

산업재해 예방정책에 대한 평가모형*

김영선¹ · 조진남² · 백재욱^{3†}

¹사회보장정보원, ²동덕여자대학교, ³한국방송통신대학교

Assessment Model for Industrial Accidents Prevention Policy*

Youngsun Kim¹ · Jinnam Jo² · Jaiwook Baik^{3†}

¹Social Security Information Service

²Department of Statistics & Information Science, Dongduk Women's University

³Department of Information Statistics, Korea National Open University

Purpose: The purpose of this study is to introduce the assessment model for industrial accidents prevention policy.

Methods: 10 years of industrial accidents data are explored through EDA approach. Case control study is tried in order to assess the effectiveness of the measures taken by Korea Safety and Health Agency, Civilian, and 'Ministry of Employment and Labor'. Propensity score matching is used to match the characteristics of the two groups compared, and then case control study is again conducted. Next, logistic and Poisson regressions are used to assess the risk factors.

Results: According to case control study involvement of 'Korea Safety and Health Agency' and 'Ministry of Employment and Labor' were not effective, but Civilian was. Propensity score matching leads to the same conclusion. Poisson regression reveals the impact of the risk factors on the industrial accidents. Industrial accidents occur more often as the number of employees grows. Mining, farming, fishing, 'transportation-storage and telecommunication' and forestry have a higher level of industrial accidents but service industry has a lower level. It is odd that more involvement of Korea Safety and Health Agency, Civilian, and Ministry of Employment and Labor means more industrial accidents.

Conclusion: 'Korea Safety and Health Agency', Civilian, and 'Ministry of Employment and Labor' seem to visit those industries with more industrial accidents.

Keywords: Industrial Accident, Industrial Accident Prevention Policy, Case Control Study, Logistic Regression, Poisson Regression

1. 서론

기업 간 경쟁이 치열해지면서 산업재해가 많이 일

어나고, 산업재해의 유형은 점점 더 다양화되고 있다. 고용노동부에서 발행한 "2014년 산업재해 현황분석"에 의하면 2014년도 산업재해보상보험법 적용사업장

* 본 연구는 산업안전보건연구원 안전보건정책연구실로부터 산업재해자료를 제공받아 수행한 것으로 이 자리를 빌려 산업안전보건 연구원에 감사의 마음을 표합니다. 또한 본 연구의 내용은 연구자의 개인적 견해이며, 산업안전보건연구원의 공식적인 견해와 다를 수도 있음을 알려드립니다.

† 교신저자 jbaik@knou.ac.kr

2016년 12월 7일 접수; 2017년 2월 8일 수정본 접수; 2017년 2월 17일 게재 확정.

2,187,391개소에 종사하는 근로자 17,062,308명 중에서 4일 이상 요양을 요하는 재해자가 90,909명이 발생하였으며, 산업재해로 인한 직접손실액은 3,926,559백만 원으로 전년대비 3.45% 증가하였으며, 직·간접손실을 포함한 경제적 손실 추정액은 19,632,795백만 원으로 전년대비 3.45%가 증가하였고, 근로손실일수는 48,398,387일로 전년대비 8.26%가 감소한 것으로 나타났다. 이와 같이 산업재해로 인해 입는 직·간접손실액이 너무 크므로 산업재해가 발생하지 않도록 미리 조치를 취할 필요가 있다.

우리나라는 1970년대 이후 성장 위주의 경제성장 전략으로 인해 근로현장에서 산업재해가 어쩔 수 없이 발생하는 것으로 인식되었다[1]. 더욱이 사회와 기업이 산업재해가 근로자 개인의 부주의에 기인한다고 여겨, 산업재해 발생의 본질을 왜곡시켜 산업재해의 문제해결을 더욱 어렵게 만들어왔다[2].

최근 들어서는 산업재해 보상내용, 보상범위 등 법제도 정비와 관련된 내용[3, 4], 자동차 관련업, 병원, 제조업, 건설업 등의 특정 직업군의 재해특성을 분석한 연구[5-9], 특정 사업장의 특성과 재해발생간의 연관성에 관한 연구[10], 특정집단의 집단적 혹은 개인적 차원의 재해원인 및 관련 예방대책[11, 12] 등 다각적인 측면을 다루고 있다.

근래에 어떤 연구자들은 7년 간(2002~2008년) 산업재해 발생현황을 이용하여, 국내 전체산업에 걸쳐 여러 재해유형별로 산업재해 발생의 상대적 위험도를 비교분석하고, 재해발생시간에 관한 수리적 모델인 순환모델을 추정하였다[1].

또 다른 연구자는 C4.5 알고리즘을 사용하여 의사결정나무를 구성하였고[13], C4.5 알고리즘을 사용하기 위한 툴로 SAS의 Enterprise Miner를 사용하였다. 이외에 CHIAD 알고리즘을 이용하여 정량적 위험성 평가기법이 부재한 제조업에서의 산업재해에 대해 정량적 평가가 가능한 특성분석(feature analysis)을 제시한 연구도 있다[14, 15].

데이터 마이닝(Data Mining) 기법은 대량의 과거 데이터로부터 예측하기 위해 활용된다. 기존의 데이터 마이닝은 이동통신사의 이탈 고객 예측, 취업고객 분석 및 예측, 의학적 진단 예측 등에 활용되어져 왔다[16]. 하지만 근래에는 데이터 마이닝 기법을 이용하여 강원도 내 제조업 재해자 총 10,536명의 자료를

바탕으로 재해형태인 사망 및 부상을 예측하기 위하여 의사결정나무(decision tree) 알고리즘인 CHAID, C4.5, CART, 로지스틱회귀(Logistic Regression), 신경망(Neural Network) 등의 다양한 기법을 적용했다[17].

한편, 산업안전보건법에서는 기본적인 산업재해와 관련한 규정을 제시하고 있는데, 이는 산재예방정책보다 앞서 지켜야 하는 규정으로 볼 수 있으며, 따라서 이러한 기초적인 규정과 제도를 지키는 것이 산업재해를 예방하는 가장 근본적인 방법이다. 이에 어떤 연구자는 t 검정 및 로지스틱 회귀모델을 이용하여 산업안전보건법의 이행여부와 산업재해 발생과의 관계를 고찰하였다[18].

