

연구보고서

빅데이터 기반 직업코호트 구축을 통한 질병 발생 연구

윤진하 · 고상백 · 안연순 · 김환철 · 강모열 · 박철용 ·
김지현 · 임성실 · 김민석 · 육지후 · 주진한 · 김은미

산업재해예방

안전보건공단

산업안전보건연구원



제 출 문

산업안전보건연구원장 귀하

본 보고서를 “빅데이터 기반 직업코호트 구축을 통한
질병 발생 연구”의 최종보고서로 제출합니다.

2017 년 11 월

연구기관: 연세대학교 산학협력단

연구기간: 2017.04.12.~ 2017.11.30.

연구책임자: 윤진하 (연세대학교 의과대학 교수)

요 약 문

연구기간

2017. 04. 12. ~ 2017.11.30.

핵심 단어

빅데이터, 직업코호트, 전국민 수진자료,
전체근로자

연구과제명

빅데이터 기반 직업 코호트 구축을
통한 질병 발생 연구

1. 연구배경

빅데이터 기반 직업관련코호트 구축은 이미 알려진 직업병의 예방과 관리 뿐 아니라 알려지지 않은 위험요인과 건강영향을 발견할 수 있다는 점에서 필수적인 연구라고 판단된다. 본 연구의 목적은 국민건강보험공단 빅데이터를 이용한 직업관련 코호트를 구축하고, 향후 근로자의 건강영향평가 또는 직업관련질환 예방 및 관리에 이용가능한지에 대한 평가 및 고찰을 하는데 있다

2. 주요 연구내용

우리나라 전체 근로자의 2002년부터 2015년까지의 건강보험공단 입원 수진내역을 바탕으로 하여 코호트를 구축하였다. 기본적인 대조군으로서 “공무원 및 사립학교 교직원”을 설정하였다.

본 연구에서도 건강보험공단의 데이터를 사용하기 위한 권한을 얻는 과정을 거쳐야 하지만 재정적 한계 등으로 인하여 그리하지 못하였다. 따라서 본 연구만을 위한 건강보험공단 빅데이터 분석은 불가능하였다. 그러나 코호트의 구축가능 여부에 대한 현실성을 확인, 검증하기 위해서는 분석 후 고찰이 반드시 필요하므로, 본 보고서에 실린 코호트

시범운영 및 활용방안은 대한직업환경의학 빅데이터위원회와 국민건강보험공단 사이의 업무협약(MOU)을 통해 이루어진 연구결과 및 세미나(2017년 8월, 2017년 11월) 결과를 참조, 인용하여 제시함으로써 빅데이터의 시범적 구축과 운용 과정에 대한 고찰을 시행하였다.

2002년부터 2015년까지의 근로자 입원수진기록 및 상병명에 대하여 근로자 10만 명 당 환자 수, 근로자 1,000명 당 입원일수, 질환이 있는 근로자 1명 당 입원일수를 지표로 설정하여 구하였으며 그 중 근로자 1,000명 당 입원일수 중 평균 순위가 상위 10위안에 속하는 질환 10가지를 선정하여 17개의 산업군별로 대조군에 대비한 연령표준화발생비를 구하였다. 코호트 구축은 2015년 사업장 정보를 기준으로 이루어졌으며, 입사년도와 퇴사년도를 확인하여 코호트 유입시점을 확인하고 퇴사자에서의 질병 발생도 확인할 수 있었다. 더불어 병·의원 종사자 코호트, 육상운수업종사자 코호트, 반도체·디스플레이·타이어 제조사업장 코호트, 분진 노출 근로자 코호트 등을 구축한 결과를 평가, 고찰하여 다양한 코호트 구축방법과 직업군, 건강영향을 확인하고 향후 이용가능성에 대하여 평가하였다. 다만, 현재 국민건강보험공단에서 직무, 직종, 직업을 유추할 수 있는 변수가 없어 직종, 직무분석과 관련된 코호트는 구축하기 어려울 것으로 판단된다.

3. 연구 활용방안

연구 결과, 본래 목적이 직업병 감시를 위한 것이 아닌 청구자료로서 구축된 데이터이므로 제한점과 한계점이 분명히 다수 존재하지만 탐색적 연구로서의 가치는 있다고 판단된다. 예를 들어 소규모사업장의 건강실태 등을 파악하여 근로자건강센터를 보다 효율적으로 활용이 가능해질 것으로 기대된다. 더불어 직업관련질환의 감시체계로서 관심 직업군의 건강영향 및 질병 발생을 모니터링 할 수 있고 아직까지 밝혀지지 않은 건강영향을 발견하는 데 이용가능성이 있다고 판단된다. 그리고 향후 누출사고 이후 노출 근로자 및 주변 주민의 건강영향을 모니터링하는 데 있어서도 빅데이터 기반 직업관련질환 코호트의 활용가

능성을 생각해볼 수 있다.

4. 본 연구의 제한점

본 연구는 이차 자료를 이용한 후향적 코호트이며, 연구의 범위가 기존에 조사된 변수에 제한된다는 단점, 원하는 변수(예: 직종)가 없을 수 있다는 제한점, 부적절한 조사결과가 포함되어 있을 가능성이 있다. 또한 2015년 사업장 정보를 기준으로 하였으므로 2015년 이전 폐업한 사업장과 해당 사업장 근로자가 누락되었다는 점, 산업군별 근로자의 지역분포에 차이가 있을 수 있음에도 불구하고 공무원과 사립학교교직원이라는 단일 대조군과 비교했다는 점이 한계점으로 있어 해석 시 세밀한 주의가 필요하다. 또한 앞서 업종분류코드가 표준산업분류와 유사하나 동일하지는 않고, 분류가 불명확하여 업종코드가 충분한 동질성도, 배타성도 확보하지 못하는 데에서 오는 한계점, 직종 및 직무 정보의 부재, 대조군 내 이질성, 그리고 대조군이 다른 직업군에 비해 갖는 독특한 입사와 퇴사의 성격 등이 제한점으로 작용하며, 진단명에 있어서도 입원 시 진단명을 사용하기는 했지만 진단의 불확실성을 투약·수술·처치 정보로 확인하지 못했다는 점 역시 제한점으로 작용한다.

따라서 본 연구의 결과를 해석을 탐색적 연구의 의미 이상으로 확대하여 해석하는 것은 과학적 오류를 범하게 되므로 옳지 못하지만, 이러한 제한점 및 한계점이 보완된다면 가치가 있다고 생각된다.

4. 연락처

연구책임자: 연세대학교 의과대학 교수 윤진하

연구상대역: 산업안전보건연구원 직업건강연구실 안정호

■ ☎ 052) 7030.878

■ E-mail: Jhoahn@kosha.or.kr

차 례

I. 서론 1

1. 연구목적 및 필요성	1
1) 직업병 연구의 배경 및 필요성	1
2. 연구목표	4
1) 빅데이터 기반 직업군 관련 코호트 설계	4
2) 우선순위 질병 선정 및 취약군 발굴	4
3) 근로자 질병 예방을 위한 코호트 활용 방안 제언	4
3. 기대효과 및 활용방안	5
1) 기대효과	5
2) 활용방안	6

II. 연구방법 7

1. 빅데이터 기반 직업군 관련 코호트 설계	7
1) 빅데이터 자료의 특성 분석	7
2. 우선순위 질병 선정 및 취약군 발굴	14
1) 우선순위 질병 선정	14
2) 취약집단의 발굴	17

3) 질병 선정 및 취약군 발굴의 전제조건	29
3. 근로자 질병 예방을 위한 코호트 활용방안 제언	31
1) 시범 운용을 통한 코호트 활용 방안	31
2) 직업 관련 코호트 시스템 구축 설계	32

III. 연구결과 34

1. 빅데이터 기반 직업군 관련 코호트 설계	34
1) 국민건강보험공단 빅데이터 자료의 특성 검토	34
2) 빅데이터 기반 구축 코호트의 특징	53
2. 우선순위 질병 선정 및 취약군 발굴	60
1) 질병 선정 및 취약군 발굴의 전제조건	60
2) 질환에 대한 고찰 및 우선순위 질병 탐색	60
3) 직업·직종·산업 분류에 따른 취약집단의 발굴	104
4) 사업장 특성에 따른 취약집단 선정	109
3. 근로자 질병 예방을 위한 코호트 구축 및 활용 제언	110
1) 시범운용을 통한 코호트 활용방안	110
2) 직업관련 코호트 시스템 구축 및 관리방안 설계	134
4. 직업관련 코호트의 향후 이용방안	136
1) 근로자건강센터의 효율적 활용	136
2) 직업관련질환 감시체계로서의 기능	137
3) 누출사고 이후 모니터링	137
5. 현 분석 제한점 및 적용의 한계	138
1) 후향적 코호트의 제한점	138

2) 업종코드 명확성	139
3) 직종 · 직무정보의 부재	141
4) 대조군의 불명확성	142
5) 진단 질병의 불명확성	143
6) 제한점에 대한 소결	144
 IV. 참고문헌	145
 V. 영문요약문	147
 VI. 부록 및 부표	150

표 차 례

<표 2-1> 국민건강보험공단 데이터 구성 내역	8
<표 2-2> 국민건강보험공단 정관 제 44조.....	9
<표 2-3> 국민건강보험공단 빅데이터 내 분석 가능 변수	10
<표 2-4> 한국표준직업분류.....	18
<표 2-5> 표준산업분류 중 대분류.....	21
<표 2-6> 직장가입자 산업 중분류별 정신질환 교차비 (남성, vs. 교육행 정).....	23
<표 2-7> 표준산업분류 중 대분류 “전기, 가스, 증기 및 수도사업” 의 하위 분류체계 및 타 분류체계와의 비교.....	27
<표 3-1> 국민건강보험공단 빅데이터 내 분석 가능 변수	34
<표 3-2> 국민건강보험공단 빅데이터 내 주소 관련 변수.....	39
<표 3-3> 국민건강보험공단 빅데이터 내 가입자와의 관계 변수	40
<표 3-4> 국민보험공단과 표준산업분류의 산업분류 비교.....	41
<표 3-5> 건보공단 업종분류와 표준산업분류 중 상이한 업종에 관한 연 구진 의견	43
<표 3-6> 국민건강보험공단 빅데이터 내 직역상세코드.....	49
<표 3-7> 업종분류코드 이용한 자동차 수리업 종사자 코드.....	57
<표 3-8> 공무원 및 교직원의 코호트 등록 현황.....	58
<표 3-9> 건강보험공단 데이터 내 상병코드의 가장 앞 1자리 알파벳.....	63
<표 3-10> ICD 상위 3자리(알파벳 1자리+숫자 2자리) 분류의 주요 질환	68
<표 3-11> 2002-2015년 남성 근로자의 다빈도 입원 상위 20개 상병...	81
<표 3-12> 2002-2015년 남성 근로자 1,000명 당 입원일수 상위 20개 상 병.....	84

〈표 3-13〉 2002-2015년 입원을 경험한 남성 근로자 1명 당 평균 입원일수 상위 20개 상병	88
〈표 3-14〉 2002-2015년 여성 근로자의 다빈도 입원 상위 20개 상병 ...	92
〈표 3-15〉 2002-2015년 여성 근로자 1,000명 당 입원일수 상위 20개 상병	96
〈표 3-16〉 2002-2015년 입원을 경험한 여성 근로자 1명 당 평균 입원일수 상위 20개 상병	100
〈표 3-17〉 2002-2015년 다빈도상병 평균 순위 상위질환	104
〈표 3-18〉 2002-2015년 산업분류별 우선순위 10개 질환의 표준화발생비, 남성	105
〈표 3-19〉 2002-2015년 산업분류별 우선순위 10개 질환의 표준화발생비, 여성	107
〈표 3-20〉 분진노출근로자의 소화기계암	111
〈표 3-21〉 병·의원 종사자의 각 질환별 표준화발생비	115
〈표 3-22〉 육상운수업 종사자의 표준화발생비	117
〈표 3-23〉 반도체 근로자의 건강영향	120
〈표 3-24〉 디스플레이 근로자의 건강영향	122
〈표 3-25〉 타이어업종 근로자의 건강영향 (암)	124
〈표 3-26〉 타이어업종 근로자의 건강영향 (뇌심혈관질환)	125
〈표 3-27〉 석유화학단지 근로자의 건강영향	127
〈표 3-28〉 암모니아 노출 관련 추정 질환의 ICD-10 코드	130
〈표 3-29〉 업종분류코드에 따른 사업장 선정 예시	140

그림 차례

[그림 2-1] 국민건강보험공단 데이터의 수집체계	7
[그림 2-2] 빅데이터 기반 직업코호트의 구축 형태	12
[그림 2-3] 국민건강보험공단 빅데이터를 활용한 코호트 체계	13
[그림 2-4] 우선순위 질병 선정 기준	14
[그림 2-5] 우선순위 질병 선정 방법	15
[그림 2-6] 직종별 10만 명당 자살 사망률 변화(남성)	19
[그림 2-7] 2015년 근로환경조사 업종별 산업재해 재해율	20
[그림 2-8] 빅데이터 기반 직업관련 코호트 시스템 구축 및 관련기관 연계	33
[그림 3-1] 2002-2015년 전체 직장가입자 수	50
[그림 3-2] 2002-2015년 직장가입자 중 총 입원일 수	50
[그림 3-3] 2002-2015년 직장가입자 1000명 당 입원일수	51
[그림 3-4] 2002-2015년 직장가입자 중 총 입원자 수	51
[그림 3-5] 2002-2015년 직장가입자 10만 명 당 입원 근로자 수	52
[그림 3-6] 2002-2015년 입원근로자 1명 당 평균 입원일수	53
[그림 3-7] 국민건강보험공단 빅데이터를 활용한 코호트 체계	54
[그림 3-8] 사업장 이름을 이용한 텍스트 분석	55
[그림 3-9] 업종분류코드 이용한 연구의 분류체계에 따른 특징	56
[그림 3-10] 빅데이터 기반 직업코호트의 구축 모식도	59
[그림 3-11] 우선순위 질병 선정 기준	61
[그림 3-12] 우선순위 질병 선정 방법	62
[그림 3-13] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 십만 명 당 환자 수	65
[그림 3-14] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 천 명당	

입원일 수	65
[그림 3-15] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 중 환자 당 입원일 수	66
[그림 3-16] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 십만 명 당 환자 수.....	66
[그림 3-17] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 천 명당 입원일 수.....	67
[그림 3-18] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 중 환자 당 입원일 수.....	67
[그림 3-19] 의료기관 추출 과정	113
[그림 3-20] WHO 병·의원 종사자의 직업성 위험요인.....	114
[그림 3-21] 분석에 사용된 석유화학단지.....	126
[그림 3-22] 암모니아에 의한 유발 증상.....	129
[그림 3-23] 각 집단별 노출시점(분기)을 기준으로 본 남성에서의 급성 인두염 외래이용률.....	131
[그림 3-24] 불산 고노출지역 남성 전체 표준화외래이용비	132
[그림 3-25] 불산 고노출지역 여성 전체 표준화외래이용비.....	132
[그림 3-26] 불산 중등도 노출지역 남성 전체 표준화외래이용비	133
[그림 3-27] 빅데이터 기반 직업관련 코호트 시스템 구축 및 관련기관 연계	135

I. 서론

1. 연구목적 및 필요성

1) 직업병 연구의 배경 및 필요성

국제노동기구(International Labour Organization, ILO)에 의하면 직업병이란 어떤 특정 직업에 종사하는 과정에서 위험요인에 노출되어 발생한 모든 질환을 의미한다. 직업병의 정의에는 두 가지 선행조건이 충족되어야 하는데, 하나는 특정 직업환경에의 노출 또는 노동활동과 질환 사이의 인과관계가 성립되어야 한다는 것이고, 다른 하나는 위험 요소에 노출된 사람들이 다른 인구집단에 비해 그 질병에 이환될 확률이 더 높아야 한다는 것이다.

이러한 직업병의 개념은 단계적으로 변화하고 있다. 과거의 직업병은 환경적 요인과 질병 발생간의 관계가 명확한 것만을 직업병이라고 했으나, 지금은 그 범위가 넓어져 환경적 요인이 질병 발생의 한 요인이 되거나 질병발생을 촉발시키는 경우도 직업병의 개념에 포함하고 있다.

최근에는 직업병의 발생과 원인 사이에 과학적인 인과관계가 명확하게 입증되지 않더라도 사회적으로 추론이 가능한 경우까지도 직업병의 개념으로 포함하는 등 그 의미가 점차 확대되고 있다.

이러한 흐름에 발맞추어 우리나라에서도 진폐증, 소음성난청, 화학물질 및 중금속 중독 등 전통적인 직업병의 범주를 넘어서서 서비스산업 종사자의 정신질환, 운수업 종사자의 비노기계 질환, 디스플레이 생산 산업 종사자의 조혈기계 질환 등 직업병 연구의 폭이 더 다양해지고 세분화되고 있다.

특정 질병 별 취약 직업군이 보고됨과 동시에 직업병 관리에 있어서도 더 이상 사후관리가 아닌, 능동적인 직업병 예방사업에 대한 사회적 요구도가 높아졌다.

따라서 이러한 흐름에 발맞추어 특정 질병에 대한 취약 집단(취약 직업군, 취약 사업장)을 과학적 근거를 기반으로 사전에 찾아내어 예방하는 적극적인 직업병 예방을 위한 노력이 필요하며, 이를 위한 정책적 접근이 요구되는 실정이다.

직업관련 코호트는 위험요인 노출과 질병발생 사이의 인과관계에 대하여 규명하기에 적합한 연구방법이다. 우리나라의 직업 코호트 중 대규모 코호트는 대부분 2000년대 초반을 전후하여 구축된 것으로 비교적 최근에 구축되었다. 직업 코호트 중 다수는 직업병의 집단 발생이 계기가 되었으며 기업의 인사자료, 암 등록, 사망통계 등으로부터 정보를 수집하였다.

현재의 직업 코호트는 정량적 노출평가 및 혼란변수의 통제가 미흡한 편이며, 연구 설계 및 연구를 위한 재정이 충분치 못한 상태에서 구축된 것이 많으며 직업병에 있어 인과관계보다는 관련성을 보이고 있다는 점에서 질적 한계점을 가진다.

대규모 직업병 발생이나 치명적인 질환이 발생하기 이전 그 위험요인과 취약군을 발견하여 능동적인 예방정책을 펼 수 있을 뿐만 아니라 이전에 알려지지 않았던 취약군과 위험요인을 발견할 수 있다는 점에서 국가 전체를 반영할 수 있는 코호트 구축은 필수적이라고 판단된다.

따라서 직업병 발생이나 사건이 발생한 이후 코호트를 구축하여 연구하는 것도 중요하지만, 근로자의 질병 예방 정책을 효과적으로 시행하고 우선 관리대상을 발견하기 위해서는 국가 전체를 포괄하는 기초자료를 바탕으로 한 코호트 구축이 필요하다.

전국민을 포괄할 수 있는 정보를 구축하기 위해서는 전국민의 의료이용 정보가 있는 건강보험공단의 빅데이터 자료 등을 활용하는 것이 그 방법이 될 수 있다. 이를 이용할 경우 암과 같은 중증 질환 뿐 아니라 경증 질환의 단계에서도 파악이 가능하므로 좀 더 다양한 건강상태에 대한 파악이 가능하며 중증 질환으로 이환되기 전 단계에 대해서도 연구가 가능하다. 이러한 다양한 연구를 통해서 얻은 결과는 직장생활 뿐만 아니라 일상생활에서도 근로자의 삶의 질 향상을 기대

할 수 있을 것으로 생각된다.

또한, 빅데이터를 기반으로 한 직업코호트 구축 기반을 마련하는 것이 중요할 뿐만 아니라 직업관련 코호트 구축 이후에 이를 활용하기 위한 방안도 고민할 필요가 있다. 따라서 직업 코호트를 이용한 연구 프로토콜을 제시하는 등 체계와 관련된 구체적 연구가 필요하다.

2. 연구 목표

1) 빅데이터 기반 직업군 관련 코호트 설계

- (1) 국민건강보험공단 빅데이터의 특성 검토
- (2) 구축될 코호트의 특징 소개

2) 우선순위 질병 선정 및 취약군 발굴

- (1) 질병의 유병률(발생률)이나 그 중증도에 대한 고찰
- (2) 전문가 자문, 기존에 조사된 통계자료 분석
- (3) 면담조사 등을 통한 우선순위 질병 탐색
- (4) 직업·직종 분류에 따른 취약집단 발굴
- (5) 산업 분류에 따른 취약집단 발굴
- (6) 사업장 특성에 따른 취약집단 선정

3) 근로자 질병 예방을 위한 코호트 활용 방안 제언

- (1) 직업 및 산업분류와 직업관련 질환의 연관분석
- (2) 사업장별 직업관련 질환 분석 및 맞춤형 건강관리
- (3) 유해물질 및 화학사고 이후 추적조사

3. 기대효과 및 활용방안

1) 기대효과

(1) 직업인 맞춤 건강관리 방안 마련

- 가) 높은 유병률 및 중증도를 가진 직업관련 질환에 대한 발견 및 취약집단 발굴
- 나) 직업, 산업, 또는 사업장별 근로자의 건강관리를 할 수 있는 과학적이고 객관적인 기반 근거를 마련

(2) 세계 최고 수준의 직업환경 코호트 구축

- 가) 전국민 데이터를 기반으로 한 코호트 구축함으로써 각 산업 또는 사업장 종사자의 전수조사를 가능하게 함.
- 나) 관찰인년을 통한 효과적인 연관성 또는 인과관계 도출

(3) 각 관련 단체의 정보 활용도 극대화

- 가) 정부 정책기관 뿐만 아니라 실제 일선에서 근로자의 건강관리를 담당하는 사업장 별 보건관리자, 근로자건강센터, 그리고 보건관리대행기관 등에서 직업관련 코호트를 이용한 구결과를 바탕으로 근로자건강관리가 보다 합리적이고 실질적으로 이루어질 수 있도록 함.

(4) 국민 참여 및 활용 데이터로 정보의 쌍방향화 가속

- 가) 건강보험 가입 유형에 따라 매년 새롭게 유입되는 근로자 및 퇴사자에 대한 추적관찰
- 나) 국민의 관심이 높은 유해물질 또는 질병에 대한 연구 가능

2) 활용방안

연구성과 활용구분 (해당항목에 (V) 표시)	1. 논문게재 및 발표 ☒ 2. 특허 출원·등록 등 ☐ 3. 후속사업추진 ☒ 4. 사업계획 반영 ☒ 5. 법·제도 등 반영 ☐ 6. 기타목적활용(교육, 연구) ☒ 7. 단행본 발간 ☐ 8. 일간지 등 보도 및 기사 게재 ☐ 9. 기타 ☐
연구성과 활용계획	1. 빅데이터를 기반으로 한 직업군 관련 코호트 설계 2. 연구 우선순위 직업관련 질환 및 취약 직업군 발굴 3. 코호트 시범 운용을 통한 활용방안 제언 4. 효율적 코호트 활용을 위해 필요한 제반사항에 대한 검토

‘국민건강정보 데이터’의 구성은 다음과 같다.

- 자격 데이터: 자격 데이터는 나이, 거주지 및 국민건강보험 관련 자격, 보험료, 소득분위, 가족관계 등과 같은 자료에 관한 자료로 자료의 규모는 데이터 구축 기준 연도인 2002년 기준 47,851,928명 (남성 23,993,697명, 여성 23,858,231명)으로 구축되어 매년 업데이트 되고 있다.
- 진료 데이터: 진료 데이터는 건강보험 가입자들의 진료내역 및 명세서, 환자정보와 요양기관 자료 등으로 구성되어 있다. 즉, 건강보험 가입자들에 대한 진료 건 단위의 자료로 개인의 해당 연도 진료기록을 모두 포함하기 때문에 그 규모가 매우 방대하다. 진료 데이터는 구체적으로 전산진료 청구명세서, 진료명세서 진료내역, 수진자 상병내역 등으로 구성되어 있다.
- 검진 데이터: 검진 데이터는 2002년부터 국민건강보험공단에서 국민들을 대상으로 실시한 일반검진 1차 및 2차, 생애전환기 건강진단 1차 및 2차, 암 검진 및 영·유아 검진 자료 등으로 구성된 데이터이다. 그 외 의료급여 데이터와 노인 장기요양 데이터가 있다.

<표 2-1> 국민건강보험공단 데이터 구성 내역

지격 및 보험료 데이터	• 자격 데이터 : 생년월일, 성별, 나이, 거주지역, 사업자/가입자, 해외출입 국자료 등 • 보험료 데이터 : 종합소득, 연금소득, 전월세, 재산세 • 자동차세 등 과세자료, 보수월액, 국가유공자 명부, 장애인 등록자료(장애등급, 장애유형 등)
진료내역데이터	• 명세서일반 : 요양기관, 상병명, 내원일수, 요양개시일, 진료과목 등 • 진료내역 : 병원 내의 처치 및 수술, 원내 약처방, 투여량, 진료비 등 • 수진자 상병내역: 부상병을 추가한 상세 • 처방전교부 상세내역 : 원외 처방약, 투약량, 투여일수, 약물정보 등
건강검진 데이터	• 일반건강검진, 생애전환기 건강진단, 5대 암 검진(위암/대장암/폐암/간암/유방암), 영유아 건강검진, 구강검진 • 문진자료(생활습관, 가족력, 기왕력 등) • 검진결과 실측 데이터
의료급여 데이터	• 생년월일, 성별, 나이, 보장기관(시도·시군구), 의료급여 유형 및 거주 지역 등 자격 자료 • 요양기관, 상병명, 내원일수, 요양일수, 요양개시일, 진료비 등 명세서 내역
노인장기요양 데이터	• 장기요양 신청내역, 노인상태상 등 인정조사자료, 등급판정자료 등 • 희망급여종류, 시설·재가급여, 특별현금급여, 급여제공기록지 등 • 의사소견서, 장기요양급여 청구자료, 심사자료, 기관 평가자료 등

국민건강보험공단은 국가, 지방자치단체, 요양기관, 보험료율 산출기관, 공공기관, 그리고 그 밖의 공공단체 등에 대하여 주민등록·가족관계등록·국세·지방세·토지·건물·출입국관리 등의 자료를 제공하도록 요청할 수 있으며, 이러한 권리는 국민건강보험법 제 96조에 명시되어 있다. 국민건강보험공단에서는 타 기관에 요청하여 얻는 자료들에 대하여 다음과 같이 국민건강보험공단 정관 제 44조에 자세히 고시하고 있다.

<표 2-2> 국민건강보험공단 정관 제 44조

1. 직장가입자의 자격취득·변동·상실 및 피부양자에 관한 사항
2. 지역가입자의 자격취득·변동·상실 및 거주지 변경에 관한 주민등록자료
3. 건강보험급여비용 및 장기요양급여비용의 지급에 관한 사항
4. 건강보험급여 및 장기요양급여의 관리에 관한 사항. 다만, 요양기관 및 장기요양기관에 대하여는 다음 각 목의 어느 하나에 해당하는 경우에 한한다.
 - 가. 건강보험급여 및 장기요양급여의 제한에 관한 사항
 - 나. 건강보험급여 및 장기요양급여의 정지에 관한 사항
 - 다. 부담이득의 징수에 관한 사항
 - 라. 구상권 행사에 관한 사항
 - 마. 가입자와 피부양자가 공단에 신고한 수진내역의 확인에 관한 사항
5. 가입자 및 피부양자의 건강검진에 관한 사항
6. 건강보험료 및 장기요양보험료 부과에 필요한 보수·소득·재산 등에 관한 자료
7. 건강보험료, 장기요양보험료, 법·「노인장기요양보험법」에 따른 그 밖의 징수금 및 징수위탁보험료등의 징수에 관한 사항
8. 장기요양급여의 제공내용 확인, 장기요양급여의 관리·평가 및 장기요양보험료의 산정 등 장기요양사업수행에 필요한 사항
9. 그 밖에 이사장이 건강보험사업, 장기요양사업 및 징수위탁근거법에 따라 위탁받은 사업에 필요하다고 인정하여 요청하는 사항

건강보험 수진자료 상 보험의 유형에 따라 직장가입자와 비직장가입자로 구분할 수 있는데, 2014년 건강보험 수진자료상 남성인 직장가입자는 8,611,742명, 비직장가입자는 10,640,890명이었으며, 여성 직장가입자는 5,387,599명, 비직장가입자는 13,023,794명 이었다.

또한 지역가입자로 분류할 때에는 거주지를 기준으로 분류할 수 있으며 이때 국민건강보험공단에서 가지고 있는 거주지 자료는 주민등록표에 기록된 내용과

일치한다. 2017년 현재 주민등록표에 기입되는 주소는 도로명주소법 시행령 제 3조에 따라 기재되며 특별시·광역시·특별자치시·도·특별자치도의 이름, 시·군·자치구의 이름, 행정구(자치구가 아닌 구)·읍·면의 이름, 도로명, 건물번호, 상세주소(동·층·호 순)의 순서로 작성되어 있다. 이러한 거주지·지역 관련 정보 외에 국민건강보험공단 빅데이터에서 분석 가능한 변수를 표로 정리하면 다음과 같다.

<표 2-3> 국민건강보험공단 빅데이터 내 분석 가능 변수

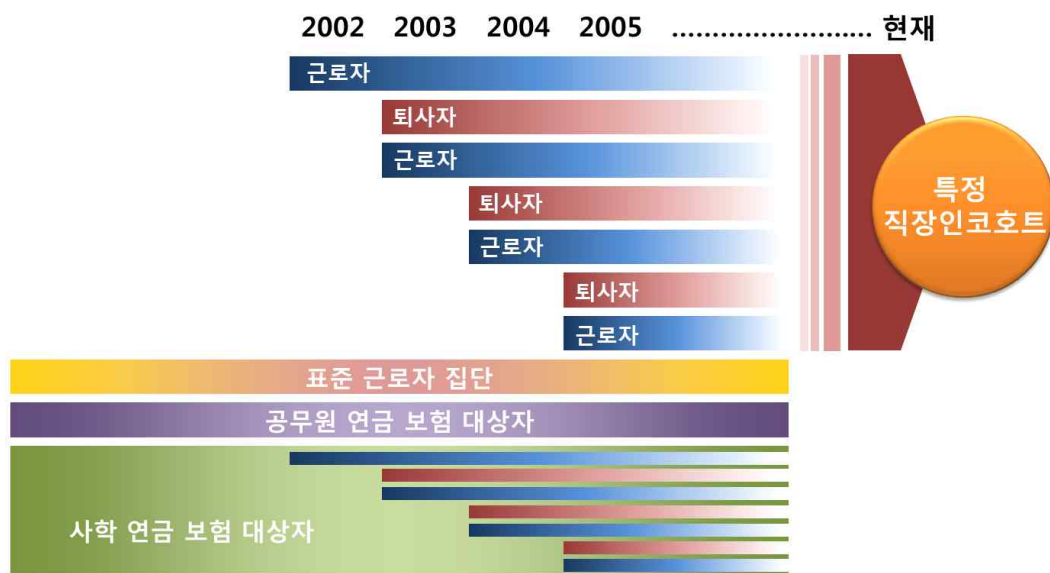
인구사회학적 요인	성	남성, 여성	
	연령	수진시점, 연말시점	
	출생	출생일자	
	사망	사망일자	
	장애	15개 장애유형, 1 ~6개의 장애등급, 장애등록일자	
	지역	수진자의 읍면동 단위, 요양기관의 읍면동 단위, 사업장의 읍면동 단위	
	보험종류	직장·지역·의료급여, 가입자·피부양자	
	보험료	보험료 납부금액, 산정기준변수 보험료 20분위, 피부양자 수 조정 보험료	
사업장	직종·업종(대분류, 세분류), 규모		
요양기관 및 검진기관 요인	요양 기관	인력현 황	요양기관인원현황, 요양기관관리의사현황 진료과목별 전문의 현황, 물리치료 인력
		시설현 황	설립구분, 요양기관종별 진료표방과목, 병실 및 병상 수 산재지정 요양기관, 물리치료기관 일반현황
		장비	요양기관 장비 상세내역
	검진 기관	인력 현황	검진기관기본현황, 검진기관위탁항목 검진기관검진유형
		시설 현황	검진기관인력, 검진기관인력교육 출장검진인력

	장비현 황	검진기관 장비확인
건강행태 및 LAB DATA	흡연	현재 흡연여부, 흡연량 · 기간
	음주	현재 음주여부, 음주량 · 기간
	신체활동	걷기, 중등도, 고강도
	과거력	고혈압, 당뇨, 이상지질혈증, 심 · 뇌혈관질환, 암의 문진 및 진료내역
	가족력	고혈압, 당뇨, 이상지질혈증, 심 · 뇌혈관질환, 암의 문 진
	인지기능	KDSQ
	우울감	
	ADL	
	신체계측	키, 몸무게, BMI, 허리둘레, 시력검사, 청력검사
	혈압	수축기혈압, 이완기혈압
	혈당	공복혈당
	지질농도	혈중 총 콜레스테롤, HDL/TG/LDL
	간 · 신기능	간기능검사(OT/PT), 신기능검사, B형간염(생애전환기)
	골밀도	골밀도검사(생애)
	암검사	내시경, 생검
질환 검사 및 치료행위	질환명	KCD6; 주상병 및 부상병
	의료이용	입원 및 외래, 응급실 경유, 입원일수, 내원일수, 처방 일수, 진료과목코드, 심결비용
	상세내역	행위수가, 치료재료, 약재 1회 투약량 및 1일 투약량, 총 투여일수/실시횟수

(2) 구축될 직업 관련 코호트의 설계 및 특징

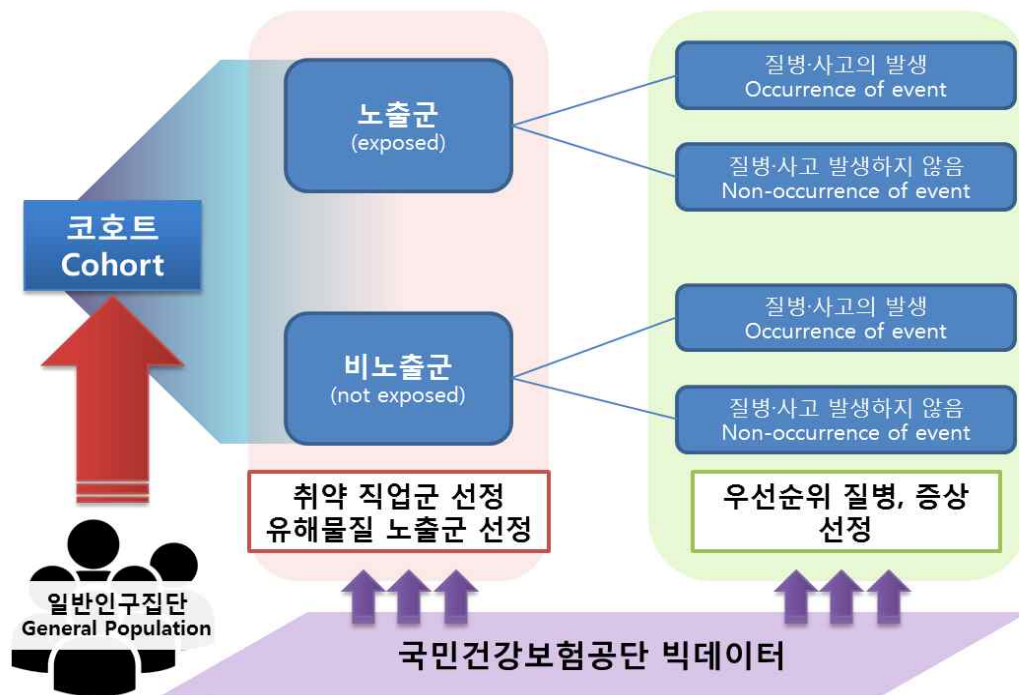
국민건강보험공단(이하 국민건강보험)의 데이터는 2002년 이후 매년 12월 31일을 기준으로 업데이트 되고 있다. 이 데이터에서는 매년 건강보험 수진자의 건강보험 가입 내역의

변화 등을 파악하는 과정을 통해 매년 새로운 근로자 및 퇴사자에 대한 정보를 파악할 수 있다. 따라서 이를 바탕으로 직업 관련 코호트를 구축할 경우 코호트 유입 시작 시점(Enroll 시점)이 다양한 역동적 코호트(Dynamic cohort)의 형태를 갖게 된다.



[그림 2-2] 빅데이터 기반 직업코호트의 구축 형태

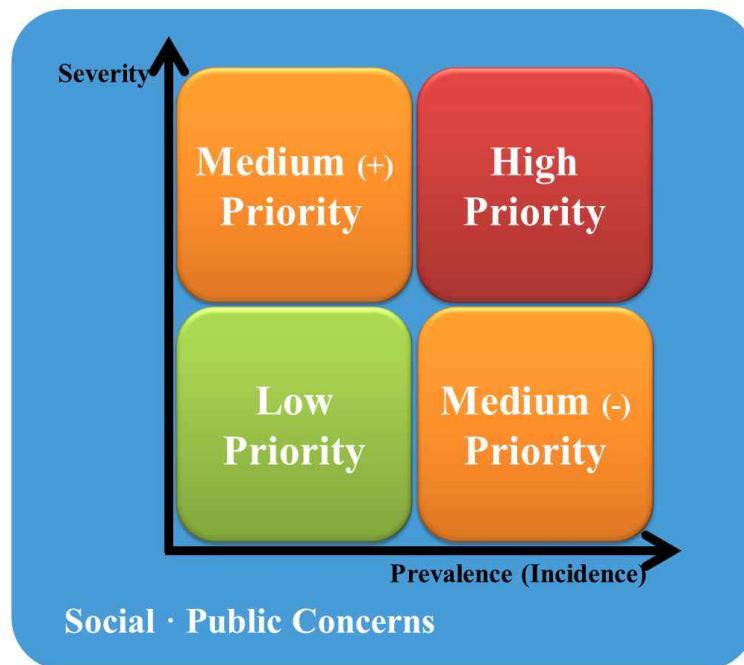
코호트를 구축하기 위해서는 연구하고자 하는 주제에 알맞은 위험요인의 선정 또는 질병의 발생을 먼저 선정할 필요가 있다. 연구하고자 하는 주제의 중심이 질병에 있는 경우 현재 질병의 유무에 따라 과거 노출기록을 살펴볼 수 있다. 또한 연구하고자 하는 주제의 중심이 유해환경 또는 유해물질 노출, 또는 특정 직업군에 있는 경우 노출 이후의 질병 또는 사고발생여부에 대하여 관찰 및 연구를 시행할 수 있다. 이처럼 연구 주제 설정에 필요한 질병의 선정 및 취약군 발굴 방법에 대해서 다음 항에서 제시하도록 하겠다.



