성을 의미하는 과분산 계수(α) 및 과분산 함수의 회귀계수(γ_k) 모두 신뢰수준 99% 이상에서 통계적 유의성을 보이고 있어 음이항 회귀모형의 적용이 타당한 것으로 분석된다.

대안모형 2의 경우 과분산 함수에 2시간 이상 연속주행 차량의 비율이 독립변수로 반영되었으며, 그 계수 값이 음수로 도출되었다. 이는 과분산계수의 추정 시 2시간 이상 연속주행 차량 비율(Z_2)을 고려함에 따라 더 작은 과분산 계수가 추정됨을 의미하며, 더 작은 과분산 계수는 모형의 통계적 신뢰성이 높다는 의미를 전달한다. Highway safety manual (AASHTO, 2010)에서는 Empirical-Bayes 기법의 적용을 Equation 4와 같이 제시하고 있으며 사고건수의 기대치($N_{expected}$)는 사고건수 예측치($N_{predicted}$)와 사고건수 관측치($N_{observed}$) 간의 가중평균을 통해 산출된다. 여기서, 가중치(ω)는 Equation 5와 같이 과분산계수(α)의 함수로써 정의되며, 더 작은 과분산계수를 갖을수록, 안전성능함수에 할당되는 가중치는 커지게 된다. 요컨대, 본 연구에서 중점 변수로 제안하는 연속주행시간은 안전성능함수의 통계적 신뢰성을 향상시킬 뿐 아니라, Empirical-Bayes 기법의 정확성을 제고하는 데 유용하다.

$$N_{expected} = \omega \times N_{predicted} + (1 - \omega) \times N_{observed} \tag{4}$$

$$\omega = 1 / [1 + \alpha \times \sum_{all\ year} N_{predicted}] \tag{5}$$

Table 4, 5, 6에서 도출된 세 가지 모형의 통계적 적합성을 비교하기 위한 방법으로 i)McFadden’s Pseudo R², ii)Baysian Information criteria(BIC), iii)log-likelihood ratio 통계치가 비교된다. McFadden’s Pseudo R²는 상수항만 존재하는 모형의 우도함수 값(LL(C))과, 설명변수와 회귀계수가 모두 반영된 전체 모형간의 우도함수 값(LL($\hat{\beta}$))의 비교를 통해 모형 전체의 통계적 적합성을 나타낸다. 최대우도법을 통해 추정되는 모형의 ρ^2 와 동일한 의미를 내포하며, Equation 6을 통해 도출된다. 모형 추정 시 설명변수가 증가할수록 우도함수 값이 증가되는 경향이 있으며 BIC 지수는 이러한 문제점을 해결하기 위해 Equation 7과 같이 추정계수가 증가되는 것에 대한 패널티를 고려할 수 있다. log-likelihood ratio는 기본모형의 우도함수 값(LL($\hat{\beta}^R$))과 개선모형의 우도함수 값(LL($\hat{\beta}^U$))에 대한 차이가 통계적으로 유의미한지를 판단하기 위해 Equation 8과 같이 카이제곱 분포의 임계값($\chi^2_{DF,\alpha}$)과 비교한다. 여기서, k는 회귀계수의 개수(상수항 포함), n은 관측치의 개수를 의미한다.

$$\text{McFadden's Pseudo } R^2 = 1 - \frac{LL(\hat{\beta})}{LL(C)} \tag{6}$$

$$\text{Baysian Information criteria(BIC)} = k \times \text{Ln}(n) - 2LL(\hat{\beta}) \tag{7}$$

$$\text{Log-likelihood ratio} = -2[LL(\hat{\beta}^R) - LL(\hat{\beta}^U)] \sim \chi^2_{DF,\alpha} \tag{8}$$

Table 7에서 제시된 바와 같이, 졸음사고건수를 예측하는데 있어 연속주행시간의 변수는 통계적 유의성을 향상시키는에 중요한 역할을 한다. 우도비 검정을 통해 모형별 우도함수 값의 차이는 통계적으로 유의한 것으로 도출되었으며, 그 결과 McFadden’s Pseudo R², BIC 지수에서 모두 연속주행시간을 고려한 모형이 우수한 것으로 도출되었다. 대안모형 3은 과분산성의 정도를 구간 길이로 추정하는 가장 통상적인 방법으로, 가장 높은 설명력을 보이고 있다. 요컨대, 대안모형 1, 2를 통해 그 중요성이 확인된 연속주행시간 변수를 대안모형 3의 독립변수로 취하고, 기존방법론과 동일하게 과분산계수를 구간거리의 함수로 설정하는 것이 사고건수기대치($N_{expected}$)의 정확성을 제고하는 것으로 분석되었다.