근래에는 우리나라의 4차(2014년) 근로환경조사 결과를 바탕으로 근로자의 사회·인구학적 특성별 압력과 상태의 수준을 파악하였고, 근로시간과 밀접한 관련이 있는 중요 변수들 간의 관계를 파악하였다[19].

정부에서는 산업재해를 줄여 보다 나은 대한민국을 만들기 위해 여러 가지 정책 활동을 펴왔다. 그 결과 예를 들어 고용노동부에서는 특정 재해에 대한 적절한 정책 활동(시정조치, 고소 등)을 취하고, 공단의 경우 사업장 단위별로 여러 가지 예방활동(예를 들어 클린사업, 기술지원 등)을 펴나가고 있다. 하지만 이런 시정조치나 클린사업 등이 제대로 효과를 발휘했는지 기존에는 주로 빈도분석을 통해 검증은 했는데, 본 연구에서는 산업재해와 관련된 과거 10년간(2006~2015년)의 데이터를 이용하여 적절한 통계분석을 실시하고자 한다. 이에 제 2장에서는 과거 10년간의 산업재해 데이터에서 중요 변수들을 선별하여 이들 변수에 대한 탐색적 자료분석을 실시한다. 다음으로 제 3장부터는 좀 더 구체적인 통계분석방법을 적용해보고, 제 4장에서는 이를 보완하는 측면에서 성향점수 매칭 방법을 적용한 후 화학 대조군 연구를 다시 실시해본다. 이어서 제 5장에서는 산업재해 여부를 Y로 놓고, 근로자수, 업종 등의 여러 위험요인의 영향력을 살펴볼 수 있는 로지스틱 회귀분석 방법을 적용해보고, 제 6장에서는 산업재해횟수를 Y로 놓고, 앞의 위험요인들의 영향력을 살펴볼 수 있는 포아송 회귀분석 방법을 실시해본다. 마지막으로 제 7장에서는 지금까지 살펴본 것을 요약하며 추후에 해야 할 일을 기술한다.

2. 산업재해 데이터에 대한 탐색적 자료분석

산업재해와 관련된 데이터는 산업안전보건연구원 에서 여러 가지의 형태로 수집하고 분석하고 있다. 본 연구에서는 이들 여러 데이터를 병합하여 하나의 파일 (본 연구에서는 c2006_date_re 파일이라고 부름)을 중심으로 산업재해를 분석하고자 한다. c2006_date_re 파일에는 2006년도에 사업을 개시한 78,743개의 사업장에서 발생한 산업재해정보가 있다. 구체적으로 사업장별 사업장관리번호, 사업개시번호, 사업개시일(date_start), 사업종료일(date_end) 및 산업재해 횟수와 각 산업재해일에 대한 정보가 있으며, 2006년부터 2015년까지 각 년도마다

- (*) 사업장상태(정상, 소멸), 업종(제조업, 건설업, 서비스업 등), 근로자수 및 재해자수(= 사고사망자수+질병사망자수+사고부상자수+질병이환자수)에 대한 정보가 있으며,
- (**) 산업안전보건공단, 민간 및 고용노동부가 각 사업장의 안전 및 보건의 확보를 위해 각 사업장에 개입한 횟수에 대한 정보가 있다.

이들 변수에 대해 탐색적 방법으로 단변량 분석을 해 볼 수 있는데, 본 연구에서는 이들 중 일부분만 보여준다. 첫 번째로 사업개시연월일(data_start)이 있는데, 월별로 사업개시 분포가 어떻게 되어 있는지 살펴본 결과 <Fig. 1>에서와 같다. 이로부터 이상하게도 1월에 다른 어느 달보다 두 배 또는 세 배까지도 사업개시를 많이 한 것으로 보인다(또는 전년도에 사업개시를 했는데, 1월에 사업 개시했다고 신고한 것일 수도 있음).

다음으로 사업장별 산업재해수를 상자그림(boxplot)으로 그려본 결과 <Fig. 2>에서와 같다. 이로부터 대부분의 사업장에서 산업재해는 일어나지 않았지만 일부 사업장에서는 여러 번 산업재해가 일어났다는 것을 알 수 있다. <Table 1>은 구체적으로 사업장별

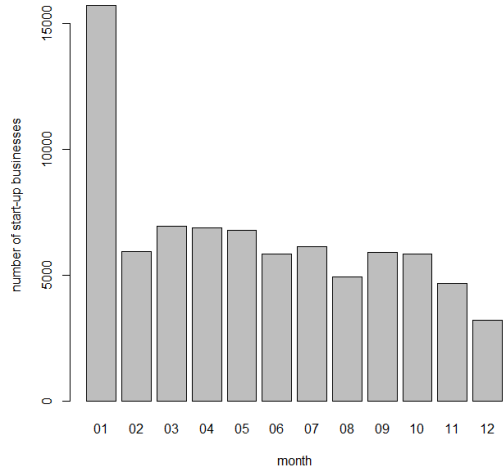


Fig. 1 Monthly number of start-up businesses

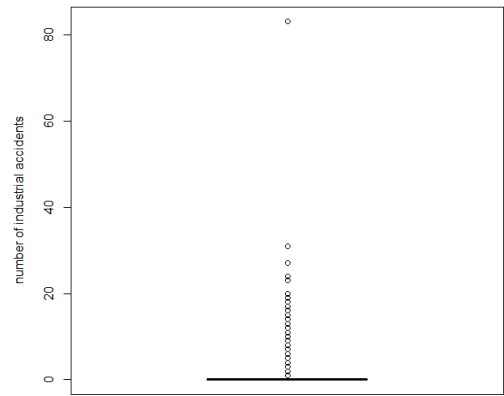


Fig. 2 Boxplot of the number of industrial accidents

산업재해가 일어난 횟수를 나타낸다. 따라서 10년간 64,408개의 사업장에서는 산업재해가 한 번도 일어나지 않았으며, 8,750개의 사업장에서는 산업재해가 한 번만 일어났다. 10년간 산업재해가 5회까지 일어난 곳은 78,156개의 사업장으로 전체의 99.35%를 점유한다. 10년간 83회의 산업재해가 일어난 사업장도 한 군데 있다.