[그림 2-3] 국민건강보험공단 빅데이터를 활용한 코호트 체계

2. 우선순위 질병 선정 및 취약군 발굴

1) 우선순위 질병 선정



[그림 2-4] 우선순위 질병 선정 기준

직업코호트를 바탕으로 연구할 수 있는 우선순위 질병을 선정하는 과정은 연구하고자 하는 직업군이나 사업장의 특성별로 다양할 수 있다. 먼저 우선순위 질병을 선정할 때는 질병의 유병률(발생률)이나 그 중증도에 대한 고찰이 필요하다. 일반적으로 질병의 중증도가 높고, 유병률이 높을 경우 그 질병에 대한 예방 및 관리 정책은 다른 질병에 비해 높은 우선순위를 갖는다고 볼 수 있다.

우선순위 질병을 선정하기 위한 조사방법으로는 문헌고찰, 전문가 자문, 기존에 조사된 통계자료 분석, 그리고 면담 및 설문조사가 있다. 각자의 조사방법에

따라 장점과 단점이 혼재하므로 연구 목적에 맞는 방법을 영리하게 선택하거나 연구 주제에 맞게 여러 조사방법을 실시하여 서로의 장·단점을 보완할 필요가 있다. 본 연구에서도 이러한 과정을 거쳐 시범 운영된 코호트를 설명하였다.



[그림 2-5] 우선순위 질병 선정 방법

(1) 전문가 자문

학회 및 연구기관에서 산업보건 및 근로자의 직업병에 대하여 연구해 온 전문가나 일선 사업장에서 근로자들의 보건관리를 해 온 전문가 등의 자문을 통하여

직업코호트 연구를 통해 그 인과성, 연관성이 입증될 필요가 있다고 생각되는 질병을 선정하는 방법이 있다. 본 연구에서도 현업에 종사하고 있는 다수의 직업환경의학과 전문의 및 교수를 연구진으로 하여 학문적으로 타당하고 실제적인 자문을 구하는 방식으로 연구대상 질병을 선정한 코호트가 있었다.

(2) 면담조사 또는 설문조사

실제 현장에서 근무하는 근로자들을 대상으로 구조적 설문지를 이용한 심층 면담조사를 진행하여 근로자의 근무환경 및 유해요인에 대하여 파악하고 건강문제에 대해 조사하는 방법을 이용할 수도 있다. 면담조사를 통해 실제 조사하는 근로자의 수는 적다하더라도, 얻을 수 있는 정보의 다양성과 깊이는 다른 연구방법에 비해 훨씬 깊다. 즉, 면담조사는 설문조사만큼 다양한 근로자들의 요구도를 반영하지는 못하지만, 설문조사에서 얻을 수 있는 정보보다 더 자세하고 실제적인 양질의 정보를 얻을 수 있다는 점이 장점이다.

또한, 특정 사업장이나 직종 종사 근로자들을 대상으로 설문조사를 실시하여 노출되는 유해물질의 종류나 공정에 대한 정보, 그리고 호발 증상이나 질병명에 대한 정보를 얻을 수 있다. 이 방법을 이용할 경우 질환으로 이환되기 이전 증상에 대해서도 파악을 할 수 있으며, 발생 빈도와 관련된 우선순위 질환뿐만 아니라 특정 산업 등에 종사하는 근로자들의 주관적 요구도 충족시킬 수 있는 질환의 선정이 가능하다는 점에서 장점을 갖지만 근로자의 기억에 의존해야 하므로 의학적 정보 등이 다소 부정확할 수 있다는 점은 한계점으로 남는다. 그리고 잘 설계된 설문조사의 경우 그 자체로도 한 연구의 주제가 될 수 있는 만큼 직업 관련 코호트를 구축하는 단계에서 설문조사를 실시하는 것은 시간적·경제적으로 비효율적인 방법이 될 수 있어 유의하여야 한다.

본 연구에서의 코호트는 다수의 연구진들이 각자의 기관에서 진행하였던 특수검진 경험과 기타 연구용역 수행과정에서의 면담, 설문조사 경험을 바탕으로 하여 연구진 회의를 통한 우선순위 질병에 대한 논의 과정을 거쳤다.

(3) 통계자료를 이용한 상병명 선정

기존에 구축되어 있는 자료를 이용하여 상병명을 선정할 수 있다. 작은 규모의 통계자료로는 사업장별 검진 자료 등을 바탕으로 사업장의 다빈도 발생 질환 등에 대하여 파악하는 방법이 있을 수 있다. 또한 산재근로자 관련 자료 등을 통하여 산재로 인정된 주 상병명 등을 바탕으로 우선순위 질병을 선정할 수도 있다. 여기서 사용할 수 있는 기존 구축 자료란 국민건강영양조사, 근로환경조사, 국민건강보험공단 표본코호트 등이 있다.

이 중 국민건강보험공단의 청구자료에 포함되어 있는 질병분류 기준은 국제 질병 분류 코드(International Statistical Classification of Diseases and Related Health Problems 10th Revision: ICD-10)를 기반으로 분류할 수 있는데, 이를 이용할 수 있는 경우 좀 더 표준화된, 객관적인 정보를 얻을 수 있다.

이미 수집된 데이터에서 질병의 유병률 또는 발생률의 빈도를 확인하는 방법을 통해 다빈도상병을 확인하는 방법이 있다. 이는 전 국민을 대상으로 한 자료에서 실제 진단받은 진단명을 확인할 수 있다는 점에서 근로자의 기억에 의존하는 방법보다 상대적으로 정확하다는 장점이 있지만 국내 의료의 특성상 보험 청구 등의 과정을 이유로 상병명에 따라서는 과다청구 혹은 과소청구의 가능성에 대해 주의할 필요가 있을 것으로 판단된다.

따라서 본 연구에서는 이러한 제한점을 다소 극복하기 위하여 상병명 사용 시 외래 상병명을 배제하고 ‘입원 상병명’을 중점으로 하여 분석을 진행한 코호트를 소개하였다. 입원 시 상병명은 실제 상병명과 청구 기록 상 상병명이 일치할 확률이 높다고 판단되며, 질병의 중증도를 고려할 때 외래진료로 치료가능한 질환보다 입원진료가 필요한 상병명이 그 중증도가 더 높다고 볼 수 있다고 기대되기 때문이다. 이러한 과정을 통해서 입원 시 상병명을 선택함으로써 상병명의 신뢰성을 높일 수 있을 것으로 생각된다.

2) 취약집단의 발굴

(1) 직업·직종 분류에 따른 취약집단 발굴

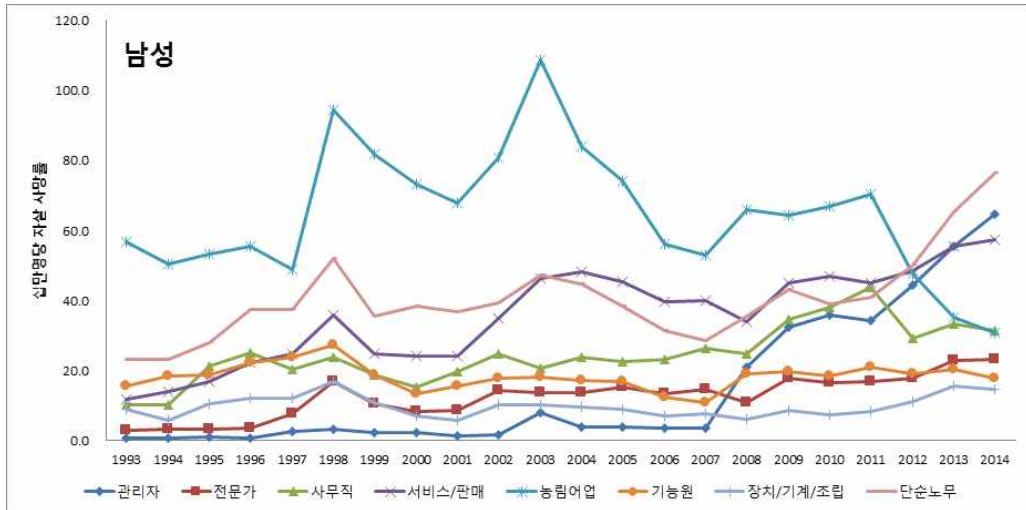
국내 직업을 분류하는 표준화된 기준으로는 한국표준직업분류가 있다. 이는 통계조사를 목적으로 수입을 위해 개인이 하고 있는 일(경제활동)을 그 수행되는 일의 형태에 따라 체계적으로 분류한 것이다. 한국표준직업분류는 여러 번의 개정절차를 걸쳐 현재 6차 개정안을 사용하고 있으며, 대분류에서는 관리자, 전문가 및 관련종사자, 사무종사자, 서비스종사자, 농림어업 숙련 종사자, 기능원 및 관련 기능 종사자, 장치·기계 조작 및 조립 종사자, 단순노무종사자, 그리고 군인으로 이루어져 있다. 그리고 해당 대분류 아래에 중분류와 세분류, 그리고 세세분류로 나뉘어 직업을 분류하고 있다.

<표 2-4> 한국표준직업분류

대분류	중분류	소분류	세분류	세세분류
1 관리자	5	15	24	77
2 전문가 및 관련 종사자	8	41	153	445
3 사무 종사자	4	9	26	57
4 서비스 종사자	4	10	33	73
5 판매 종사자	3	4	13	38
6 농림어업 숙련 종사자	3	5	12	29
7 기능원 및 관련 기능 종사자	9	20	73	201
8 장치·기계조작 및 조립종사자	9	31	65	235
9 단순노무 종사자	6	12	24	48
A 군인	1	2	3	3
	52	149	426	1,206

한국표준직업분류를 사용한 통계자료의 예로는 통계청 사망원인통계를 예로 들 수 있다. 이를 이용한 연구로 예를 들면 윤진하 등 (2016)의 직장인 정신 건강 증진 방안 연구에서 직종별 자살 사망률 변화를 보고한 바 있었다. 이에 따르면 단순노무직, 관리자, 서비스/판매직은 최근 자살 사망률이 지속적으로 증가하는

양상을 보였으며 농림어업 종사자는 반대로 자살 사망률이 감소하는 양상을 보였다.



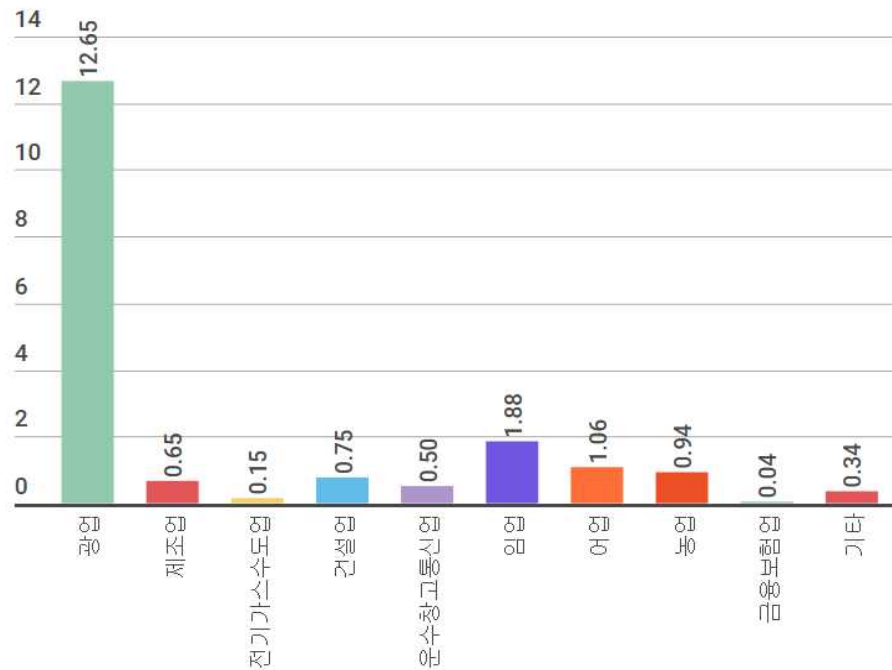
[그림 2-6] 직종별 10만 명당 자살 사망률 변화(남성)

그러나 이러한 한국표준직업분류가 있음에도 불구하고, 몇몇 이차자료에서는 한국표준직업분류에 의한 직업 조사가 이루어지지 않아 데이터 분석이 불가능하였다. 예를 들어 국민건강보험공단의 건강보험 수진자료에서도 산업별 종사자는 구분할 수 있으나 한국표준직업분류에 따른 직업별 종사자는 구분할 수 없었다. 근로자의 근무환경을 결정짓는 데에는 어떤 산업에 종사하느냐도 중요하지만 종사하고 있는 직무의 특성 역시 중요한 요소이므로 직업 또는 직무에 대한 정보에 제한점이 있다는 것은 본 데이터를 기반으로 한 코호트의 단점이 될 수 있다.

(2) 산업 분류에 따른 취약집단 발굴

취약집단을 발굴하기 위해서는 직업분류 외에도 종사하는 산업에 따라 분류하는 방법을 선택할 수도 있다. 어떤 산업에 종사하는지에 따라 근무환경은 서로 크게 달라질 수 있기 때문이다. 실제로도 종사 산업 종류에 따라 재해율 등에서 차이가 나타나기도 한다. 2015년 근로환경조사 결과에 따르면 전체 90,129명의 재해자가 조사되었는데 그 중 산업별로 재해율을 구한 결과 광업의 재해율이

12.65%로 가장 높았으며 임업 종사자 1.88%, 어업 종사자 1.06% 순으로 조사되었다.



[그림 2-7] 2015년 근로환경조사 업종별 산업재해 재해율

산업을 분류하는 기준으로는 한국표준산업분류가 개발되어 있다. 이는 산업관련 통계자료의 정확성, 비교성을 확보하기 위하여 작성된 것으로서 1963년 제정되기 시작하였으며 그 기초는 유엔의 국제표준산업분류(1차 개정, 1958년)에 있다. 한국표준산업분류의 대분류로는 현재 총 21가지가 있으며 아래 <표 2-5>와 같다.

<표 2-5> 표준산업분류 중 대분류

A	농업, 임업 및 어업
B	광업
C	제조업
D	전기, 가스, 증기 및 수도사업
E	하수, 폐기물 처리, 원료재생 및 환경복원업
F	건설업
G	도매 및 소매업
H	운수업
I	숙박 및 음식점업
J	출판, 영상, 방송통신 및 정보서비스업
K	금융 및 보험업
L	부동산 및 임대업
M	전문, 과학 및 기술 서비스업
N	사업시설관리 및 사업지원서비스업
O	공공 행정, 국방 및 사회보장 행정
P	교육 서비스업
Q	보건업 및 사회복지 서비스업
R	예술, 스포츠 및 여가관련 서비스업
S	협회 및 단체, 수리 및 기타 개인 서비스업
T	가구 내 고용활동 및 달리 분류되지 않은 자가소비 생산활동
U	국제 및 외국기관

윤진하 등 (2016)의 직장인 정신 건강 관련 연구에서 국민건강보험공단 표본 코호트를 이용하여 산업 중분류별 관심 정신질환의 위험군을 찾는 연구를 진행한 바 있다. 조사년도의 건강보험 직장가입자 중에서 다수를 차지하는 교육·행정 분야 종사자를 기준 집단으로 하여 건강보험의 소득 20분위와 연령을 보정한 정신질환의 교차비를 남·여 층화하여 분석하였다. 이 중 남성에 대한 연구결과를 예로 들면 기분장애에 있어서는 국제 및 기타 외국기관 종사자, 금속광업 중

사자 , 사회복지 서비스업 종사자가 교차비가 높게 나타났으며 불안장애에서는 국제 및 기타 외국기관, 사회복지 서비스업 종사자가 높은 교차비를 보이는 등 각 정신질환별로 위험도가 높은 산업군을 발견할 수 있었다.

<표 2-6> 직장가입자 산업 중분류별 정신질환 교차비 (남성, vs. 교육행정)

기분장애	알코올 남용	불안장애	외상후 스트레스장애	적응장애	수면장애	모든 정신질환
산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)
국제 및 기타 외국기관 1.76 (1.52-2.04)	광업 지원 서비스업 22.94 (3.06-171.89)	국제 및 기타 외국기관 1.57 (1.35-1.81)	폐기물 수집운반, 처리 및 원료재생업 3.28 (1.38-7.82)	광업 지원 서비스업 case 1 47.28 (6.31-354)	보건업 2.33 (2.25-2.42)	국제 및 기타 외국기관 1.70 (1.55-1.86)
금속광업 1.532 (1.06-2.22)	임업 1.77 (1.14-2.74)	사회복지 서비스업 1.12 (1.04-1.21)	농업 3.23 (1.36-7.70)	담배 제조업 case 3 5.97 (1.91-18.66)	영상·오디오 기록물 제작 및 배급업 1.47 (1.23-1.76)	사회복지 서비스업 1.21 (1.16-1.27)
사회복지 서비스업 1.27 (1.18-1.37)	숙박업 1.73 (1.39-2.21)		창작, 예술 및 여가관련 서비스업 2.82 (1.10-7.22)	국제 및 외국기관 2.304 (1.49-3.56)	의료용 물질 및 의약품 제조업 1.36 (1.19-1.54)	보건업 1.21 (1.18-1.24)

24...빅데이터 기반 직업 코호트 구축을 통한 질병 발생 연구

기분장애	알코올 남용	불안장애	외상후 스트레스장애	적응장애	수면장애	모든 정신질환
산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)	산업중분류 aOR (95% CI)
	수상 운송업 1.72 (1.35-2.21)		공공행정, 국방 및 사회보장 행정 2.48 (1.71-3.59)		사회복지 서비스업 1.16 (1.10-1.28)	의료용 물질 및 의약품 제조업 1.13 (1.04-1.22)
	폐기물 수집운반, 처리 및 원료재생업 1.72 (1.52-1.95)		- .		음식점 및 주점업 1.16 (1.10-1.23)	공공행정, 및 사회보장 행정 1.05 (1.03-1.07)
	스포츠 및 오락관련 서비스업 1.62 (1.26-2.09)		- .		소매업; 자동차 제외 1.04 (1.00-1.08)	

이러한 표준산업분류가 있음에도 불구하고, 정부 및 기관에서 구축된 빅데이터에는 서로 다른 산업분류코드를 사용하는 경우가 있다. 국민건강보험공단 데이터 역시 표준산업분류코드와는 다소 다른 코드로 직장 가입자가 종사하는 산업을 기재하고 있으며 산업재해근로자 데이터에서는 또 다른 산업분류 코드를 사용하고 있다.

예를 들어 생활용수 공급업에 종사하는 사람은 국민건강보험공단에서는 코드 번호 54101에 해당하며, 통계청의 표준산업분류코드에서는 36010으로, 산재근로자 데이터에서는 30003으로 조사된다. 이처럼 국가 데이터마다 산업분류가 코딩이 다른 것은 보다 복잡적이고 다양한 통계분석을 위한 데이터 연계에 한계점으로 작용하고 있다.

각 데이터 별로 직업 또는 종사 산업에 대한 정보 부재를 상호 보완하는 방법으로서는 데이터의 직업·직종·업무·사업장 관련 정보를 서로 연계하는 방법을 적용할 수 있지만, 이처럼 각 빅데이터마다 각기 나름의 코딩방식을 사용하는 상황에서는 무작정 데이터를 통합하는 것은 타당한 분석 방법이 아니다. 각자의 코딩방식을 통합하고 표준화하기 위해서는 코딩명에 따른 대략적인 분류 외에도 차이가 나는 부분에 대한 직접조사(방문·면담·전화조사 등)를 통하여 분류 방식의 신뢰도 및 일치도, 그리고 일관성에 대하여 점검할 필요가 있어 이 역시 시간적·경제적으로 많은 비용이 소모되는 일이다. 따라서 산업분류코드의 통합 및 표준화는 산업명과 산업내용으로 미봉책 수준에서 맞추는 것을 넘어서서 장기적 안목을 갖고 이를 객관적이고 일관성 있는 표준화 방안에 대한 연구가 필요할 것이다.

따라서 모든 자료를 아우를 수 있는 객관적이고 표준화된 자료가 부족하고, 직업에 대한 정보 역시 한계점이 있는 상황에서 직업 관련 코호트를 효과적으로 구축하고 취약군을 찾기 위해서는 연구에 착수하기 전 현재 산업분류를 어떻게 적용하고 분류해야 실제 근무환경을 최대한 구분할 수 있을지에 대한 고찰이 필요하다. 따라서 본 연구에서는 현 수준에서 설계할 수 있는 합리적인 산업분류를

찾기 위하여 다수의 직업환경의학과 전문의 등으로 이루어진 연구진 내부에서 충분한 논의 및 내부 자문 과정을 거쳐 서로 호환 가능한 분류표를 정리하는 과정을 거쳤으나, 향후 타당한 분류를 위해서는 이에 대한 체계적 연구가 필요하다.

<표 2-7> 표준산업분류 중 대분류 “전기, 가스, 증기 및 수도사업”의 하위 분류체계 및 타 분류체계와의 비교

표준산업분류						산재 코드	건강보험공단 분류코드	
	중분류 코드	소분류 코드	소분류 이름	산업코드	산업내용			
전기, 가스, 증기 및 공기조절 공급업	35	351	전기업	351	전기업	30001	50400 /60400	전기업
				3511	발전업		54011	발전업
				35111	원자력 발전업			
				35112	수력 발전업			
				35113	화력 발전업			
				35119	기타 발전업			
				3512	송전 및 배전업		54012	송전 및 배전업
				35120				
	352	가스 제조 및 배관공급업	352	가스 제조 및 배관공급업	21101,	50401	가스제조 및	

						30002		공급업
				3520			54020	가스제조 및 배관공급업
				35200				
		353	증기, 냉온수 및 공기조절 공급업	353	증기, 냉온수 및 공기조절 공급업	30003	50402 /54030	증기 및 온수공급업
				3530				
				35300				
수도사업	36	360	수도사업	360	수도사업		50403	수도사업
				3601	생활용수 공급업			
				36010				
				3602	산업용수 공급업	30003	54102	산업용수 공급업
				36020				

(3) 사업장 특성에 따른 취약집단 선정

직업분류나 산업분류 외에 사업장 특성에 따른 코호트 구축방법을 생각해 볼 수 있다. 예를 들어 사업장 이름을 텍스트 분석하여 산업분류에서도 분류하기 어려운 사업장 코호트 구축이 가능하며, 그 외 사업장별 직장가입자 수를 확인하여 사업장 규모를 확인할 수 있다.

이외에 근무환경과 유해물질 노출 여부는 산업 분류를 통해 동질성을 유추하는 방법이 대안이지만, 현재 자료로는 결과의 현실성이 부족하다. 다만, 추후 유해물질 노출 여부 파악이 어렵다는 단점은 특수건강검진과 사업장 작업환경 측정 자료가 융합될 경우 상당 부분 극복할 수 있다. 이에 대한 추가 논의가 진행 중이다.

3) 질병 선정 및 취약군 발굴의 전제조건

우선순위 질병 선정 과정에서 취약집단의 발굴에 이르기까지 국민건강보험공단의 데이터베이스 활용하여 다양한 정보를 얻을 수 있다. 다만, 전 국민의 국민건강보험공단 자료를 얻기 위해서는 각 연구기관 혹은 진행되는 각 연구마다 필요한 데이터 이용에 대한 비용을 지불하고 이용할 수 있는 권리를 획득하여야 하며, 이는 앞으로의 맞춤형 직업 관련 코호트를 구축하고 관리하며 그 연구결과를 발표하는 데 있어 필수적인 단계라 볼 수 있다.

이번 연구에서도 건강보험공단의 데이터를 사용하기 위한 권한을 얻는 과정을 거쳐야 하지만 재정적 한계 등으로 인하여 그리하지 못하였다. 그러나 코호트의 구축가능 여부에 대한 현실성을 확인, 검증하기 위해서는 분석 후 고찰이 반드시 필요하므로 시범적 운용을 수행하는 것은 필수적인 과정이다. 따라서 본 보고서에 실린 코호트 시범운영 및 활용방안은 대한직업환경의학회 빅데이터위원회와 국민건강보험공단 사이의 업무협약(MOU)을 통해 이루어진 연구결과 및 두 차례의 세미나(2017년 8월, 2017년 11월) 결과를 이용하여 실시한 고찰 내용을 제시하였다.

그리고 데이터 마련은 앞으로의 빅데이터 기반 직업 코호트를 구축하는 데 있어서는 반드시 필요한 과정이므로, 추후 연구원에서 빅데이터 기반 직업 코호트에 있어 중추적 역할을 수행하기 위해서는 충분한 논의를 거쳐 데이터 이용권한 획득을 위한 절차를 시행해야 할 것으로 판단된다.

3. 근로자 질병 예방을 위한 코호트 활용방안 제언

1) 시범 운용을 통한 코호트 활용 방안

빅데이터를 기반으로 한 직업 관련 코호트 구축은 수많은 유해요인 및 환경과 직업 관련 질환과 증상들의 관계를 규명하고 개선하기 위해서 단순히 일회성에 그치지 않고 계속해서 활용 가능한 시스템을 구축하는 것이 필요하다. 이러한 직업관련 코호트의 활용방안으로는 다음과 같은 예로 설명할 수 있다.

(1) 직업 및 산업 분류와 직업관련 질환의 연관분석

특정 유해물질이나 유해환경을 공유한다고 판단되는 직업분류 또는 산업분류가 있을 경우, 이러한 분류에 따라 직업 관련 질환 사이의 연관 분석을 시행할 수 있다. 예를 들어 광업 또는 탄광에서 근무한 경험이 있는 종사자는 다른 산업 종사 근로자와는 달리 특징적으로 결정형 유리규산이나 석면, 용접 흙, 분진 등에 노출되는 환경을 공유한다고 생각할 수 있다. 또 농업 또는 어업 종사자의 경우 다른 산업 종사자에 비해 자외선 노출이 더 많이 이루어진다고 판단할 수 있어 산업 분류별로 특징적인 근무환경이 있는 경우에는 산업분류와 직업관련 사이 연관 분석을 시행함으로써 그 사이의 연관성, 더 나아가서는 인과관계에 대하여 연구할 수 있다. 본 연구에서는 분진에 노출되는 환경을 전제할 수 있는 ‘진폐 진단 환자’ 및 ‘광업’종사자를 대상으로 코호트를 구축하여 소화기계 암에 대하여 진행한 연구를 설명하였다.

(2) 사업장 별 직업관련 질환 분석 및 맞춤형 건강관리

빅데이터를 이용한 직업관련 코호트 구축을 통해 학문적인 성취를 달성하는 것 외에도 실제 근로자들의 건강관리 및 근무환경 개선을 위한 기반자료로서 코호트를 이용할 수 있다. 먼저, 사업장 별 근로자의 의료이용을 분석하는 것을 통해 관리가 어려웠던 소규모사업장의 근로자 건강관리 및 질환 예방을 용이하게

할 수 있다. 소규모사업장의 경우 보다 큰 규모의 사업장에 비하여 상대적으로 열악한 근무환경에 노출되는 경우가 많으며 실제 재해율도 높게 나타나는 경향이 있다. 이러한 곳에 빅데이터를 이용할 경우 소규모사업장에 종사하는 근로자 하나하나에 대해서도 조사가 가능하므로 근로자 건강관리에 있어 사각지대를 줄일 수 있을 것으로 기대된다. 또한 근로자 개개인에 대한 파악을 통하여 각 사업장별로 알맞은 맞춤형 건강관리 방안을 제시해줄 수 있다.

(3) 유해물질 및 화학사고 이후 추적조사

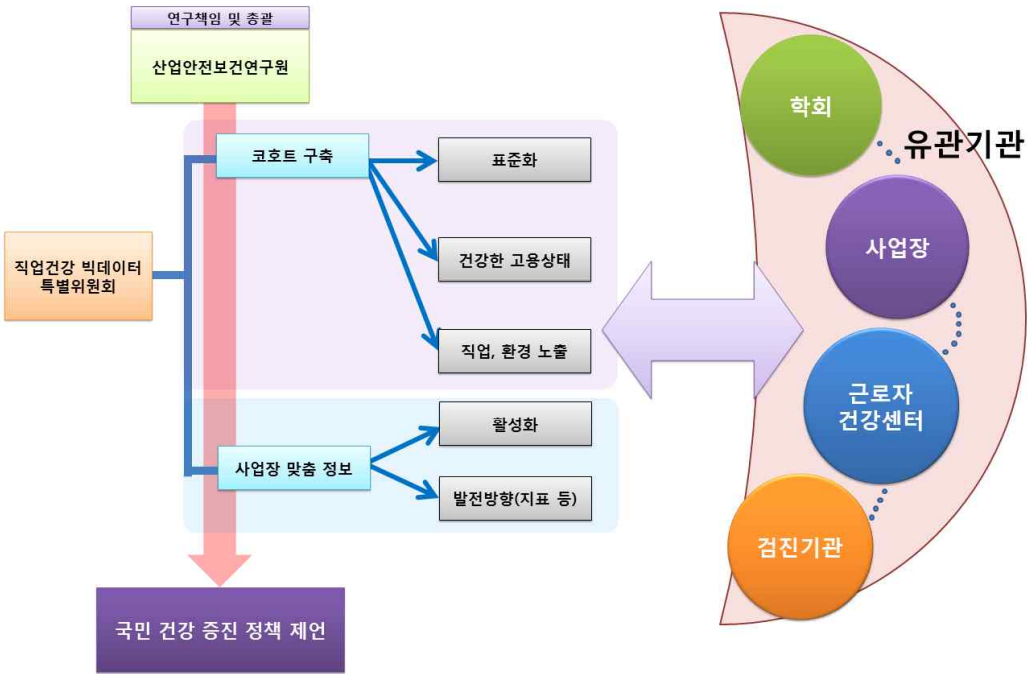
이처럼 빅데이터를 바탕으로 한 코호트 구축은 근로자의 근무환경이나 직장에서의 유해물질 노출에 따른 정보를 제공하고 맞춤형 건강관리 방안을 제공할 수 있을 뿐만 아니라 환경 노출과 관련해서도 활용할 수 있다는 장점이 있다. 유해물질 노출사고, 화학사고 등 사업장을 넘어서서 방대한 시·공간적 범위를 갖는 사건이 발생할 경우, 그 건강영향에 대해서 모두 직접 조사하기는 거의 불가능하다. 특히 그러한 증상이 대면조사를 통해 얻어지기 어려운 비 특이적인 증상이거나 경한 증상인 경우 더더욱 그 건강영향에 대하여 평가하는데 제약이 많아진다. 그러나 빅데이터 기반 코호트를 이용한다면 노출사고 이후 사고 발생지점으로부터의 거리 등을 이용하여 상대적으로 손쉽게 유해물질 노출에 의한 건강영향을 정확한 진단명을 토대로 판단할 수 있으며 추후 추적관찰도 용이하다는 점에서 압도적인 장점을 가질 것으로 기대할 수 있다.

2) 직업 관련 코호트 시스템 구축 설계

국민건강보험공단의 건강보험 수진자료를 이용한 방대한 데이터를 효과적으로 이용하여 향후 지속적인 직업 관련 특성과 직업 관련 질환 사이의 연관성 및 인과관계에 대한 연구를 진행하기 위해서는 체계적인 시스템 설계가 필요하다.

앞에서 설명한 방법 등을 이용하여 연구자들은 직업 관련 코호트를 통한 연구를 진행하기 위해 연구대상집단(target group) 또는 연구대상질환(target disease)을 선정할 수 있으며, 내부의 논의 및 사전조사 과정을 통해 연구주제를 결정하

고 구체적 연구 범위를 설정하여 코호트를 설계할 수 있다.



[그림 2-8] 빅데이터 기반 직업관련 코호트 시스템 구축 및 관련기관 연계

III. 연구 결과

1. 빅데이터 기반 직업군 관련 코호트 설계

1) 국민건강보험공단 빅데이터 자료의 특성 검토

(1) 이용가능한 변수

국민건강보험공단 빅데이터 자료는 크게 자격 및 보험료 데이터, 진료내역 데이터, 건강검진 데이터, 의료급여 데이터, 노인 장기요양 데이터로 이루어져 있으며 각각의 데이터 내 이용 가능한 변수에 대한 자세한 내용은 다음 표와 같다.

<표 3-1> 국민건강보험공단 빅데이터 내 분석 가능 변수

연말 시점 자격 및 보험료 (BFC)	STD_YYYYMM	기준연월
	INDI_DSCM_NO	개인식별아이디
	SEX_TYPE	성
	YEND_STD_AGE_ID	연령
	GAIBJA_TYPE	가입자구분
	REL_CD	가입자와의 관계
	SEDAE_CNT	증 인원수
	RADDR_ADDR_CD	실거주동읍면주소
	ROAD_NM_CD	실거주도로명주소
	CALC_CTRB	보험료
	YEND_POTM_CTRB_20CLS_LEVL	보험료20분위
	N_HGRNK_TGT_YN	차상위계층 여부
	OBST_REG_DT	장애인등록일자

	OBST_GRADE_CD	장애등급
	OBST_PTTN_CD	장애유형
	NATION_CD	국가코드
	FIRM_ADDR_CD	사업장시군구주소
	INDTP_CD	사업장업종세분류
	NATION_MRTPS_CD	국가유공자코드
	RRNO_ERR	주민번호로직 오류 여부
	JIKEYOK_DTL_CD	지역상세코드
사망 (DTH)	RVSND_INDI_DSCM_ID	개인식별아이디
	DEATH_DT	사망일자
공통 (COM)	CMN_KEY	공통키(진료내역연계)
	INDI_DSCM_NO	개인식별ID
	SEX_TYPE	성
	SUJIN_POTM_AGE_ID	수진시점 연령
	YEND_POTM_AGE_ID	연말시점 연령
	GAIBJA_TYPE_CD	수진시점 가입자구분
	OBST_GRADE_CD	수진시점 장애등급
	OBST_PTTN_CD	수진시점 장애유형
	SUJINJA_HM_ADDR_CD	수진시점 환자 시군구코드
	SUJIN_POTM_CALC_CTRB	수진시점 산정보험료
	SUJIN_POTM_CTRB_20CLS	수진시점 보험료20분위수
	SUJIN_POTM_JIKEYOK_CD	수진시점 지역코드
	DISP_SUBJ_TYPE	표시과목
	SUJIN_POTM_YOYANG_ADDR_CD	수진시점 요양기관시도시군구코드
	SICK_SYM1	주상병
	SICK_SYM2	부상병

	SICK_SYM3	3상병코드
	SICK_SYM4	4상병코드
	SICK_SYM5	5상병코드
	MDCARE_SYM	요양기관 기호
	MDCARE_STRT_DT	요양개시일자
	FORM_CD	서식코드
	SPCF_SYM_TYPE	특정기호구분
	EDC_ADD_RT	심결가산율
명세서 (T20)	CMN_KEY	공통키(진료내역연계)
	MCARE_SUBJ_CD	진료과목코드
	HSPTZ_PATH_TYPE	입원경로구분
	VSHSP_DD_CNT	입내원일수
	MDCARE_DD_CNT	요양일수
	TOT_PRSC_DD_CNT	총처방일수
	MCARE_RSLT_TYPE	진료결과구분
	OPRTN_YN	수술여부
	ED_RC_TOT_AMT	심결요양급여비용총액
	EDC_SBA	심결본인부담금
	EDC_INSUR_BRDN_AMT	심결보험자부담금
	FST_HSPTZ_DT	최초입원일
	IOPF_TYPE_CD	상해외인구분
	OFIJ_TYPE	공상구분
	EDC_DRG_NO	심결DRG번호
진료내역 (T30)	CMN_KEY	공통키(진료내역연계)
	MCARE_DESC_LN_NO	줄번호
	CLA_CD	명세서항코드
	ITEM_CD	명세서목코드
	CLSF_PTTN_CD	분류유형코드

	MCARE_DIV_CD	분류코드
	MCARE_DIV_CD_NM2	분류코드명_2
	UPRC	단가
	DD1_MQTY_FREQ	일일투여량또는실시횟수
	TOT_MCNT	총투여일수또는실시횟수
	AMT	금액
	TIME1_MDCT_CPCT	1회투약량
	MDCN_UD	약제상한차액
	MCEXP_TYPE_CD	5대분류
	EFMDC_CLSF_NO	약효분류번호
	WK_COMPN_CD	주성분 코드
상병내역 (T40)	CMN_KEY	공통키(진료내역연계)
	MCARE_SUBJ_CD	진료과목코드
	MCEX_SICK_SYM	상병기호
	DORC_SICK_NM	상병명
	OPRTN_YN	수술여부
	SPCF_SYM_TYPE	특정기호구분
	DETAIL_TMSG_SUBJ_CD	상세전문과목코드
	SICK_CLSF_TYPE	상병분류구분코드
처방전 교부상세 (T60)	CMN_KEY	공통키(진료내역연계)
	MCARE_DESC_LN_NO	줄번호
	CLSF_PTTN_CD	분류유형코드
	MCARE_DIV_CD	분류코드(60)
	MPRSC_TIME1_TUYAK_CPCT	1회투약량
	MPRSC_DD1_TUYAK_CPCT	1일투약량
	TOT_MCNT	총투여일수또는실시횟수
	UPRC	단가
	AMT	금액

	GNL_NM_CD	일반명코드
요양기관 (INST)	YKIHO	요양기관기호
	PAYMT_YOYANG_CLSFC_CD	급여요양기관종별코드
	FDAT_TYPE_CD	설립구분코드
	HIRA_DISP_SUBJ_CD	심사평가원표시과목코드
	YOYANG_DETAIL_DISP_SUBJ_CD	요양기관세부표시과목코드
	SI_DO_SI_GUN_GU_CD	시도시군구코드
	CNT_MDR_TOT	전체 의사 수
	CNT_MDR_SPC	전문의 수
	CNT_DDR_TOT	전체 치과의사 수
	CNT_DDR_SPC	치과전문의 수
	CNT_HDR_TOT	전체 한의사 수
	CNT_HDR_SPC	한의원전문의 수
	CNT_NRS_TOT	전체 간호사 수
	CNT_NRS_AID_TOT	전체 간호조무사 수
	CNT_PHRM_TOT	전체 약사 수
	CNT_ROOM_TOT	전체 병실 수
	CNT_BED_TOT	전체 병상 수
	CNT_ROOM_INP	입원병실 수
	CNT_BED_INP	입원병상 수
	CNT_ROOM_OP	수술실 수
	CNT_BED_OP	수술병상 수
	CNT_ROOM_ICU_ADLT	성인/소아중환자실 수
	CNT_BED_ICU_ADLT	성인/소아중환자실병상 수
	CNT_ROOM_NICU	신생아중환자실 수
	CNT_BED_NICU	신생아중환자실병상 수

여기서 보험가입자 및 피부양자의 소재지, 사업장소재지 등은 동·읍·면 단위까지 숫자로 치환되어 코딩되었다. 주소와 관련하여 사용할 수 있는 구체적 변수

및 적용예시는 다음 <표 3-2>와 같다.