Table 7. Statistical indices for estimated 4 models

	Base model	Alternative model 1	Alternative model 2	Alternative model 3
Number of observations (n)	649	649	649	649
Number of parameter (k)	3	4	6	6
McFadden’s Pseudo R ²	0.165	0.178	0.192	0.209
Log-likelihood function (LL($\hat{\beta}$))	-505.340	-497.578	-488.055	-478.699
Baysian Information criteria	1030.107	1021.058	1014.963	996.251
Log-likelihood ratio	LLR = 15.52 > ($\chi^2_{1,0.01}=6.63$) LLR = 19.05 > ($\chi^2_{2,0.01}=9.21$)			

마지막으로, 종속변수의 집계범위에 따른 모형간 통계적 적합성이 Table 8에 제시된다. 종속변수가 서로 다른 경우 BIC 지수 및 log-likelihood ratio의 비교는 불가하여 McFadden’s Pseudo R²를 통해 비교한 결과 3년간 사고건수를 누적한 모형의 통계적 적합성이 우수한 것으로 분석되었다, 비교모형(1-1, 2-1)의 경우 연속주행시간 자료와 시간적 정합성을 갖는다는 장점이 있지만, 분석단위(평균연장 1.1km)내에서 발생한 연간 사고건수를 졸음 및 주시태만으로 한정함에 따라 집계된 사고의 변동 폭이 감소하여 모형의 설명력이 상대적으로 감소하는 것으로 분석되었다.

그러나 대안모형 및 비교모형 모두 신뢰수준 95% 이상에서 2시간 이상 연속주행차량 비율과 졸음 및 주시태만 사고건수간의 정의 상관관계를 보이고 있어 본 연구결과의 일관성은 유지되는 것으로 분석되었다.

Table 8. Parameter estimation result for comparison models (1-1, 2-1)

Variable	Alternative 1	Comparison 1-1	Variable	Alternative 2	Comparison 2-1
N_{SPF}	β_0	-3.705**	β_0	-3.668**	-4.323**
	β_1	7.47×10^{-5} **	β_1	6.98×10^{-5} **	4.870×10^{-5} **
	β_2	3.40×10^{-4} **	β_2	3.98×10^{-4} **	4.08×10^{-4} **
	β_3	4.974**	β_3	4.801**	4.299*
α	1.031**	1.827**	Function of α γ_0	1.722**	2.456**
			γ_1	-11.341**	-13.730**
McFadden's Pseudo R^2	0.178	0.140	McFadden's Pseudo R^2	0.192	0.159

※Statistical significant at the 95%(*) / 99%(**) level of confidence

연속주행시간 자료의 활용방안

전술된 바와 같이 연속주행 시간의 자료는 졸음사고건수를 예측하는 데 중요한 설명인자로 활용될 수 있으며, 2시간 이상 연속 주행한 차량의 비율이 높은 링크일수록 졸음사고의 발생 확률이 증가하는 것을 알 수 있다. 최근 국토교통부에서는 여객자동차 운수사업법 시행령·시행규칙 및 화물자동차 운수사업법 시행령·시행규칙의 개정을 통해 연속주행시간에 따른 휴식시간의 의무화(4시간 연속운전 후 30분 이상 휴식)를 추진하였다. 또한 교통안전법의 개정을 통해 최소휴게시간, 연속주행시간에 대한 단속이 가능하도록 제도적 기반을 마련하였다. 교통안전공단에서는 사업용 자동차에 의무적으로 부착된 디지털운행기록계(digital tachograph)의 로그파일을 추출하여 최소휴게시간 준수 여부를 판단할 수 있는 기계적 장치를 개발하고 실용화를 추진하고 있다.