Table 1 Number of industrial accidents

# of industrial accidents	0	1	2	3	4	5	6	7	8	9	10	11	12
frequency	64,408	8,750	2,892	1,202	569	335	173	116	65	42	36	15	20
# of industrial accidents	13	14	15	16	17	18	19	20	23	24	27	31	83
frequency	9	12	3	3	6	1	2	1	1	1	1	2	1

앞의 산업재해수는 다른 말로 재해자수와 동일하다. 그런데 재해자수는 다음과 같이 표현할 수도 있다.

$$\text{재해자수} = \text{사고사망자수} + \text{질병사망자수} + \text{사고부상자수} + \text{질병이환자수}$$

따라서 앞의 오른쪽에 있는 각각의 변수에 대해 <Fig. 2>에서와 같은 상자그림을 그릴 수도 있다. 마찬가지로 이들 각각의 변수에 대해 <Table 1>과 같은 표를 그릴 수도 있다.

2006년도에 사업을 개시한 사업장들의 업종별 분포를 2006, 2007, ..., 2015년도에 조사한 결과 <Table 2>와 같다. 이로부터 ‘기타의 사업(서비스업)’의 사업장이 어느 연도이든 가장 많고, 그 다음으로 제조업이 많다는 것을 알 수 있다.

2006년도에 사업을 개시한 사업장의 근로자수를 2006, 2007, ..., 2015년도 별로 요약통계(summary sta-

tistics)를 구한 결과 <Table 3>과 같다. 이로부터 2006년에 사업장의 근로자수는 적게는 0명이며, 많게는 5,292명이나 중앙값(median)은 2명이므로 전체 사업장들 중에서 근로자수가 2명인 사업장이 가장 많으며, 평균적으로 근로자는 4.408명이라는 것을 알 수 있다.

많은 사업장의 경우 근로자수는 매년 변할 수 있다. 어떤 사업장은 그 변동폭이 상당히 크다. 예를 들어 어떤 사업장은 제조업으로 사업개시를 2006년 4월 1일에 시작하여 2006년에는 근로자를 5,492명 투입했지만 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015년도에는 근로자를 각각 1, 1, 1, 2, 3, 3, 3, 3명씩만 투입했다. 한편, 이 사업장의 경우 사업개시 후 2011일이 지난 후인 2011년 10월 3일에 산업재해가 한 번 발생했다(2명의 근로자 중 1명이 산업재해가 발생한 것임). 이에 해당 사업장의 근로자수는 매년 근로자수의 평균치로 잡는 것이 적절할 것이다.

Table 2 Yearly business category of 2006 start-ups

business category	year									
	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
con	667	575	552	793	791	669	603	461	402	336
mining	55	55	53	52	51	44	41	36	34	30
banking and insurance	415	419	422	422	422	373	367	333	303	268
service	58,880	58,741	58,597	58,143	58,114	51,052	50,317	44,872	40,671	38,207
agriculture	284	284	284	288	296	280	280	264	241	219
fishing	110	111	111	112	112	100	99	86	82	74
transportation, storage, telecommunication	2,489	2,489	2,490	2,487	2,483	2,199	2,178	1,989	1,829	1,675
forestry	77	72	69	64	61	54	53	43	42	38
electricity, gas, waterwork	47	46	47	48	49	44	45	43	42	41
manufacturing	15,642	15,714	15,800	15,834	15,928	14,727	14,570	13,350	12,410	11,413
no response	-	160	241	423	359	9,124	10,113	17,189	22,610	28,365

Table 3 Summary statistics for the yearly number of workers

year	minimum	1 st quartile	median	mean	3 rd quartile	maximum	# missing
2006	0.000	1.000	2.000	4.408	3.000	5492.000	0
2007	0.000	1.000	2.000	4.748	4.000	1962.000	160
2008	0.000	1.000	2.000	5.340	4.000	2441.000	241
2009	0.000	1.000	2.000	5.656	4.000	4104.000	423
2010	0.000	1.000	2.000	5.848	5.000	4104.000	359
2011	0.000	1.000	2.000	6.183	5.000	4271.000	9,124
2012	0.000	1.000	2.000	6.677	5.000	4271.000	10,113
2013	0.000	1.000	2.000	6.871	6.000	4271.000	17,189
2014	0.000	1.000	2.000	6.987	6.000	3347.000	22,610
2015	0.000	1.000	3.000	7.356	6.000	3347.000	28,365

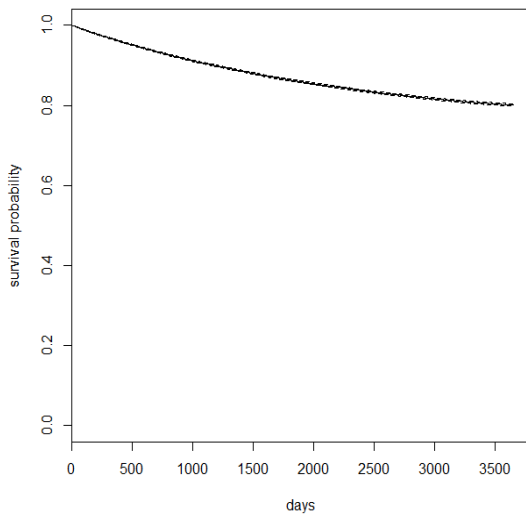


Fig. 3 Survival probability for the period till the first industrial accident

다음으로 2006년도에 사업을 개시한 78,666개의 사업장에서 첫 번째 산업재해가 발생하기까지 걸린 시간을 구할 수 있다. 그런데 어떤 사업장은 사업개시(date_start)부터 사업종료(date_end) 또는 관측이 끝나는 2015년 12월 31일까지 산업재해가 발생하지 않았다. 이 경우 해당 사업장에 대해서는 생존분석에서 일컫는 중도중단시간(censored time)을 얻게 된다. 이와 같이 특정 사건이 관측되지 않은 중도중단시간도 있는

경우 통상적으로 카플런-마이어(Kaplan-Meier) 방법에 의하여 시간에 따른 survival probability를 구한다.