<표 3-2> 국민건강보험공단 빅데이터 내 주소 관련 변수

	예 1) 서울시 종로구 삼청동	예 2) 부산광역시 중구 중앙동
SIDO_NM	서울특별시	부산광역시
SSG_NM	종로구	중구
DONG_NM	삼청동	중앙동
SIDO_CD	11	26
SIDO_SGG_CD	11110	26110
SIDO_SGG_DONG_CD	11110540	26110510
ADDR_CD	1111054000	2611051000

보험가입자의 피부양자로 건강보험 혜택을 받고 있는 대상자의 경우 가입자와의 관계를 'REL_CD' 변수를 통하여 알 수 있으며, 다음 <표 3-3>과 같은 내용을 파악할 수 있다.

<표 3-3> 국민건강보험공단 빅데이터 내 가입자와의 관계 변수

코드	설명	코드	설명	코드	설명	코드	설명
01	본인	30	누이	59	장조모	87	처숙모
02	처	31	형부	60	고모	88	처사촌
03	부	32	매	61	고모부	89	처고모
04	모	33	매부	62	대고모	90	처고모부
05	자녀	34	시누이	63	당숙	91	처고종
06	자부	35	종매	64	당숙모	92	처이모
07	남편	36	종매부	65	서모	93	처이모부
08	사위	37	제부	66	서조모	94	처이종
10	시부	38	조카	67	이질	95	처가친척
11	시모	39	동서	68	당질	96	시가친척
12	장인	40	손	69	질부	97	외가친척
13	장모	41	손부	70	외숙	98	친척
14	백부	42	손서	71	외숙모	99	동거인
15	백모	43	증손	72	외종	A1	계부
16	숙부	44	증손부	73	이모	A2	계모
17	숙모	45	고손	74	이모부	A3	생부
18	양부	46	고손부	75	이종	A4	생모
19	양모	47	고종	76	외손	A5	계자
20	형	48	오빠	77	외손부	A6	생자
21	형수	50	조부	78	외조부	A7	기타
22	제	51	조모	79	외조모	A8	처조부
23	제수	52	시조부	80	처형	A9	처조모
24	시숙	53	시조모	81	처제	B1	생시모
25	올케	54	증조부	82	처남	B2	생장모
26	종형	55	증조모	83	처남댁	B3	생조모
27	종형수	56	고조부	84	처조카	B4	생시조모
28	종제	57	고조모	85	처질부	B5	생처조모
29	종제수	58	장조부	86	처숙부	YR	연보용 기타처리

업종분류코드는 다음 <표 3-4>와 같이 총 17가지의 업종과 ‘기타’항목으로 이루어져 있으며 이는 ‘표준산업분류’ 최근 개정 내용과는 다소 차이가 있다. 따라서 연구 과정 중 결과 해석 시, 그리고 다른 빅데이터와 연계 시 산업분류 사용에 주의가 필요하다.

<표 3-4> 국민보험공단과 표준산업분류의 산업분류 비교

건강보험공단 코딩 자료		표준산업분류 (10차 개정)	
코드	설명	코드	설명
01	농업, 수렵업 임업	A	농업, 임업 및 어업
02	어업	B	광업
03	광업	C	제조업
04	제조업	D	전기, 가스, 증기 및 수도사업
05	전기, 가스, 수도사업	E	하수, 폐기물 처리, 원료재생 및 환경복원업
06	건설업	F	건설업
07	도·소매 및 소비자용품 수리업	G	도매 및 소매업
08	숙박 음식점업	H	운수업
09	운수, 창고, 통신업	I	숙박 및 음식점업
10	금융, 보험업	J	출판, 영상, 방송통신 및 정보서비스업
11	부동산, 임대, 사업서비스	K	금융 및 보험업
12	공공국방 및 사회보장행정	L	부동산 및 임대업
13	교육서비스업	M	전문, 과학 및 기술 서비스업
14	보건, 사회복지사업	N	사업시설관리 및 사업지원서비스업
15	기타공공사회, 개인서비스	O	공공 행정, 국방 및 사회보장 행정
16	가사서비스업	P	교육 서비스업
17	국제, 기타외국기관	Q	보건업 및 사회복지 서비스업
99	기타	R	예술, 스포츠 및 여가관련 서비스업
		S	협회 및 단체, 수리 및 기타 개인 서비스업
		T	가구 내 고용활동 및 달리 분류되지 않은 자가소비 생산활동
		U	국제 및 외국기관

이렇게 현재 국내에서 사용되고 있는 산업분류는 표준산업분류, 통계청의 산업분류, 산재 데이터에서의 산업분류 등 서로 분류 방법은 유사하나 내용은 상이한 여러 분류방법이 혼재되어 있다. 따라서 구축된 데이터들을 통합하기 위해서는 이러한 서로 다른 분류방법에 대한 통합 지침이 필요하다. 본 연구진들은 각 산업분류표를 모두 대조하여 정리하였으며 건강보험공단 업종분류코드와 표준산업분류(제 9차) 사이에 완벽히 일치한다고 보기 어려운 항목에 대하여 충분한 논의를 거쳐 다음 <표 3-5>와 같이 의견을 취합하였다.

<표 3-5> 건보공단 업종분류와 표준산업분류 중 상이한 업종에 관한 연구진 의견

건보공단 분류코드	건보공단 업종분류	연구원 A	연구원 B	연구원 C	연구원 D
96412	소포송달업	49401 택배업(H)/ 61100 우편업(J)	49401 택배업(H)/ 61100 우편업(J)		61100 우편업
110721	컴퓨터 설비 자문업	6202 컴퓨터시스템 통합 자문 구축 및 관리업(J)	62022 컴퓨터시설 관리업	620 컴퓨터 프로그래밍, 시스템 통합 및 관리업/ 62021 컴퓨터시스템 통합 자문 및 구축 서비스업	620 컴퓨터 프로그래밍, 시스템 통합 및 관리업/ 62021 컴퓨터시스템 통합 자문 및 구축 서비스업
106603	비생명보험업	65122 보증 보험업(K)	65122 보증 보험업(K)	6512 손해 및 보증보험업	6512 손해 및 보증보험업
110720	정보처리 및 기타 컴퓨터운용 관련업	62090기타 정보기술 및 컴퓨터 및 컴퓨터운영 관련 서비스업 (J)	자료처리업		62090기타 정보기술 및 컴퓨터 및 컴퓨터운영 관련 서비스업 (J)
110729	기타 정보처리 및 컴퓨터 운용 관련업	62090기타 정보기술 및 컴퓨터 및 컴퓨터운영 관련 서비스업 (J)	자료처리업 or 컴퓨터시설관리업		62090기타 정보기술 및 컴퓨터 및 컴퓨터운영 관련 서비스업 (J)
110740	기타 사업관련서비스업	759 기타 사업지원 서비스업 (N)	759 기타 사업지원 서비스업 (N)		759 기타 사업지원 서비스업 (N)

110749	분류되지 않은 사업관련서비스업	7599 그외 기타 사업지원 서비스업 (N)	7599 그외 기타 사업지원 서비스업 (N)		7599 그외 기타 사업지원 서비스업 (N)
117423	지주회사	71520 비금융 지주회사(M)	사업 및 경영상담업		7152 비금융지주회사
40222	상업인쇄 및 인쇄관련서비스업	181 인쇄 및 인쇄관련 산업 (C)	인쇄업 or 인쇄관련산업	1811 인쇄업	181 인쇄 및 인쇄관련 산업 / 1811 인쇄업 / 18129 기타 인쇄관련 산업
40300	사무, 계산 및 회계용 기계 제조업	분류불가	기타 일반목적용 기계제조업		2918 사무용 기계 및 장비 제조업
40321	전자관 및 기타 전자제품 제조업	2629 기타 전자부품 제조업/ 26291 전자관제조업	2629 기타 전자부품 제조업		2629 기타 전자부품 제조업
85524	다과점	56191 제과점업/ 56220 비알콜 음료점업	56220 비알콜 음료점업		5622/56220 비알콜음료점업
90630	여행알선 및 운수관련서비스업	752 여행사 및 기타 여행보조 서비스업/ 75290 기타 여행보조 및 예약서비스업	752 여행사 및 기타 여행보조 서비스업		752 여행사 및 기타 여행보조 서비스업

42213	오디오 기록매체 출판업	592 오디오물 출판업 및 원판 녹음업 / 59201 음악 및 기타 오디오물 출판업	“영상·오디오 기록물 제작 및 배급업” 이랑 가장 유사해보이네요	592 오디오물 출판 및 원판 녹음업	
40293	분류되지 않은 가정용 가구 제조업	32099 그 외 기타 가구 제조업	32011 침대 및 내장가구 제조업	292 특수목적용 기계 제조업	32011 침대 및 내장가구 제조업
70523*	비식용 식품일반소매업	471 종합 소매업 /472 음·식료품 및 담배 소매업	47219 기타 식료품 소매업	478 기타 상품 전문 소매업	41190 그 외 기타 종합 소매업
70525	특수소매업	478 기타 상품 전문 소매업 /479 무점포 소매업	478 기타 상품 전문 소매업	478 기타 상품 전문 소매업	478 기타 상품 전문 소매업
42330	핵연료 가공업	2012 기초 무기화학물질 제조업/ 16102 연탄 및 기타 석탄 가공품 제조업 / 06210 우라늄 및 토륨 광업(;))		192 석유정제품제조업	20129 기타 기초 무기화학물질 제조업
42434	기록용 매체 및 관련화학제품	204 기타 화학제품 제조업 266 마그네틱 및 광학	266 마그네틱 및 광학 매체 제조업	2049	20423 인쇄잉크 제조업

	제조업	매체 제조업			
138091	사무관련 교육기관	85699기타 기술 및 직업훈련학원	8569 그 외 기타 교육기관	8550 일반 교습학원	8569 그 외 기타 교육기관
139030/15903 0/160930	공공장소 청소 및 유사 서비스업	742 건물·산업설비 청소 및 방제 서비스업	7421 건물 및 산업설비 청소업	742 건물, 산업설비 청소 및 방제 서비스업	742 건물, 산업설비 청소 및 방제 서비스업
140990/40900	위생 및 유사서비스업	742 건물·산업설비 청소 및 방제 서비스업 (*실제 건보 업종코드 확인시 정화, 미화,위생 방역 관련 사업장 및 관리소 속해있음)	742 건물, 산업설비 청소 및 방제 서비스업	742 건물, 산업설비 청소 및 방제 서비스업	742 건물, 산업설비 청소 및 방제 서비스업
60930/150930 /40930	기타 서비스업	(*세부항목 확인 필요함)		969 그 외 기타 개인 서비스업	
150920/40920	오락, 문화 및 운동관련사업	91 스포츠 및 오락관련 서비스업 / 912 유원지 및 기타 오락관련 서비스업	912 유원지 및 기타 오락관련 서비스업	911 스포츠 서비스업	91 스포츠 및 오락관련 서비스업
150921	영화, 방송 및 기타 공연관련사업	59141 영화관 운영업 (591 영화, 비디오물,	591 영화, 비디오물, 방송프로그램 제작 및	591 영화, 비디오물, 방송프로그램 제작 및	591 영화, 비디오물, 방송프로그램 제작 및

		방송프로그램 제작 및 배급업) 90110 공연시설 운영업 (901 창작 및 예술관련 서비스업)	배급업	배급업	배급업
150924	운동경기 및 기타 오락관련사업	91 스포츠 및 오락관련 서비스업 / 912 유원지 및 기타 오락관련 서비스업	91191 스포츠 클럽 운영업	9129 그 외 기타 오락관련 서비스업	91 스포츠 및 오락관련 서비스업
160950/16950 0/110950	가사서비스업	97000 가구 내 고용활동	그외 기타 개인 서비스업	970 가구내 고용활동	970 가구내고용활동
158734	공연관련산업	90110 공연시설 운영업 90191 공연기획업 90192 공연 및 제작관련 대리업 (901 창작 및 예술관련 서비스업)	90110 공연시설 운영업	9019 기타 창작 및 예술관련 서비스업	9019 기타 창작 및 예술관련 서비스업
158832	경기전문 종사업	9111 경기장 운영업	91191 스포츠 클럽 운영업	911 스포츠 서비스업	911 스포츠 서비스업

본 분류가 현재로서는 최선일 것으로 생각되나, 실제로 분석에 응용하기에는 제한점이 있으므로 사용에 주의하여야 하고, 보다 정확한 데이터 통합을 위해서는 이에 대한 추가 연구과정이 필요하다.

그리고 본 연구가 진행되는 동안 2017년 6월을 기준으로 제 10차 표준산업분류가 발표되었다. 대분류 항목에는 차이가 없으나 하위 분류기준에서 차이 및 변동사항이 발생하는 만큼 앞으로의 코호트를 구축하는데 있어서 표준산업분류가 변동하는 것을 어떻게 반영하여 건보공단 업종분류코드에 적용시킬 수 있을지 일관된 변환 및 통합방법에 대해서도 향후 고찰과 연구가 필요하다.

또한 건보공단의 업종분류코드는 존재하지만 코드 아래에 해당 사업장이 하나도 존재하지 않는 업종도 있을 수 있다. 따라서 잘 사용하지 않는 업종분류코드에 대한 검토 및 점검, 수정작업을 거치는 과정이 정기적으로 필요할 것으로 여겨진다. 또한 새롭게 등장하는 업종 및 산업분류 변화를 반영하기 위하여 건보공단 업종분류코드에 대한 지속적인 보완작업은 반드시 함께 진행되어야 할 것이다.

현재 제 9차 표준산업분류를 기준으로 건보공단 업종분류코드와 서로 상이한 업종분류에 대하여 연구진 회의를 거친 결과, 세분류 또는 세세분류에서는 의견이 갈리더라도 중분류 수준에서는 연구진 간에 의견이 어느 정도 합치되는 경향을 보였다. 따라서 표준산업분류별로 큰 수준부터(대분류) 작은 수준의 순서대로 ‘당장 사용가능한 업종코드’, ‘사용가능성이 있는 업종코드’, ‘사용이 불가능한 업종코드’로 나누는 작업이 필요할 것으로 판단된다. 실제 업종코드 외 하위 사업장의 일치도를 점검하고 그 수를 파악하는 과정은 대규모의 일이 될 것이나 이러한 과정을 통해 향후 건보공단 데이터를 기반으로 한 직업코호트를 이용하는 데 유용한 도구를 획득할 수 있을 것으로 생각된다.

둘째로는 현재 건강보험공단의 코드로는 업종에 대한 파악 이외에 직무분석이 되지 않는다는 제한점이 있다. 같은 제조회사에서 근무하는 동료라 하더라도 생산직 근로자와 회계 담당 근로자는 맡은 바 업무의 특성이 다르지만 둘 다 제조

업 종사자로 집계된다. 따라서 보다 정확한 직무분석 및 향후 연구를 위해서는 향후 데이터 수집 및 구축 시 직업 분류와의 병행이 필요하다.

수진자가 건강보험의 지역가입자인지, 공무원, 교직원 가입자인지, 직장가입자인지 등에 대한 정보를 얻기 위해서는 다음 <표 3-6>과 같이 직역코드를 이용할 수 있다.

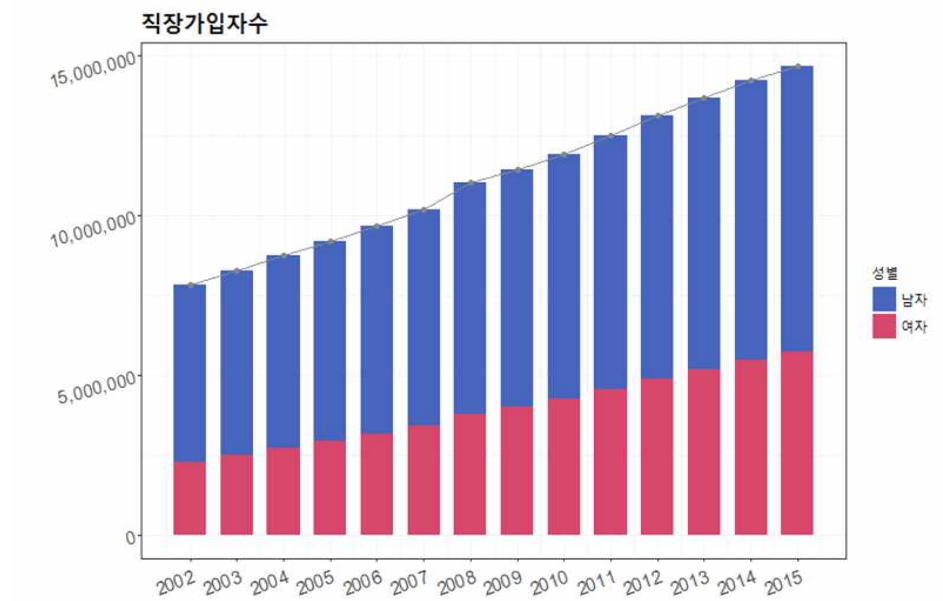
<표 3-6> 국민건강보험공단 빅데이터 내 직역상세코드

직역코드	직역상세코드	설 명
1	11	지역(농어촌)
1	12	지역(도시)
1	13	지역(NONE)
5	53	공교(공무원)
5	54	공교(교직원)
7	7	일반직장
9	9	의료급여
Z	Z	기타
{	{	NONE

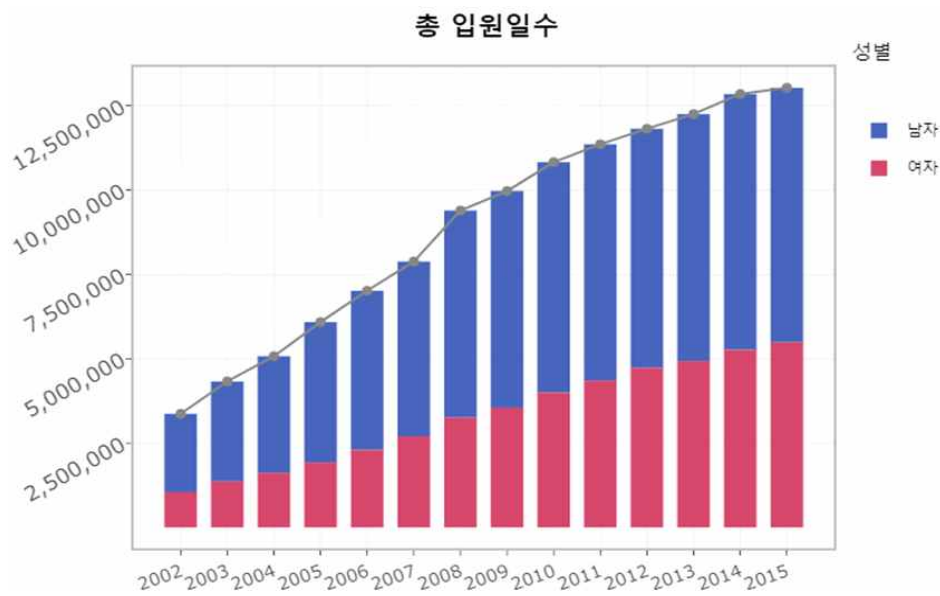
(2) 빅데이터를 통한 전체근로자 현황 파악

현재 우리나라 전체의 건강보험공단 가입자 유형 및 입원내역을 검토한 결과 총 근로자 수, 총 근로자 중 입원환자 수, 총 입원일수, 근로자 10만 명 당 입원일수, 입원한 근로자 1명 당 입원일수를 연도별로 구하였으며 다음 [그림 3-1], [그림 3-2], [그림 3-3], [그림 3-4], [그림 3-5]와 같다.

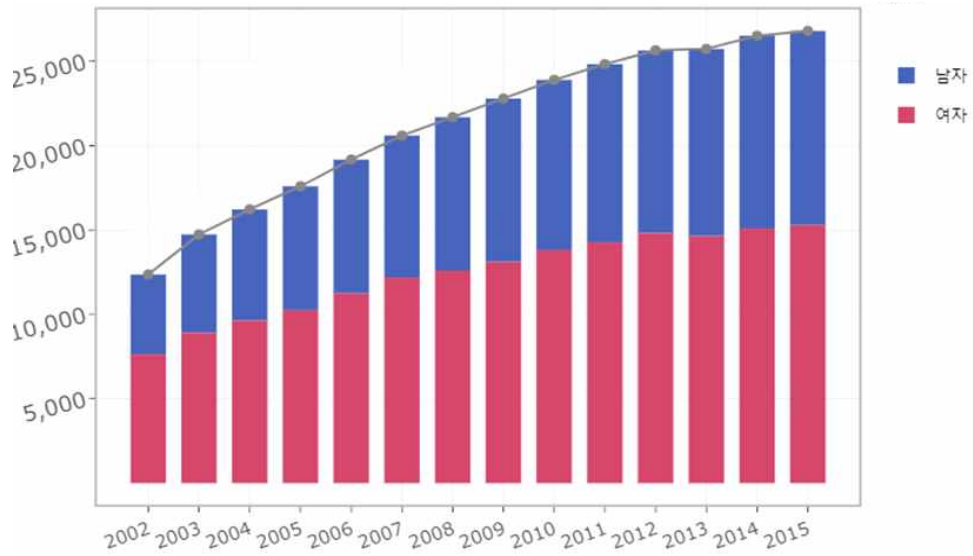
이를 위해 해당년도의 총 근로자 수는 해당년도 건강보험 가입상태가 직장가입자 또는 공무원, 교직원 가입자 ‘본인’인 사람을 추출하였다. 이에 해당하는 총 근로자 수는 2002년 약 7,829,548 명에서 2015년 약 14,675,579 명으로 약 87.4 % 증가하였다.



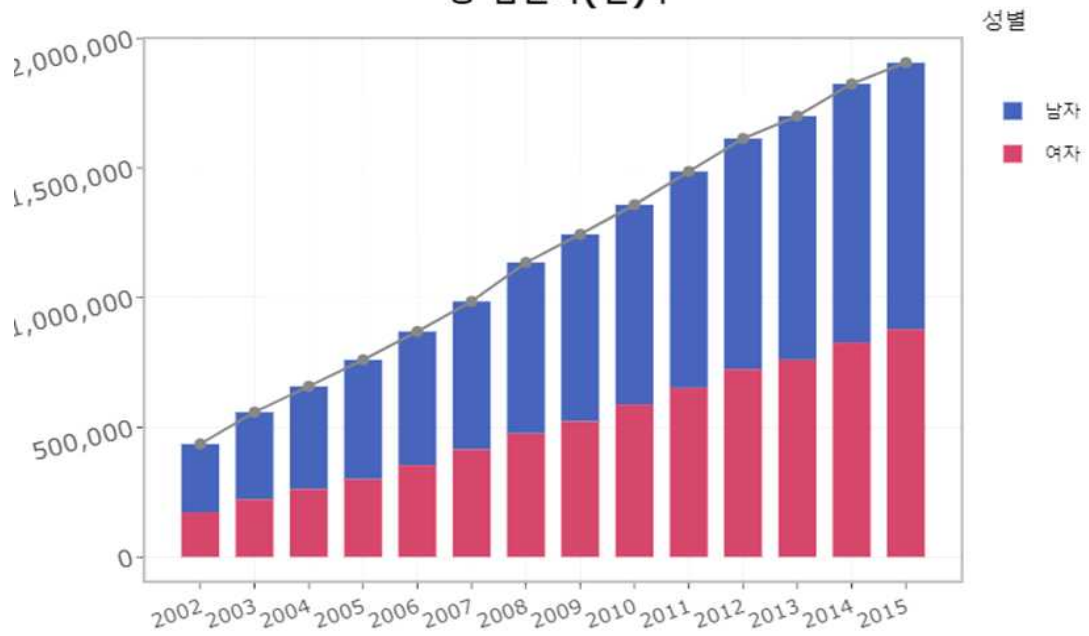
[그림 3-1] 2002-2015년 전체 직장가입자 수



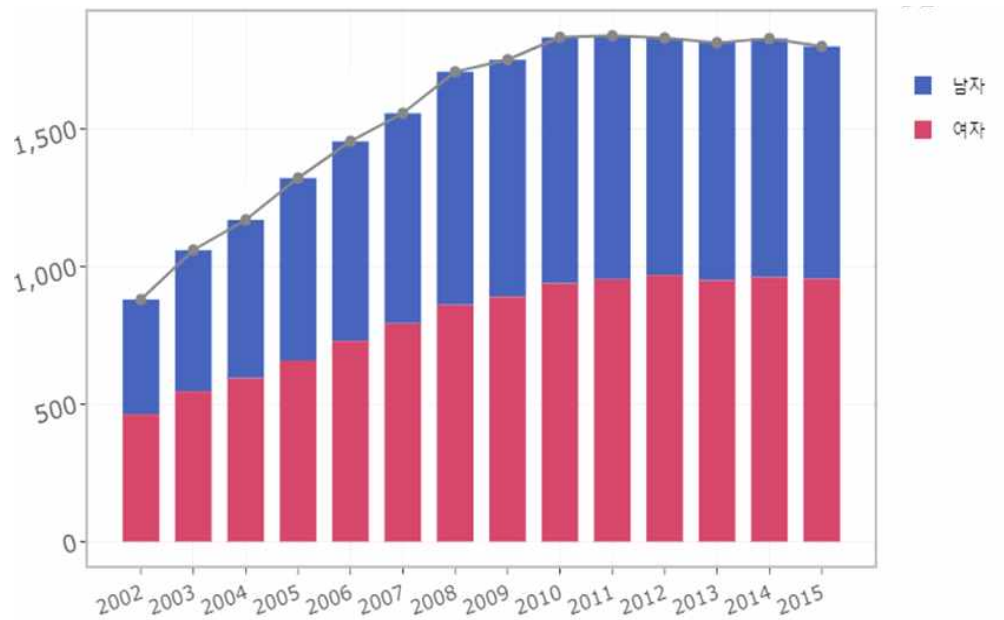
[그림 3-2] 2002-2015년 직장가입자 중 총 인원일 수



[그림 3-3] 2002-2015년 직장가입자 1000명 당 인원일 수
총 인원자(건)수

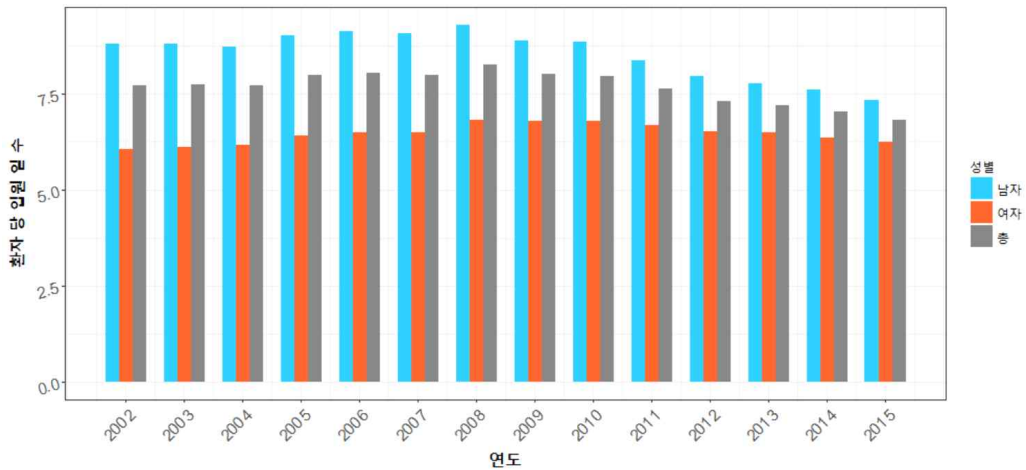


[그림 3-4] 2002-2015년 직장가입자 중 총 인원자 수



[그림 3-5] 2002-2015년 직장가입자 10만 명 당 입원 근로자 수

근로자 중 총 입원환자 수 및 입원일수 역시 매해 증가하는 양상이었으나 근로자의 수가 2002년보다 2015년 약 87% 가량 증가한 것을 감안하여 정확한 비교를 위해 표준화할 필요가 있으므로 근로자 10만 명 당 입원자수를 함께 조사하였다. 그 결과, 근로자 수의 증가를 보정하였을 때에도 총 입원자수는 증가하는 경향을 보였다.



[그림 3-6] 2002-2015년 입원근로자 1명당 평균 입원일 수

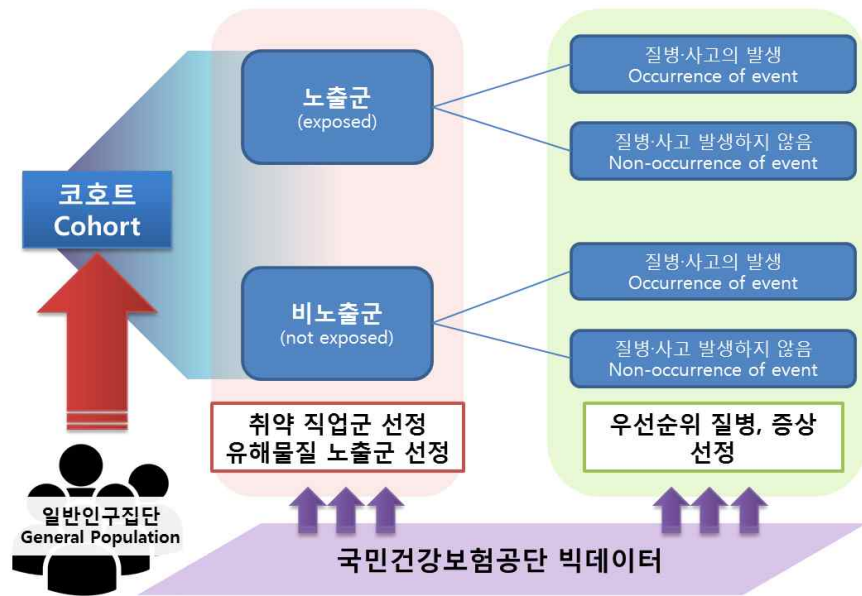
입원한 근로자 1명 당 입원일수는 2008년 이전에는 그 정도가 거의 비슷하였으나 2008년 이후 입원일수가 점차 감소하는 것을 확인할 수 있었다.

2) 빅데이터 기반 구축 코호트의 특징

(1) 대조군의 선정

일반적인 코호트 구축 과정과 마찬가지로 빅데이터 기반 구축 코호트 역사 연구를 위한 대조군의 선정과 연구대상자가 되는 특정 취약 직업군의 선정이 필요하다. 대조군을 선정하는 방법은 여러 가지가 있으나 본 연구에서는 여러 업종 종사 근로자에 있어서 적용될 수 있는 공통적인 대조군으로서 공무원 및 사립학교 교직원을 선정하였다.

대조군을 선정하기 위하여 공무원 및 교직원을 추출하는 과정에서 가입자 유형(GAIBJA_TYPE) 및 직역코드 및 직역상세코드에서 공무원 및 교직원을 추출하는 방식으로 대조군을 설정하였다.



[그림 3-7] 국민건강보험공단 빅데이터를 활용한 코호트 체계

(2) 연구대상 근로자의 선정

빅데이터를 기반으로 코호트를 구축하기 위해서 해당 업종에 종사하는 근로자를 찾는 과정이 필요하다. 현재 국민건강보험공단 데이터에서 직업·직장관련 정보를 추출할 수 있는 변수는 사업장이름(FIRM_NM)과 사업장코드(FIRM_CD), 업종코드(INDTP_CD)가 대표적이다. 직종코드(CLFY_CD)도 존재하지만 이는 표준직업분류와 같은 분류와는 달리, 건강검진 실시 주기를 분류하기 위하여 사용하는 코드로 이를 이용하여 직무 등을 파악하기에는 무리가 있어 실제로 이용 가능한 변수는 사업장 이름(FIRM_NM), 사업장 코드(FIRM_CD), 업종코드(INDTP_CD)가 유용하다.

사업장이름을 이용하여 텍스트분석을 실시하는 방법은 특정한 고유이름을 검색어로 사용하는 방법이다. 예를 들어 근로복지공단에 재직 중인 사람을 추출하고 싶다면 ‘근로복지공단’이라는 연속된 6음절의 낱말을 내포하고 있는 사업장을 검색하여 추출할 수 있다. 여기서 단어선정을 통한 텍스트 분석이 가능한 이유는

‘근로복지공단’ 산하 기관 모두가 해당 단어를 사업장 이름으로 기재하고 있을 것 이란 ‘동질성’을 기대할 수 있음과 동시에 ‘근로복지공단’과 무관한 기관이 해당 단어를 사업장 이름 내에 포함하고 있을 가능성이 낮다는 ‘배타성’이 담보되기 때문이다.

법인 또는 사업장의 고유이름과 같이 완벽한 동질성과 배타성을 확보하지 못 하더라도, 동질성과 배타성 이 두 가지가 높은 업종의 경우에는 적절한 ‘포함’, ‘제 외’ 연산자를 사용하여 검색할 수 있다. 예를 들어 병·의원의 경우, 병원이나 의 원에서는 모두 의료행위를 하고 있을 것을 기대할 수 있음과 동시에 의료행위를 하지 않는 다른 사업장에서 ‘병원’이나 ‘의원’을 사업장 이름으로 사용할 가능성이 낮다는 약간의 배타적 특성을 가지고 있기 때문이다. 그럼에도 불구하고 ‘동물병 원’이나 ‘○○도의원회관’과 같이 예외가 있을 수 있으므로 추출된 사업장 이름을 검토하면서 제외조건을 선정하여 제외시키는 과정이 필요하다.

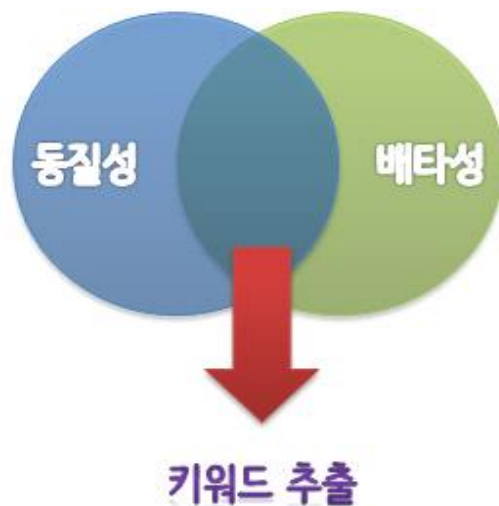


그림 3-8] 사업장 이름을 이용한 텍스트 분석

만약 이렇게 텍스트분석을 통하여 검출하기가 어려운 경우에는 산업분류를 사

용하여 ‘유사한 업무를 하는 사람’을 추정할 수 있다. 예를 들어 자동차정비소의 경우에는 일반적으로 생각할 수 있는 ‘카센타’, ‘카센터’ 외에도 각 기업 브랜드마다 프랜차이즈 고유명을 사업장 이름으로 사용하는 경우가 있어 텍스트분석을 통해 추출하기에는 굉장히 많은 포함조건과 제외조건이 필요하다. 따라서 이 경우에는 산업분류를 이용하여 ‘자동차 수리, 정비’에 종사하는 근로자를 추출한 뒤 사업장 이름을 검토하고 연구진의 논의를 거쳐 제거하는 과정을 거치는 것이 효율적이다.

이때 사용하는 사업장 분류방법으로는 사업장(세분류)코드(5-6자리 숫자)의 앞 1-2자리를 추출하여 대분류 수준에서 각각의 산업을 분류하는 방법과 사업장 분류코드 전체를 사용하는 방법이 있다.

업종분류코드를 이용하여 업종을 추출할 경우에는 연구하고자 하는 주제에 알맞은 수준의 분류코드를 사용하는 것이 중요하다. 상위 산업분류코드를 사용할 경우에는 너무 다양한 업종의 근로자가 해당 코드 내에 포함되어 이질성이 증가하고, 적정 수준보다 하위 분류체계를 사용할 경우에는 실제 근로자 수보다 해당 코드 내 근로자 수가 적어 통계적 분석의 불확정성이 증가하는 단점이 있기 때문이다. 따라서 업종분류코드를 이용할 경우에는 연구에 적합한 수준의 업종분류코드를 선택하기 위한 충분한 고찰이 필요하다.



[그림 3-9] 업종분류코드 이용한 연구의 분류체계에 따른 특징

그리고 사업장의 업종분류코드를 이용하여 분석할 때에는 추출한 사업장의 실제 이름 등을 최대한 확인하여 정확성을 높이는 과정이 필요하다. 예를 들어 자동차 수리업 종사자를 추출하고 싶어 아래 <표 3-7>과 같은 코드를 이용해 추출했을 때, 하위 항목으로 주유소 종사자 또는 자동차 판매 종사자가 함께 포함될 가능성이 있다. 이러한 오류를 최소화하기 위하여 추출 과정에서 사업장 이름을 검토하여 예외가 되는 사업장들을 줄여나가는 과정이 필수적이라고 판단된다. 업종분류코드를 이용한 직업코호트의 이질성과 불확정성, 그리고 이러한 분류 오류의 문제 때문에 업종분류코드를 기반으로 하여 직업코호트를 확립하는 것은 아직 무리가 있을 것으로 판단되므로 분석을 위해서는 사업장 이름, 업종코드 등을 통합적으로 고려하여 연구를 진행하는 것이 현재 수준으로서 가장 바람직한 방법이 될 것이다.