사업용 자동차의 디지털운행기록계 장착은 2009년 교통안전법의 개정을 통해 의무화 되었으나 급출발, 급정거 및 급회전 등의 이벤트 자료를 근거로 운전자의 안전운전 여부를 판단/평가하는 과정에는 제도적 근거(안전운전 불이행의 기준)와, 관측의 정확성(기계적 오류 및 관측오차) 측면에서의 모호성이 존재한다. 그러나 연속주행시간 단속의 경우 관련 기준이 제도화되었고, 차량의 이동/정차로 이원화된 신호만을 통해 관련 수치를 계산할 수 있어 단속의 정확성을 담보할 수 있어 최종적으로 디지털운행기록계 장착의 실효성이 제고될 것으로 판단된다. 더불어 디지털운행기록계 자료는 사업용 자동차의 차종정보가 포함되어있어, 본 연구에서 활용한 자료와 더불어 세부적인 빅데이터 분석이 가능하다.

연속주행시간은 운전자의 피로도를 간접적으로 측정할 수 있는 정략적 지표가 되기 때문에 국내 뿐 아니라 국외에서도 화물차 운전자의 연속주행시간, 최소휴게시간, 하루 최대 근무시간 등에 대한 규제를 시행해 왔다. 이러한 규제는 운수업 종사자들의 노동기본권을 보장하기 위한 국제노동기구(international labour organization) 제 153호 협약(1979)에 정책적 근거를 두고 있으나(Korea Transportation Safety Authority, 2013), 연속주행시간이 갖는 학술적 근거에 대한 접근은 미미한 상태이다. 본 연구는 졸음사고와 연속주행시간간의 통계적 상관성을 규명함에 따라 2017년부터 국내에서 본격적으로 시행되는 연속주행시간 단속에 대한 근거를 제공할 것으로 기대한다.

이러한 연속주행시간 자료는 추후 교통망 네트워크 측면에서 정규휴게소 및 졸음쉼터의 입지를 결정하는 주요인자로 활용될 수 있으며 노면요철포장, 도출형 차선, LED 경광등, 졸음사고 방지 현수막 등 운전자에게 경각심을 고취시킬 수 있는 도로안전시설물의 배치 전략에 적용할 수 있다. 또한 연속주행시간이 특정 임계점을 초과할 경우 네비게이션에서 휴식을 권고하는 안내 메시지(현시점 연속주행시간, 가까운 휴게소 위치 및 다음 휴게소까지의 잔여거리 등)가 자동 송출되는 서비스가 가능할 것으로 판단되며, 실시간 네비게이션 서비스와 자동차 보험 서비스간의 연계를 통해, 안전운전을 이행하는 운전자에게 인센티브를 제공하는 방법 또한 가능할 것으로 판단된다.

결론

본 연구는 연속주행시간이 졸음사고에 미치는 영향을 검토하기 위해 모수적 통계기법을 적용하였다. 분석을 위해 네비게이션의 주행궤적 자료를 활용하였으며 링크 단위별로 2시간 이상 연속주행 차량의 비율을 집계하여 사고 건수를 예측하는 모형의 독립변수로 적용하였다. 그 결과 연속주행 차량이 많은 링크일수록 졸음 및 주시태만 사고가 높게 나타나는 것으로 분석 되었으며, 연속주행변수를 반영하지 않는 모형에 비해 통계적 적합성이 향상되었다. 또한, 과분산계수의 다양성을 반영한 모형을 적용할 경우 사고예측치의 추정오차를 가장 최소화 할 수 있는 것으로 분석되었다.

본 연구는 실제 차량의 주행궤적 자료를 이용하였기 때문에 연속주행시간에 대한 정확성을 담보할 수 있고, 추정된 안전성능함수에 대한 다각적 통계검증을 실시했다는 측면에서 기존연구와의 차별성이 존재한다. 본 연구결과는 화물 및 여객 운수사업자의 연속주행시간 단속에 학술적 근거를 뒷받침 해줄 것으로 판단되며 향후 도로안전시설물의 입지선정 및 실시간 경고 서비스의 제공에도 활용 가능할 것으로 기대 한다.