<Fig. 3>은 2006년도에 사업을 개시한 사업장들에 대해 Kaplan-Meier 방법을 적용하여 첫 번째 산업재해가 발생하기까지 걸리는 시간에 따른 생존확률(survival probability)이 어떻게 변하는지 보여준다. 이로부터 1년(365일)간의 시간이 지나면 생존확률이 0.963이 되어 전체 사업장의 3.7% 정도는 산업재해를 당하게 되고, 이 수치는 2년(730일)간의 시간이 지나면 0.932로 떨어져 2년 정도 지나면 6.8% 정도의 사업장이 산업재해를 당하게 된다.

사업장들은 산업안전공단, 민간 및 고용노동부로부터 여러 가지 지도를 받을 수 있다. <Table 4>는 2006년부터 2015년까지 10년간 2006년도에 사업 개시한 사업장들이 산업안전공단, 민간 및 고용노동부로부터 지도받은 횟수를 나타낸다. 이로부터 일반적으로 민간이 산업안전보건공단이나 고용노동부보다 사업장에 더 많이 지도했음을 알 수 있다. 예를 들어, 10년간 지도횟수가 1회인 경우를 살펴보면 민간은 17,920개의 사업장에 대해 지도를 나갔지만 산업안전보건공단과 고용노동부는 각각 7,275개 및 3,435개의 사업장에 대해 지도를 나갔다. 주어진 데이터에서는 총 78,743개의 사업장 중에서 31,139(= 78,743-47,604)개의 사업장은 해당 질문에 대해 무응답 상태이다.

Table 4 Number of involvement from 'Korea Safety and Health Agency', Civilian, and 'Ministry of Employment and Labor'

# of involvement	Agency		
	Korea Safety and Health Agency	Civilian	Ministry of Employment and Labor
0	333,29	5,206	42,927
1	7,275	17,920	3,435
2	3,384	13,094	749
3	1,756	7,829	272
4	985	3,089	123
5	465	421	45
6	239	43	13
7	96	2	18
8	45	0	13
9	18	0	6
10	12	0	3
Sum	47,604	47,604	47,604

3. 환자 대조군 연구

환자 대조군 연구(case control study)는 의학통계 분야에서 새로운 처치(new treatment)를 기존의 표준적인 처치(standard treatment)와 비교해 효과가 있는지 파악할 때 쓰는 방법으로 각 처치에 대해 특정 사건이 일어난 건수가 다음과 같을 때 OR(odds ratio)를 구하여 새로운 처치의 효과를 검증한다. 일반적으로 오즈비(OR: odds ratio)는 다음과 같이 해석할 수 있다.

오즈비>>1: 표준적인 처치 그룹이 새로운 처치 그룹에 비해 해당 사건을 겪을 가능성이 훨씬 더 크다.

오즈비≈1: 두 그룹이 해당 사건을 겪을 가능성은 비슷하다.

오즈비<<1: 표준적인 처치 그룹이 새로운 처치 그룹에 비해 해당 사건을 겪을 가능성이 더 적다.

하지만 오즈비<<1인 경우 오즈비>>1인 경우와 같이 직접적으로 해석할 수는 없다고 한다. 그래서 특정 사건이 일어날 가능성이 많은 그룹을 첫 번째 행에 둔다 (<http://www.biochemia-medica.com/content/odds-ratio-calculation-usage-and-interpretation>).

산업재해에 대한 데이터에서도 환자 대조군 연구를 적용할 수 있다. 산업재해를 당했는지 또는 당하지 않았는지는 앞에서 ‘사건이 발생했는지 또는 발생하지 않았는지’로 나타낼 수 있고, 표준적인 처치와 새로운 처치는 ‘산업안전보건공단, 민간 및 고용노동부로부터 개입이 없었는지 또는 있었는지’로 나타낼 수 있다. 이런 경우 구해지는 오즈비로 산업안전보건공단(Korea Safety and Health Agency, KSHA), 민간(Civilian), 고용노동부(Ministry of Employment and Labor, MEL)의 역할이 유효했는지 평가할 수 있다.

우선 산업안전보건공단이 사업장에 대한 지도가 유효했는지 보기 위해 환자 대조군 연구에서와 같은 표를 만들면 <Table 5>에서와 같은 결과를 얻을 수 있다 (참고로 공단 지도횟수가 0인 사업장은 33,329개이고, 공단 지도횟수가 0보다 큰 사업장은 14,275개이다. 이 결과로부터 산업안전보건공단의 오즈비가 0.135로 1보다 훨씬 작으므로 산업안전보건공단의 지도는 유효했다고 보기 어렵다. 반면, <Table 6>으로부터 오즈비가 1보다 크므로 민간의 지도는 어느 정도 유효했다고 할 수 있다. 하지만 <Table 7>로부터 오즈비가 1보다 훨씬 작으므로 고용노동부의 지도는 유효했다고 보기 어렵다.

Table 5 Effectiveness table for the involvement of ‘Korea Safety and Health Agency’

	# of industrial accidents > 0	# of industrial accidents = 0	
KSHA involvements = 0	4,318	29,011	total 33,329
KSHA involvements > 0	7,494	6,781	total 14,275
odds	4,318/7,494 = 0.576		odds ratio = 0.135

Table 6 Effectiveness table for the involvement of civilian

	# of industrial accidents > 0	# of industrial accidents = 0	
Civilian involvements = 0	2,144	3,062	total 5,206
Civilian involvements > 0	9,668	32,730	total 42,398
odds	2144/9668 = 0.223		odds ratio = 2.37

Table 7 Effectiveness table for the involvement of ‘Ministry of Employment and Labor’

	# of industrial accidents > 0	# of industrial accidents = 0	
MEL involvements = 0	9,019	33,908	total 42,927
MEL involvements > 0	2,793	1,884	total 4,677
odds	9,019/2,793 = 3.229		odds ratio = 0.18

4. 성향점수 매칭 후의 환자 대조군 연구

앞의 환자 대조군 연구에서 새로운 처치 그룹(예를 들어, 공단의 경우 지도횟수 > 0인 사업장들)과 기존의 표준적인 처치 그룹(예를 들어, 공단의 경우 지도횟수 = 0인 사업장들)은 다른 여러 속성(예를 들어 사업장규모(근로자수), 업종 등) 즉, 공변수들의 측면에서 동질적이어야 하는데, 관측연구(observational study)에서는 그렇지 못하다. 그래서 각 사업장에 대해 다음에서와 같이 성향점수(propensity score)를 구하여, 두 그룹 간 성향점수를 매칭하여 산업재해가 얼마나 많이 차이가 나는지 살펴보는 것이 적절할 것이다.