<표 3-7> 업종분류코드 이용한 자동차 수리업 종사자 코드

70500	자동차판매, 수리 및 차량연료 소매업
70502	자동차수리업
150502	
159221	
159222	이륜자동차 수리업

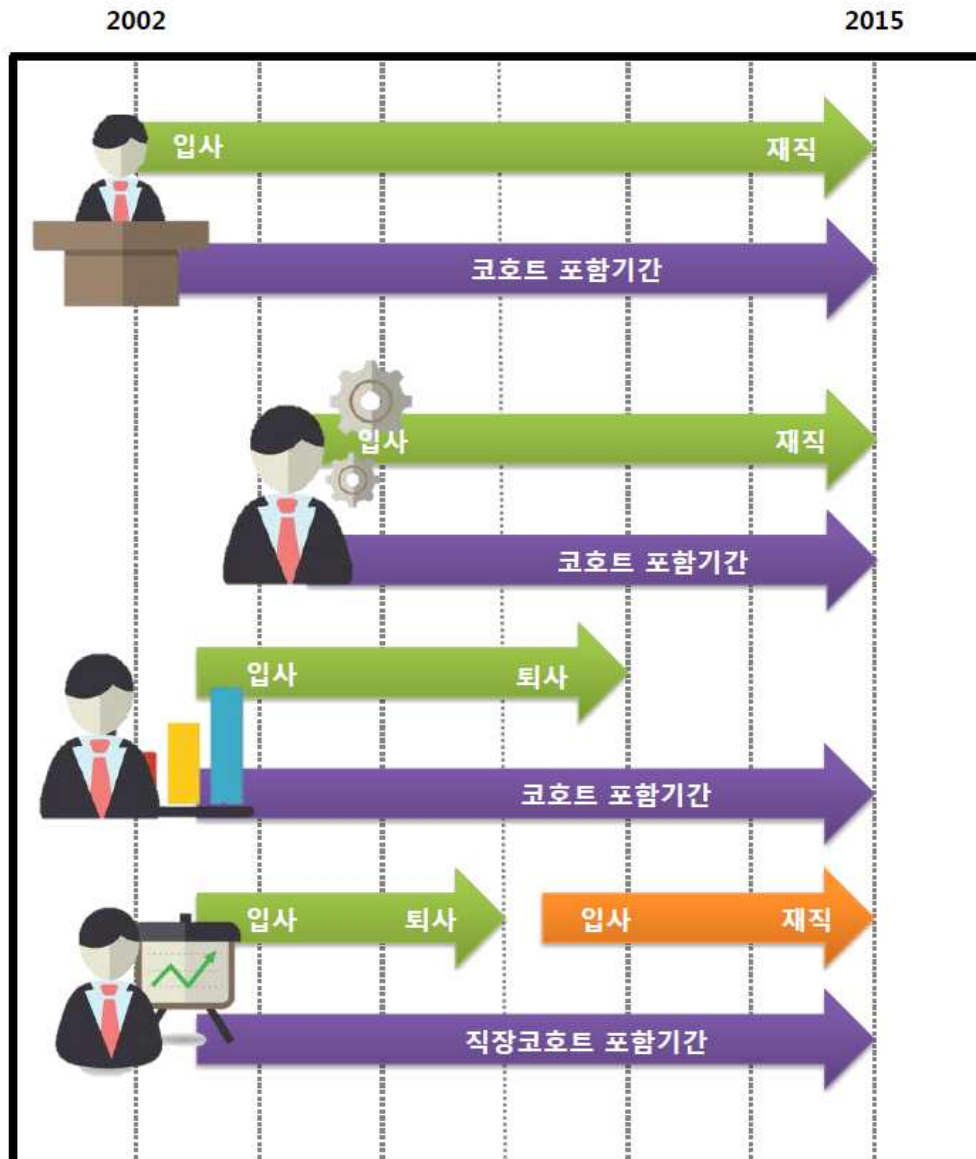
(3) 다이나믹코호트의 구축

이렇게 기준이 되는 직업·산업군을 추출한 뒤, 해당 사업장에서 직장가입자로 가입되었던 첫 해를 입사년도, 해당 사업장에서 마지막으로 직장가입자로 가입되었던 해를 퇴사년도로 설정하여 코호트의 유입 시점을 설정하고 퇴사 시점을 확인한다. 같은 방식으로 직역상세코드에서 공무원 및 사립학교 교직원 가입자에 대해서도 입사와 퇴사를 설정하여 다이나믹 코호트를 구축할 수 있다. 각 연도별로 공무원 및 교직원의 입사, 퇴사를 반영한 코호트 등록(enrol) 현황은 다음 <표 3-8>과 같다.

<표 3-8> 공무원 및 교직원의 코호트 등록 현황

연도	코호트 등록 현황			
	0	1	2	합계
2002	1,964,886	328,529	0	2,293,415
2003	2,018,018	355,227	25	2,373,270
2004	2,063,232	383,580	82	2,446,894
2005	2,087,881	424,039	146	2,512,066
2006	2,149,864	424,405	306	2,574,575
2007	1,112,140	1,498,985	332	2,611,457
2008	1,068,110	1,537,618	77,829	2,683,557
2009	984,540	1,570,075	167,490	2,722,105
2010	836,378	1,662,282	254,796	2,753,456
2011	671,713	1,691,564	406,678	2,769,955
2012	509,950	1,699,617	567,394	2,776,961
2013	337,585	1,707,075	733,367	2,778,027
2014	165,722	1,728,609	870,016	2,764,347
2015	0	1,616,779	1,153,596	2,770,375

다만 현재 구축된 코호트에서는 하드웨어 및 원 데이터(raw data)의 용량 및 컴퓨터의 성능으로 인한 물리적 분석 한계 때문에 자격 및 보험료가 실시간 반영하기 어려워 근로자의 근무 정보에서 누락되는 부분이 발생하고, 이직 및 휴직의 기간을 반영하기 어렵다는 것이 제한점으로 작용한다. 따라서 추후 정확한 직업 코호트를 구축하는 과정에서는 효과적인 분석방법을 고민하는 것과 동시에 이러한 물리적 제한점에 대한 제반적 대책이 필요할 것으로 생각된다.



[그림 3-10] 빅데이터 기반 직업코호트의 구축 모식도

2. 우선순위 질병 선정 및 취약군 발굴

1) 질병 선정 및 취약군 발굴의 전제조건

우선순위 질병 선정 과정부터 취약집단의 발굴에 이르기까지 국민건강보험공단의 데이터베이스 활용하여 다양한 정보를 얻을 수 있다. 다만, 전국민의 국민건강보험공단 자료를 얻기 위해서는 각 연구기관 혹은 진행되는 각 연구마다 필요한 데이터 이용에 대한 비용을 지불하고 이용할 수 있는 권리를 획득하여야 하며, 이는 앞으로의 맞춤형 직업 관련 코호트를 구축하고 관리하며 그 연구결과를 발표하는 데 있어 필수적인 단계라 볼 수 있다.

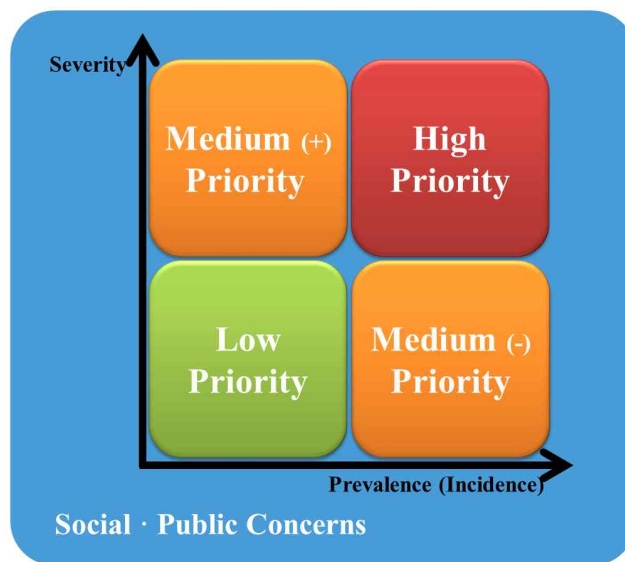
이번 연구에서도 건강보험공단의 데이터를 사용하기 위한 권한을 얻는 과정을 거쳐야 하지만 재정적 한계 등으로 인하여 그리하지 못하였다. 따라서 본 연구를 위한 건강보험공단 빅데이터 분석은 불가능하였다. 그러나 코호트의 구축가능 여부에 대한 현실성을 확인, 검증하기 위해서는 분석 후 고찰이 반드시 필요하므로 시범적 운용을 수행하는 것은 필수적인 과정이다. 따라서 본 보고서에 실린 코호트 시범운영 및 활용방안은 대한직업환경의학 빅데이터위원회와 국민건강보험공단 사이의 업무협약(MOU)을 통해 이루어진 연구결과 및 세미나(2017년 8월, 2017년 11월) 결과를 참조, 인용하여 제시함으로써 빅데이터의 시범적 구축과 운용 과정에 대한 고찰을 시행하였다.

데이터 마련은 앞으로의 빅데이터 기반 직업 코호트를 구축하는 데 있어서는 반드시 필요한 과정이므로, 추후 연구원에서 빅데이터 기반 직업 코호트에 있어 중추적 역할을 수행하기 위해서는 충분한 논의를 거쳐 데이터 이용권한 획득을 위한 절차를 시행해야 할 것으로 생각된다.

2) 질환에 대한 고찰 및 우선순위 질병 탐색

직업코호트를 바탕으로 연구할 수 있는 우선순위 질병을 선정하는 과정은 연

구하고자 하는 직업군이나 사업장의 특성별로 다양할 수 있다. 먼저 우선순위 질병을 선정할 때는 질병의 유병률(발생률)이나 그 중증도에 대한 고찰이 필요하다. 일반적으로 질병의 중증도가 높고, 유병률이 높을 경우 그 질병에 대한 예방 및 관리 정책은 다른 질병에 비해 높은 우선순위를 갖는다고 볼 수 있다.



[그림 3-11] 우선순위 질병 선정 기준

우선순위 질병을 선정하기 위한 조사방법으로는 문헌고찰, 전문가 자문, 기존에 조사된 통계자료 분석, 그리고 면담 및 설문조사가 있다. 각자의 조사방법에 따라 장점과 단점이 혼재하므로 연구 목적에 맞는 방법을 영리하게 선택하거나 연구 주제에 맞게 여러 조사방법을 실시하여 서로의 장·단점을 보완할 수 있다.



[그림 3-12] 우선순위 질병 선정 방법

본 연구의 코호트에서는 고찰 및 회의를 통하여 수집된 의견을 바탕으로, 통계적 분석을 기반으로 한 효과적인 우선순위상병 파악 방법을 선택하였다. 가장 객관적이고 현실 반영적이라 생각된 통계적 분석을 통한 상병 탐색은 대한직업환경의학회와 국민건강보험공단의 업무협약을 통해 이루어진 것으로, 해당 기관의 세미나에서 발표한 자료를 참조, 인용하였다.

먼저 2002년부터 2015년까지 남성 및 여성 근로자에게서 상병코드 2자리(알파벳 1자리, 숫자 1자리)를 기준으로 하여 입원의 원인이 된 질환에 대하여 전반적으로 조사하였다. 외래의 상병명을 선택하지 않고 입원의 상병명을 선택한 이유는, 행위별수가제가 기본적으로 시행되고 있는 우리나라 의료 환경에서 외래 상

병명보다는 입원 상병명이 보다 실제와 일치할 가능성이 높다고 판단하였기 때문이다. 또한 외래 진료로 치료가 가능한 상병명보다는 입원하여 치료하여야 하는 질환이 질병 자체의 심각성(severity)이 높다고 추정할 수 있기 때문에 우선순위 상병 조사에서는 입원 시 주상병명의 빈도에 대하여 먼저 파악하는 것을 목표로 하였다.

먼저 전반적인 상병명 탐색을 위하여 가장 상위수준의 상병코드부터 사용하여 분석을 실시하였다. 현재 국내에서 사용 중이며 건강보험공단 데이터 내 상병코드의 가장 앞 1자리 알파벳은 다음 <표 3-9>와 같은 질환 분류를 의미한다.

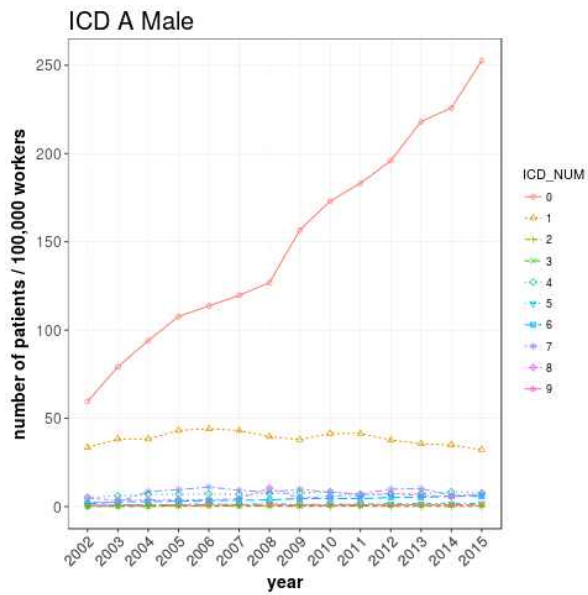
<표 3-9> 건강보험공단 데이터 내 상병코드의 가장 앞 1자리 알파벳

코드	주요 질환
A	특정 감염성 및 기생충성 질환
B	특정 감염성 및 기생충성 질환
C	악성 신생물
D	기타 신생물 및 혈액 질환
E	내분비, 영양 및 대사 질환
F	정신 및 행동 장애
G	신경계통의 질환
H	눈 및 눈 부속기, 귀 및 유도의 질환
I	순환계통의 질환
J	호흡계통의 질환
K	소화계통의 질환
L	피부 및 피하조직의 질환
M	근골격계통 및 결합조직의 질환
N	비뇨생식계통의 질환
O	임신, 출산 및 산후기
P	출생전후기에 기원한 특정 병태
Q	선천기형, 변형 및 염색체 이상

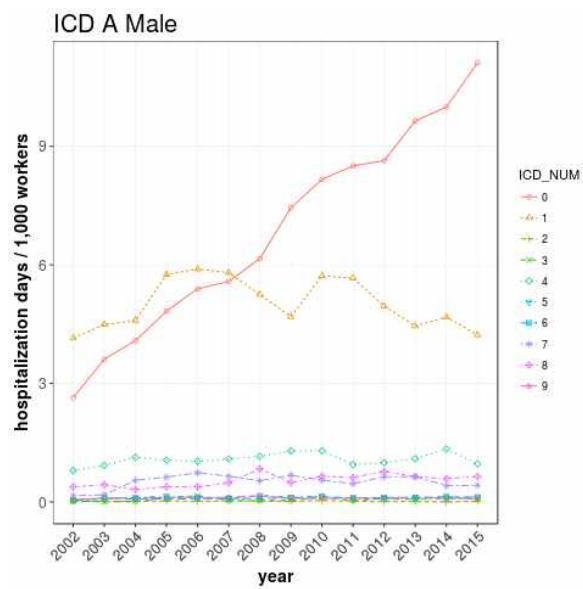
R	달리 분류되지 않은 증상, 징후와 임상 및 검사의 이상 소견
S	손상, 중독 및 외인에 의한 특정 기타 결과
T	손상, 중독 및 외인에 의한 특정 기타 결과
U	특수목적 코드
V	질병이환 및 사망의 외인
W	질병이환 및 사망의 외인
X	질병이환 및 사망의 외인
Y	질병이환 및 사망의 외인
Z	건강상태 및 보건서비스 접촉에 영향을 주는 요인

그리고 질환을 앓고 있는 근로자 수 추이는 최근의 근로자 수 증가 및 환자 수 증가에 의한 효과를 보정하고 표준화하여 비교하기 위하여 전체 환자 수뿐만 아니라 근로자 1,000명 당 입원 환자 수, 환자 1명 당 입원일수를 지표로 사용하여 전체적 추이를 분석하였다.

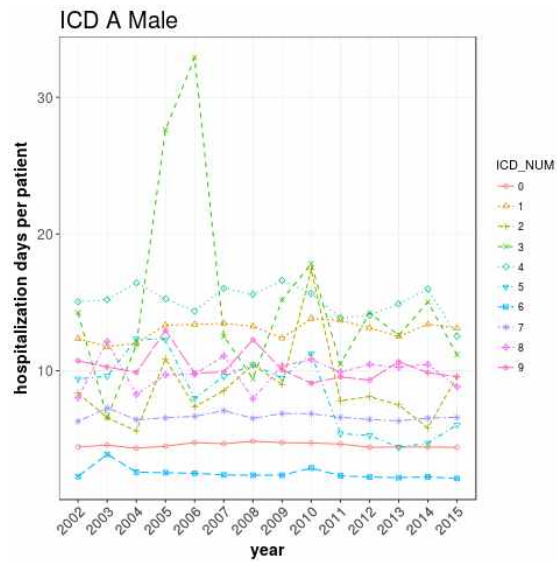
각 ICD 상위 코드 A로 시작하는 해당 코드의 추세는 다음 [그림 3-13], [그림 3-14], [그림 3-15], [그림 3-16], [그림 3-17], [그림 3-18]과 같다. 해당 자료는 남성 및 여성근로자에게서 2002년부터 2015년 사이 전체 상병코드별 발생추이를 그래프로 살펴본 결과로 이를 통하여 연도별 각 질환의 발생추이를 파악할 수 있다. 이러한 추세에 영향을 미치는 요인으로는 실제 질환의 발생률 및 유병률이 달라지기 때문일 수도 있지만 진단기준의 변화, 검사기기의 발달, 건강보험 적용 여부 등의 다른 외부효과에 의해서 달라지는 경향도 존재한다. ICD B부터 Z까지 다른 질환들의 추이를 살펴 본 그림은 부록에 첨부하였다.



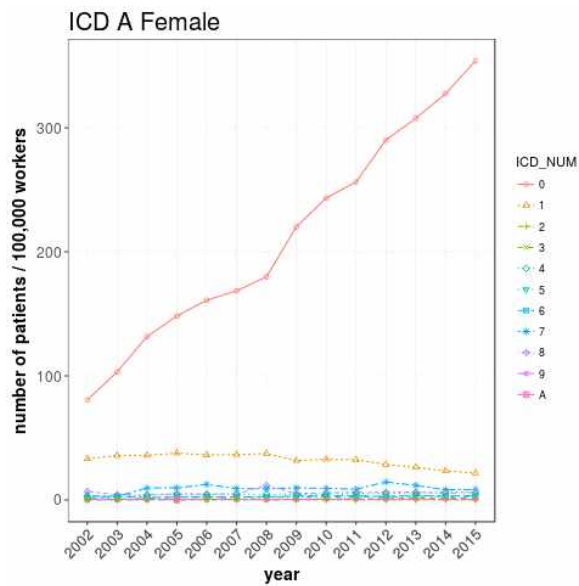
[그림 3-13] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 십만 명 당 환자 수



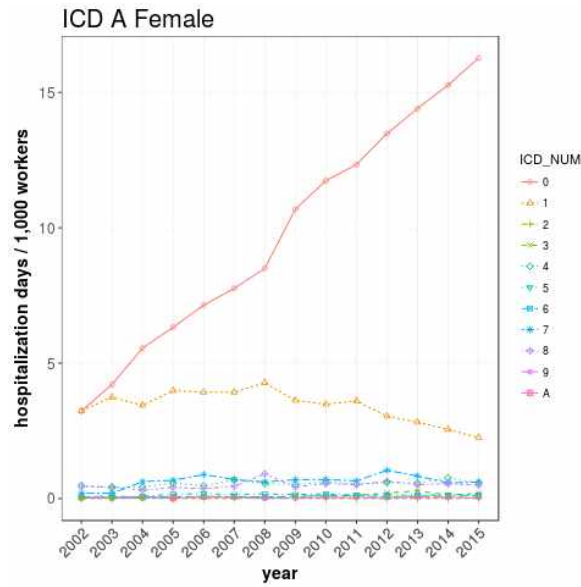
[그림 3-14] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 천 명당 입원일 수



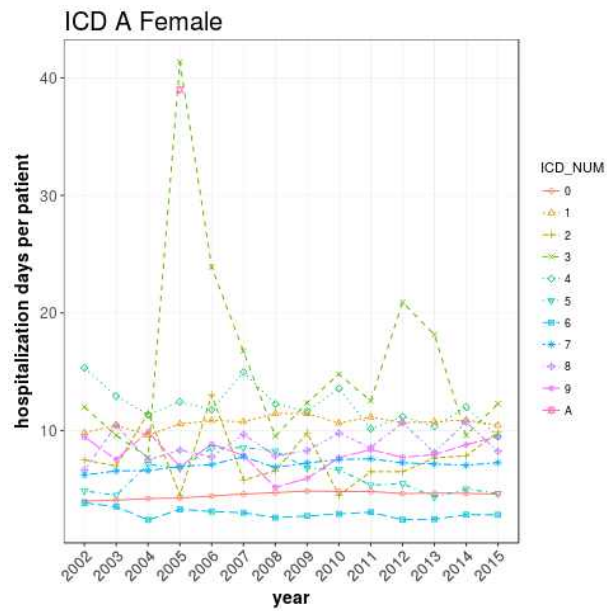
[그림 3-15] ICD 상위 A 코드 별 하위 코드의 남성 직장가입자 중 환자 당 입원일 수



[그림 3-16] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 십만 명 당 환자 수



[그림 3-17] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 천 명당 입원일 수



[그림 3-18] ICD 상위 A 코드 별 하위 코드의 여성 직장가입자 중 환자 당 입원일 수

위의 결과를 좀 더 자세히 보기 위해 ICD 상위 3자리(알파벳 1자리+숫자 2자리)를 분류한 한글명은 다음 <표 3-10>과 같다.

**<표 3-10> ICD 상위 3자리(알파벳 1자리+숫자 2자리) 분류의
주요 질환**

코드	주요 질환
A00-A09	장감염질환
A15-A19	결핵
A20-A28	특정 동물매개의 세균성 질환
A30-A49	기타 세균성 질환
A50-A64	주로 성행위로 전파되는 감염
A65-A69	기타 스피로헤타질환
A70-A74	클라미디아에 의한 기타 질환
A75-A79	리케차병
A80-A89	중추신경계통의 바이러스감염
A92-A99	절지동물매개의 바이러스열 및 바이러스출혈열
B00-B09	피부 및 점막병변이 특징인 바이러스감염
B15-B19	바이러스감염
B20-B24	인체면역결핍바이러스병
B25-B34	기타 바이러스 질환
B35-B49	진균증
B50-B64	원충질환
B65-B83	연충증
B85-B89	이감염증, 진드기증 및 기타 감염
B90-B94	감염성 및 기생충성 질환의 후유증
B95-B98	세균성, 바이러스성 및 기타 감염체
B99	기타 감염성 질환
C00-C75	림프, 조혈 및 관련 조직을 제외한 특정 부위의, 원발성이라고 정해졌거나 또는 가정되는 악성 신생물

C00-C14	입술, 구강 및 인두
C15-C26	소화기관
C30-C39	호흡기 및 흉곽 내 기관
C40-C41	골 및 관절연골
C43-C44	피부
C45-C49	종피 및 연조직
C50	유방
C51-C58	여성생식기관
C60-C63	남성생식기관
C64-C68	요로
C69-C72	눈, 뇌 및 중추신경계통의 기타 부분
C73-C75	갑상선 및 기타 내분비선
C76-C80	불명확한, 이차성 및 상세불명 부위의 악성 신생물
C81-C96	림프, 조혈 및 관련 조직의, 원발성이라고 정해졌거나 또는 가정되는 악성 신생물
C97	독립된(원발성) 여러 부위의 악성 신생물
D00-D09	제자리신생물
D10-D36	양성 신생물
D37-D48	행동양식 불명 및 미상의 신생물
D50-D53	영양성 빈혈
D55-D59	용혈성 빈혈
D60-D64	무형성 및 기타 빈혈
D63*	달리 분류된 만성 질환에서의 빈혈
D65-D69	응고결함, 자반 및 기타 출혈성 병태
D70-D77	혈액 및 조혈기관의 기타 질환
D77*	달리 분류된 질환에서의 혈액 및 조혈기관의 기타 장애
D80-D89	면역메커니즘을 침범한 특정 장애
E00-E07	갑상선 장애
E10-E14	당뇨병

E15-E16	포도당조절 및 체장내분비의 기타 장애
E20-E35	기타 내분비선장애
E40-E46	영양실조
E50-E64	기타 영양결핍
E65-E68	비만 및 기타 과영양
E70-E90	대사장애
E35*	달리 분류된 질환에서의 내분비선의 장애
E90*	달리 분류된 질환에서의 영양 및 대사 장애
F00-F09	증상성을 포함하는 기질성 정신장애
F10-F19	정신활성물질 사용에 의한 정신 및 행동 장애
F20-F29	조현병, 분열형 및 망상장애
F30-F39	기분[정동]장애
F40-F48	신경증성, 스트레스-연관 및 신체형 장애
F50-F59	생리적 장애 및 신체적 요인들과 수반된 행동증후군
F60-F69	성인 인격 및 행동의 장애
F70-F79	정신지체
F80-F89	정신발달장애
F90-F98	소아기 및 청소년기에 주로 발병하는 행동 및 정서 장애
F99	상세불명의 정신장애
F00*	알츠하이머병에서의 치매
F02*	달리 분류된 기타 질환에서의 치매
G00-G09	중추신경계통의 염증성 질환
G10-G14	일차적으로 중추신경계통에 영향을 주는 계통적 위축
G20-G26	추체외로 및 운동 장애
G30-G32	신경계통의 기타 퇴행성 질환
G35-G37	중추신경계통의 탈수초질환
G40-G47	우발적 및 발작성 장애
G50-G59	신경, 신경근 및 신경총 장애
G60-G64	다발신경병증 및 말초신경계통의 기타 장애

G70-G73	신경근접합부 및 근육질환
G80-G83	뇌성마비 및 기타 마비증후군
G90-G99	신경계통의 기타 장애
G01*	달리 분류된 세균성 질환에서의 수막염
G02*	달리 분류된 기타 감염성 및 기생충성 질환에서의 수막염
G05*	달리 분류된 질환에서의 뇌염, 척수염 및 뇌척수염
G07*	달리 분류된 질환에서의 두 개 내 및 척추 내 농양 및 육아종
G13*	달리 분류된 질환에서의 일차적으로 중추신경계통을 침범하는 계통적 위축
G22*	달리 분류된 질환에서의 파킨슨증
G26*	달리 분류된 질환에서의 추체외로 및 운동 장애
G32*	달리 분류된 질환에서의 신경계통의 기타 퇴행성 장애
G46*	뇌혈관질환에서의 혈관증후군
G53*	달리 분류된 질환에서의 뇌신경장애
G55*	달리 분류된 질환에서의 신경근 및 신경총 압박
G59*	달리 분류된 질환에서의 단일신경병증
G63*	달리 분류된 질환에서의 다발신경병증
G73*	달리 분류된 질환에서의 신경근접합부 및 근육의 장애
G94*	달리 분류된 질환에서의 뇌의 기타 장애
G99*	달리 분류된 질환에서의 신경계통의 기타 장애
H00-H06	눈꺼풀, 눈물계통 및 안와의 장애
H10-H13	결막의 장애
H15-H22	공막, 각막, 홍채 및 섬모체의 장애
H25-H28	수정체의 장애
H30-H36	맥락막 및 망막의 장애
H40-H42	녹내장
H43-H45	유리체 및 안구의 장애
H46-H48	시신경 및 시각경로의 장애
H49-H52	안근, 양안운동, 조절 및 굴절의 장애

H53-H54	시각장애 및 실명
H55-H59	눈 및 눈부속기의 기타 장애
H03*	달리 분류된 질환에서의 눈꺼풀의 장애
H06*	달리 분류된 질환에서의 눈물계통 및 안와의 장애
H13*	달리 분류된 질환에서의 결막의 장애
H19*	달리 분류된 질환에서의 공막 및 각막의 장애
H22*	달리 분류된 질환에서의 홍채 및 섬모체의 장애
H28*	달리 분류된 질환에서의 백내장 및 수정체의 기타 장애
H32*	달리 분류된 질환에서의 맥락망막장애
H36*	달리 분류된 질환에서의 망막장애
H42*	달리 분류된 질환에서의 녹내장
H45*	달리 분류된 질환에서의 유리체 및 안구의 장애
H48*	달리 분류된 질환에서의 시신경 및 시각경로의 장애
H58*	달리 분류된 질환에서의 눈 및 눈부속기의 기타 장애
H60-H62	외이의 질환
H65-H75	중이 및 유도의 질환
H80-H83	내이의 질환
H90-H95	귀의 기타 장애
H62*	달리 분류된 질환에서의 외이의 장애
H67*	달리 분류된 질환에서의 중이염
H75*	달리 분류된 질환에서의 중이 및 유도의 기타 장애
H82*	달리 분류된 질환에서의 현기증후군
H94*	달리 분류된 질환에서의 귀의 기타 장애
I00-I02	급성 류마티스열
I05-I09	만성 류마티스심장질환
I10-I15	고혈압성 질환
I20-I25	허혈심장질환
I26-I28	폐성 심장병 및 폐순환의 질환
I30-I52	기타 형태의 심장병

I60-I69	뇌혈관질환
I70-I79	동맥, 소동맥 및 모세혈관의 질환
I80-I89	달리 분류되지 않은 정맥, 림프관 및 림프절의 질환
I95-I99	순환계통의 기타 및 상세불명의 장애
I32*	달리 분류된 질환에서의 심장막염
I39*	달리 분류된 질환에서의 심내막염 및 심장판막장애
I41*	달리 분류된 질환에서의 심근염
I43*	달리 분류된 질환에서의 심근병증
I52*	달리 분류된 질환에서의 기타 심장장애
I68*	달리 분류된 질환에서의 뇌혈관장애
I79*	달리 분류된 질환에서의 동맥, 소동맥 및 모세혈관의 장애
I98*	달리 분류된 질환에서의 순환계통의 기타 장애
J00-J06	급성 상기도감염
J09-J18	인플루엔자 및 폐렴
J20-J22	기타 급성 하기도감염
J30-J39	상기도의 기타 질환
J40-J47	만성 하부호흡기질환
J60-J70	외부요인에 의한 폐질환
J80-J84	주로 간질에 영향을 주는 기타 호흡기질환
J85-J86	하기도의 화농성 및 괴사성 병태
J90-J94	흉막의 기타 질환
J95-J99	호흡계통의 기타 질환
J17*	달리 분류된 질환에서의 폐렴
J91*	달리 분류된 병태에서의 흉막삼출액
J99*	달리 분류된 질환에서의 호흡장애
K00-K14	구강, 침샘 및 턱의 질환
K20-K31	식도, 위 및 십이지장의 질환
K35-K38	충수의 질환
K40-K46	탈장

K50-K52	비감염성 장염 및 결장염
K55-K64	장의 기타 질환
K65-K67	복막질환
K70-K77	간의 질환
K80-K87	담낭, 담도 및 췌장의 장애
K90-K93	소화계통의 기타 질환
K23*	달리 분류된 질환에서의 식도의 장애
K67*	달리 분류된 감염성 질환에서의 복막의 장애
K77*	달리 분류된 질환에서의 간장애
K87*	달리 분류된 질환에서의 담낭, 담도 및 췌장의 장애
K93*	달리 분류된 질환에서의 기타 소화기관의 장애
L00-L08	피부 및 피하조직의 감염
L10-L14	수포성 장애
L20-L30	피부염 및 습진
L40-L45	구진비늘 장애
L50-L54	두드러기 및 홍반
L55-L59	피부 및 피하조직의 방사선-관련 장애
L60-L75	피부부속물의 장애
L80-L99	피부 및 피하조직의 기타 장애
L14*	달리 분류된 질환에서의 수포성 장애
L45*	달리 분류된 질환에서의 구진비늘 장애
L54*	달리 분류된 질환에서의 홍반
L62*	달리 분류된 질환에서의 손발톱장애
L86*	달리 분류된 질환에서의 각피증
L99*	달리 분류된 질환에서의 피부 및 피하조직의 기타 장애
M00-M25	관절병증
M00-M03	감염성 관절병증
M05-M14	염증성 다발관절병증
M15-M19	관절증

M20-M25	기타 관절장애
M30-M36	전신결합조직장애
M40-M54	등병증
M40-M43	변형성 등병증
M45-M49	척추병증
M50-M54	기타 등병증
M60-M79	연조직장애
M60-M63	근육장애
M65-M68	윤활막 및 힘줄장애
M70-M79	기타 연조직장애
M80-M94	골병증 및 연골병증
M80-M85	골밀도 및 구조장애
M86-M90	기타 골병증
M91-M94	연골병증
M95-M99	근골격계통 및 결합조직의 기타 장애
M01*	달리 분류된 감염성 및 기생충 질환에서의 관절의 직접 감염
M03*	달리 분류된 질환에서의 감염후 및 반응성 관절병증
M07*	건선성 및 장병증성 관절병증
M09*	달리 분류된 질환에서의 연소성 관절염
M14*	달리 분류된 기타 질환에서의 관절병증
M36*	달리 분류된 질환에서의 결합조직의 전신장애
M49*	달리 분류된 질환에서의 척추병증
M63*	달리 분류된 질환에서의 근육장애
M68*	달리 분류된 질환에서의 윤활막 및 힘줄의 장애
M73*	달리 분류된 질환에서의 연조직장애
M82*	달리 분류된 질환에서의 골다공증
M90*	달리 분류된 질환에서의 골병증
N00-N08	사구체질환
N10-N16	신세뇨관-간질질환

N17-N19	신부전
N20-N23	요로결석증
N25-N29	신장 및 요관의 기타 장애
N30-N39	비뇨계통의 기타 질환
N40-N51	남성생식기관의 질환
N60-N64	유방의 장애
N70-N77	여성골반기관의 염증성 질환
N80-N98	여성생식관의 비염증성 장애
N99	비뇨생식계통의 기타 장애
N08*	달리 분류된 질환에서의 사구체장애
N16*	달리 분류된 질환에서의 신세뇨관-간질장애
N22*	달리 분류된 질환에서의 요로의 결석
N29*	달리 분류된 질환에서의 신장 및 요관의 기타 장애
N33*	달리 분류된 질환에서의 방광장애
N37*	달리 분류된 질환에서의 요도장애
N51*	달리 분류된 질환에서의 남성생식기관의 장애
N74*	달리 분류된 질환에서의 여성골반염증장애
N77*	달리 분류된 질환에서의 외음질의 궤양 및 염증
O00-O08	유산된 임신
O10-O16	임신, 출산 및 산후기에서의 부종, 단백뇨 및 고혈압성 장애
O20-O29	주로 임신에 관련된 기타 산모장애
O30-O48	태아와 양막강 그리고 가능한 분만문제에 관련된 산모관리
O60-O75	진통 및 분만의 합병증
O80-O84	분만
O85-O92	주로 산후기에 관련된 합병증
O94-O99	달리 분류되지 않은 기타 산과적 병태
P00-P04	산모의 요인과 임신, 진통 및 분만의 합병증에 의해 영향을 받은 태아 및 신생아
P05-P08	임신기간 및 태아성장에 관련된 장애

P10-P15	출산외상
P20-P29	출생전후기에 특이한 호흡 및 심혈관 장애
P35-P39	출생전후기에 특이한 감염
P50-P61	태아 및 신생아의 출혈성 및 혈액학적 장애
P70-P74	태아 및 신생아에 특이한 일과성 내분비 및 대사 장애
P75-P78	태아 및 신생아의 소화계통장애
P80-P83	태아 및 신생아의 외피 및 체온조절에 관련된 병태
P90-P96	출생전후기에 기원한 기타 장애
P75*	낭성 섬유증에서의 태변장폐색
Q00-Q07	신경계통의 선천기형
Q10-Q18	눈, 귀, 얼굴 및 목의 선천기형
Q20-Q28	순환계통의 선천기형
Q30-Q34	호흡계통의 선천기형
Q35-Q37	구순열 및 구개열
Q38-Q45	소화계통의 기타 선천기형
Q50-Q56	생식기관의 선천기형
Q60-Q64	비뇨계통의 선천기형
Q65-Q79	근골격계통의 선천기형 및 변형
Q80-Q89	기타 선천기형
Q90-Q99	달리 분류되지 않은 염색체이상
R00-R09	순환계통 및 호흡계통의 증상 및 징후
R10-R19	소화계통 및 복부의 증상 및 징후
R20-R23	피부 및 피하조직의 증상 및 징후
R25-R29	신경계통 및 근골격계통의 증상 및 징후
R30-R39	비뇨계통의 증상 및 징후
R40-R46	인지, 지각, 정서상태 및 행위에 관한 증상 및 징후
R47-R49	말하기 및 음성에 관한 증상 및 징후
R50-R69	전신증상 및 징후
R70-R79	진단명 없는 혈액 검사상의 이상소견

R80-R82	진단명 없는 요검사상의 이상소견
R83-R89	진단명 없는 기타 체액, 물질 및 조직 검사상의 이상소견
R90-R94	진단명 없는 진단영상 및 기능 검사상 이상소견
R95-R99	불명확하고 원인불명의 사인
S00-S09	머리의 손상
S10-S19	목의 손상
S20-S29	흉부의 손상
S30-S39	복부, 아래등, 요추 및 골반의 손상
S40-S49	어깨 및 위팔의 손상
S50-S59	팔꿈치 및 아래팔의 손상
S60-S69	손목 및 손의 손상
S70-S79	고관절 및 대퇴의 손상
S80-S89	무릎 및 아래다리의 손상
S90-S99	발목 및 발의 손상
T00-T07	여러 신체부위를 침범한 손상
T08-T14	몸통, 사지 또는 신체부위의 상세불명 부분의 손상
T15-T19	자연개구를 통해 들어온 이물의 영향
T20-T32	화상 및 부식
T33-T35	동상
T36-T50	약물, 약제 및 생물학적 물질에 의한 중독
T51-T65	출처가 주로 비의약품인 물질의 독성효과
T66-T78	외인의 기타 및 상세불명의 영향
T79	외상의 특정 조기합병증
T80-T88	달리 분류되지 않은 외과적 및 내과적 치료의 합병증
T90-T98	손상, 중독 및 외인에 의한 기타 결과의 후유증
V01-X59	사고
V01-V99	운수사고
V01-V09	운수사고에서 다친 보행자
V10-V19	운수사고에서 다친 자전거 탑승자

V20-V29	사고운수사고에서 다친 모터사이클 탑승자
V30-V39	운수사고에서 다친 삼륜자동차 탑승자
V40-V49	운수사고에서 다친 승용차 탑승자
V50-V59	운수사고에서 다친 픽업트럭 또는 밴 탑승자
V60-V69	운수사고에서 다친 대형화물차 탑승자
V70-V79	운수사고에서 다친 버스 탑승자
V80-V89	기타 육상 운수사고
V90-V94	수상 운수사고
V95-V97	항공 및 우주 운수사고
V98-V99	기타 및 상세불명의 운수사고
W00-X59	불의의 손상의 기타 외인
W00-W19	낙상
W20-W49	무생물성 기계적 힘에 노출
W50-W64	생물성 기계적 힘에 노출
W65-W74	불의의 익사 및 익수
W75-W84	기타 불의의 호흡 위험
W85-W99	전류, 방사선, 및 극단적 기온 및 압력에 노출
X00-X09	연기, 불 및 불꽃에 노출
X10-X19	열 및 가열된 물질과의 접촉
X20-X29	독액성 동물 및 식물과의 접촉
X30-X39	자연의 힘에 노출
X40-X49	유독성 물질에 의한 불의의 중독 및 노출
X50-X57	과잉노력, 여행 및 결핍
X58-X59	기타 및 상세불명의 요인에 대한 사고피폭
X60-X84	고의적 자해
X85-Y09	가해
Y10-Y34	의도 미확인 사건
Y35-Y36	법적 개입 및 전쟁행위
Y40-Y84	내과적 및 외과적 치료의 합병증

Y40-Y59	치료용으로 사용시 유해작용을 일으키는 약물, 약제 및 생물학적 물질
Y60-Y69	외과적 및 내과적 치료 중 환자의 재난
Y70-Y82	진단 및 치료용으로 사용시 유해사건과 관련된 의료장치
Y83-Y84	처치 당시에는 재난에 대한 언급이 없었으나 환자의 이상반응 또는 이후 합병증의 원인이 된 외과적 및 내과적 처치
Y85-Y89	질병이환과 사망의 외인의 후유증
Y90-Y98	달리 분류된 질병이환율 및 사망원인에 관련된 보조요인
Z00-Z13	검사 및 조사를 위해 보건서비스와 접하고 있는 사람
Z20-Z29	전염성 질환과 관련되어 잠재적인 건강위험이 있는 사람
Z30-Z39	생식에 관련된 상황에서 보건서비스와 접하고 있는 사람
Z40-Z54	특정 처치 및 건강관리를 위하여 보건서비스와 접하고 있는 사람
Z55-Z65	사회경제적 및 정신사회적 상황에 관련된 잠재적 건강위험이 있는 사람
Z70-Z76	기타 상황에서 보건서비스와 접하고 있는 사람
Z80-Z99	가족 및 개인 기왕력과 건강상태에 영향을 주는 특정 병태에 관련된 잠재적 건강 위험을 가진 사람
U00-U19	병인이 불확실한 신종질환의 임시적 지정이나 응급사용
U82-U85	항균제 및 항암제 내성
U99	재발한 악성 신생물
U22-U32	한의병명
U50-U79	한의병증
U95-U98	사상체질병증

질병의 전반적인 추세분석 외에 직업코호트 구축 후 우선적 연구가 필요하다고 생각되는 우선순위질병을 선정하기 위하여 다빈도상병을 분석하는 것을 선택하였음, 현재 구축된 건강보험공단 데이터에서 2002년부터 2015년까지 중 남성 근로자 10만 명 당 해당질환 입원근로자 수, 남성근로자 1,000명 당 해당질환 입원 환자 수, 해당질병 입원근로자 1명 당 입원일수의 지표를 통해 분석한 상위 20개 다빈도 입원 상병은 다음 <표 3-11>과 같다.