본 연구는 자료 구득상의 문제로 연속주행시간 변수의 시간적 범위가 제한되어, 해당 링크의 대표성 확보에 대한 한계점을 가지고 있다. 또한 주행궤적자료에 부여된 고유 ID로는 차종구분이 되지 않아, 일반 승용차, 화물차, 여객 버스에 대한 세부적인 추가 분석이 필요하다. 더불어 다양한 시나리오(2시간 이상 연속주행차량 비율은 낮으나 졸음 및 주시태만 사고가 빈번한 노선 등)를 대상으로 공간적 범위의 확장이 필요하다. 예를 들어 연속주행시간이 비교적 낮은 도시부 고속도로(외곽순환고속도로 등)를 대상으로 안전성능함수를 구축하고 추정된 회귀계수를 본 연구와 비교한다면 졸음사고 유발과 관련된 노선 및 공간적 특성 또한 규명이 가능할 것으로 판단된다. 마지막으로 연속주행시간에 대한 임계값을 다양화하고, 사고발생시간을 주/야간으로 구분하여 졸음사고 건수의 상대적 비교를 시행하면 본 연구결과의 신뢰성을 향상시킬 수 있을 것으로 기대한다.

REFERENCES

- AASHTO (2010), Highway Safety Manual, American Association of State Highway and Transportation Officials, Washington, D.C.
- Beak S. K., Chang H. H., Kang J. G. (2005), Estimation of Freeway Traffic Accident Rate Using Traffic Volume and Trip Length, J. Korean Soc. Transp., 23(2), Korean Society of Transportation, 95-105.
- Brimley B. K., Saito M., Schultz G. G. (2012), Calibration of “Highway Safety Manual” Safety Performance Function: Development of New Models for Rural Two-lane Two-way Highways, In Transportation Research Record: Journal of the Transportation Research Board, 2279, 82-89.
- Brown I. D. (1997), Driver fatigue, Human factor, 36-2, 298-314.
- Geedipally S., Lord D., Park B. J. (2009), Analyzing Different Parameterizations of the Varying Dispersion Parameter as a Function of Segment Length, In Transportation Research Record: Journal of the Transportation Research Board, 2103, 108-118.
- Hauer E. (2001), Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation, Accident analysis and prevention, 33, 799-808.
- Kang M. W., Son B. S., Dho C. U., Kang J. K. (2002), Development of Accident Prediction Models at Freeway Curve Section Based on Geometric Characteristics, Journal of Korean Society of Civil Engineering, 28(6-D), 1077-1088.
- Kim D. (2015), An Identification Method for Hazardous Roadway Locations Considering Varying Dispersion of Traffic Crashes, Seoul National University Ph.D. dissertation.

- Kim D., Kim D., Lee C. (2014), Safety Performance Functions Reflecting Categorical Impact of Exposure Variables for Freeways, In *Transportation Research Record, Journal of the Transportation Research Board*, 2398, 67-74.
- Korea Expressway Corporation Research Institute (2014), *Expressway Safety Manual*, Korea Expressway Corporation Research Institute (2014-03-326, 021).
- Korea Expressway Corporation Research Institute (2016), *A Study on Reduction of Drowsy Driving by Expanding Break Time*, Korea Expressway Corporation Research Institute (2016-49-534, 9607).
- Korea Transportation Safety Authority (2012), *Research on Driving Time Limitation Policy for Commercial Vehicle*.
- Korea Transportation Safety Authority (2013), *Report on Survey and Analysis for Commercial Vehicle Driving*.
- Lee S. C., Hwang Y. D., Oh J. S. (2008), The Effect of the Workload in the Truck Drivers' Fatigue, *Korean Journal of Industrial and Organizational Psychology*, 21(2), 367-381.
- Lee S., Chang H., Cheon S., Baek S., Lee Y. (2016), Development of Traffic Accident Index Considering Driving Behavior of a Data Based, *J. Korean Soc. Transp.*, 34(4), Korean Society of Transportation, 341-353.
- Mitra S., Washington S. (2007), On the Nature of Overdispersion in Motor Vehicle Crash Prediction Models, *Accident Analysis and Prevention*, 39(3), 459-468.
- Olkin I., Gleser L. J., Derman C. (1980), *Probability Models and Applications*, MacMillan Publishing Co., Inc., New York, N.Y.
- STATA Corp. (2009), *STATA version 11.0*.
- Washington S. P., Karlaftis M. G., Mannering F. L. (2003), *Statistical and Econometric Methods for Transportation Data Analysis*, Chapman & Hall/CRC, Boca Raton, Florida.
- Yoon B. J. (2007), Development of Traffic Accident Forecasting Models for Trumpet IC Ramp Types of Freeway, *Journal of Korean Society of civil engineering*, 27(1-D), 81-87.