산업안전보건공단의 지도가 유효했는지 보기 위해 성향점수는 다음과 같이 로지스틱 회귀분석 모델을 이용하여 구한다. 우선 다음과 같이 변수를 지정한다.

- $X = 1$, 새로운 처치 그룹(공단 지도횟수 > 0)
- $X = 0$, 기존의 표준적 처치 그룹(공단 지도횟수 = 0)
- Y = 산업재해 수
- Z 는 공변수로서
- Z_1 = 사업장규모(근로자수)
- Z_2 = 업종

이런 경우 사업장 z^* 에 대한 성향점수는 다음과 같이 구한다.

$$P(X=1|z^*) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 z_1^* + \beta_2 z_2^*)}}$$

이제 앞의 성향점수가 같도록(또는 가능하면 같도록) 새로운 처치 그룹과 기존의 표준적인 처치 그룹을 정하고(예를 들어 1대 1로 매칭 실시) 난 후 X 와 Y 의 관계를 살펴본다. 성향점수가 같도록 두 그룹을 정하여 표를 다시 한 번 구해보면 <Table 8>과 같이 나온다(참고로 공단 지도횟수가 0인 사업장과 0보다 큰 사업장의 수는 각각 14,275개이다).

이로부터 성향점수 매칭 결과 산업재해를 당한 그룹의 경우 공단 지도를 받지 않은 사업장이 공단 지도를 받은 사업장에 비해 그 비율이 0.352로 성향점수 매칭 전의 0.576보다 작아졌지만 산업재해를 당하지 않은 그룹의 경우 공단지도를 받지 않은 사업장이 공단지도를 받은 사업장에 비해 그 비율이 1.716으로 성향점수 매칭 전의 4.278보다 훨씬 더 줄어들어, 전체적으로 오즈비는 0.204로 성향점수 매칭 전의 0.135보다 더 커졌다. 하지만 아직도 오즈비가 1보다 훨씬 작게 나와 산업안전보건공단의 지도 효과가 의문시 된다.

다음으로 민간의 지도가 유효했는지 보기 위해 앞에서와 같은 방법으로 성향점수를 구해, 두 그룹 간 성향점수가 같도록(또는 가능하면 같도록) 새로운 처치 그룹과 기존의 표준적인 처치 그룹을 정하여(예를 들어 1대 1로 매칭 실시) X 와 Y 의 관계를 보기 위해 표를 다시 한 번 구해보면 <Table 9>에서와 같이 나온다(참고로 민간 지도횟수가 0인 사업장과 0보다 큰 사업장의 수는 각각 5,206개이다).

Table 8 Effectiveness table for the involvement of ‘Korea Safety and Health Agency’(After applying PSM method)

# accidents involvement	# of industrial accidents > 0	# of industrial accidents = 0	total
KSHA involvements = 0	2,628	11,647	total 14,275
KSHA involvements > 0	7,494	6,781	total 14,275
odds	0.352	1.716	odds ratio = 0.204

Table 9 Effectiveness table for the involvement of civilian(After applying PSM method)

# accidents involvement	# of industrial accidents > 0	# of industrial accidents = 0	total
Civilian involvements = 0	2,144	3,062	total 5,206
Civilian involvements > 0	530	4,676	total 5,206
odds	4.045	0.655	odds ratio = 6.18

Table 10 Effectiveness table for the involvement of ‘Ministry of Employment and Labor’ (After applying PSM method)

# accidents involvement	# of industrial accidents > 0	# of industrial accidents = 0	total
MEL involvements = 0	1,909	2,768	total 4,677
MEL involvements > 0	2,793	1,884	total 4,677
odds	0.683	1.470	odds ratio = 0.465

이로부터 성향점수 매칭 결과 산업재해를 당한 그룹의 경우 민간 지도를 받지 않은 사업장이 민간 지도를 받은 사업장에 비해 그 비율이 4.045로 성향점수 매칭 전의 0.223보다 훨씬 더 커졌으며, 산업재해를 당하지 않은 그룹의 경우 민간 지도를 받지 않은 사업장이 민간 지도를 받은 사업장에 비해 그 비율이 0.655로 성향점수 매칭 전의 0.094보다 더 커졌으며, 전체적으로 오즈비는 6.18로 성향점수 매칭 전의 2.37보다 훨씬 더 커졌다. 따라서 오즈비가 1보다 훨씬 더 크게 나오며, 더욱이 성향점수 매칭 전보다 더 커졌으므로 민간의 지도 효과는 두드러지는 것으로 보인다.

마지막으로 성향점수를 매칭하여 고용노동부의 지도가 유효했는지 보기 위해 구한 표는 <Table 10>에서와 같이 나온다(참고로 고용노동부 지도횟수가 0인 사업장과 0보다 큰 사업장의 수는 각각 4,677개이다).

이로부터 성향점수 매칭 결과 산업재해를 당한 그룹의 경우 고용노동부 지도를 받지 않은 사업장이 고용노동부 지도를 받은 사업장에 비해 그 비율이 0.683으로 성향점수 매칭 전의 3.229보다 훨씬 작아졌지만 산업재해를 당하지 않은 그룹의 경우 고용노동부 지도를 받지 않은 사업장이 고용노동부 지도를 받은 사업장에 비해 그 비율이 1.470으로 성향점수 매칭 전의 17.998보다 훨씬 더 줄어들어, 전체적으로 오즈비는 0.465로 성향점수 매칭 전의 0.18보다 더 커졌다. 하지만 아직도 오즈비가 1보다 훨씬 작게 나와 고용노동부의 지도 효과가 의문시 된다.