<표 3-11> 2002-2015년 남성 근로자의 다빈도 입원 상위 20개 상병

2002			2003			2004		
1	I8	643.83081	1	I8	726.70020	1	I8	749.97469
2	K6	238.84799	2	K6	296.39373	2	K6	321.35835
3	J3	228.65523	3	J3	278.58831	3	J3	300.52849
4	C1	179.32441	4	M5	230.13817	4	M5	263.14114
5	M5	177.65553	5	C1	200.37594	5	S8	232.76911
6	C2	160.19504	6	S8	191.37806	6	C1	220.26787
7	H2	153.93223	7	C2	184.62964	7	I2	199.35489
8	K3	148.44106	8	I2	179.09248	8	K3	197.85873
9	I2	138.62515	9	K3	177.29291	9	C2	197.47638
10	S8	135.37710	10	H2	168.20851	10	H2	187.43549
11	K7	102.30446	11	K7	129.67334	11	S6	152.94071
12	I6	97.38753	12	K2	125.27822	12	K7	148.76809
13	K2	92.90128	13	I6	125.01867	13	I6	143.44841
14	S0	84.66452	14	S0	117.59541	14	K2	141.18755
15	S6	84.23384	15	S6	112.35242	15	S0	138.12873
16	D1	78.63500	16	K8	95.68903	16	K8	112.44467
17	K8	78.24021	17	D1	90.98244	17	D1	107.65696
18	C3	75.47668	18	E1	82.78053	18	E1	102.18767
19	E1	67.20404	19	C3	82.62479	19	S3	98.26441
20	J0	64.45845	20	A0	79.16407	20	A0	94.02529
2005			2006			2007		
1	I8	781.6289	1	I8	802.1514	1	I8	795.0823
2	K6	340.3065	2	K6	360.1336	2	M5	405.9675
3	J3	328.3492	3	M5	358.0569	3	K6	378.5137
4	M5	312.6611	4	J3	341.3355	4	S8	355.4685
5	S8	276.2629	5	S8	321.5067	5	J3	336.9055

6	C1	240.0401	6	C1	249.1754	6	C1	269.5539
7	I2	216.2848	7	I2	241.8838	7	I2	246.1990
8	C2	211.2149	8	K3	220.5013	8	S6	240.7732
9	K3	209.0466	9	S6	214.4096	9	C2	226.5008
10	S6	190.2815	10	C2	213.1328	10	K3	216.3862
11	H2	189.8670	11	H2	194.7346	11	H2	214.3958
12	K2	162.3650	12	S0	175.3211	12	S0	189.1094
13	K7	161.7592	13	I6	167.4296	13	I6	180.7052
14	S0	161.6954	14	K7	166.2605	14	K7	169.2195
15	I6	148.1438	15	K2	163.7992	15	K2	165.1058
16	D1	137.7169	16	D1	154.6462	16	D1	161.2281
17	K8	128.6612	17	S3	138.2017	17	S3	159.3408
18	S3	117.0227	18	K8	134.6944	18	K8	140.5272
19	E1	113.9616	19	E1	116.7269	19	S9	120.6815
20	A0	107.6482	20	A0	113.6811	20	A0	119.6346
2008			2009			2010		
1	I8	781.3010	1	I8	771.8965	1	I8	758.6547
2	M5	471.3032	2	M5	553.2909	2	M5	642.3312
3	S8	400.9382	3	S8	432.7559	3	S8	458.7893
4	K6	387.2489	4	K6	394.2550	4	K6	413.2179
5	J3	346.3464	5	J3	356.3312	5	J3	353.2501
6	C1	271.8001	6	S6	283.2882	6	S6	295.9783
7	S6	268.9711	7	C1	282.1878	7	C1	286.5421
8	I2	254.0398	8	I2	258.4216	8	I2	263.4161
9	C2	233.7127	9	C2	236.6013	9	H2	241.3634
10	H2	225.1707	10	H2	229.2876	10	C2	240.3818
11	K3	215.1659	11	S0	218.7532	11	S3	225.3702
12	S0	202.6081	12	K3	217.8809	12	I6	224.0614
13	I6	200.5519	13	S3	216.8745	13	S0	222.6218

14	S3	184.7374	14	I6	199.4021	14	K3	211.1439
15	D1	175.8227	15	D1	190.2365	15	D1	204.3513
16	K2	172.3590	16	B1	185.8752	16	K8	174.3674
17	K7	170.9238	17	K2	178.4138	17	A0	172.9932
18	K8	151.3282	18	K7	169.9192	18	K2	172.6660
19	S9	136.0242	19	K8	162.6726	19	S4	165.2191
20	F1	129.9247	20	A0	156.6741	20	K7	157.9816
2011			2012			2013		
1	I8	783.1981	1	M5	884.9931	1	M5	977.3320
2	M5	763.8307	2	I8	769.2238	2	I8	740.6437
3	S8	467.7105	3	S8	468.7469	3	S8	479.8181
4	K6	434.8251	4	K6	451.3116	4	K6	455.7989
5	J3	367.4751	5	J3	362.7964	5	J3	346.6860
6	S6	304.3194	6	S6	311.8653	6	S6	317.6530
7	C1	295.2358	7	C1	286.1198	7	H2	263.1260
8	I2	262.5273	8	I2	271.1422	8	I2	262.8547
9	C2	257.3222	9	H2	251.2491	9	M7	260.8610
10	H2	252.1551	10	C2	251.0058	10	C1	256.1775
11	I6	236.5020	11	I6	243.7056	11	I6	252.2726
12	K3	220.9121	12	K3	219.3350	12	C2	238.0805
13	D1	214.8732	13	M7	213.8233	13	A0	218.0959
14	S3	212.5865	14	D1	213.7017	14	M2	216.6448
15	S0	208.8722	15	S0	208.2265	15	K3	215.7601
16	K8	184.0723	16	S3	206.6813	16	S3	212.8579
17	A0	183.1627	17	M2	200.7681	17	S4	210.8406
18	M2	177.7050	18	A0	195.9378	18	K8	201.5326
19	S4	174.9508	19	S4	192.4337	19	S0	201.3320
20	K2	170.3901	20	K8	191.9592	20	D1	199.7984
2014			2015					

1	M5	1064.4436	1	M5	1099.4859
2	I8	694.8563	2	I8	652.8799
3	S8	474.6948	3	S8	474.4279
4	K6	458.2128	4	K6	459.3917
5	J3	335.9365	5	M7	338.2853
6	S6	333.1780	6	S6	335.2624
7	M7	311.9574	7	J3	311.0679
8	H2	280.5155	8	H2	297.0618
9	I2	272.5950	9	I6	276.6068
10	I6	259.1575	10	I2	266.2954
11	C1	244.4038	11	A0	252.6363
12	C2	234.1368	12	M2	245.6277
13	M2	229.2609	13	C1	232.0918
14	A0	225.8615	14	S3	226.0908
15	S3	219.9325	15	C2	224.3778
16	S4	212.9963	16	S4	211.5025
17	K8	205.4535	17	K8	209.6887
18	K3	204.8011	18	K3	204.0124
19	D1	196.1480	19	S9	193.1299
20	S0	195.9992	20	M4	192.5030

추가적으로, 2002년부터 2015년까지의 남성 근로자 1,000명 당 입원일수가 많은 상위 20개 상병에 대하여서도 구하였으며 이는 <표 3-12>와 같다.

<표 3-12> 2002-2015년 남성 근로자 1,000명 당 입원일수 상위 20개 상병

2002			2003			2004		
1	I8	24.517	1	S8	28.833	1	S8	34.008
2	C1	20.869	2	I8	27.198	2	M5	28.353

3	S8	20.363	3	M5	24.748	3	I8	26.806
4	M5	20.339	4	I6	22.603	4	I6	24.086
5	C2	19.530	5	C1	21.762	5	C1	23.852
6	I6	18.403	6	C2	21.678	6	C2	22.156
7	K7	14.162	7	K7	17.235	7	K7	19.789
8	E1	11.240	8	E1	13.521	8	E1	16.642
9	K3	9.433	9	S0	12.336	9	S0	13.971
10	S0	9.400	10	K2	11.809	10	K2	13.707
11	I2	9.158	11	K3	11.193	11	S6	12.336
12	K6	8.973	12	K6	10.860	12	K3	12.159
13	K2	8.687	13	I2	10.331	13	F1	11.420
14	C3	8.425	14	J3	10.148	14	S3	11.363
15	J3	8.286	15	F1	9.482	15	I2	11.242
16	K8	7.631	16	S6	9.440	16	K6	11.192
17	S3	7.364	17	S3	8.868	17	J3	10.960
18	S6	7.089	18	C3	8.810	18	I1	10.657
19	S9	6.679	19	K8	8.794	19	S9	10.209
20	F1	5.438	20	I1	8.477	20	K8	9.714
2005			2006			2007		
1	S8	41.759	1	S8	49.714	1	S8	53.081
2	M5	36.062	2	M5	41.257	2	M5	46.114
3	I8	27.714	3	I6	29.614	3	I6	31.122
4	I6	27.025	4	I8	27.958	4	C1	28.009
5	C1	26.236	5	C1	26.827	5	I8	27.230
6	C2	24.149	6	C2	25.379	6	C2	27.144
7	K7	21.462	7	K7	22.148	7	K7	23.061
8	E1	19.904	8	F1	19.597	8	F1	21.669
9	S0	16.160	9	E1	18.746	9	S6	20.168
10	S6	15.963	10	S6	18.187	10	E1	18.134

11	F1	15.478	11	S0	17.448	11	S3	18.035
12	K2	15.156	12	S3	16.398	12	S0	17.959
13	S3	13.693	13	K2	15.518	13	S9	16.612
14	K3	13.112	14	S9	15.415	14	K2	14.246
15	S9	12.398	15	K3	13.703	15	K3	13.343
16	J3	12.019	16	J3	12.943	16	K6	12.854
17	K6	11.898	17	K6	12.502	17	J3	12.677
18	I2	11.868	18	I2	12.106	18	I2	12.515
19	I1	11.577	19	K8	11.327	19	S4	11.831
20	K8	10.974	20	I1	11.240	20	M4	11.822
2008			2009			2010		
1	S8	59.513	1	S8	60.969	1	M5	64.852
2	M5	53.865	2	M5	59.984	2	S8	62.566
3	I6	37.309	3	I6	34.830	3	I6	39.691
4	C1	28.914	4	C1	27.850	4	C1	28.775
5	C2	27.615	5	C2	26.140	5	C2	27.310
6	I8	26.711	6	I8	26.002	6	F1	25.548
7	F1	24.481	7	S3	24.514	7	I8	25.252
8	S6	23.411	8	F1	23.983	8	S3	25.046
9	K7	22.389	9	S6	23.922	9	S6	24.932
10	S3	21.489	10	K7	21.045	10	S0	21.011
11	S0	19.960	11	S0	20.685	11	S9	20.297
12	E1	19.495	12	S9	19.905	12	S4	20.285
13	S9	18.635	13	B1	18.571	13	K7	19.086
14	S4	15.395	14	E1	18.141	14	G8	18.594
15	G8	14.945	15	S4	18.056	15	E1	18.183
16	K2	14.799	16	G8	16.024	16	M4	15.901
17	M4	13.618	17	M4	15.034	17	M2	15.262
18	M2	13.424	18	M2	14.150	18	S2	13.725

19	K3	13.188	19	K2	14.105	19	K2	13.592
20	I2	13.186	20	J3	13.643	20	K6	13.340
2011			2012			2013		
1	M5	70.833	1	M5	70.670	1	M5	71.525
2	S8	63.738	2	S8	60.523	2	S8	58.983
3	I6	42.366	3	I6	42.104	3	I6	43.779
4	C1	29.314	4	C2	28.224	4	C2	26.799
5	C2	28.496	5	C1	27.761	5	C1	25.938
6	F1	26.875	6	F1	26.503	6	S6	25.782
7	S6	25.410	7	S6	25.589	7	F1	24.241
8	I8	25.329	8	I8	24.090	8	S4	23.169
9	S3	22.924	9	G8	21.588	9	I8	22.141
10	G8	20.959	10	S4	21.455	10	G8	21.827
11	S4	20.582	11	S3	20.984	11	S3	20.810
12	S0	20.278	12	S9	19.313	12	M7	20.637
13	S9	18.930	13	S0	18.849	13	S9	19.989
14	M4	17.588	14	M7	18.685	14	S0	18.360
15	M2	15.866	15	M2	17.129	15	M2	17.649
16	M7	15.378	16	M4	16.623	16	M4	17.038
17	K7	15.210	17	K6	13.397	17	K8	13.609
18	K8	13.568	18	K8	13.242	18	K6	12.977
19	E1	13.438	19	K7	13.065	19	S2	12.949
20	K6	13.364	20	S2	12.947	20	K7	12.624
2014			2015					
1	M5	74.725	1	M5	69.866			
2	S8	55.786	2	S8	53.353			
3	I6	44.836	3	I6	48.149			
4	C2	26.772	4	S6	25.604			
5	S6	26.470	5	C2	25.019			

6	C1	25.307	6	C1	24.373
7	M7	22.851	7	G8	23.116
8	F1	22.669	8	M7	23.053
9	G8	22.129	9	F1	22.024
10	S4	22.010	10	S9	20.128
11	S9	20.608	11	S4	20.062
12	I8	20.042	12	S3	18.864
13	S3	20.032	13	I8	18.477
14	S0	18.372	14	M2	18.044
15	M2	18.186	15	S0	17.051
16	M4	17.396	16	M4	16.835
17	K8	13.613	17	K8	13.743
18	K6	12.598	18	J1	12.708
19	I2	12.598	19	K6	12.514
20	S2	12.398	20	I2	12.347

다음으로, 2002년부터 2015년 중 남성 근로자에게서 입원을 경험한 근로자 1명 당 평균 입원일수를 구하여 상위 20개 상병에 대하여 다음 <표 3-13>과 같이 기재하였다.

<표 3-13> 2002-2015년 남성 근로자 의 입원을 경험한 근로자 1명 당 평균 입원일수 상위 20개 상병

2002			2003			2004		
1	G8	36.713	1	J7	52.000	1	G8	35.692
2	W2	28.000	2	V8	45.000	2	F0	27.108
3	R7	27.929	3	G8	36.792	3	B4	24.622
4	L8	26.676	4	Z2	26.667	4	E7	24.033
5	F0	25.080	5	F7	25.800	5	B3	23.847
6	F2	24.055	6	F0	24.650	6	F7	23.606

7	B3	23.558	7	B3	23.919	7	L6	22.505
8	B4	23.396	8	E7	23.800	8	F9	22.438
9	F6	20.529	9	F2	23.232	9	F2	22.362
10	C9	20.493	10	G9	21.205	10	F6	21.344
11	L6	19.714	11	I1	20.506	11	Q0	21.200
12	E7	19.298	12	F5	20.500	12	I1	20.862
13	G9	19.099	13	J6	20.422	13	C9	19.364
14	S7	19.034	14	S7	20.293	14	S7	18.841
15	M8	19.026	15	B2	20.243	15	F3	18.687
16	I6	18.897	16	R7	19.786	16	K0	17.743
17	I1	18.789	17	G6	19.596	17	M8	17.628
18	F3	18.454	18	F6	19.543	18	L1	17.556
19	F9	18.375	19	C9	19.049	19	F1	17.115
20	F8	18.000	20	L8	18.784	20	X0	17.000
2005			2006			2007		
1	A	155.000	1	A3	32.906	1	Q9	98.000
2	G8	33.606	2	G8	32.056	2	Y0	48.000
3	Y1	28.000	3	F8	29.000	3	U0	30.500
4	A3	27.636	4	B4	26.364	4	F8	30.333
5	F0	25.745	5	F0	24.975	5	G8	29.419
6	F7	25.070	6	L3	23.494	6	J7	27.000
7	F2	23.401	7	F2	23.445	7	F0	23.823
8	E7	22.884	8	F7	22.922	8	C9	23.296
9	F6	22.653	9	F6	22.516	9	F2	23.145
10	B3	21.944	10	C9	22.282	10	L1	21.895
11	G1	21.189	11	E7	22.124	11	F6	21.489
12	B4	21.184	12	W5	22.000	12	F7	20.593
13	C9	20.763	13	B3	20.313	13	G9	20.120
14	G9	20.582	14	E6	20.000	14	B3	19.353

15	I1	19.977	15	G9	19.740	15	F3	19.112
16	L1	19.714	16	M8	19.276	16	F1	18.691
17	F3	18.827	17	I1	19.097	17	B4	18.579
18	F9	18.821	18	X8	19.000	18	M8	18.569
19	M8	18.607	19	G6	18.911	19	G1	18.241
20	I6	18.242	20	F3	18.888	20	G6	17.933
2008			2009			2010		
1	P9	44.000	1	G8	26.815	1	G8	27.311
2	ER	38.000	2	U8	26.493	2	F0	23.274
3	G8	29.144	3	F0	23.300	3	F2	22.106
4	L1	25.875	4	E6	23.000	4	F7	21.591
5	U8	25.777	5	F7	22.972	5	U8	20.864
6	C9	25.614	6	F2	22.155	6	F6	20.578
7	F0	23.386	7	C9	21.303	7	C9	20.200
8	F2	22.986	8	F6	20.765	8	F9	20.048
9	Q9	22.667	9	L3	19.987	9	G6	19.877
10	F7	21.824	10	P2	19.200	10	G9	19.796
11	F8	21.143	11	G6	19.061	11	E5	19.351
12	B3	19.866	12	F1	18.866	12	F1	19.108
13	F6	19.602	13	F3	18.837	13	F3	18.956
14	G9	19.549	14	Q0	18.686	14	L1	18.438
15	F3	19.494	15	G9	17.993	15	A3	17.836
16	B2	19.200	16	M8	17.941	16	E7	17.756
17	G6	18.968	17	B3	17.803	17	M8	17.738
18	F1	18.842	18	I6	17.467	18	I6	17.714
19	M8	18.807	19	E5	16.917	19	B3	17.481
20	I6	18.603	20	A4	16.604	20	A2	17.412
2011			2012			2013		
1	G8	27.775	1	Q9	34.000	1	G8	28.351

2	U8	27.723	2	U8	31.107	2	U8	27.949
3	F2	22.780	3	F8	29.375	3	P7	27.000
4	Q0	22.130	4	G8	27.991	4	F8	23.263
5	F6	21.425	5	F2	22.840	5	F2	23.078
6	F8	20.667	6	J7	22.571	6	P2	21.500
7	E5	20.607	7	E4	21.571	7	F7	20.323
8	F0	19.820	8	F7	20.582	8	F0	19.742
9	C9	19.552	9	F0	20.553	9	F6	19.714
10	E4	19.364	10	F6	19.572	10	F1	19.332
11	F1	19.235	11	E5	19.416	11	C9	17.689
12	F7	19.221	12	F1	19.214	12	F3	17.558
13	G9	18.754	13	C9	18.574	13	I6	17.354
14	F3	17.934	14	F3	18.087	14	G6	16.984
15	I6	17.913	15	I6	17.277	15	T9	16.358
16	G6	17.280	16	B4	17.011	16	B4	16.333
17	I3	16.623	17	G9	16.368	17	M8	16.227
18	B4	16.492	18	M8	15.800	18	I3	15.847
19	M8	16.371	19	M0	15.535	19	F9	15.674
20	T3	15.981	20	Q0	15.459	20	M0	15.267
2014			2015					
1	P1	60.000	1	G8	28.974			
2	G8	28.105	2	F8	27.000			
3	U8	24.977	3	U8	23.269			
4	P2	23.750	4	F2	22.157			
5	F8	23.316	5	B4	21.534			
6	F2	23.098	6	Z8	21.500			
7	F7	20.113	7	F7	21.133			
8	F0	19.120	8	F6	20.352			
9	F1	19.046	9	F1	19.341			

10	Q0	18.971	10	U0	19.145
11	F6	18.908	11	F3	18.093
12	C9	18.292	12	G9	17.927
13	F3	17.986	13	F0	17.906
14	B6	17.667	14	Q0	17.814
15	I6	17.301	15	C9	17.499
16	G9	17.286	16	I6	17.407
17	E5	17.177	17	T3	16.444
18	G6	16.784	18	F9	15.455
19	T3	16.509	19	Z6	15.000
20	A4	15.989	20	P2	15.000

같은 방법으로, 여성 근로자에서도 2002년부터 2015년까지 근로자 10만 명 당 해당질환 입원근로자 수, 근로자 1,000명 당 입원일 수, 입원한 근로자 1명 당 평균 입원일수를 통한 상위 20개 다빈도상병 분석 결과는 다음 <표3-14>, <표3-15>, <표3-16>과 같다.

<표 3-14> 2002-2015년 여성 근로자의 다빈도 입원 상위 20개 상병

2002			2003			2004		
1	O8	2934.16028	1	O8	3241.08342	1	O8	3118.11879
2	I8	566.46844	2	I8	661.10053	2	I8	755.27177
3	D2	342.58380	3	D2	425.47227	3	D2	440.10280
4	O4	168.89931	4	O4	231.29081	4	O4	273.02129
5	K3	165.53196	5	K3	197.19359	5	K3	216.29676
6	N8	161.01262	6	N8	194.90437	6	N8	211.40100
7	C5	155.65144	7	O0	171.65079	7	O2	190.78729
8	O0	149.62565	8	C5	171.32950	8	C5	183.38844

9	O6	143.20109	9	O2	170.80740	9	O0	177.83010
10	J3	143.06817	10	J3	162.37346	10	J3	171.83003
11	O2	133.36488	11	O6	151.89127	11	O6	165.13058
12	M5	111.43278	12	M5	137.79454	12	M5	164.68885
13	K6	108.64142	13	K6	127.35252	13	N1	150.55374
14	O3	104.47654	14	N1	125.90670	14	K6	146.24694
15	N7	86.66502	15	A0	103.21538	15	A0	131.67010
16	N1	85.20288	16	N7	100.00245	16	Z3	124.01358
17	A0	80.50631	17	O3	99.07873	17	N7	107.96433
18	J0	71.20178	18	Z3	93.97821	18	O3	107.66985
19	Z3	69.47380	19	S8	77.55210	19	S8	90.77396
20	C1	65.92922	20	K2	71.72866	20	K2	90.22181
2005			2006			2007		
1	O8	3058.9043	1	O8	3272.3505	1	O8	3736.4548
2	I8	779.6611	2	I8	836.6565	2	I8	816.1575
3	D2	464.9669	3	D2	495.3507	3	D2	522.5922
4	O4	292.5822	4	O4	318.0609	4	O4	344.6266
5	K3	228.1943	5	N8	252.6768	5	M5	284.7860
6	N8	227.3741	6	N3	245.9129	6	C5	255.2474
7	C5	204.5786	7	C5	230.9562	7	N8	253.0746
8	O2	201.3660	8	K3	228.2570	8	K3	225.0041
9	M5	199.1104	9	M5	224.7639	9	O2	217.2231
10	J3	195.2827	10	O2	212.7922	10	N3	210.0880
11	O0	189.8828	11	O0	199.3597	11	O0	199.3414
12	N1	168.7961	12	J3	196.5970	12	J3	196.9924
13	O6	160.5255	13	N1	184.8793	13	N1	185.9228
14	A0	148.2562	14	O6	171.2880	14	O6	180.8137
15	K6	144.5310	15	A0	160.9675	15	A0	168.5402
16	S8	118.6597	16	K6	142.0732	16	S8	151.0696

17	Z3	114.4560	17	S8	133.5627	17	S3	143.3473
18	N7	114.3193	18	N7	120.2573	18	K6	140.1468
19	S3	106.0828	19	S3	120.1620	19	C7	139.2072
20	K2	105.7069	20	K2	119.3364	20	K2	121.4136
2008			2009			2010		
1	O8	3482.7210	1	O8	3285.3790	1	O8	3424.1644
2	I8	825.8846	2	I8	810.0722	2	I8	772.1530
3	D2	570.9894	3	D2	597.9546	3	D2	632.6588
4	M5	341.8693	4	M5	421.3779	4	M5	531.4221
5	C5	290.9273	5	C5	307.1607	5	C5	356.5609
6	N8	272.6072	6	N8	301.1955	6	N8	323.1208
7	O4	264.6500	7	O4	236.4549	7	S8	257.1807
8	K3	223.4893	8	C7	228.2589	8	O6	256.3817
9	J3	210.7736	9	S3	225.1509	9	C7	252.9507
10	O6	210.7207	10	S8	224.2486	10	S3	247.4518
11	N3	204.7198	11	K3	220.8148	11	A0	243.3864
12	O2	194.0397	12	A0	220.1130	12	O4	235.3260
13	S8	189.9950	13	O6	219.9877	13	K3	212.4138
14	S3	189.1226	14	J3	213.6214	14	J3	203.3429
15	N1	188.7525	15	N3	209.2603	15	O2	199.1364
16	O0	184.6813	16	O2	188.5323	16	N1	185.6241
17	A0	179.8700	17	N1	182.3665	17	O0	185.0836
18	C7	177.8873	18	O0	181.9655	18	N3	184.6606
19	K6	143.0448	19	J0	176.8023	19	M2	176.6002
20	J0	136.0392	20	K6	149.0563	20	J0	158.9285
2011			2012			2013		
1	O8	3303.8274	1	O8	3323.7488	1	O8	2811.9002
2	I8	790.5344	2	M5	814.8396	2	M5	892.5661
3	D2	666.6382	3	I8	778.8875	3	I8	743.8790

4	M5	658.7933	4	D2	674.2046	4	D2	695.2797
5	C5	404.1858	5	C5	426.7574	5	C5	448.9702
6	N8	327.9066	6	N8	320.8460	6	N8	335.3407
7	C7	291.5529	7	C7	306.2072	7	A0	307.6605
8	S8	275.2057	8	A0	290.1148	8	O6	306.4277
9	O6	261.2252	9	S8	284.9553	9	S8	305.9654
10	A0	256.3167	10	O6	266.3241	10	C7	302.0359
11	S3	236.3320	11	M2	254.4902	11	M7	270.1757
12	N1	222.6583	12	S3	231.6004	12	M2	269.9253
13	O4	218.6482	13	M7	212.8054	13	S3	235.6188
14	K3	211.5046	14	K3	212.1298	14	K3	205.0107
15	J3	209.3790	15	O4	210.3690	15	J3	195.3024
16	M2	207.6260	16	J3	205.2506	16	M4	188.8495
17	J1	184.3981	17	N1	189.8747	17	N1	187.3855
18	N3	181.0235	18	M4	181.2347	18	N3	172.2452
19	O2	175.1728	19	N3	174.2736	19	O4	167.5837
20	O0	165.5748	20	J1	163.2382	20	H2	164.5017
2014			2015					
1	O8	2313.8303	1	O8	2138.4046			
2	M5	1007.8056	2	M5	1087.0564			
3	I8	710.2253	3	D2	686.9532			
4	D2	696.8300	4	I8	652.4289			
5	C5	443.0673	5	O6	500.6296			
6	O6	441.6073	6	C5	454.1618			
7	O3	366.4549	7	O3	435.9333			
8	N8	351.3076	8	M7	364.0293			
9	M7	330.4664	9	N8	360.8259			
10	A0	327.5099	10	A0	354.0185			
11	S8	308.1652	11	S8	313.6269			

12	M2	288.6562	12	M2	309.9359
13	S3	242.5391	13	S3	255.7555
14	C7	234.3085	14	J1	249.2092
15	M4	214.7995	15	M4	231.4509
16	J1	214.7265	16	H2	217.3312
17	K3	200.9479	17	M1	203.5946
18	N1	193.3378	18	N1	200.3389
19	J3	190.3631	19	S9	195.2725
20	H2	185.8736	20	K3	192.9047

<표 3-15> 2002-2015년 여성 근로자 1,000명 당 입원일수 상위 20개 상병

2002			2003			2004		
1	O8	134.981	1	O8	148.673	1	O8	144.988
2	D2	21.954	2	D2	26.601	2	D2	27.201
3	I8	19.597	3	I8	22.312	3	I8	24.746
4	C5	16.093	4	C5	17.453	4	C5	19.014
5	M5	12.521	5	O4	14.911	5	M5	18.125
6	O4	10.656	6	M5	14.809	6	O4	17.949
7	K3	9.941	7	K3	11.773	7	S8	13.052
8	N8	9.083	8	S8	11.247	8	K3	12.800
9	O6	8.230	9	N8	10.846	9	N8	11.282
10	C1	8.035	10	O2	9.297	10	N1	10.370
11	S8	8.023	11	N1	8.623	11	O2	10.296
12	O2	7.273	12	O6	8.432	12	C1	10.017
13	O3	6.913	13	S3	7.819	13	S3	9.478
14	I6	6.469	14	C1	7.814	14	I6	9.358
15	N1	6.316	15	I6	7.404	15	O6	9.278
16	S3	6.005	16	F2	7.118	16	K2	7.324

17	O0	5.290	17	O3	6.499	17	O3	7.246
18	J3	5.145	18	O0	6.443	18	O0	6.528
19	F2	4.978	19	J3	6.008	19	F2	6.522
20	N7	4.867	20	N7	5.785	20	N7	6.079
2005			2006			2007		
1	O8	141.801	1	O8	151.513	1	O8	170.584
2	D2	28.341	2	D2	29.496	2	M5	32.334
3	I8	25.979	3	I8	27.012	3	D2	30.572
4	M5	23.112	4	M5	26.711	4	I8	26.231
5	C5	20.327	5	C5	24.744	5	C5	25.938
6	O4	20.277	6	O4	23.200	6	O4	25.164
7	S8	17.733	7	S8	20.177	7	S8	22.266
8	S3	13.645	8	S3	15.384	8	S3	17.888
9	K3	13.639	9	K3	13.708	9	K3	13.397
10	N8	11.926	10	O2	12.781	10	N1	13.169
11	N1	11.423	11	N8	12.620	11	O2	13.059
12	O2	11.344	12	N1	12.464	12	N8	12.450
13	I6	10.907	13	I6	11.748	13	I6	12.357
14	C1	9.824	14	C1	10.079	14	C1	11.324
15	K2	9.170	15	K2	9.707	15	M4	10.839
16	O6	8.903	16	S9	8.913	16	S9	10.770
17	M4	7.613	17	C7	8.774	17	C7	10.705
18	S9	7.429	18	O6	8.758	18	M2	9.907
19	J3	7.043	19	N3	8.607	19	O6	9.638
20	C7	6.844	20	M4	8.520	20	K2	9.373
2008			2009			2010		
1	O8	161.249	1	O8	150.705	1	O8	156.398
2	M5	39.344	2	M5	45.590	2	M5	53.197
3	D2	32.189	3	D2	32.452	3	S8	34.859

98...빅데이터 기반 직업 코호트 구축을 통한 질병 발생 연구

4	C5	29.952	4	S8	31.540	4	C5	34.546
5	S8	28.188	5	C5	30.433	5	D2	33.505
6	I8	26.559	6	S3	26.843	6	S3	29.301
7	S3	23.881	7	I8	25.937	7	I8	24.251
8	O4	19.102	8	I6	16.762	8	M2	19.307
9	I6	14.891	9	O4	16.323	9	C7	18.021
10	O6	14.790	10	C7	16.236	10	O6	17.864
11	M4	13.928	11	M2	16.166	11	M1	17.199
12	S9	13.847	12	O6	15.311	12	M4	16.691
13	N1	13.755	13	S9	15.179	13	O4	16.592
14	M2	13.639	14	M4	14.734	14	I6	16.381
15	C7	13.377	15	M1	14.163	15	S9	15.830
16	K3	13.280	16	N8	14.105	16	N8	14.861
17	N8	12.978	17	N1	13.013	17	N1	13.145
18	C1	12.375	18	K3	12.992	18	C1	12.804
19	M1	11.638	19	B1	12.387	19	K3	12.273
20	O2	11.623	20	C1	11.733	20	A0	11.739
2011			2012			2013		
1	O8	150.059	1	O8	152.073	1	O8	125.647
2	M5	58.096	2	M5	61.866	2	M5	62.554
3	C5	38.815	3	C5	43.167	3	C5	45.872
4	S8	37.837	4	S8	36.682	4	S8	38.750
5	D2	34.009	5	D2	32.882	5	D2	31.745
6	S3	26.897	6	M2	26.143	6	M2	26.597
7	I8	24.308	7	S3	24.490	7	M1	24.191
8	M2	21.768	8	I8	23.217	8	S3	23.984
9	C7	19.997	9	C7	21.357	9	O6	23.765
10	I6	19.307	10	M1	21.083	10	M7	21.900
11	O6	18.509	11	O6	19.798	11	C7	21.747

12	M4	18.153	12	I6	19.588	12	I8	21.247
13	M1	17.987	13	M4	19.558	13	M4	19.367
14	S9	16.417	14	M7	19.202	14	I6	19.181
15	O4	15.108	15	S9	16.712	15	S9	17.858
16	M7	14.522	16	N8	14.143	16	A0	14.403
17	J1	14.262	17	O4	14.057	17	N8	13.697
18	N8	14.230	18	C1	13.645	18	N1	13.266
19	N1	13.906	19	A0	13.482	19	S5	12.640
20	C1	13.532	20	N1	13.172	20	C1	12.621
2014			2015					
1	O8	94.659	1	O8	85.952			
2	M5	65.384	2	M5	64.053			
3	C5	46.962	3	C5	49.977			
4	S8	37.057	4	S8	36.884			
5	O6	33.192	5	O6	36.131			
6	D2	29.768	6	D2	28.625			
7	M2	27.171	7	M2	28.037			
8	M1	26.245	8	M1	27.988			
9	M7	24.975	9	O3	27.964			
10	O3	23.854	10	M7	25.606			
11	S3	22.396	11	I6	23.457			
12	I6	21.416	12	S3	20.768			
13	M4	20.197	13	M4	20.543			
14	I8	19.543	14	S9	19.419			
15	S9	19.339	15	I8	17.649			
16	C7	18.111	16	J1	16.416			
17	A0	15.274	17	A0	16.269			
18	J1	14.068	18	C7	14.725			
19	N8	13.609	19	N1	14.294			

20	N1	13.448	20	C1	13.787
----	----	--------	----	----	--------

**<표 3-16> 2002-2015년 입원을 경험한 여성근로자의 1명 당
평균 입원일수 상위 20개 상병**

2002			2003			2004		
1	G8	39.192	1	G8	35.723	1	E5	74.500
2	F6	25.833	2	F2	24.481	2	M	31.000
3	F2	24.531	3	P9	24.471	3	G8	29.746
4	L8	24.400	4	J6	22.636	4	G3	27.564
5	C9	24.191	5	C9	22.625	5	F2	21.449
6	F0	24.000	6	M8	22.125	6	C9	21.421
7	M8	20.635	7	G6	21.133	7	I1	18.284
8	S7	19.588	8	G3	20.740	8	M8	17.723
9	I0	18.321	9	M1	19.749	9	S7	17.608
10	I6	18.320	10	B3	19.037	10	C0	17.541
11	M1	17.951	11	I1	17.275	11	G9	17.376
12	G7	17.750	12	L8	17.217	12	B4	17.176
13	L4	17.667	13	F3	17.189	13	K0	16.773
14	F3	17.203	14	I7	16.807	14	Q7	16.756
15	G3	17.189	15	I6	16.431	15	I6	16.648
16	I7	16.764	16	M4	15.629	16	M1	16.480
17	M4	16.754	17	S7	15.247	17	F3	15.870
18	J8	16.404	18	E7	15.200	18	B6	15.600
19	I3	16.107	19	G9	15.110	19	G6	15.161
20	B4	15.750	20	C4	14.696	20	M4	14.920
2005			2006			2007		
1	A3	41.333	1	N4	44.000	1	G8	28.079
2	AA	39.000	2	G8	30.629	2	C9	23.749
3	G8	33.062	3	U0	30.500	3	F2	22.975

4	R7	23.500	4	C9	27.131	4	F6	21.115
5	C9	23.008	5	J6	26.071	5	M8	19.147
6	S7	21.277	6	L1	25.727	6	G9	18.406
7	F2	21.167	7	A3	23.909	7	M1	17.994
8	G6	21.041	8	F2	23.229	8	L1	17.091
9	G3	20.829	9	F6	21.035	9	A3	16.750
10	F7	19.471	10	G6	20.831	10	C0	16.699
11	M8	19.060	11	F5	19.380	11	I6	16.497
12	I1	18.528	12	S7	18.595	12	F3	16.492
13	F0	17.975	13	M1	18.391	13	M4	16.444
14	I6	17.554	14	B6	18.375	14	F0	15.982
15	M1	17.442	15	F0	17.864	15	G3	15.795
16	F3	17.162	16	M8	17.712	16	S7	15.526
17	W0	17.000	17	E5	17.667	17	F5	15.000
18	F5	16.981	18	F3	17.172	18	A4	14.975
19	F6	16.952	19	G9	16.428	19	S8	14.739
20	W5	16.500	20	I6	16.341	20	I1	14.576
2008			2009			2010		
1	G8	28.534	1	G8	26.635	1	G8	26.161
2	C9	25.930	2	F6	21.628	2	C9	22.605
3	B4	24.409	3	U8	21.304	3	L1	21.231
4	W5	24.000	4	C9	21.190	4	F2	21.146
5	G1	23.696	5	F2	20.949	5	F6	18.865
6	F6	22.690	6	M1	18.042	6	U8	18.692
7	F2	22.386	7	F0	17.884	7	G6	17.520
8	F0	20.403	8	E5	17.667	8	M1	17.459
9	M1	18.887	9	L3	17.328	9	M8	17.110
10	M8	18.624	10	I6	16.940	10	F3	16.668
11	I6	17.450	11	F3	16.403	11	S7	16.554

12	F3	17.243	12	M8	16.009	12	F7	16.400
13	Q0	16.958	13	F7	16.000	13	I6	15.453
14	G6	16.734	14	G6	15.856	14	F0	15.218
15	M4	16.673	15	F1	15.366	15	G3	15.207
16	S7	16.446	16	E7	15.352	16	A3	14.794
17	E5	16.143	17	C0	15.155	17	F5	14.618
18	F5	16.055	18	S7	14.855	18	S2	14.595
19	U8	15.912	19	M4	14.719	19	L9	14.533
20	L4	15.786	20	S2	14.507	20	G1	14.417
2011			2012			2013		
1	F8	51.000	1	U8	29.687	1	G8	28.391
2	G8	27.982	2	G8	26.907	2	U8	26.823
3	J7	27.000	3	F8	25.500	3	F2	21.630
4	U8	24.629	4	F2	22.064	4	F7	20.889
5	F2	21.474	5	A3	20.917	5	E5	20.125
6	F6	20.465	6	F9	20.846	6	C9	19.735
7	G6	20.331	7	E5	19.429	7	F6	19.197
8	C9	20.180	8	C9	19.059	8	Q0	18.846
9	F7	19.955	9	F7	18.000	9	A3	18.159
10	G1	17.926	10	F6	16.702	10	F3	16.328
11	Z8	17.875	11	F3	16.686	11	E9	16.000
12	M1	16.670	12	G6	16.343	12	M1	15.626
13	F3	16.321	13	G3	15.788	13	F9	15.455
14	S7	16.076	14	M1	15.760	14	F1	15.348
15	I6	15.460	15	F1	15.490	15	P2	15.000
16	F0	14.991	16	I6	15.203	16	F5	14.673
17	C0	14.839	17	C0	15.140	17	S7	14.630
18	S2	14.688	18	W5	15.000	18	I3	14.241
19	G9	13.966	19	I3	14.189	19	I6	14.238

20	M8	13.789	20	S7	13.987	20	C0	13.735
2014			2015					
1	P1	40.500	1	G8	28.587			
2	G8	27.242	2	P1	23.000			
3	F2	21.615	3	U8	21.897			
4	C9	20.080	4	F2	20.556			
5	U8	19.558	5	F7	19.894			
6	P2	19.500	6	F6	19.639			
7	F1	16.630	7	C9	19.037			
8	F3	16.339	8	F8	18.800			
9	G1	15.178	9	F9	18.556			
10	I6	14.698	10	F1	16.265			
11	M1	14.516	11	F3	16.253			
12	G6	14.423	12	F0	14.973			
13	S7	14.083	13	I6	14.433			
14	C4	14.016	14	G3	14.224			
15	I3	13.934	15	M1	13.747			
16	F6	13.923	16	I0	13.478			
17	F5	13.626	17	G6	13.368			
18	M0	13.367	18	M0	13.239			
19	F0	13.043	19	G9	13.141			
20	S2	13.019	20	S7	13.058			

이를 통하여 남자 및 여자에서 근로자 1,000명 당 입원일수가 많았던 상병명에 대해서 2002년부터 2015년 사이 각 연도별 순위의 합을 구하고, 이를 상위 20위 안에 들었던 횟수로 나누어 평균 순위를 구하여 중복을 제외하고 우선분석대상 상위 10개 질환을 선정하였다. 선정한 결과는 다음 <표 3-17>과 같다.