5. 로지스틱 회귀분석

로지스틱 회귀분석(logistic regression)은 반응변수가 두 개의 범주로 나누어지는 경우 독립변수(들)의 영향력을 파악하기 위해 적용하는 통계적 분석방법이다. 본 연구에서 반응변수 y 와 독립변수 x 를 나열하면 다음과 같다.

y (산업재해) = 0, 산업재해 수 = 0인 경우
 = 1, 산업재해 수 > 0인 경우

x (위험요인들)

x_1 : 근로자수(0, 1, 2, ...의 값을 취하는 이산형 데이터)

x_2 : 업종(범주형 데이터)

x_3 : 공단의 지도 여부(0 for none, 1 for ‘greater than 0’)

x_4 : 민간의 지도 여부(0 for none, 1 for ‘greater than 0’)

x_5 : 고용노동부의 지도 여부(0 for none, 1 for ‘greater than 0’)

이 경우 로지스틱 회귀분석 모델은 다음과 같다.

$$P(Y=1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}$$

본 연구에 주어진 데이터에 대해 로지스틱 회귀분석을 적용해본 결과 <Table 11>과 같다. 이로부터 근로자수, 업종들 중 일부(농업, 어업, ‘운수·창고 및 통신업’, 임업), 그리고 안전보건공단, 민간 및 고용노동부의 지도는 산업재해와 통계적으로 유의한 관계에 있음을 알 수 있다. 근로자수의 경우 회귀계수가+이므로 근로자수가 증가할수록 산업재해가 일어날 가능성은 많아진다는 것을 알 수 있다.

업종들 중 농업, 어업, ‘운수·창고 및 통신업’ 및 임업은 회귀계수가+로 유의하므로 건설업에 비해 산업재해가 일어날 가능성이 많다는 것을 알 수 있다. 한편, 광업, ‘금융 및 보험업’, ‘기타의 사업(서비스업)’, ‘전기·가스 및 상수도업’ 및 제조업은 회귀계수가 유의하지 않으므로 건설업에 비해 산업재해가 많게도 또는 적게도 일어나지 않는다.

마지막으로 산업안전보건공단, 민간 및 고용노동부의 지도여부는 산업재해 발생과 통계적으로 유의한 관계에 있다. 그런데 회귀계수가 모두 양(+)의 값을 가지므로 각 기관의 지도가 이루어지는 사업장이

Table 11 Results from logistic regression

Coefficients:		Estimate	Std. Error	z value	Pr(> z)
(Intercept)		-2.499882	0.191827	-13.032	< 2e-16***
no_employees		0.021432	0.001198	17.892	< 2e-16***
business_category: mining		0.507220	0.652469	0.777	0.436931
business_category: banking &		-0.326385	0.738009	-0.442	0.658308
business_category: service		-0.291096	0.191960	-1.516	0.129408
business_category: agriculture		1.727666	0.334954	5.158	2.5e-07***
business_category: no response		0.046503	0.202668	0.229	0.818516
business_category: fishing		2.130921	0.815660	2.613	0.008988**
business_category: transportation &		0.400318	0.209226	1.913	0.055706
business_category: forestry		2.743195	0.793211	3.458	0.000544***
business_category: electricity &		0.578424	0.622818	0.929	0.353033
business_category: manufacturing		0.179196	0.191813	0.934	0.350188
KSHA		1.686901	0.029704	56.791	< 2e-16***
Civilian		0.688759	0.037763	18.239	< 2e-16***
MEL		1.055597	0.039753	26.554	< 2e-16***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

산업재해가 일어날 가능성이 더 크다는 것을 알 수 있다. 산업안전보건공단, 민간 및 고용노동부가 산업재해가 보다 많이 일어나는 사업장에 더욱 더 많은 지도를 하는 것으로 판단된다.

이 경우 포아송 회귀분석 모델은 다음과 같다.

$$P(Y=y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}$$

$$y = 0, 1, 2, \dots$$

6. 포아송 회귀분석

포아송 회귀분석(Poisson regression)은 산업재해수와 같이 반응변수의 값이 0, 1, 2, ...과 같이 이산형인 경우 독립변수(들)의 영향력을 파악하기 위해 적용하는 통계적 분석방법이다. 본 연구에서 반응변수 y 와 독립변수 x 를 나열하면 다음과 같다.

$$y(\text{산업재해 수}) = 0, 1, 2, \dots$$

x (위험요인들)

x_1 : 근로자수 (0, 1, 2, ...의 값을 취하는 이산형 데이터)

x_2 : 업종(범주형 데이터)

x_3 : 공단의 지도횟수 (0, 1, 2, ...의 값을 취하는 이산형 데이터)

x_4 : 민간의 지도횟수 (0, 1, 2, ...의 값을 취하는 이산형 데이터)

x_5 : 고용노동부의 지도횟수(0, 1, 2, ...의 값을 취하는 이산형 데이터)

본 연구에 주어진 데이터에 대해 포아송 회귀분석을 적용해본 결과 <Table 11>에서와 같다. 이로부터 로지스틱 회귀분석의 결과와 마찬가지로 근로자수, 업종들 중 일부(광업, 기타의 사업(서비스업), 농업, 어업, '운수·창고 및 통신업', 임업 및 '전기·가스 및 상수도업'), 그리고 안전보건공단, 민간 및 고용노동부의 지도는 산업재해와 통계적으로 유의한 관계에 있음을 알 수 있다. 근로자수의 경우 회귀계수가 +이므로 근로자수가 증가할수록 산업재해가 일어날 가능성은 많아진다는 것을 알 수 있다. 한편, 업종들 중 산업재해와 통계적으로 유의한 관계에 있지 않은 것은 '금융 및 보험업'과 제조업뿐이다. 다음으로 산업재해와 통계적으로 유의한 관계에 있는 업종은 광업, '기타의 사업(서비스업)', 농업, 어업, '운수·창고 및 통신업', 임업 및 '전기·가스 및 상수도업'인데, '기타의 사업(서비스업)'과 '전기·가스 및 상수도업'만 회귀계수가 -이고, 나머지는 +이므로 광업, 농업, 어업, '운수·창고 및 통신업' 및 임업은 건설업에 비해 산업재해가 많이 일어날 위험성이 크고, '기타의

Table 12 Results from Poisson regression

Coefficients:		Estimate	Std. Error	z value	Pr(> z)
(Intercept)		-9.381e-01	6.133e-02	-15.296	< 2e-16***
no_employees		2.444e-03	8.071e-05	30.278	< 2e-16***
business_category: mining		1.284e+00	2.053e-01	6.257	3.92e-10***
business_category: banking &		-7.195e-01	5.805e-01	-1.239	0.215
business_category: service		-6.445e-01	6.201e-02	-10.394	< 2e-16***
business_category: agriculture		9.901e-01	1.120e-01	8.843	< 2e-16***
business_category: no response		-6.500e-02	7.281e-02	-0.893	0.372
business_category: agriculture		1.287e+00	2.130e-01	6.040	1.54e-09***
business_category: transportation &		5.444e-01	7.224e-02	7.537	4.82e-14***
business_category: forestry		1.802e+00	1.408e-01	12.803	< 2e-16***
business_category: electricity &		-1.366e+00	3.076e-01	-4.440	8.99e-06***
business_category: manufacturing		1.128e-02	6.158e-02	0.183	0.855
KSHA		2.835e-01	4.087e-03	69.361	< 2e-16***
Civilian		8.838e-02	5.999e-03	14.732	< 2e-16***
MEL		2.361e-01	5.677e-03	41.592	< 2e-16***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

사업(서비스업)과 ‘전기·가스 및 상수도업’에서는 건설업에 비해 산업재해가 일어날 가능성이 적다.

마지막으로 로지스틱 회귀분석 결과와 마찬가지로 산업안전보건공단, 민간 및 고용노동부의 지도는 산업재해와 통계적으로 유의한 관계에 있다. 그런데 각각의 회귀계수가 양(+)의 값을 가지므로 각 기관의 높은 지도횟수와 산업재해 발생과 밀접한 관련이 있다. 산업안전보건공단, 민간 및 고용노동부가 산업재해가 보다 많이 일어나는 사업장에 대해 더욱 더 많은 지도를 하는 것으로 판단된다.

7. 결론 및 추후 연구

본 연구에서는 정부정책이 효과가 있었는지 파악하기 위하여 과거 10년간의 산업재해 데이터를 살펴보았다. 이들 데이터로부터 중요 변수들을 먼저 도출한 결과 산업재해자수(=사고사망자수+질병사망자수+사고부상자수+질병이환자수), 사업장 상태(정상 또는 소멸), 업종, 근로자수, 순차적인 산업재해 발생시점 이외에 과거 10년간 산업안전보건공단, 민간 및 고용노동부의 지도여부 등이 나왔으며, 이들 각 변수에 대해 탐색적 자료분석 방법으로 데이터를 분석하였다.

환자 대조군 연구에서는 산업재해가 일어나는 것

을 환자(case)로 보고 산업재해가 일어나지 않은 것을 대조군(control)으로 보고, 이런 산업재해의 발생 여부를 산업안전보건공단, 민간 및 고용노동부의 지도 여부와 관련이 있는지 살펴보았다. 그 결과 산업안전보건공단과 고용노동부의 경우 오즈비가 1보다 아주 작게 나와 공단 지도와 고용노동부 지도의 유효성을 의심할 수밖에 없었다. 하지만 이들 기관은 통상 산업재해가 일어난 곳을 집중적으로 지도하므로 오즈비가 1보다 작게 나올 수밖에 없는 것으로 판명된다. 한편, 민간의 경우 오즈비가 1보다 크므로 민간의 지도는 어느 정도 유효했다고 판단된다.

한편, 본 연구에 사용된 데이터는 관측연구로부터 나온 데이터이므로 비교되는 그룹 간 관심요인 이외에 다른 요인의 효과가 혼돈(confounding)되어 나타날 수 있다. 이에 성향점수 매칭 기법을 활용하여 다른 요인들의 효과를 배제시킨다. 그 결과 산업안전보건공단과 고용노동부는 성향점수 매칭 이전보다 사업장에 대한 지도의 효과가 더 있는 것으로 판명되나 아직도 오즈비가 1보다 작으므로 이들의 지도의 효과는 아직도 의문시된다. 앞에서 언급했듯이 산업안전보건공단과 고용노동부는 산업재해가 일어나는 곳을 찾아가 지도를 하는 것으로 판명된다.

다음으로 산업재해의 발생여부를 y 변수로 놓고, x (위험요인들)(x_1 : 근로자수, x_2 : 업종, x_3 : 공단의 지도

여부, x_4 : 민간의 지도여부, x_5 : 고용노동부의 지도여부)의 영향력을 파악하고자 로지스틱 회귀분석을 활용하였다. 그 결과 근로자수는 많을수록 산업재해가 일어날 가능성이 많으며, 업종별로는 농업, 어업, '운수·창고 및 통신업' 및 임업은 건설업에 비해 산업재해가 일어날 가능성이 많다는 것을 알 수 있다. 그리고 산업안전보건공단, 민간 및 고용노동부의 지도와 산업재해 발생과 밀접한 관련이 있는데, 이는 산업안전보건공단, 민간 및 고용노동부가 산업재해가 보다 많이 일어나는 사업장에 더욱 더 많은 지도를 하기 때문인 것으로 판단된다.

다음으로 산업재해 발생횟수를 y 변수로 놓고 x (위험요인들)(x_1 : 근로자수, x_2 : 업종, x_3 : 공단의 지도횟수, x_4 : 민간의 지도횟수, x_5 : 고용노동부의 지도횟수)의 영향력을 파악하고자 포아송 회귀분석을 활용하였다. 그 결과 근로자수는 많을수록 산업재해가 일어날 가능성이 많으며, 업종별로는 광업, 농업, 어업, '운수·창고 및 통신업', 및 임업은 건설업에 비해 산업재해가 더 많이 일어나고, '기타의 사업(서비스업)'은 건설업에 비해 산업재해가 덜 일어난다. 산업안전보건공단, 민간 및 고용노동부의 역할은 앞의 다른 분석에서와 똑같은 결과를 얻을 수 있다.

본 연구에서 이용한 탐색적 자료분석 방법, 환자 대조군 연구, 로지스틱 회귀분석 및 포아송 회귀분석 방법은 신뢰성 분야에도 그대로 적용될 수 있다. 예를 들어 어떤 아이템의 수명(반응변수)에 영향을 미치는 여러 요소(독립변수)들이 있을 때 독립변수들과 반응변수 간의 관계를 앞의 여러 가지 분석방법으로 살펴볼 수 있다.