<표 3-17> 2002-2015년 다빈도상병 평균 순위 상위질환

남			여		
상병 코드	상병명	평균 순위	상병 코드	상병명	평균 순위
S8	무릎 및 아래다리의 손상	1.57	O8	분만	1
M5	기타 등병증	1.79	M5	기타 등병증	3
I6	뇌혈관질환	3.43	D2	양성 신생물	3.64
C1	입술, 구강 및 인두, 소화기관중 일부 기관의 암	4.57	C5	유방 또는 여성생식기관의 암	3.93
C2	소화기관 중 일부의 암	5.14	S8	무릎 및 아래다리의 손상	5.64
I8	달리 분류되지 않은 정맥, 림프관 및 림프절의 질환	6.21	I8	달리 분류되지 않은 정맥, 림프관 및 림프절의 질환	6.79
S6	손목 및 손의 손상	9.21	O4	분만 문제에 관련된 산모관리	8.82

3) 직업 · 직종 · 산업 분류에 따른 취약집단의 발굴

2002년부터 2015년까지의 다빈도상병을 바탕으로 대조군인 ‘공무원, 사립학교 교직원’에 대하여 17개 산업군의 표준화발생비를 남성, 여성에서 각각 구하였으며 그 결과는 다음 <표 3-18>, <표 3-19>와 같다. 대조군에 비하여 유의하게 표준화발생비가 높은 경우는 음영으로 표시하였다. 즉, 표준화발생비가 높은 업종군은 대조군에 비하여 해당질환에 대해 상대적으로 취약한 집단이라 생각할 수 있을 것이라 기대된다.

<표 3-18> 2002-2015년 산업분류별 우선순위 10개 질환의 표준화발생비, 남성

	C1	C2	C5	D2	I6	I8	M5	O4	O8	S8
01. 농업, 수렵업, 임업	418 1.06 (0.95-1.15)	315 1.14 (1-1.25)	2 1.69 (0.28-5.04)	93 0.94 (0.74-1.12)	586 1.23 (1.12-1.31)	1760 0.83 (0.78-0.86)	1762 1.04 (0.99-1.09)	0 . (.-)	0 . (.-)	1208 1.04 (0.98-1.09)
02 어업	315 0.91 (0.8-1)	263 1.09 (0.95-1.21)	0 0 (-0.04-1.42)	107 1.25 (1-1.46)	664 1.59 (1.45-1.69)	1708 0.93 (0.88-0.97)	1441 0.98 (0.93-1.03)	0 . (.-)	0 . (.-)	728 0.74 (0.68-0.79)
03 광업	493 1.04 (0.93-1.12)	409 1.23 (1.1-1.34)	3 2.16 (0.51-5.34)	135 1.24 (1.02-1.43)	47 1.31 (1.21-1.4)	2208 0.97 (0.92-1)	1795 0.98 (0.93-1.02)	0 . (.-)	0 . (.-)	1077 0.92 (0.85-0.96)
04 제조업	29045 1 (0.99-1.01)	20535 1.04 (1.02-1.05)	82 0.86 (0.66-1.02)	12177 1.06 (1.04-1.07)	40633 1.06 (1.05-1.07)	273938 1.05 (1.05-1.06)	20516 0.97 (0.96-0.97)	0 . (.-)	0 . (.-)	111897 0.72 (0.72-0.73)
05 전기, 가스, 수도사업	1386 1.01 (0.95-1.05)	1034 1.1 (1.02-1.15)	4 0.92 (0.27-2.04)	469 1.08 (0.97-1.16)	1664 0.95 (0.9-0.99)	9835 1.03 (1-1.04)	7194 0.93 (0.9-0.95)	0 . (.-)	0 . (.-)	4633 0.87 (0.84-0.89)
06 건설업	10632 1.02 (1-1.04)	7885 1.1 (1.07-1.12)	31 0.96 (0.63-1.28)	3103 1 (0.96-1.03)	15279 1.18 (1.16-1.19)	64358 0.94 (0.94-0.95)	55514 1.01 (1-1.02)	0 . (.-)	0 . (.-)	30824 0.81 (0.8-0.82)
07 도소매 및 소비자용품 수리업	9313 0.99 (0.97-1.01)	6510 1.01 (0.98-1.03)	29 0.92 (0.59-1.25)	3549 0.98 (0.94-1.01)	13375 1.09 (1.07-1.11)	77571 0.94 (0.93-0.95)	69515 1.05 (1.04-1.06)	0 . (.-)	0 . (.-)	35319 0.72 (0.71-0.73)
08 숙박음식점업	984 0.98 (0.91-1.03)	675 0.98 (0.9-1.04)	1 0.3 (0.02-1.29)	438 1 (0.9-1.09)	1512 1.14 (1.07-1.19)	10213 1.03 (1.01-1.05)	8884 1.14 (1.11-1.16)	0 . (.-)	0 . (.-)	4426 0.71 (0.69-0.73)
09 운수, 창고, 통신업	15227 1.04 (1.02-1.05)	11352 1.11 (1.09-1.13)	34 0.78 (0.52-1.04)	3446 0.96 (0.93-0.99)	24265 1.36 (1.34-1.38)	77379 1.03 (1.02-1.03)	57834 0.95 (0.94-0.96)	0 . (.-)	0 . (.-)	31094 0.78 (0.77-0.79)
10 금융, 보험업	3274	2392	12	998	4236	26059	15506	0	0	9791

106...빅데이터 기반 직업 코호트 구축을 통한 질병 발생 연구

보험업	0.98 (0.94-1.01)	1.05 (1-1.09)	1.09 (0.55-1.74)	0.79 (0.73-0.83)	0.96 (0.93-0.99)	0.91 (0.9-0.92)	0.67 (0.66-0.68)	· (·-)	· (·-)	0.6 (0.59-0.61)
11 부동산, 임대, 사업서비스	25526 1.02 (1.01-1.03)	17359 0.98 (0.97-1.01)	58 0.79 (0.58-0.97)	4828 0.88 (0.85-0.9)	34103 1.16 (1.14-1.17)	108116 0.93 (0.92-0.93)	81042 0.87 (0.86-0.88)	0 · (·-)	0 · (·-)	43013 0.66 (0.65-0.67)
12 공공국방 및 사회보장행정	4588 1.26 (1.22-1.29)	3849 1.5 (1.44-1.54)	14 1.3 (0.69-2)	624 0.87 (0.79-0.93)	5527 1.32 (1.28-1.35)	13442 0.89 (0.87-0.9)	9200 0.76 (0.74-0.77)	0 · (·-)	0 · (·-)	6070 0.74 (0.72-0.76)
13 교육서비스업	3495 1.14 (1.09-1.17)	2733 1.25 (1.2-1.29)	8 0.91 (0.39-1.61)	485 0.89 (0.8-0.96)	3884 1.1 (1.06-1.13)	10019 0.89 (0.87-0.91)	7039 0.78 (0.76-0.8)	0 · (·-)	0 · (·-)	4421 0.74 (0.71-0.76)
14 보건 사회복지사업	3128 0.95 (0.92-0.98)	2189 0.96 (0.91-0.99)	3 0.30 (0.07-0.75)	730 0.84 (0.77-0.89)	4300 1.07 (1.04-1.1)	15646 0.83 (0.81-0.84)	12981 0.86 (0.84-0.87)	0 · (·-)	0 · (·-)	8007 0.76 (0.74-0.78)
15 기타공공사회, 개인서비스	11085 1.05 (1.02-1.06)	8145 1.02 (1-1.04)	21 0.63 (0.38-0.9)	2248 0.92 (0.88-0.96)	45974 1.2 (1.18-1.21)	48253 0.94 (0.93-0.95)	40298 0.98 (0.97-0.99)	0 · (·-)	0 · (·-)	22367 0.79 (0.78-0.8)
16 가사서비스업	1978 1.06 (1.01-1.1)	1303 0.98 (0.92-1.02)	4 0.78 (0.23-1.72)	195 0.81 (0.69-0.92)	2531 1.22 (1.16-1.26)	4122 0.92 (0.89-0.94)	2828 0.79 (0.76-0.81)	0 · (·-)	0 · (·-)	1476 0.7 (0.66-0.73)
17 국제, 기타 외국기관	432 1.06 (0.95-1.15)	331 1.15 (1.02-1.26)	2 1.73 (0.29-5.17)	58 0.79 (0.58-0.98)	505 1.07 (0.96-1.15)	1497 1.01 (0.95-1.05)	1096 0.92 (0.86-0.96)	0 · (·-)	0 · (·-)	594 0.79 (0.72-0.84)

<표 3-19> 2002-2015년 산업분류별 우선순위 10개 질환의 표준화발생비, 여성

	C1	C2	C5	D2	I6	I8	M5	O4	O8	S8
01. 농업, 수렵업, 임업	81 0.93 (0.72-1.11)	35 0.73 (0.49-0.96)	180 0.75 (0.63-0.85)	697 0.86 (0.79-0.92)	190 1.29 (1.09-1.45)	902 0.98 (0.91-1.04)	880 1.19 (1.1-1.26)	223 0.88 (0.76-0.98)	2523 0.89 (0.85-0.91)	367 1.15 (1.02-1.26)
02 어업	41 0.86 (0.59-1.1)	29 1.25 (0.81-1.69)	126 0.82 (0.66-0.94)	670 1.02 (0.93-1.09)	101 1.29 (1.02-1.51)	851 1.25 (1.16-1.32)	628 1.22 (1.11-1.3)	248 1.09 (0.94-1.21)	2145 0.78 (0.74-0.81)	239 1.18 (1.02-1.31)
03 광업	24 0.9 (0.56-1.26)	21 1.45 (0.87-2.07)	52 0.72 (0.52-0.9)	256 0.88 (0.76-0.98)	61 1.36 (1.01-1.67)	349 1.02 (0.9-1.11)	276 1.09 (0.95-1.2)	115 0.83 (0.67-0.96)	1371 0.87 (0.82-0.91)	112 1.06 (0.85-1.24)
04 제조업	6182 0.95 (0.92-0.97)	3139 1 (0.96-1.03)	15900 0.77 (0.76-0.78)	90434 0.95 (0.94-0.95)	12188 1.13 (1.11-1.15)	127695 1.22 (1.21-1.23)	85146 1.13 (1.12-1.14)	36537 0.86 (0.85-0.87)	415730 0.87 (0.86-0.87)	31151 1.03 (1.02-1.04)
05 전기, 가스, 수도사업	54 0.76 (0.56-0.95)	27 0.85 (0.54-1.16)	206 0.88 (0.75-0.99)	1306 1 (0.94-1.05)	106 0.94 (0.75-1.11)	1535 1.09 (1.03-1.14)	985 1.04 (0.97-1.09)	748 1.04 (0.95-1.1)	7638 0.93 (0.91-0.95)	372 1.08 (0.96-1.17)
06 건설업	883 0.95 (0.88-1)	493 1.03 (0.93-1.11)	2347 0.88 (0.84-0.91)	12064 0.93 (0.91-0.94)	1827 1.2 (1.14-1.25)	15939 1.09 (1.07-1.1)	11859 1.16 (1.14-1.18)	6783 0.94 (0.92-0.96)	74350 0.92 (0.91-0.93)	4323 1.09 (1.05-1.12)
07 도소매 및 소비자용품 수리업	2564 0.94 (0.9-0.97)	1194 0.92 (0.86-0.97)	7683 0.87 (0.84-0.88)	43588 0.94 (0.93-0.95)	4877 1.1 (1.06-1.13)	54369 1.08 (1.07-1.09)	40520 1.18 (1.16-1.19)	54369 1.08 (1.07-1.09)	220628 0.82 (0.81-0.82)	13474 1.03 (1.01-1.04)
08 숙박음식점업	971 0.92 (0.85-0.97)	508 0.92 (0.83-0.99)	2341 0.76 (0.72-0.78)	10189 0.92 (0.9-0.94)	2148 1.18 (1.12-1.22)	14051 1.07 (1.05-1.08)	13355 1.31 (1.28-1.32)	2809 0.7 (0.67-0.72)	35382 0.79 (0.78-0.8)	5213 1.17 (1.13-1.2)
09 운수, 창고, 숙박업	626	302	1928	11098	1235	13470	9452	5539	64398	3268

108...빅데이터 기반 직업 코호트 구축을 통한 질병 발생 연구

통신업	0.98 (0.89-1.04)	1 (0.88-1.1)	0.94 (0.89-0.97)	0.97 (0.95-0.99)	1.2 (1.12-1.25)	1.07 (1.05-1.08)	1.13 (1.1-1.14)	0.85 (0.83-0.87)	0.86 (0.86-0.87)	1.04 (1-1.07)
10 금융, 보험업	731 1.05 (0.97-1.12)	281 1.02 (0.89-1.12)	2429 0.98 (0.94-1.01)	18760 1.02 (1.01-1.03)	1181 1.13 (1.05-1.18)	21168 1.04 (1.03-1.06)	11213 0.91 (0.89-0.93)	11952 0.96 (0.94-0.97)	134991 0.98 (0.98-0.99)	3528 0.84 (0.81-0.86)
11 부동산, 임대, 사업서비스	5058 0.92 (0.89-0.94)	2994 0.95 (0.91-0.98)	10865 0.77 (0.76-0.78)	55449 0.94 (0.93-0.95)	11046 1.18 (1.16-1.2)	74534 1.04 (1.03-1.05)	56656 1.11 (1.1-1.11)	25984 0.8 (0.79-0.81)	302969 0.81 (0.81-0.82)	21996 1.02 (1.01-1.03)
12 공공국방 및 사회보장행정	508 1.13 (1.02-1.22)	232 0.98 (0.85-1.1)	1090 0.85 (0.79-0.89)	5002 0.93 (0.9-0.95)	917 1.24 (1.15-1.31)	5910 0.99 (0.96-1.01)	4701 1.07 (1.03-1.09)	2160 0.92 (0.87-0.95)	23030 0.83 (0.82-0.84)	1940 1.1 (1.04-1.14)
13 교육서비스업	1994 1.02 (0.97-1.06)	861 0.95 (0.88-1.01)	5593 0.85 (0.83-0.87)	30858 0.94 (0.93-0.95)	3174 1.02 (0.98-1.05)	34242 1.02 (1.01-1.03)	25214 1.06 (1.04-1.07)	12531 0.91 (0.89-0.92)	140338 0.89 (0.88-0.89)	8915 1.01 (0.99-1.03)
14 보건 사회복지사업	3158 0.95 (0.91-0.98)	1576 0.94 (0.89-0.98)	8424 0.85 (0.83-0.87)	44050 0.93 (0.92-0.94)	6716 1.22 (1.18-1.24)	59422 1.09 (1.08-1.1)	48201 1.28 (1.27-1.29)	27428 1.1 (1.09-1.11)	272454 1.02 (1.01-1.02)	17484 1.17 (1.15-1.19)
15 기타공공사회, 개인서비스	2891 0.94 (0.9-0.97)	1680 0.96 (0.91-1)	6034 0.76 (0.74-0.78)	28563 0.95 (0.94-0.96)	6328 1.21 (1.18-1.24)	38793 1.07 (1.06-1.08)	31712 1.17 (1.16-1.19)	12221 0.84 (0.82-0.85)	138731 0.82 (0.82-0.83)	12953 1.12 (1.09-1.13)
16 가사서비스업	261 0.98 (0.85-1.08)	143 0.87 (0.72-1)	424 0.7 (0.63-0.76)	1669 0.95 (0.9-0.99)	567 1.23 (1.12-1.32)	2346 1.02 (0.98-1.06)	1926 1.06 (1.01-1.1)	712 0.96 (0.88-1.02)	7598 0.89 (0.87-0.91)	905 1.05 (0.97-1.11)
17 국제, 기타 외국기관	40 0.84 (0.58-1.09)	17 0.64 (0.36-0.94)	120 0.96 (0.78-1.12)	366 0.84 (0.74-0.91)	86 1.09 (0.85-1.3)	484 0.99 (0.89-1.07)	297 0.76 (0.66-0.83)	114 0.74 (0.59-0.86)	1340 0.65 (0.61-0.68)	123 0.74 (0.6-0.86)

4) 사업장 특성에 따른 취약집단 선정

300인 이상의 사업장은 만성질환과 근골격계 질환의 유병률을 산출하여 배포하는 작업을 건강보험공단에 요청하여 실제 진행 중이다. 다만, 연구진 회의 결과 특정 사업장의 질병 분포를 무분별하게 공개하는 것은 근로자의 개인정보를 침해할 위험이 있다고 판단되었다. 300인 이하의 사업장의 경우에는 개별적인 유병 근로자를 추적할 수 있을 위험성이 더 높으므로 실제로 무분별한 유병률 공개는 개인정보 보호의 측면에서 위험성이 높다. 따라서 이러한 300인 이하의 사업장에서는 개인의 유병정보를 공개하는 것보다는 개선 가능한 만성질환과 근골격계 질환 관리에 있어 300인 이상 사업장에서 얻어진 정보를 이용할 수 있다고 판단했다. 특히, 각 진단별로 5명 이하의 유병인구를 가지고 있는 질환의 경우 개인을 추정할 수 있으므로 개인정보를 보호하는 차원에서 특정 사업장 별 특정질환 유병 근로자에 대한 조사 및 자료 공개는 실익과 인권 모두를 고려해야 한다고 판단된다.

3. 근로자 질병 예방을 위한 코호트 활용 방안 제언

1) 시범 운용을 통한 코호트 활용 방안

(1) 직업 및 산업분류와 직업관련 질환의 연관분석

가) 분진노출 근로자 코호트

직업 및 산업분류와 직업관련 질환으로서 잘 알려진 질환으로는 방사선 피폭에 의한 혈액암, 염화비닐에 의한 폐암, 타르에 의한 폐암, 크롬에 의한 폐암, 벤젠에 의한 조혈기계암, 석면에 의한 악성중피종과 폐암 등이 잘 알려져 있다. 분진노출에 의한 진폐증 역시 잘 알려진 직업성 질환으로 여기서 진폐증이란 분진 흡입에 동반한 폐내 분진의 축적과 그에 의한 폐의 조직반응으로 나타나는 것으로 여기서 분진이란 고체상태의 무생물 입자로서 에어로졸(aerosol)을 이야기한다. 진폐증은 원인이 되는 물질에 따라 석면폐증, 규폐증, 탄광부진폐증 등으로 나눌 수 있으며 폐결핵, 폐기종, 특발성 폐섬유화증 등이 합병증으로 발생할 수 있다. 1999년부터는 진폐증 환자에게서 폐암 발생률이 높다는 연구 결과를 근거로 하여 광업 종사 경험이 있는 진폐증 환자에게서 원발성 폐암이 발생한 경우 직업성 암으로 인정하고 있다.

한편으로 1960년대 미국에서 탄광부진폐증에서 위암 발생률이 높다는 연구가 나오기 시작한 이래로 전 세계적으로 진폐를 포함, 다양한 분진에 노출된 근로자에서 위암 사망률 및 발생률이 높다는 보고가 있었지만 국내 연구에서는 근거가 부족한 실정이다. 따라서 진폐 진단환자 또는 광업 종사자를 대상으로 식도를 시작으로 전반적인 소화기계 암의 표준화발생비를 구하기 위한 코호트가 구축되었다.

먼저 분진노출 과거력이 있는 대상자를 선정하기 위하여 ‘진폐를 진단받은 환자’ 또는 ‘광업에 종사한 경력이 있는 자’를 선정하였으며 일차적인 대조군으로는 ‘공무원 및 사립학교 교직원’을 선택하였다. 상병코드는 ICD 3자리를 기준으로 선

택하여 구하였으며, 구강암은 일부 두경부암을 포함 'C00-C14', 식도암은 'C15', 위암은 'C16', 소장암은 'C17', 대장암은 'C18', S자결장암은 'C19', 직장암은 'C20', 항문암은 'C21'로 구하였다.

<표 3-20> 분진노출근로자의 소화기계암

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C00 - C14 구강암	남	656	14.59	13.34	15.55
	여	13	5.74	2.96	9.00
C15 식도암	남	1075	14.99	13.97	15.75
	여	5	4.58	1.83	8.41
C16 위암	남	9968	13.92	13.61	14.15
	여	864	9.87	9.13	10.43
C17 소장암	남	195	17.01	14.41	19.11
	여	13	21.85	11.25	34.24
C18 대장암	남	3811	10.53	10.14	10.81
	여	696	9.93	9.10	10.56
C19 S자결장암	남	704	12.82	11.75	13.63
	여	126	21.04	17.10	24.33
C20 직장암	남	2624	13.77	13.16	14.21
	여	417	12.66	11.31	13.71
C21 항문암	남	99	18.19	14.39	21.43
	여	22	36.46	21.99	51.56

연구결과 남자 및 여자 모두에게서 대조군에 비교하여 표준화발생비가 모두 높은 것으로 나타났다. 분진이 호흡기계 뿐 아니라 소화기계를 통하여 소화기계 점막을 통해 노출이 일어날 수 있으므로 소화기계암 표준화발생비가 높은 원인을 추정해볼 수 있다. 개개인마다 각 소화기관 별로 암의 병리조직학적 소견이 상이하고 각 장기의 담당하는 기능이 이동, 소화액 분비, 흡수 등으로 다양하며 음식물이 지나는데 소요되는 시간(bowel transit time)이 다르기 때문에 각 기관 별 암 표준화발생비가 서로 다른 경향을 보일 것으로 생각된다. 더불어 근로자의 건강행태(흡연력 및 음주력, 운동력, 식습관 등) 및 기타 위험요인(가족력, 체질량 지수 등)이 소화기계암에 미치는 영향이 무시하기 어려우므로 향후 이러한 제한

점을 반영하기 위해 건강대조군 뿐만 아니라 질병대조군, 위험대조군을 설정하는 등 다양한 접근방식으로 추가적인 연구를 진행할 필요가 있다.

나) 병·의원 종사자 코호트

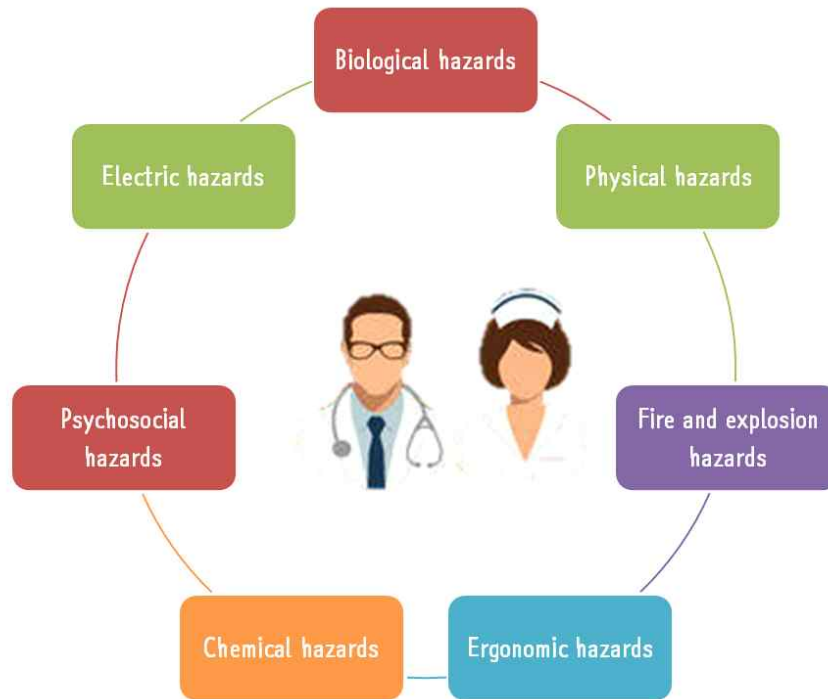
병원 및 의원에 종사하는 사람은 대부분 의료인 또는 의료보조인력일 것으로 기대할 수 있으며 전국적으로 병의원에 종사하는 사람들은 근무하고 있는 사업장이 다르다 하더라도 근무형태(교대제) 및 주요 업무 환경과 같이 업무의 특성이 어느 정도 유사하여 동질성(homogeneity)을 담보할 것이라 기대할 수 있다. 따라서 병의원 종사자의 경우 빅데이터를 기반으로 한 직업코호트로 건강영향을 파악하는 것은 타당하고 효과적인 방법이라 볼 수 있으므로 직업과 직업관련질환의 연관 분석에서 시범적으로 실시하기에 적합한 연구 주제라 판단하였다.

보건 및 사회복지사업 종사자가 건보공단 업종분류코드 중 대분류 '14'에 해당하고 그 하위항목으로 의료기관 종사자를 좁힐 수 있으나 실제로 건보공단 업종분류코드를 이용하여 선정할 경우 의료기 상사 등을 비롯하여 이질적인 업종에 종사하는 종사자가 포함(enrol)되는 경향을 보였다. 따라서 보다 효과적인 대상자 추출을 위해서는 사업장 명을 사용하여 사업장을 선정하고, 2002년부터 2015년 데이터에서 해당 사업장에 근무할 예정이거나, 근무 중이거나, 퇴직한 근로자를 추적하여 코호트를 구축하였다. 여기에서도 대조군은 공무원 및 사립학교 교직원 이 설정되었다.



[그림 3-19] 의료기관 추출 과정

사업장 명에 대한 텍스트 분석을 위하여 포함조건으로는 ‘병원’, ‘의원’, ‘의료원’, ‘보건소’를 선정하여 총 51,374개 사업장이 추출되었으나 그 중 ‘의회의원’ ‘의회의원 회관’ 등을 제외하기 위하여 ‘의회’, ‘협회’를 제외조건으로 선정하여 12개를 제외하였다. 또한 추가적으로 일반적인 병, 의원과 상이한 근무형태를 가질 것이라 생각될 수 있는 한방병원 및 한의원을 제외하여 38,971개 사업장을 추출하였다. 그 후 ‘동물병원’을 제외하기 위해 제외조건으로 ‘동물병원’을 추가 설정한 결과 1,626개 사업장이 제외되었고, 병원 내 어린이집 또는 장례식장과 같은 부대 시설 266곳을 제외하여 최종적으로 37,079개 사업장을 대상으로 분석을 진행하였다.



[그림 3-20] WHO 병·의원 종사자의 직업성 위험요인

WHO에서는 보건의료 또는 병원 종사자가 노출되는 직업적 위험요인으로 위의 [그림 3-20]과 같은 요소들을 제시하고 있다. 이러한 요소들을 바탕으로 할 때 본 연구에서 검토할 질환으로는 내부 회의를 통해 근골격계질환 및 유방암을 선정하였으며 이에 대한 분석을 위해 다음 <표 3-21>과 같은 상병코드를 선택하였다.

<표 3-21> 병·의원 종사자의 각 질환별 표준화발생비

		실제 관찰수	표준화 발생비	표준화발생비 범위	
				LL	UL
M50 경추추간판탈출	남	290	1.16	1.01	1.28
	여	1380	1.62	1.52	1.69
M51 흉요추추간판탈출	남	1367	1.39	1.31	1.46
	여	4942	1.74	1.68	1.78
M75 어깨질환, 회전근개질환	남	230	0.92	0.79	1.02
	여	1095	1.47	1.37	1.54
G56 수근관증후군	남	36	1.32	0.89	1.73
	여	493	2.54	2.29	2.73
S22 흉추척추체 압박골절	남	284	2.10	1.83	2.32
	여	655	2.07	1.89	2.21
S32 요추흉추체 압박골절	남	94	1.37	1.08	1.62
	여	625	1.76	1.60	1.88
F43 급성스트레스증후군	남	15	1.22	0.66	1.85
	여	64	0.77	0.58	0.95
F32 주요우울증(single)	남	100	2.94	2.33	3.46
	여	333	1.81	1.59	1.97
F33 주요우울증(recurrent)	남	12	2.91	1.45	4.64
	여	60	2.12	1.57	2.62

분석결과 병·의원 종사자에서 근골격계 질환 발생비가 공무원 및 사립학교 교직원에 비해서 높은 경향을 보이고 있었다. 경추추간판탈출증(남자 1.16(1.01-1.28), 여자 1.62(1.52-1.69)), 흉요추추간판탈출증(남자, 1.39(1.31-1.46), 여자 1.74(1.68-1.78)), 흉추척추체압박골절(남자 2.10(1.83-2.32), 여자 2.07(1.89-2.21)), 요추흉추체압박골절 질환(남자 1.37(1.08-1.62), 여자, 1.76(1.60-1.88))의 경우 남, 여 모두에서 대조군에 비해 높게 나타났다. 어깨질환

의 경우는 여성에서만 대조군에 비해 유의하게 높은 표준화발생비(1.47, 1.37-1.4)를 보였다.

물론 위에서 진폐 및 광업 종사자 코호트와 마찬가지로 병·의원 코호트 역시 개인의 건강행태 및 가족력과 같은 기타 다른 요인들이 반영되지 않아 연구결과에 절대적인 의미를 부여하기는 어렵지만 높은 표준화발생비를 보인 질환에 대해서는 적극적인 감시 및 관리를 시행할 필요가 있을 것으로 판단된다.

다) 육상 운수업 종사자 코호트

현재 코호트를 구축되었으나 좀 더 검토가 필요한 코호트로는 ‘육상 운수업 종사자 코호트’가 있다. 육상운수업 종사자는 건강보험공단 사업장(세분류)코드를 기준으로 ‘육상운송 및 파이프라인운송업(90600)’, ‘도시간철도운송업(90601)’, ‘기타 육상운송업(90602)’를 선택하였다. 육상운수업 종사자의 경우 근무환경의 일반적인 유해인자로 오래 앉는 자세, 자외선, 지속되는 진동 등을 건강유해인자로 생각하여 우선 분석할 상병명으로는 피부암(C43-C44), 방광암(C67), 전립선암(C61), 경추추간판탈출증(M50), 요추추간판탈출증(M51)을 선정하였다. 건설업의 경우는 상당부분이 지역가입자라는 연구진 의견에 따라 업종 정보를 통한 단순 구축은 신뢰성이 떨어진다고 판단하였다. 그 결과는 다음 <표 3-22>와 같다.

<표 3-22> 육상운수업 종사자의 표준화발생비

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C43 악성흑색종	남	118	1.20	0.97	1.40
	여	8	0.87	0.37	1.53
C44 기타 피부암	남	302	0.83	0.73	0.91
	여	15	0.88	0.48	1.34
C43+C44	남	417	0.91	0.87	0.99
	여	23	0.88	0.54	1.24
C431-434 두경부 피부암	남	18	1.24	0.71	1.82
	여	2	1.28	0.21	3.80
C67 방광암	남	1,324	1.07	1.01	1.12
	여	19	1.16	0.67	1.68
C61 전립선암	남	1,646	0.58	0.55	0.61
	여	-	-	-	-
M50 경추추간판탈출증	남	9,967	0.91	0.89	0.92
	여	1,175	1.05	0.98	1.10
M51 요추추간판탈출증	남	31,767	0.94	0.93	0.95
	여	114	1.14	1.10	1.19

그 결과 남자에서의 방광암(SIR 1.0, 1.01-1.12)과 여성에서의 요추추간판탈출증(SIR 1.14, 1.10-1.19)이 유의하게 나타났으나 다른 질환에서는 유의하지 않았다. 특히 직관적으로 높을 것이라 기대되었던 전립선암은 오히려 대조군에 비해 유의하게 낮게 발생하는 것으로 나타났다. 이는 다른 위험요인(건강행태 등)을 반영할 수 없었던 본 코호트의 제한점 때문일 수도 있지만 우선적으로는 코호트 검출 과정에 대한 검토가 필요하다. 단순히 사업장분류코드를 가지고 검출할 때에는 내부의 이질성과 동질성, 그리고 불확정성에 대한 고민이 필요한데 아직 육상운수업근로자 코호트는 그 단계를 거치지 않은 코호트로, 앞으로 최종보고서

수룩까지는 다각적인 방면으로 해당 코호트를 고찰, 수정 및 보완해나갈 필요가 있다고 생각된다.

또한 건강유해사업장으로서 자동차 정비소를 고려하였으나 실제 코호트를 구축하기 위해 시도해본 결과, 여러 어려움이 있었다. 먼저 사업장(세분류)코드로 하기에는 주유소직원, 각 자동차제조회사의 판매원, 부품 판매원, 중고차 딜러들이 정비소 근로자와 구분되지 않아 사업장(세분류)코드로 자동차 정비소 근로자를 추출하기에는 어려운 점이 있었다.

한편으로 사업장 이름에서 자동차 정비를 검출하기 위해서는 키워드 선정에 있어 어려운 점이 많았다. 실제로 자동차정비 업무를 담당하지만 각 자동차제조회사에 따라 자체적인 정비소 체인 이름이 다양하고, 사업장 이름에 ‘정비’라는 단어가 명시되어 있지 않아 선정조건에 자체적 정비소 브랜드 이름을 사용하여 했다. 그러한 애로사항들을 최대한 반영한 뒤에도 불구하고 정비 대신에 ‘얼라이먼트’, ‘오토카’ 등의 외래어 및 독창적인 합성어 등이 혼재되어 수많은 포함·제외조건을 설정하고 난 이후에도 여전히 누락되는 사업장이 존재하여 효과적인 코호트 구축이 불가능하였다. 자동차 정비소 직원이 디젤엔진배출물질 및 각종 유류제품, 유기용제 등 유해물질 및 환경에 노출될 것이라 생각되는 만큼 코호트 구축과정을 재검토하여 다시 한 번 코호트 구축 가능여부에 대해 고려할 필요가 있을 것으로 생각된다.

라) 반도체 및 디스플레이 관련 종사자 코호트

건강보험공단에서 제공하는 이러한 산업분류코드나 사업장 이름을 사용하는 방법을 제외하고, 각 산업 협회에서 제공하는 회원 사업장 목록을 이용한 코호트 구축도 가능하다. 이 방법으로 사업장을 추출하여 근로자를 구축할 경우에는 사업장의 생산 활동 및 근로 환경이 좀 더 균질성을 갖는다는 장점이 있다. 이러한 방법을 이용하여 반도체 및 디스플레이 관련 사업장에 대해 각각 한국반도체산업협회 및 한국디스플레이산업협회에서 제공하는 협회 회원사 목록을 통한 추출을 실시하였다.

반도체 및 디스플레이 산업은 우리나라의 정보통신기술 품목의 주축을 담당하는 주요 생산품이다. 고도의 기술이 집약되어 만들어지는 반도체 및 디스플레이는 공정은 제조 공정 과정에서 나름의 특성을 갖는다. 방진시설을 비롯한 여러 특수한 시설 등이 필요하고 여러 유기용제가 사용되며 제품의 크기는 작지만 여러 소재를 사용한 복잡한 공정을 거쳐야 반도체 또는 디스플레이 생산이 가능하다. 이러한 전문적인 과정에서 근로자는 생산 전반 과정에 걸쳐 세밀하고 섬세하며 정확한 기술을 요구받으며, 다양한 잠재적 건강위험요인에 필연적으로 노출될 가능성이 있다.

그러나 제조 과정에 사용되는 수많은 물질의 종류를 생각할 때 작업장 유해요인에 대한 완벽한 사전교육이나 예방은 어려우며 이전 연구 결과조차 충분치 못한 실정이다. 그러한 이유로 사업장 내 사고 또는 특이 질환, 산재 등이 발생할 경우 인과관계에 대한 파악이 어려우며, 인과관계 파악에도 많은 노력이 필요하다. 이러한 제한점을 국민건강보험공단 데이터베이스를 활용한 코호트를 구축함으로써 건강 위해성에 대한 즉각적이고 빠른 분석 및 파악이 가능하며, 비용 및 시간적 측면에서 효용성이 높다고 보인다.

본 연구에서 반도체 사업 종사 사업장을 선택하기 위해서 한국반도체산업협회에 등록되어 있는 회사명을 근거로 하여 회사이름에서 해당 회사 이름을 검출하였다. 그 과정에서 동명의 사업장을 구별하기 위해 사업장의 주소 수준까지 확인하는 과정을 거쳤다. 그 과정에서 6곳의 소사업체, 4곳의 테스트/패키징업체, 79곳의 설계업체, 34곳의 재료업체, 79곳의 장비업체, 29곳의 부품품업체, 10곳의 설비업체가 조사대상이 되었으며, 3곳의 사업장은 주소가 불명확하여 감별이 어려워 누락되었다. 최종적으로 241곳의 반도체 사업장에 2002년부터 2015년 사이 한 번이라도 근무한 적이 있는 사람을 비교군으로 하여 공무원 및 사립학교 교직원을 대조군으로 설정하고, 국내 암 중 유병률 순위로 10대 암을 골지로 한 표준화 발생비 및 다발성경화증에 대하여 <표 3-23>과 같이 조사하였다.

그 결과, 다른 질환에 있어서는 통계적으로 유의한 차이가 없었으나 C91-95 상병코드로 분석한 백혈병에 대하여 반도체 산업에 종사하는 여성 근로자가 대조군에 비하여 백혈병의 표준화발생비가 2.57(1.49-3.73)으로 높게 나타났다.

<표 3-23> 반도체 근로자의 건강영향

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C15 식도암	남	11	0.60	0.29	0.97
	여	0	0		10.47
C16 위암	남	416	1.09	0.97	1.18
	여	25	0.80	0.50	1.11
C18-21 대장암	남	270	1.05	0.91	1.16
	여	17	0.87	0.49	1.29
C22 간암	남	193	0.90	0.76	1.01
	여	6	1.11	0.41	2.14
C23-24 담낭 및 기타 담도암	남	33	1.10	0.73	1.46
	여	5	2.78	0.93	5.68
C25 췌장암	남	32	0.82	0.54	1.09
	여	1	0.37	0.03	1.62
C33-34 폐암	남	150	1.17	0.97	1.34
	여	9	0.91	0.41	1.55
G35 다발성경화증	남	6	0.78	0.29	1.50
	여	3	1.00	0.24	2.48
C50 유방암	남	0	0	-	-
	여	107	0.97	0.78	1.14
C53 자궁경부암	남	0	-	-	-
	여	22	1.47	0.88	2.07
C56 난소암	남	0	-	-	-
	여	11	0.66	0.32	1.07
C61 전립선암	남	77	1.11	0.85	1.33
	여	0	-	-	-
C82-86 비호지킨 림프종	남	66	0.99	0.74	1.20
	여	10	1.48	0.69	2.46
C91-95 백혈병	남	49	0.95	0.68	1.20
	여	19	2.57	1.49	3.73

반도체 관련 사업장에 대한 표준화발생비 분석과 마찬가지로 디스플레이 사업장에 대하여서도 사업장 이름을 사용한 근로자 코호트 구축을 시도하였다. 디스플레이 사업장 목록은 한국디스플레이산업협회를 통하여 목록을 얻었으며, 반도체

체 사업장과 마찬가지로 사업장 주소 등을 추적하여 동명의 사업장을 가려내는 작업을 반복하였다. 총 158개 사업장 중 패널 제조 사업장 2곳, 디스플레이 장비 관련 79곳, 소재 제조 17곳, 부품 및 재료 분야 제조 6곳, 모듈 1곳, 부품분야 28곳, 설비 및 기타 25곳으로 주소 확인 과정을 거쳐 동명의 사업장을 구별하는 작업을 거쳤다. 디스플레이 사업장의 경우에도 대조군은 공무원 및 사립학교 교직원으로 설정하였고, 다발성 경화증의 발병여부를 포함한 10대 암을 중심으로 다음 <표 3-24>와 같이 분석을 시행하였다.

그 결과 자궁경부암을 제외한 다른 질환에서는 대조군에 비해 표준화발생비가 유의하게 높지 않았으나, 자궁경부암의 경우 디스플레이 산업에 종사하는 여성 근로자가 공무원 및 사립학교의 여성 근로자에 비해 자궁경부암(C53)의 표준화발생비가 2.00(1.16-2.90)으로 높게 나타났다.

<표 3-24> 디스플레이 근로자의 건강영향

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C15 식도암	남	8	0.87	0.37	1.54
	여	0	0.00	-	9.08
C16 위암	남	217	1.13	0.96	1.26
	여	30	1.41	0.91	1.89
C18-21 대장암	남	123	0.93	0.76	1.08
	여	17	1.12	0.63	1.65
C22 간암	남	94	0.88	0.69	1.04
	여	5	1.17	0.39	2.39
C23-24 담낭 및 기타 담도암	남	14	0.93	0.49	1.44
	여	2	1.16	0.19	3.47
C25 췌장암	남	22	1.11	0.67	1.57
	여	2	0.88	0.15	2.63
C33-34 폐암	남	79	1.21	0.93	1.46
	여	4	0.48	0.14	1.06
C50 유방암	남	0	0.00	-	1.16
	여	65	0.83	0.62	1.01
C53 자궁경부암	남	0	-	-	-
	여	19	2.00	1.16	2.90
C56 난소암	남	0	-	-	-
	여	8	0.69	0.29	1.22
C61 전립선암	남	46	1.30	0.92	1.65
	여	0	-	-	-
C82-86 비호지킨 림프종	남	37	0.99	0.67	1.29
	여	4	0.88	0.26	1.96
C91-95 백혈병	남	23	0.76	0.46	1.06
	여	3	0.64	0.15	1.60
G35 다발성경화증	남	6	1.27	0.47	2.44
	여	3	1.65	0.39	4.09

마) 타이어 제조 사업장 근로자의 건강 영향 연구

반도체 및 디스플레이 관련 사업장 외에 추가적으로 타이어 제조 사업장에서 근로자의 건강영향에 대한 분석을 실시하였다. 지난 2006년에 국내 한 타이어회사에서 근무하는 15명의 근로자가 한 해 동안 15명이 사망하는 사건이 발생한 적도 있었으며, 그 외 타이어 제조 공장 근로자에게서 유해물질에 의한 건강 위협성이 추정될 때마다 정부 및 관련기관의 역학조사가 실시된 바 있었지만 명확하

게 밝혀진 부분은 없었다.

본 연구에서는 타이어공장 근로자를 국내 3대 타이어회사를 선정하여 분석하였다. 텍스트분석을 통해 타이어 제조회사를 찾고, 한 기업 당 여러 지역에 산재하여 있는 공장을 이름 및 주소를 확인하여 연구 대상자를 선택하였다. 그 뒤, 타이어공장 근로자의 전반적인 건강 상태에 대하여 확인 및 평가하기 위하여 10대 암 및 뇌심혈관계 질환에 대한 분석을 다음 <표 3-25>, <표 3-26>과 같이 실시하였다.

그 결과, 타이어 제조 산업에 종사하는 여자 근로자는 대조군에 비해 표준화발생비가 유의하게 높은 질환이 관찰되지 않았으나 남성 근로자의 경우에는 대조군에 비하여 위암(C16)의 표준화발생비가 1.35(.1-1.55)로 높게 나타났으며 고혈압(I10-I15) 역시 공무원 및 사립학교 교직원 대조군보다 표준화발생비가 높게 나타났다(1.41,1.11-1.67).

<표 3-25> 타이어업종 근로자의 건강영향 (암)

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C15 식도암	남	7	1.17	0.47	2.15
	여	0	0	-	31.76
C16 위암	남	129	1.35	1.1	1.55
	여	4	1.26	0.37	2.79
C18-21 대장암	남	50	0.75	0.54	0.95
	여	4	1.39	0.4	3.07
C22 간암	남	44	0.75	0.53	0.96
	여	0	0	-	1.74
C23-24 담낭 및 기타 담도암	남	14	1.58	0.83	2.44
	여	0	0	-	3.48
C25 췌장암	남	14	1.27	0.67	1.97
	여	1	1.97	0.14	8.55
C33-34 폐암	남	36	0.97	0.65	1.27
	여	0	0	-	0.83
C50 유방암	남	1	1.99	0.14	8.63
	여	8	0.63	0.27	1.11
C53 자궁경부암	남	0	-	-	-
	여	0	0	-	1.21
C56 난소암	남	0	-	-	-
	여	3	1.94	0.46	4.82
C61 전립선암	남	23	0.93	0.57	1.31
	여	0	-	-	-
C82-86 비호지킨 림프종	남	11	0.81	0.39	1.32
	여	0	0	-	2.51
C91-95 백혈병	남	8	0.89	0.38	1.58
	여	0	0	-	3.37

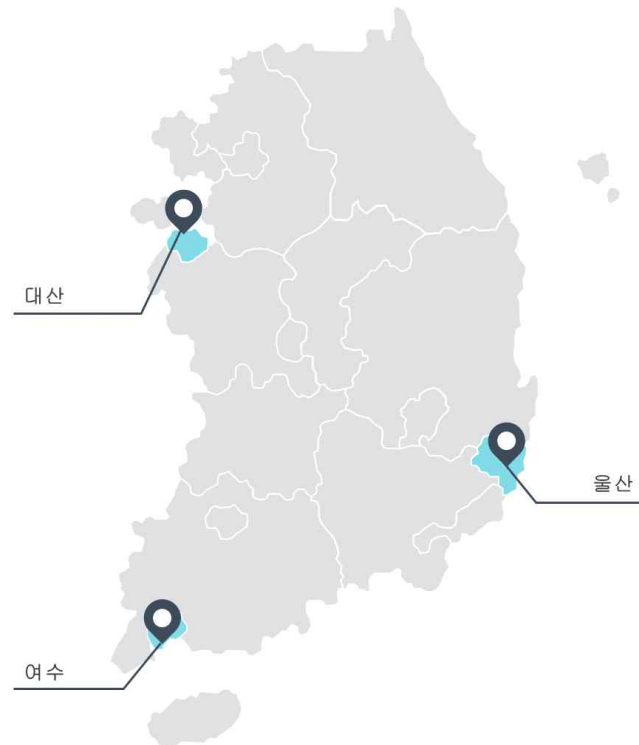
<표 3-26> 타이어업종 근로자의 건강영향 (뇌심혈관질환)

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
I10-15 고혈압	남	95	1.41	1.11	1.67
	여	3	1.09	0.26	2.7
I20-25 허혈성 심장질환	남	254	0.79	0.68	0.87
	여	4	0.57	0.17	1.27
I60-69 뇌혈관질환	남	104	0.89	0.71	1.04
	여	9	1.28	0.57	2.19

바) 석유화학단지 내 석유화학 산업 종사 근로자 코호트

사업장의 산업분류코드나 이름을 사용하는 것 외에도 사업장의 주소지를 함께 이용한 코호트를 구축할 수도 있다. 대규모 산업단지에서 근무하는 여러 사업장의 근로자들은 상대적으로 유사한 근무환경과 업무를 수행할 뿐만 아니라 비슷한 주거환경을 공유하는 등 서로 긴밀한 연관성을 갖는다. 그러므로 이러한 산업단지를 기반으로 한 근로자 코호트는 직업환경 및 거주환경에 대해 높은 동질성을 가지고 있을 것이라 생각되어 보다 유용한 활용이 가능하리라 기대된다.

따라서 본 연구진은 주변의 주민 건강 영향 연구를 위하여 관심 산업단지로 울산, 여수, 대산 3개 지역을 선정하여 해당 지역 내 사업장에서 종사한 근로자의 건강영향에 대하여 살펴보았다. 해당 지역은 석유화학 산업과 관련된 여러 사업장이 모여 있는 대규모 산업단지가 있는 곳으로, 지역을 선정한 후 해당 지역 내 존재하는 사업장은 한국석유화학협회에서 확인가능한 회원사 정보를 바탕으로 사업장 이름과 소재지를 사용하였다.



[그림 3-21] 분석에 사용된 석유화학단지

이러한 방법으로 분석에 사용된 사업장은 울산에 30개, 여수에 총 19곳, 대산에 총 5곳의 사업장이 있었다. 총 54곳의 사업장 내 근무하는 근로자들의 건강영향을 대조군으로는 공무원 및 사립학교 교직원으로 설정하여 10대 암 위주로 다음 <표 3-27>와 같이 분석하였다. 그 결과, 여성 석유화학 산업 종사자에서 폐암(3.27, 1.21-6.30)과 비호치킨림프종(.56, 2.22-10.22), 백혈병(5.17, 2.06-9.49)에서 대조군에 비해 질병 발생의 위험도가 상대적으로 높은 것으로 나타났다.

<표 3-27> 석유화학단지 근로자의 건강영향

	성별	실제 관찰수	표준화발생비	표준화발생비 범위	
				LL	UL
C15 식도암	남	8	0.51	0.22	0.90
	여	0	0.00	-	55.95
C16 위암	남	299	1.10	0.96	1.21
	여	10	1.69	0.79	2.82
C18-21 대장암	남	187	1.01	0.85	1.14
	여	6	1.65	0.61	3.17
C22 간암	남	138	0.84	0.69	0.97
	여	1	1.03	0.07	4.45
C23-24 담낭 및 기타 담도암	남	23	0.96	0.59	1.35
	여	1	2.99	0.21	12.99
C25 췌장암	남	39	1.30	0.89	1.68
	여	0	0.00	-	2.88
C33-34 폐암	남	94	0.94	0.74	1.11
	여	6	3.27	1.21	6.30
C50 유방암	남	0	0.00	-	1.04
	여	23	1.10	0.67	1.55
C53 자궁경부암	남	0	-	-	-
	여	5	1.75	0.58	3.57
C56 난소암	남	0	-	-	-
	여	5	1.62	0.54	3.32
C61 전립선암	남	62	0.99	0.74	1.22
	여	0	-	-	-
C82-86 비호지킨 림프종	남	32	0.83	0.54	1.10
	여	7	5.56	2.22	10.22
C91-95 백혈병	남	30	1.17	0.76	1.58
	여	7	5.17	2.06	9.49

(2) 사업장 별 직업관련 질환 분석 및 맞춤형 건강관리

구축된 코호트를 통해 사업장 특성에 따른 분석이 가능한지 여부를 연구하겠다. 사업장 특성에는 산업의 종류 및 근무환경, 유해물질 노출 여부 및 사업장 규모가 있다.

상기 특성을 파악할 수 있는 자료가 있는지 확인하였다. 이에 따라 현재 분석 가능한 것으로는 산업의 종류 사업장 이름, 사업장 규모가 유일하였다.

이외에 근무환경과 유해물질 노출 여부는 산업 분류를 통해 동질성을 유추하는 방법이 대안이지만, 현재 자료로는 결과의 현실성이 부족하다. 다만, 추후 유해물질 노출 여부 파악이 어렵다는 단점은 특수건강검진과 사업장 작업환경 측정 자료가 융합될 경우 상당 부분 극복할 수 있다. 이에 대한 추가 논의가 진행 중이다.

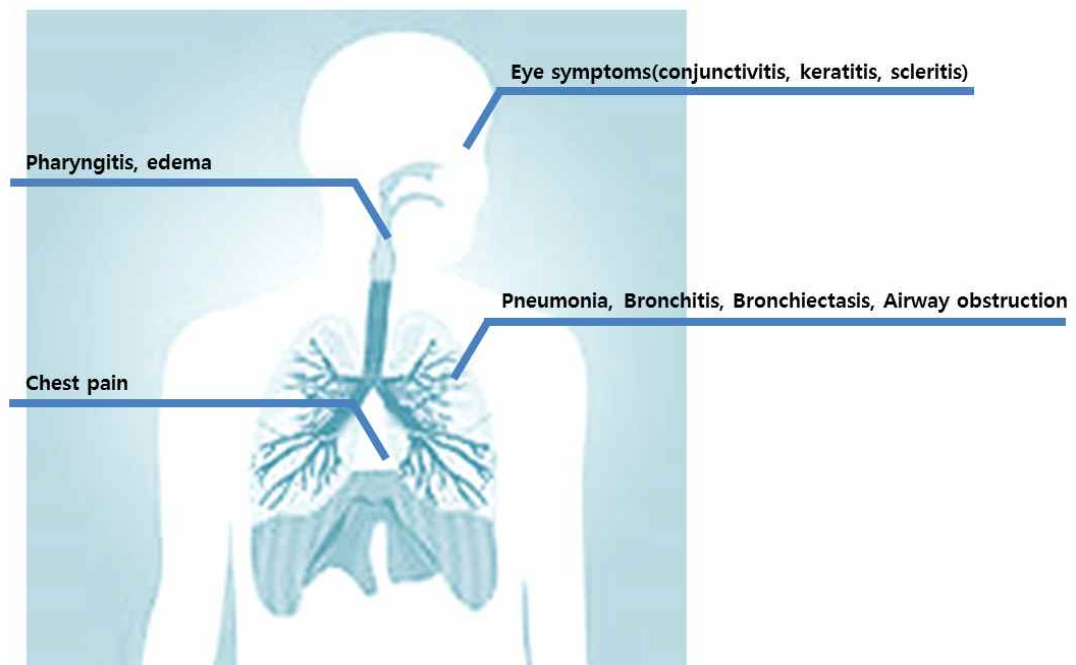
300인 이상의 사업장은 만성질환과 근골격계 질환의 유병률을 산출하여 배포하는 작업을 건강보험공단에 요청하여 실제 진행 중이다. 다만, 연구진 회의 결과 특정 사업장의 질병 분포를 무분별하게 공개하는 것은 근로자의 개인정보를 침해할 위험이 있다고 판단되었다. 300인 이하의 사업장의 경우에는 개별적인 유병 근로자를 추적할 수 있을 위험성이 더 높으므로 실제로 무분별한 유병률 공개는 개인정보 보호의 측면에서 위험성이 높다. 따라서 이러한 300인 이하의 사업장에서는 개인의 유병정보를 공개하는 것보다는 개선 가능한 만성질환과 근골격계 질환 관리에 있어 300인 이상 사업장에서 얻어진 정보를 이용할 수 있다고 판단했다. 특히, 각 진단별로 5명 이하의 유병인구를 가지고 있는 질환의 경우 개인을 추정할 수 있으므로 개인정보를 보호하는 차원에서 특정 사업장 별 특정질환 유병 근로자에 대한 조사 및 자료 공개는 실익과 인권 모두를 고려해야 한다고 판단된다.

(3) 유해물질 및 화학사고 이후 추적조사

거의 매년 뉴스에서는 각종 누출사고 등에 대한 보도가 이어지고는 한다. 특히 일부 누출이 아닌 유해물질의 대규모 누출사고와 같은 경우에는 직접 사고현장에 있었던 근로자뿐만 아니라 사고지역 주변 주민들의 건강영향에 대하여서도 감시체계를 작동할 필요가 있다. 이를 통해 유해물질 노출로 인해 어떠한 건강영향이 어느 정도의 기간 동안 지속되었는지 등에 대하여 좀 더 용이하게 파악할 수 있을 것이라 생각되며 효과적인 사후관리의 기반적인 근거로서 추적조사 결과가 사용될 수 있을 것이라 생각된다.

그 중, 암모니아 누출 사고 이후의 암모니아 노출자에 대한 수진내역 연구를

예로 들 수 있다. 2014년 2월 13일 오후 1시 경기도 남양주시 소재 한 공장에서 냉매인 암모니아 가스가 누출되어 체류하던 중, 원인 미상의 점화원에 의해 폭발이 발생하여 4명의 사상자가 발생했던 사고가 있었다.



[그림 3-22] 암모니아에 의한 유발 증상

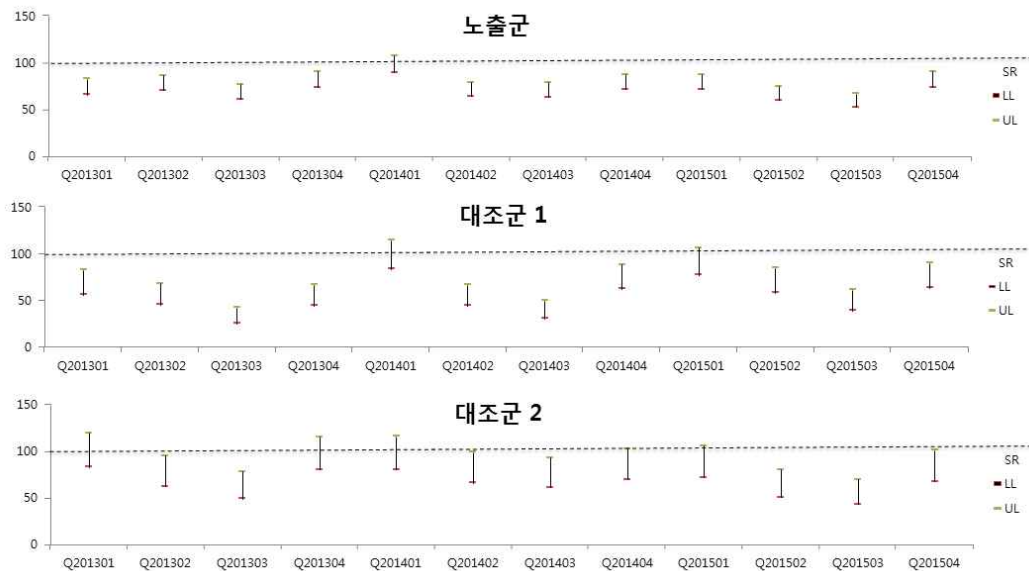
암모니아는 눈, 피부 및 상기도 점막을 자극하여 급성기 증상으로는 눈의 작열감, 인후통, 쉼 목소리, 기침, 타액 분비 증가 등의 점막 자극 현상을 유발할 수 있으며 그 외 두통, 가슴통증이나 오심, 구토를 호소하기도 한다. 이러한 암모니아 노출의 임상적 증상들을 파악한 뒤, 질병분류기호 ICD-10코드 중 이러한 증상과 연관된 진단명을 연구진의 논의를 통하여 선택하였다. 선택한 진단명은 다음 <표 3-28>과 같다.

<표 3-28> 암모니아 노출 관련 추정 질환의 ICD-10코드

ICD-10 코드	관련 질환명
J02	Acute pharyngitis
H10 - H16	Abnormalities of Conjunctiva and Sclera
R00 - R09	Abnormalities of Cardio-Respiratory tract (Hemorrhage, dysarrhythmia etc.)
F40 - F48	Psychiatric diseases (anxiety disorder, obsessive-compulsive disorder, etc.)

그 뒤 국민건강보험공단의 빅데이터를 이용해서 집단별 노출시점(분기)을 기준으로 위험도를 구하였으며, 고노출집단의 각 분기를 기준으로 한 분석도 시행하였다. 먼저 집단별 노출시점을 구하기 위해서 노출군과 대조군 두 곳을 거리에 따라 선정하였다. 대조군 1과 대조군 2로부터 노출장소까지의 거리는 비슷하였다. 그리고 암모니아 누출사고가 발생한 2014년 1분기를 기준기간으로 선정하여 이 시점에서의 지역별 외래이용률을 100으로 추산하였으며 질병분류코드에 따라 각 분기별 외래이용률을 분석하였다.

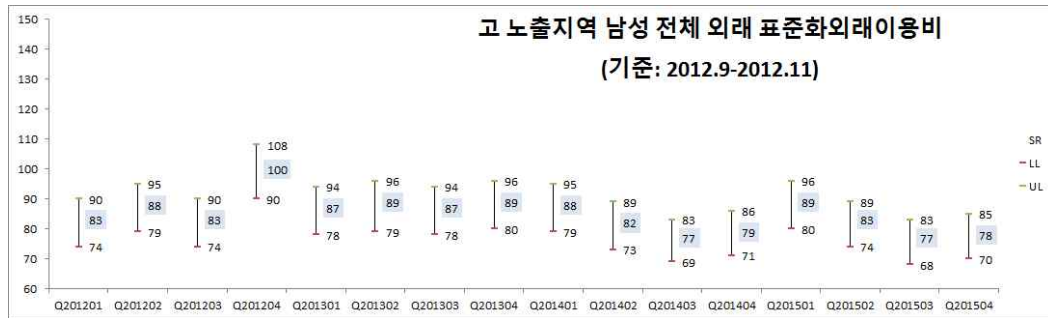
이 중 급성 인두염(Acute pharyngitis)로 인한 남자에서의 외래이용률을 예로 들어 살펴볼 때 노출군에서는 사건이 발생한 2013년부터 2015년까지 총 12개 분기의 기간 중 사건이 발생한 2014년 1분기에 가장 높은 외래이용률을 보였다. 대조군 1과 대조군 2에서도 해당기간 높은 외래이용률을 보였으나 각각의 집단 모두에서 다른 분기에서도 2014년 1분기만큼 높은 급성인두염 외래이용률을 보였다. 따라서 노출군에서는 노출 사건 직후 기간 동안 급성인두염으로 인한 외래이용률 증가가 다른 대조군들에 비해 두드러졌다.



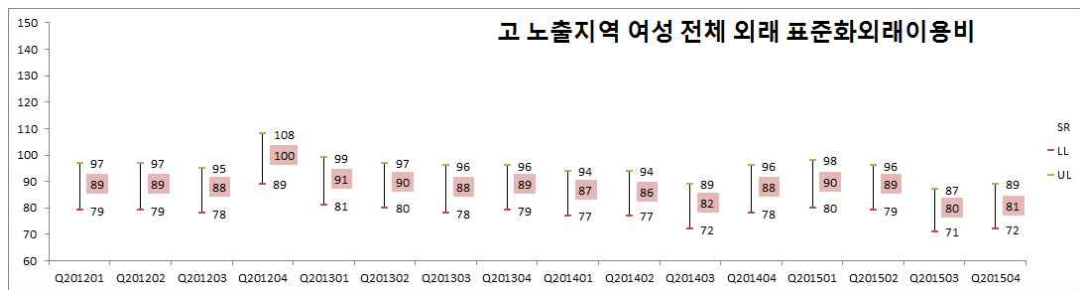
[그림 3-23] 각 집단별 노출시점(분기)을 기준으로 본 남성에서의 급성인두염 외레이용률

암모니아 누출 사고의 예 외에도 2012년 구미 불산 누출 사고 발생에 대해서도 유사한 방법으로 코호트를 구축할 수 있었다. 먼저 지역에 대한 정보를 이용하여 고·중·저노출군을 분류하였다. 노출수준에 따라 구분한 지역마다 불산 사고 발생일인 2012년 9월 27일이 포함된 2012년 9월 - 11월 외레이용자 수를 기준으로 하였으며, 그 기간의 전 9개월, 후 3년 동안의 전체 외래 및 선택 상병 코드에 대한 표준화외레이용비를 산출하였다.

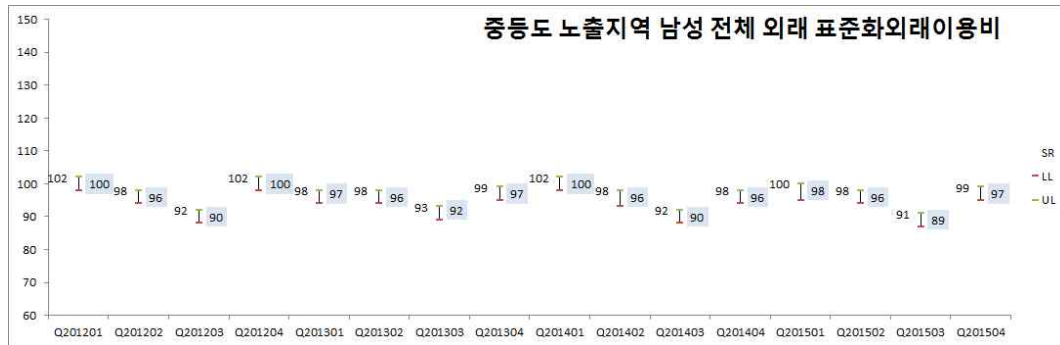
불산 사고 발생 기준 기간을 중심으로 고노출지역의 경우 사고 전·후 모두 기준기간보다 전체 표준화외레이용비가 낮았다. 중등도 노출지역의 경우 남성에서는 사고발생 전과 발생 후에 기준기간과 표준화외레이용비가 동일한 기간이 각각 1회씩 있었으나 유의미한 증가기간은 없었고 2014년 1분기를 제외하고는 모두 기준기간보다 감소하였다. 여성의 경우는 기준기간보다 외레이용이 증가한 기간이 없었고 모두 기준기간보다 낮았다.



[그림 3-24] 부산 고노출지역 남성 전체 표준화외래이용비



[그림 3-25] 부산 고노출지역 여성 전체 표준화외래이용비



[그림 3-26] 불산 중등도 노출지역 남성 전체 표준화외래이용비

저노출지역에서는 남성은 화학사고 전인 2012년 1분기의 표준화외래이용비가 기준기간보다 높았으며, 사고 이후의 기간 동안은 표준화외래이용비가 기준기간보다 동일하거나 낮았다. 여성에서는 기준기간보다 표준화외래이용비가 높은 기간이 사고 전·후에 모두 존재하여 일관성을 찾을 수 없었다.

결론적으로 고노출지역은 남·여 모두 불산누출시점 부근인 기준기간 동안 표준화외래이용비가 가장 높게 나타났으며 그보다 더 이전 이후 시점에서는 표준화외래이용비가 기준기간보다 낮아졌다. 중등도 노출지역에서 남성의 경우 기준기간 이전 이후에 표준화외래이용비가 기준기간과 비슷했던 적이 있었으나 전반적으로 낮게 나타났으며 여성에서는 사고 전후 모두 표준화외래이용비가 기준기간보다 낮았다. 저노출지역은 기준기간 기준으로 표준화외래이용비가 높고 낮은 기간이 혼재되어 일관된 해석을 하기 어려운 점이 있었다. 즉, 고농도 노출지역과 중등도 노출지역의 경우 불산 누출 사고 발생 시 외래이용이 상대적으로 가장 증가하였다가 이후 원래 상태로 환원되어 사고로 인한 건강영향이 주민의 의료이용에 지속적인 영향을 미치지 않았다.

2) 직업관련 코호트 시스템 구축 및 관리방안 설계

추후 국민건강보험공단의 건강보험 수진자료를 이용한 방대한 데이터를 효과적으로 이용하여 향후 지속적인 직업 관련 특성과 직업 관련 질환 사이의 연관성 및 인과관계에 대한 연구를 진행하기 위해서는 기본적인 체계가 필요하다.

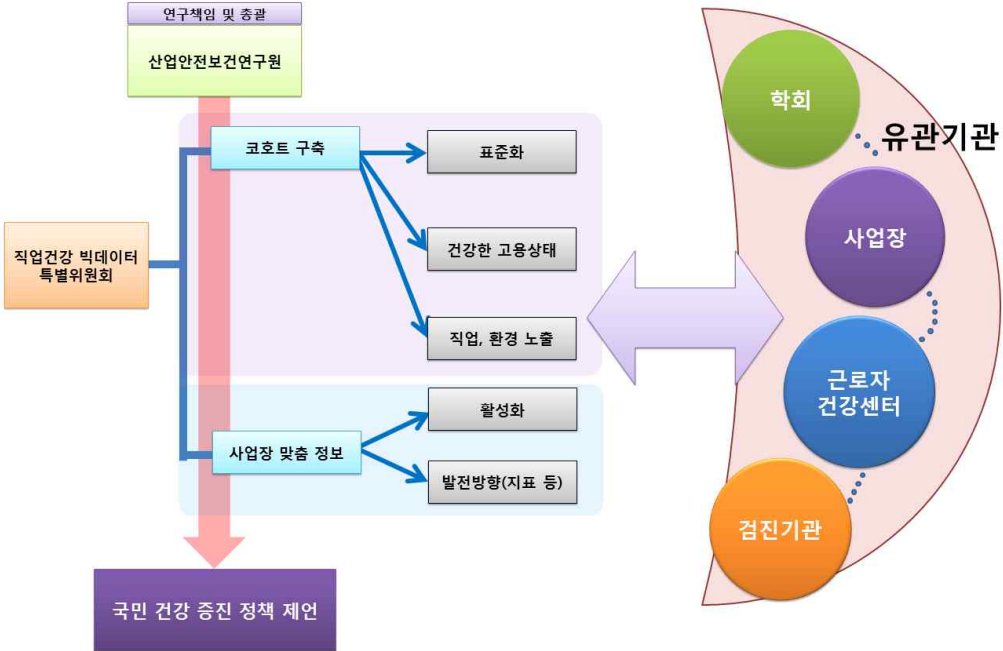
(1) 직업관련 코호트 구축을 위한 데이터 마련

향후 다양한 맞춤형 코호트를 구축하여 연구를 진행하고자 한다면, 장기적 관점에서 국민건강보험공단 데이터 이용 권한을 얻는 과정이 필수적이다. 중요한 건강유해요인이 어떤 산업에서, 사업장에서 발생할 지 미리 알고 있기 어렵기 때문에 언제 발생할지 모르는 직업병 발생이나 사고에 적절히 대응하고, 더 나아가 보다 본격적이고 주도적 연구를 진행하기 위해서는 다소 복잡할지라도 장기적인 안목을 가지고 전국민 데이터를 이용하기 위한 여러 금전적, 행정적 절차를 사전에 진행하는 것이 요구된다.

본 연구에서는 국민건강보험공단 데이터를 이용할 권한을 부여받지 못하였다. 따라서 본 연구에서는 직업성 질환 관리 및 예방에 대한 건강보험공단 데이터베이스의 활용가능성을 직접적 분석이 아닌, 다른 방법으로 시행하였다. 그리하여 2017년 8월과 11월 세미나 및 심포지움 자리에서 발표된 대한직업환경의학회와 국민건강보험공단의 업무협약을 통한 맞춤형 직업관련코호트 연구결과를 본 연구에서 소개하고 그 의미를 해석하여 앞 장에서와 코호트의 활용가능성을 평가하였다. 추후 보다 본격적이고 주도적인 연구를 진행하기 위해서는 장기적 안목을 가지고 데이터를 구매, 준비할 필요가 있다. 초기 절차 및 비용 문제는 단기간에 해결되는 것이 아니므로, 사전적 의미의 준비가 필요하다.

이러한 기본적 데이터베이스를 바탕으로 시행될 다양한 연구주제들의 총괄적 관리는 산업안전보건연구원의 차원에서 이루어질 예정이며, 연구주제를 선정하기 위한 방법으로는 내부의 직업건강빅데이터 특별위원회(가칭) 등을 설립하여 심의를 통해 선정된 주제에 대해서 알맞은 코호트를 구축하고 관련 연구를 진행

할 수 있다. 이러한 과정에서 이미 국민건강보험공단과 업무협약을 통해 다수의 건강보험데이터 분석 경험이 있는 유관 학회 등의 자문 및 협조를 얻는 것도 필요하다



[그림 3-27] 빅데이터 기반 직업관련 코호트 시스템 구축 및 관련기관 연계

4. 직업관련 코호트의 향후 이용방안

연구 결과, 본래 목적이 직업병 감시를 위한 것이 아닌 청구자료로서 구축된 데이터이므로 제한점과 한계점이 분명히 다수 존재하지만 탐색적 연구로서의 가치는 있다고 판단된다. 따라서 탐색적 연구로서의 가치에 주목하여 본 연구진의 연구 및 고찰한 결과, 건강보험공단 청구데이터를 이용한 직업 관련 코호트 구축은 다음과 같은 이익을 기대할 수 있을 것이라 생각된다.

1) 근로자건강센터의 효율적 활용

상기 구조대로 구축될 경우 첫째, 소규모사업장의 건강 현황에 대한 실태 파악을 통해 근로자건강센터의 보다 효과적인 역할 수행이 가능할 것으로 기대할 수 있다. 앞서 조사한 바에 따르면 5인 미만 사업장에서의 직장가입자 수는 현재 약 88만 여 명으로, 약 62.3%의 비율을 차지하고 있으며 보건관리자의 선임 의무가 없는 50인 미만의 중소기업에 종사하고 있는 근로자 수 (직장가입자 수)는 약 97.45%로 절대 다수를 차지하고 있다. 그러나 개업과 폐업의 전환이 빠르고 보건관리자가 부재하는 등 소규모사업장의 보건관리는 어려운 측면이 존재하므로, 현재까지의 소규모사업장 근로자의 건강관리는 대기업 종사자나 특정 연구대상자를 추출해 연구한 지식을 기반으로 하여 이루어지는 경향이 있었으나, 본 코호트를 통하여 소규모사업장의 건강실태에 대해 보다 구체적으로 추산할 수 있을 것으로 기대된다. 이러한 활용을 통해 근로자건강센터는 소규모 사업장 근로자 중에서도 어떤 직업군이 어떠한 질환에 취약한 지, 어떠한 점에 더 중점적으로 접근해야 하는지에 대한 정보를 얻을 수 있어 근로자건강센터의 보다 효율적인 운영이 가능할 것으로 기대된다. 다만, 소규모사업장일수록 사업장 내 개인을 특정할 수 있는 가능성이 높아지므로 이에 대한 주의가 필요하며, 각 개별 소규모사업장의 건강영향에 대해 살펴보는 것보다는 각 지역별 근로자건강센터가 관할하는 지역별 소규모사업장, 업종별 소규모사업장의 건강영향을 분석하는 방향을 고

려하는 것이 바람직하다고 생각된다.