본 연구에서는 중요 변수에 대한 탐색적 자료분석에서 해당 변수 하나에 대해서만 분포 또는 표를 그렸다. 하지만 중요 변수들 2개 또는 3개간의 관계를 파악하기 위해 그림이나 표를 그릴 수도 있을 것이다.

또한 본 연구에서는 로지스틱 회귀분석과 포아송 회귀분석 방법을 사용하여 위험요인들의 영향력을 살펴보았는데, 각 사업장에서 산업재해가 일어날 때까지의 시간을 염두에 두고 생존분석 방법을 적용할 수도 있을 것이다. 특히 각 사업장에서 1번 이상의 산업재해가 생길 수 있다는 것을 감안하여 주어진 산업재해 데이터를 반복사건(recurrent event) 데이터의 형태로 바꾸어 Cox의 비례위험모형을 적용하여 위험요인들의 영향력을 파악해볼 수 있다.

한편, 지금까지 산업재해 데이터베이스에 들어간 정보는 재해가 일어날 때까지 걸린 시간, 재해자가 속한 사업장 정보(업종, 근로자수 등) 및 '공단, 민간, 고용노동부의 총지도횟수' 등이다. 이외에 재해자에 대한 정보, 사고와 관련된 상황 정보, '공단, 민간, 고용노동부의 정책활동 내역 및 실시일', 사업장 지리정보 등 산업재해와 관련된 정보 등을 산업재해 데이터베이스로 구축할 수 있다면 보다 다양한 분석을 할 수 있을 것이다.

끝으로 지금까지의 분석에서는 산업안전보건공단, 민간 및 고용노동부의 지도가 효과가 있었는지 판단할 때 과거 10년간 누적지도횟수를 가지고 판단하였다. 하지만 총 관측기간을 각 기관의 지도 전과 후로 나누어 산업재해의 발생여부 또는 빈도 등을 통계적인 방법으로 비교한다든지 또는 의학통계에서 반복측정치(repeated measure)에 대한 효과 분석방법을 이용하여 각 기관의 지도가 효과가 있었는지 판단할 수도 있을 것이다.

References

- [1] Kim, H. Y. and Heo, T. Y. (2010). "An Analysis of relative injury risk by industry and estimation of a circular distribution model for industrial injury". Seoul City Research, Vol. 11, No. 1, pp. 127-138.
- [2] Ju, J. H. (1997). "An analysis of factors and structure affecting industrial accidents in Korea". Kyungsoong University Ph.D. Dissertation.
- [3] Kim, H. S. (2008). "A Study on the relations between industrial accident insurance and the automobile insurance". Labour Law, No. 26, pp. 303-325.
- [4] Park, J. S. (2006). "Occupational accidents due to colleague's abusive act and right to reimbursement of industrial accidents compensation insurance". Labour Law, Vol. 22, pp. 363-386.
- [5] Kim, S. K. (1998). "A status of the report for industrial injuries and illnesses at an automobile related plant". Annals of Occupational and Environmental Medicine, Vol. 10, No. 4, pp. 562-570.
- [6] Lee, C. J., Jun, Y. U., Choi, Y. H., and Jo, A. (2002). "Causes and preventive measures for low back pain

- industrial accidents suffered by automobile assembly workers". Korea Ergonomics Society Conference Proceeding, pp. 119-123.
- [7] Lee, J. C., Shin, S. W., and Lee C. S. (2007). "Accident analysis of middle-aged & advanced-aged construction workers". Korea Architecture Association Conference Proceeding (Structural System), Vol. 27, No. 1, pp. 797-800.
- [8] Lee, K. S. and Jung, B. Y. (2008). "An analysis of industrial accidents in small-scale fiber business". Korea Ergonomics Society Spring Conference Proceeding, pp. 252-255.
- [9] Park, H. J. (2007). "Research on industrial disaster in hospital". Korea Ergonomics Society Fall Conference Proceeding, pp. 492-495.
- [10] Lee, G. S. et al. (2006). "Relationship between Injury Occurrence and Workplace Organization in Small-sized Manufacturing Factories". Korea Industrial Medicine Society, Vol. 18, No. 2, pp. 73-86.
- [11] Lee, S. W., Kim, K. S., and Kim, T. W. (2008). "The status and characteristics of industrial accidents for migrant workers in Korea compared with native workers". Korea Industrial Medicine Society, Vol. 20, No. 4, pp. 351-361.
- [12] Kim, H. H. et al. (2009). "An analysis of characteristics of musculoskeletal disorders risk factors". Korea Ergonomics Society, Vol. 28, No. 3, pp. 17-25.
- [13] Leem, Y. M., Kwag, J. K., and Hwang, Y. S. (2005). "A feature analysis of industrial accidents using C4.5 algorithm". Korea Safety Society, Vol. 20, No. 4, pp. 130-137.
- [14] Leem, Y. M. and Hwang, Y. S. (2006). "Data Analysis of Industrial Accidents in Manufacturing Industries Using CHIAD Algorithm". Korea Safety Management Society/ Korea Cyber Terrorism Information Transfer Society Spring Conference Proceeding, pp. 45-50.
- [15] Song, J. M. and Yoon, S. U. (2004). "A study on split selection algorithms in decision tree". Yonsei University Master's Thesis.
- [16] Lee, K. N. and Lee, H. C. (2003). "A Study on the combined decision tree (C4.5) and neural network algorithm for classification of mobile telecommunication customer". Korea Intelligence Information System Society, Vol. 9, No. 1, pp. 139-155.
- [17] Leem, Y. M. and Ryu, C. H. (2006). "A comparison of data mining techniques for predicting model of industrial accidents". Korea Industrial Management System Society Conference Proceeding, pp. 107-113.
- [18] Jung, W. I. and Jun, Y. I. (2014). "Working conditions and industrial accidents in accordance with safety and health environment in the workplace". Korea Crisis Management Society, Vol. 10, pp. 323-344.
- [19] Kim, Y. S., Lee, J. H., and Baik, J. W. (2015). "2014 Korea Working Conditions Survey Data Analysis". Journal of Applied Reliability, Vol. 15, No. 3, pp. 181-191.