2) 직업관련질환 감시체계로서의 기능

둘째, 직업관련코호트는 감시체계로서의 기능을 담당할 수 있다. 업종분류코드별 구축된 코호트나 앞서 소개한 병·의원 종사자 코호트, 반도체·디스플레이·타이어 제조 사업장 종사자 등 각 코호트별 산재 또는 외상, 건강영향에 대한 장기적인 감시체계로서 기능할 수 있으며 다른 직업·업종군보다 특별히 높은 발생률 혹은 유병률을 보이는 질환을 모니터링할 수 있다. 이러한 감시체계는 새로운 건강영향을 발견하고 직업관련질환을 조기 관리를 가능하게 함으로써 장기적으로는 근무 환경을 개선하고 근로자의 건강을 증진시킬 수 있을 것으로 기대된다.

3) 누출사고 이후 모니터링

셋째, 구축된 코호트를 바탕으로 직업 뿐만 아니라 누출 사고 이후의 건강영향 모니터링이 가능하다. 일회성이지만 유해물질에 고농도 노출이 의심되는 누출 사고 발생 시, 노출 근로자나 주변 지역 주민들의 건강영향을 추적함으로써 건강영향에 대한 의학적 확인뿐만 아니라 향후 추적 관찰 계획이나 관련 질환을 감시하는 데 유용한 방법으로 사용할 수 있으며 보상 범위나 기간을 설정하는 데 있어서도 타당한 근거를 제시해 줄 수 있을 것으로 생각된다.

5. 현 분석 제한점 및 적용의 한계

본 연구의 분석 목적은, 향후 산업보건 및 직업병, 직업관련 질환의 예방 및 관리를 위해 건강보험공단 청구자료를 활용할 수 있는지 여부를 확인하는 데 있다. 따라서, 이러한 활용가능성에 대한 탐색적 분석을 차용한 것¹⁾으로, 본 연구의 결과를 직업병 발생의 원인과 결과로서 이해하는 것은 큰 오류가 발생할 소지가 있다. 그러므로 본 연구 결과를 활용하는 데 있어서는 아래와 같은 제한점과 한계점을 명확히 하여 과학적으로 타당한 선에서 이해하는 과정이 반드시 수반되어야 한다.

1) 후향적 코호트로서의 제한점

본 연구는 이미 구축된 건강보험공단자료, 즉 이차 자료를 이용한 연구이다. 이차자료를 사용하기 때문에 이를 바탕으로 한 직업관련 코호트 구축 시 별도의 조사비용이나 데이터 구축비용 없이 이미 만들어진 전국민 데이터를 사용할 수 있는 것이 장점이다. 그러나 본래 직업성 질환에 대한 연구나 의학적 연구 등을 위해 수집된 자료가 아니므로 연구의 범위가 기존에 조사된 변수에 제한된다는 단점이 있고, 미처 조사되지 못하거나 부적절한 조사결과가 있을 수 있다는 제한점이 있다. 더욱이 건강보험공단 데이터베이스를 이용하여 구축되는 코호트는 후향적 코호트의 한계를 갖게 된다.

현재의 각 산업군 및 특정 산업종사자에 대한 코호트 분석에서는 대조군의 경우와 특정 비교군이 입사와 퇴사를 비롯한 여러 사회·경제적 변수, 지역적 특성 등이 다름에도 불구하고 모두 같은 후향적 코호트 구축 방법을 이용하였다. 본 분석에서 퇴사자를 추적 관찰하였지만 코호트 구축 방법은 각 업종마다, 분석마다 세밀하지 못하였으므로 향후 연구에 있어서는 각 가설에 맞는 코호트 구축방

1) 본 연구는 건강보험청구자료를 직접 분석하지 않았다. 보고서에 수록된 연구결과는 윤진하 등이 2017년 8월에 시행한 세미나 발표자료를 2차 가공한 것이다.

법을 설계하여 사용하여야 한다.

또한, 본 연구에서의 코호트는 2015년의 사업장 정보를 기준으로 한 후향적 코호트로, 그 특성상 2015년 이전에 폐업한 사업장의 근무 근로자에 대해서는 직업력에 대한 정보 누락이 발생한다. 대조군인 공무원과 사립학교 교직원에 비해서 일반 사업장은 2014년 이전 폐업율이 높으므로 근로자의 종사 기간이 실제보다 단축되는 경향을 갖게 되며, 특히 소규모사업장의 경우 상대적인 높은 폐업율로 인해 이런 경향이 더 크게 나타날 것으로 생각된다.

따라서 건강보험공단 데이터를 이용하여 직업성 질환 및 근로자의 건강영향에 대하여 얻어진 연구결과를 해석할 때에는 자료의 근본적 한계점에 대하여 이해하고 결과를 해석할 필요가 있다.

2) 업종코드의 불명확성

앞서 논의한 것처럼, 현재 국내 업종코드는 불명확성을 담보하고 있다. 최근 표준산업분류가 21개 대분류로 이루어져 있는 것과 달리 건강보험공단 업종분류코드는 대분류 총 17개로 과거의 표준산업분류와 유사한 형태를 가지고 있지만 사업장이 정확히 분류되었다고 보기 어렵다. 건강보험공단 데이터베이스 내 업종분류코드는 일반적으로 사업주의 자가보고에 의해 조사된 결과로 같은 업종코드를 가지고 있다고 해서 같은 산업군이라 생각하기 어렵고, 다른 업종코드를 가지고 있다고 해서 다른 산업군이라 확인하기도 어렵다.

업종분류코드에서 가장 하위분류인 6자리 코드를 사용하여 업종을 선택하였을 때, “140851” 의료업, “140852” 수의업, “148511” 병원, “148512” 의원 “148513” 공중보건의료업, “148520” 수의업으로 총 63,599개 사업장을 확인할 수 있는데, 하위 항목에는 의료기관 뿐만 아니라 기공소, 협회, 자활센터 등 이질적인 사업장 이름도 다음 <표 3-29>와 같이 섞여 있는 것을 확인할 수 있었다. 즉, 업종분류코드를 사용할 경우 병의원 및 기타 의료관련기관의 구분이 불가능하였으며 양·한방병원의 구분도 불가능하다는 단점이 있었고, 교육기관을 겸한 의료기관의 경우 보건, 의료기관 종사자의 업종 대분류인 ‘14’의 하위항목이 아닌 다른 항목

에 기재된 경우가 있어 이에 대한 파악이 불가능했다. 또한, 오분류의 존재에 대해서는 사업장 이름을 목록에서 검토할 때 확인 가능하지만, 사업장 수가 방대하므로 그 규모를 명확하게 파악하는 것은 불가능했다.

<표 3-29> 업종분류코드에 따른 사업장 선정 예시

업종분류 코드	업종명	사업장 수 (개)	사업장 예
140851	의료업	211	치과기공소, 의학연구소, 의료재단 사무소, 의료기상사, 의료기 생산업체, 자활센터, 사회복지시설
140852	수의업	2	유지재단, 주식회사
148511	병원	4723	병원, 의원, 일부 재단, 의과대학 법인, 병원 내 약국
148512	의원	5610	의원, 치과, 한의원, OO대한방병원
148513	공중보건의료 업	972	대한적십자사, 결핵협회, 보건소, 공공의료원, 협회, 기공소, 약사회, 약국
148520	수의업	1522	동물병원, 수의약품 관리소, 유기동물보호소

즉, 현재 업종분류코드는 이질성과 이질성을 모두 가지고 있어 업종분류코드를 이용하여 얻어진 결과를 해석할 때에는 주의가 필요하다.

또한 이러한 불확실한 업종분류코드 대신 사업장 이름을 통해 분석을 실시할 경우에는, 사업장의 공통 특성을 아우를 수 있는 공통 사업장명을 유한 개 선정하는 과정이 어려웠다. 예를 들어 차량 수리업체에 대하여 검색하고자 할 때에도 자동차 제조사마다 수리업체 프랜차이즈의 명칭이 다양하여 공통점이 없어 공통 검색어의 선정이 불가능했으며 기타 다른 차량 수리업체의 경우에도 ‘공업사’ 등 비특이적인 사업장 이름을 가지고 있어 텍스트분석을 통한 사업장 추출의 한계를 보여주었다. 또한 이름 및 업종분류코드를 사용해서도 동명의 기업이 다수 존재하는 경우에는 사업장 확인이 어려웠다. 사업장 소재지 등을 사용하여 파악할 수는 있지만 실제 소재지와 건강보험공단에 신고된 소재지가 다를 가능성이 있

어 완전한 방법은 아니다.

따라서, 현재의 분석을 역학조사 등 인과관계를 입증하는 용도로 사용하는 것은 과학적이지 못하다. 즉, 현재의 업종분류에 따른 건강영향 조사 결과를 탐색적인 연구 결과로서 향후 연구방향에 대한 제안을 위한 도구 이상의 의미를 부여하기는 어렵다. 업종분류에 의한 건강영향조사를 현재보다 유의미한 것으로 얻기 위해서는 향후 업종분류코드를 체계화하여 정리할 필요가 있다. 본 연구진에서 검토한 결과, 통계청의 표준산업분류, 국민건강보험공단의 업종분류코드, 산재 데이터에서의 업종분류 각각이 각자의 장·단점 및 한계를 가지고 있었으며 이름을 검토했을 때 상당한 공통부분을 확인할 수 있었지만, 여전히 서로 통합, 호환 불가능한 부분도 분명 존재하였다.

이런 이유에서, 업종분류코드에 대한 대분류, 중분류, 소분류를 갖춘 체계 마련이 필요하다. 그리고 업종 이름과 분류코드가 적절히 분류되었는지를 확인할 수 있도록 사업장에 대한 전반적인 검토가 필요하다. 그리고 업종분류코드 마련이 어렵다면, 표준산업분류 등의 다른 분류를 차용하는 방법 등에 대한 면밀한 고찰이 필요하다. 현재 수준에서의 사업장 분류로는 정확성과 동질성 면에서 부족한 점이 있으므로, 탐색적 의미 이상의 연구를 진행하기 어렵다. 따라서 전국민 전수 데이터라는 이름에 걸맞은 코호트를 구축하기 위해서는 근로자 추출 시 적절한 동질성과 정확성을 확보할 수 있도록 업종 분류의 구축이 필요하다.

3) 직종 및 직무 정보의 부재

또한, 업종분류의 불확정성 외에 근로자의 직무, 직종을 확인할 수 없다는 점도 본 코호트의 제한점이다. 현재 국민건강보험공단에서 조사된 근로자의 정보는 업종에 대한 정보는 획득할 수 있으나, 이 근로자가 해당 사업장에서 ‘관리자’인지, ‘전문직’인지, ‘생산직’인지에 대한 정보는 획득하기 어렵다. 직종코드가 존재하기는 하지만, 이는 정기적인 건강검진에서 대상자와 비대상자를 구분하기 위한 수단이지 표준직업분류와 같은 직업을 구분하는 코드로서의 기능은 없다.

의학적 연구, 특히 직업병 및 직업관련질환의 기본원칙은 대부분 근로자의 노

출력에 근거하고 있다. 동일 사업장이라고 하더라도 근로자의 맡은 직무에 따라 혹은 부서에 따라 유해물질에 대한 노출력은 매우 상이하다. 본 연구에서는 데이터에서 직종 및 근무부서 분류가 불가능하였으므로 노출력에 근거한 원인 및 결과의 판단은 불가능하였다.

직종 분류가 불가능하다는 점에 있어서 고용노동부 및 관련 부처의 협력을 통해 특수건강검진 자료 또는 사업장의 작업환경측정자료와 같은 다른 데이터를 융합하여 해당 사업장의 실제적인 노출력을 확인하는 방법을 생각할 수 있지만, 이는 관계부처의 협력 및 개인정보보호법과 관련한 여러 절차 등이 필요하다. 따라서, 현재로서는 근로자의 직종 및 직무분석이 불가능하므로 사업장 내 근로자들의 노출력이 서로 상이할 수 있으므로 연구에 있어 과대추정되거나 과소추정되는 등 결과에 편의(bias)가 발생할 수 있어 연구결과 해석 및 이용 시 주의하여야 한다.

4) 대조군의 불명확성

각 직업군별 건강영향 및 질환의 위험도를 산출하기 위해서 공무원과 사립학교 교직원을 대조군으로 선정하였다. 공무원과 사립학교 교직원이 다른 직종에 비해 고용 안정성 측면에서 상대적으로 동질성을 갖고 건강한 집단일 것이라는 기대 하에 대조군으로 설정하기는 하였다. 그러나 실제로는 공무원과 사립학교 교직원 내에도 업종의 다양성이 존재한다. 공무원 및 사립학교 교직원 중에서도 일반 공무원과 교육 공무원은 업무의 성격과 내용이 다르고, 근무 부처에 따라서 맡은 직무가 다양하지만 이를 하나의 집단으로 산정하였다.

또한, 타 직종대비 입사와 퇴사의 기본적 성격이 상이하다. 첫째로 입사 시점에 대해 생각할 때 입사는 임용의 과정을 거치므로 타 산업군과 최초 근무 시점의 연령에 차이가 발생하며, 임용 방법 중 진입장벽의 정도에 따라 근로자의 교육수준 등 다른 사회·경제적인 요인에 있어서도 차이가 존재할 수 있다. 퇴사의 경우 대부분이 정규직이므로 타 직종의 퇴사와 동일시하여 분석하는 경우, 건강근로자 효과 등으로 대변되는 선택편의에서 자유로울 수 없다는 제한점을 갖는다.

또한, 공무원 및 사립학교 교직원의 지역적 분포와 특이 사업장 또는 업종별 근로자의 지역별 분포는 필연적으로 서로 차이가 발생한다. 따라서 현재 연구된 코호트에서 얻어진 건강영향결과는 이러한 단순한 지역적 차이에서 자유로울 수 없다. 하지만 본 연구는 건강보험공단 데이터가 활용가능한 지에 대한 탐색적 연구로 종합적 분석을 실시한 것이다. 향후 비교군 선정과 위험도 산출을 보다 명확히 하고 위와 같은 제한점을 통제하기 위해서는 각각의 분석마다 명확한 가설을 세우고 해당 가설에 적합한 대조군을 선정하여 분석하는 과정이 필요하다.

5) 진단 질병의 불명확성

본 연구에서는 국제질병분류(ICD)를 기준으로 코호트 구축을 시범 운영한 사례를 제시하였다. 다양한 코호트의 전체적인 시범운영을 위해 진단 질병의 불명확성을 제한점으로 가지고 있다. 현재 구축된 코호트에서는 입원한 근로자의 입원 시 주진단상병의 상병코드를 사용하여 질병을 분석하였다. 입원 진단명을 사용할 경우 외래 진단명을 가지고 분석할 때보다 질병의 경중 상 실제 유병 여부와 진단명이 일치할 확률이 높다고 기대할 수 있지만, 상대적으로 경하여 외래 내원으로 오는 근로자를 놓칠 수 있다는 점과, 입원 시 진단명을 사용하였지만 여전히 불명확성이 완전히 제거되지 못했다는 제한점이 있다. 예를 들어 의료진의 실수나 의무기록의 오류로 환자의 주 진단명이 잘못 기재되어 있는 경우가 있을 수 있다. 또한, 진단적 목적으로 처음 예상되었던 진단을 기재하고 처치를 시행하였다가 다른 질환을 진단받고 재입원하게 되는 경우, 이전 기록이 수정되지 않는다면 근로자는 기록상 두 질병을 모두 가진 환자로 조사된다. 따라서 보다 진단명의 불명확성을 제거하고 이에 대한 신뢰성을 높이기 위해서는 단순한 상병코드에 따라 유병 여부를 판단하기보다는 환자의 시술, 수술, 투약 내역 등을 검토하여 해당 질환을 실제로 가지고 있는지에 대한 확인을 함께 시행하여야 한다. 이를 위해 추가적 자료를 종합하여 분석하는 과정은 본 연구에서는 자료 수집의 한계상 수행하지 않았다. 다만, 본 연구에서 사용된 코호트에서는 입원 시 진단명을 사용하여 오류의 범위를 최소화 하고자 하였다. 그 외에도 선별검사와

실제 진단 및 치료과정에서의 차이 및 오류의 발생은 질병의 성격과 경과에 따라 다를 수 있고, 같은 질병이라도 어떤 의료진을 만나느냐에 따라서도 다양할 수 있으며 해당 근로자가 어떠한 사회적 집단에 속해있는지 그 집단의 사회적 특성에 따라서도 다를 수 있다. 따라서 본 연구 결과를 해석할 때에는 단순한 표준화 발생비 뿐만 아니라 각 질병과 집단의 특성을 고려한 결과 해석이 필요하다.

6) 제한점에 대한 소결

본 연구 결과를 활용하고자 하는 이용자가 해석하려는 가설과, 본 연구의 연구자가 세운 가설이 상이할 경우 해석의 오류가 발생할 수 있다. 다시 강조하지만 본 연구의 분석 목적은 향후 산업보건 및 직업병, 직업관련질환을 예방하고 관리하는데 있어 건강보험공단 청구자료를 활용할 수 있는지 여부를 알아보기 위해 탐색적 분석을 차용한 것이다. 상기 논의한 제한점에도 불구하고, 코호트 구축의 활용방안을 최대한 종합적으로 검토할 수 있도록 다수의 결과들을 본 보고서에 수록, 소개하였다.

현재 연구결과는 후향적 코호트이며, 조사되지 못한 사회·경제적 요인, 지역적 요인, 업종 분류의 불명확성, 직종 및 직무에 대한 정보 부재, 진단 질병의 불명확성 등 과학적으로 인과를 입증하기에는 여러 제한점이 존재하였다. 따라서 본 연구의 결과를 해석 범위를 향후 산업보건 및 직업병 예방과 관리에 빅데이터를 기반으로 한 코호트 연구를 활용가능성에 대한 평가, 혹은 시범적 코호트 운영을 통한 탐색적 연구 그 이상으로 확대하여 해석하는 것은 과학적 오류를 범하게 되므로 옳지 못하다.

참고문헌

- 강성규, 안연순, 정호근. 1990년대 한국의 직업성 암. 대한산업의학회지. 2001;13:351-9.
- 안연순. 우리나라의 직업 환경 코호트. 대한직업환경의학회 학술대회 논문집. 76-117.
- 이원진. 직업/환경 코호트 연구의 설계와 원칙. 대한직업환경의학회 학술대회 논문집. 2009:68-72.
- 한경아. 대학병원 보건의료인의 근골격계 증상 관련요인. 한국산화기술학회 논문지, 2015;16(3):1928-1936.
- 김영실, 박재용, 박상연. 종합병원 간호사의 직무스트레스와 근골격계 증상과의 관련성. 근관절건강학회지. 2009;16(1):13-25.
- 윤진하. 건강보험공단 직업관련 빅데이터 코호트 소개. 대한직업환경의학회 가을학술대회. 심포지움 발표자료. 2017
- 윤진하. 김지현. 건강보험공단 직업코호트 활용방안. 대한직업환경의학회 가을학술대회. 심포지움 발표자료. 2017
- 윤진하. 김지현. 임성실 등. 직업코호트의 구축과 그 결과 현황; 국민건강보험공단 데이터베이스를 기반으로. 빅데이터 기반 직업·산업군 중심 코호트 구축연구 세미나. 세미나 발표자료. 2017.
- Laplaud, D. A., Confavreux, C. Etiology of multiple sclerosis. La Revue du praticien, 2006;56(12):1306-1312.
- Smith, K. J. Pathophysiology of multiple sclerosis. La Revue du praticien, 2006;56(12):1299-1303.
- Gérin, M., Siemiatycki, J., Déry, M., et al. Associations between

several sites of cancer and occupational exposure to benzene, toluene, xylene, and styrene: Results of a case control study in Montreal. American journal of industrial medicine, 1998;34(2): 144-156.

Zahm, S. H., Blair, A. Occupational cancer among women: where have we been and where are we going?. American journal of industrial medicine, 2003;44(6):565-575.

Nichols, L., Sorahan, T. Cancer incidence and cancer mortality in a cohort of UK semiconductor workers, 1970 - 2002. Occupational Medicine, 2005;55(8):625-630.

Lim, S. R., Schoenung, J. M. Human health and ecological toxicity potentials due to heavy metal content in waste electronic devices with flat panel displays. Journal of Hazardous Materials, 2010;177(1):251-259.

de Vocht, F., Sobala, W., Wilczynska, U., Kromhout, H., Szeszenia-Dabrowska, N., Peplonska, B. Cancer mortality and occupational exposure to aromatic amines and inhalable aerosols in rubber tire manufacturing in Poland. Cancer epidemiology, 2009;33(2): 94-102.

Peters, J. M., Monson, R. R., Burgess, W. A., Fine, L. J. Occupational disease in the rubber industry. Environmental health perspectives, 1976;17:31.

Summary

Title

Research for disease incidence, prevalence and health effects of workers by retrospective cohort, based on NHIS database

1. Objectives

Current retrospective cohort establishment is regarded as a fundamental research for managing known occupational diseases and investigating unknown risk factors and health effects. Therefore, the main purpose of this research is evaluating and considering possibilities of using National Health Insurance Service database for the researches of workers' health.

2. Methods

The common control group was selected as the public workers and teachers from private schools, from 2002 to 2015. For choosing the target diseases, researchers found the 20 most prevalent diseases from 2002 to 2015, by demonstrating inpatient number by 100,000 overall workers, hospital days by 1,000 overall workers, and hospital days by one inpatient workers. Then, choosing 10 diseases which reported the highest number of hospital days by 1,000 overall workers. Whole workers were divided into 17 industrial groups according to the industrial classification code assigned by NHIS, and demonstrated by age-standardized incidence ratio. Tracing the type of insurance, researchers could figure out the workers' starting point(=enroll point) and retiring point. Additionally,

researchers demonstrated and established many different occupational cohort such as health-workers cohort, land-transportation cohort, semiconductor-production workers' cohort, display-production workers' cohort, tire-production workers' cohort, dust-exposed workers' cohort. Researchers also evaluated the possibilities of those cohort by considering standardized incidence/prevalence ratios of popular diseases.

Results

In the established cohort, there were several interesting results from the cohort. musculo-skeletal diseases among health workers' were relatively highly reported than control groups. Also, cancers on GI tracts showed high standardized incidence ratios in dust-exposed workers' cohort and some other high standardized incidence/prevalence ratios among other cohorts. While considering usual working environments, those phenomenon could be explained by occupational and biological plausibility. However, this study has a lot of limitations at this time so far, the one who wants to use this results should be cautious to apply this. This is the retrospective cohort studies based on secondary database, the range of research is quite limited in the structured database. Also, basic information for workplace was obtained in 2015, therefore, this research has own weak point because it could not reflect the workers who worked in the company which closed before 2015.

Classification of industry were not verified enough yet. 17 different occupational groups have so much common part between standardized industrial classification, but also there are so much different part. Moreover, most of the workplace information and industrial types reported by the owner of enterprises, there are some vagueness and inaccuracies. It should be proved by future researches, but before then, people should carefully interpreted this research results.

The absence of occupational and job title information is another limitation of this study. Even if workers who work in the same place, they could be exposed to hazards in different levels by what kinds of role they have. However, there is no information about workers' job titles or working departments, it is unable to estimate the level of exposure accurately in the current cohort study. Not only in the workers' aspects, but there's a limitation in medical aspects. In the current cohort, researchers used inpatient disease code to ensure to reflect actual disease prevalence, but it still could not guarantee the actual diagnosis and the cause of admission. Therefore, for ensuring the diagnosis and the admission, further researches should match the diagnosis term with procedure codes, surgery codes, or medicine codes.

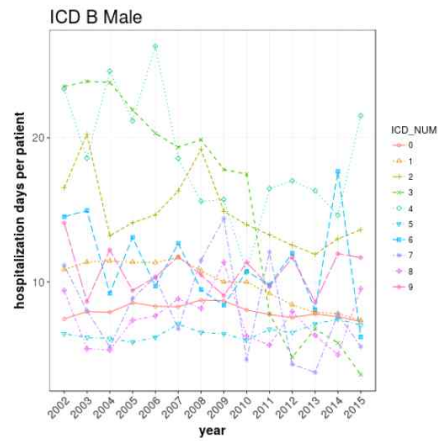
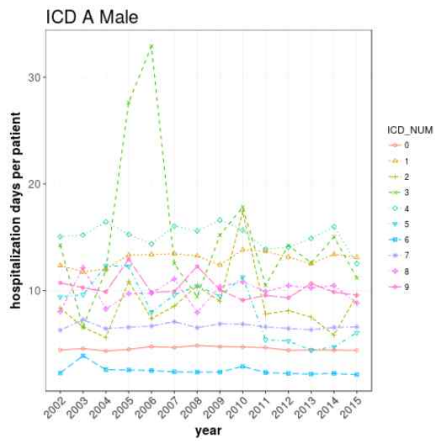
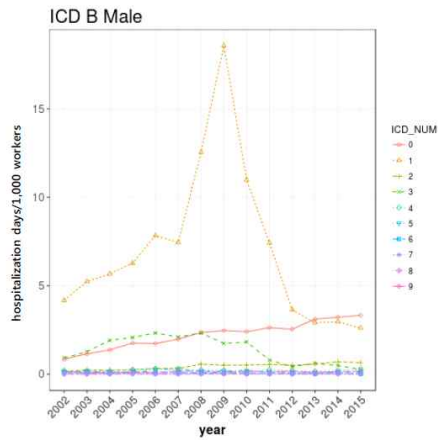
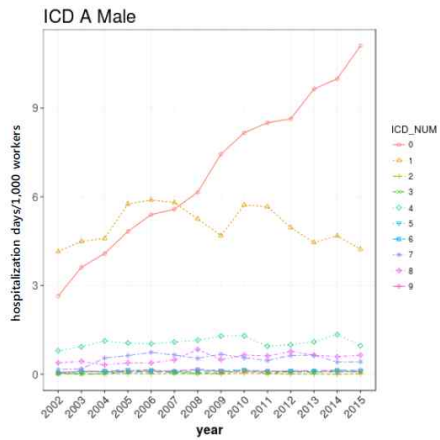
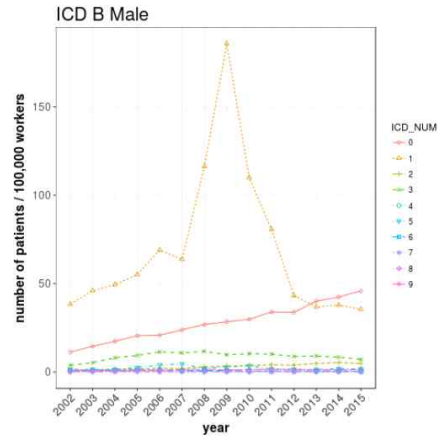
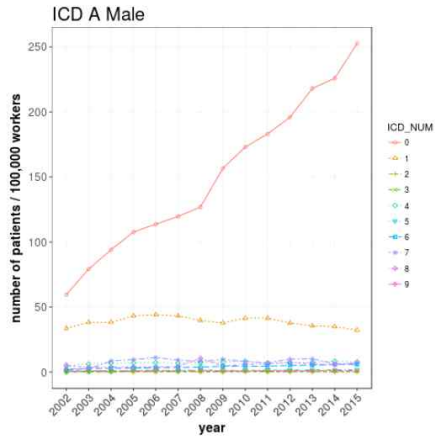
4. Conclusion

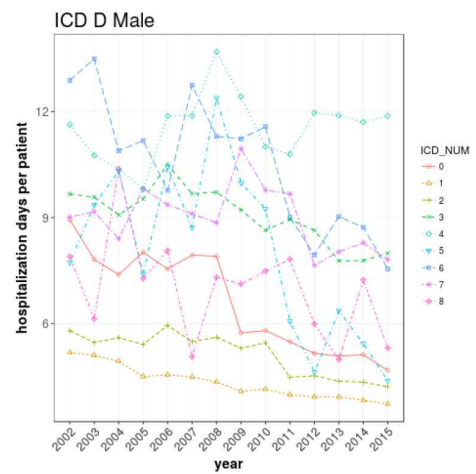
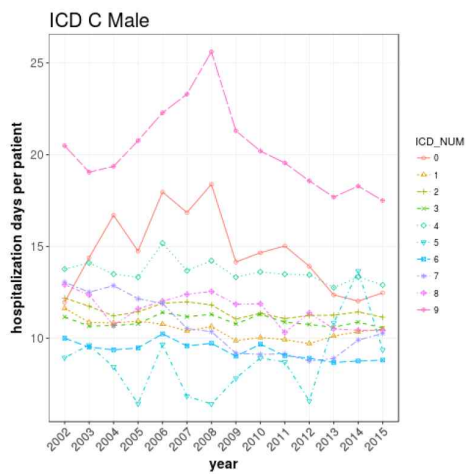
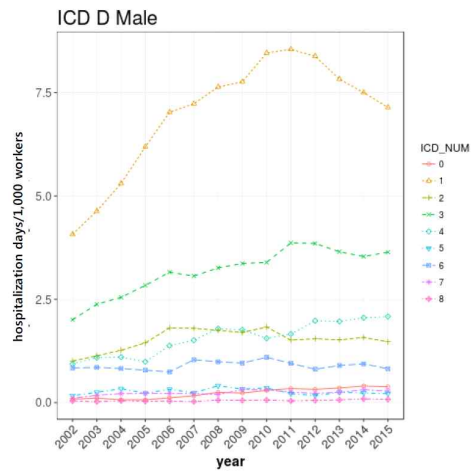
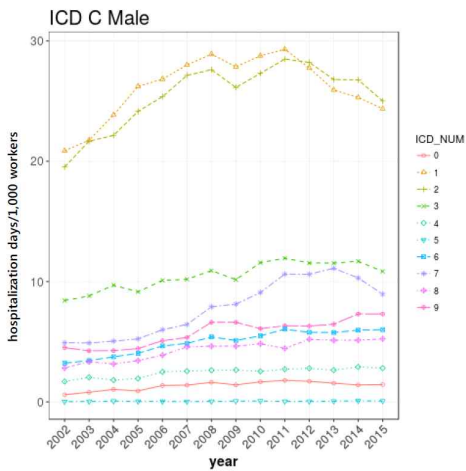
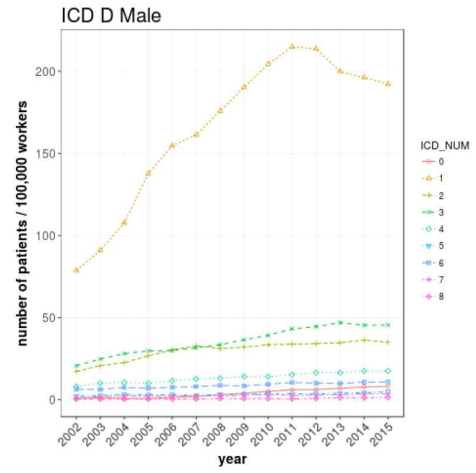
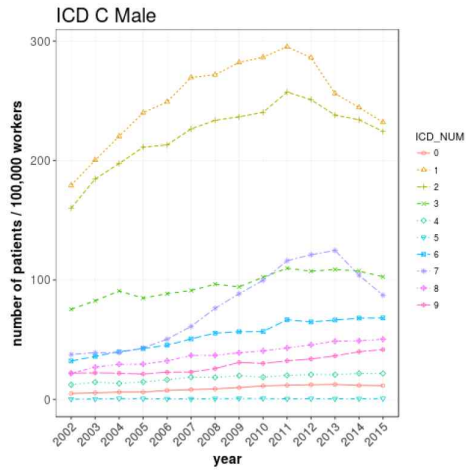
Using big data on investigation of workers' health effects is quite difficult methods, and it also needs some delicate data handling methods. However, considering its benefits and necessities, it is worthy of investigation. In the current stage, there are many limitations found in previously established cohort, but the researchers of this study concluded that it is still useful for exploring health risks of workers and finding new, potential risk factors of workers. In other words, the results of current cohort will be confined in exploratory results, but with careful interpretation and supplementation of current limitations, it could be applied with great potentiality and usability.

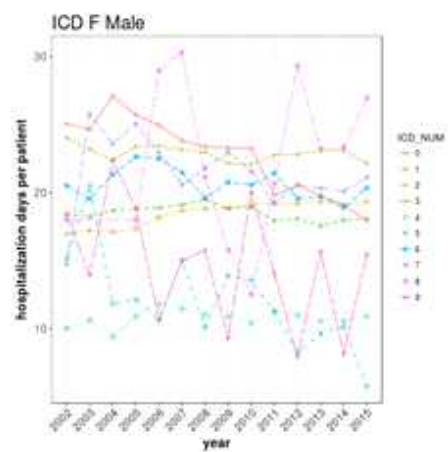
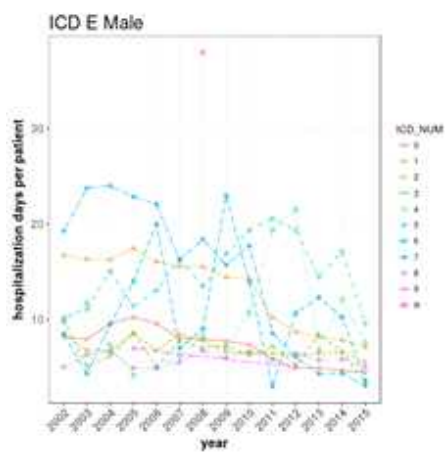
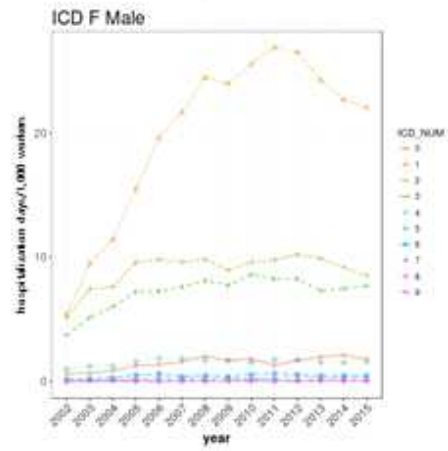
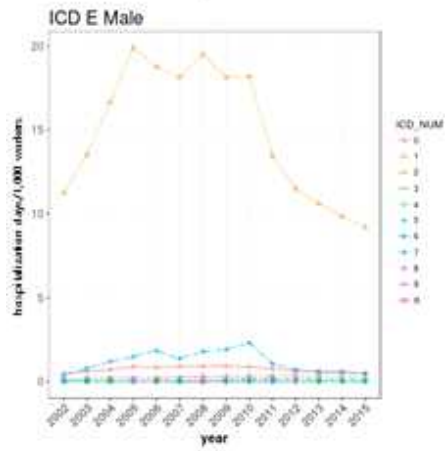
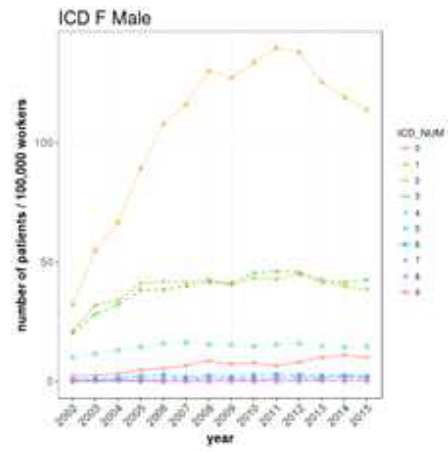
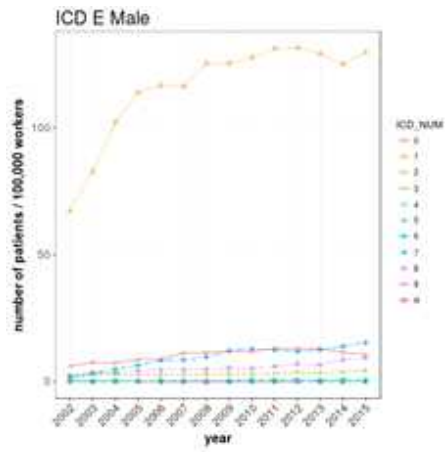
Keywords

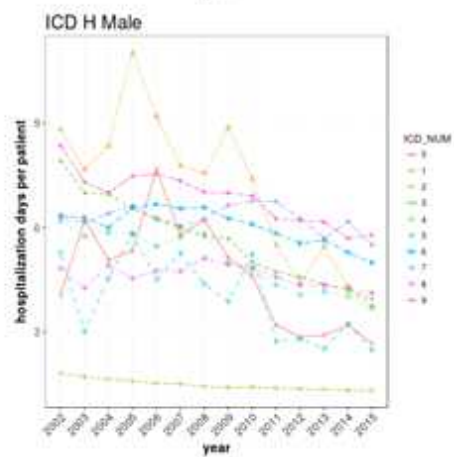
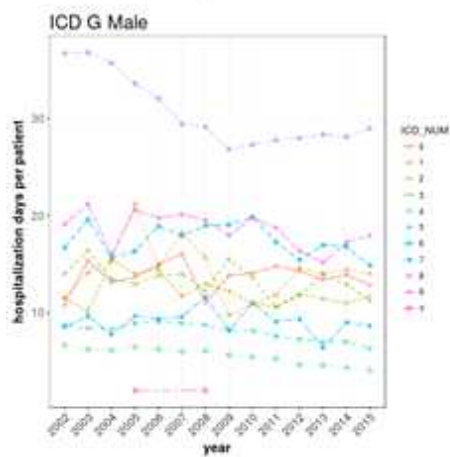
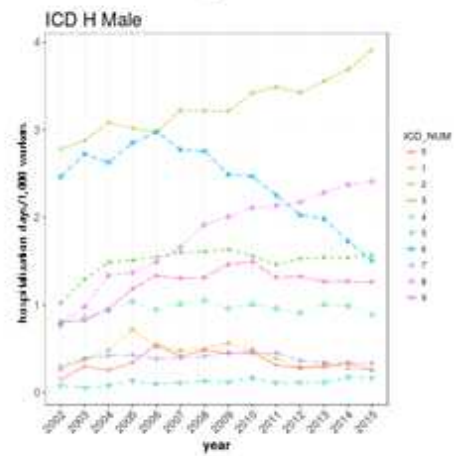
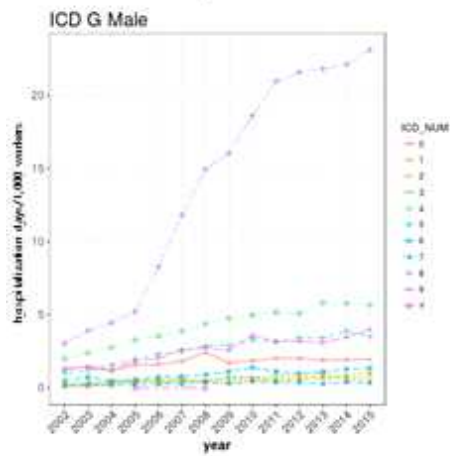
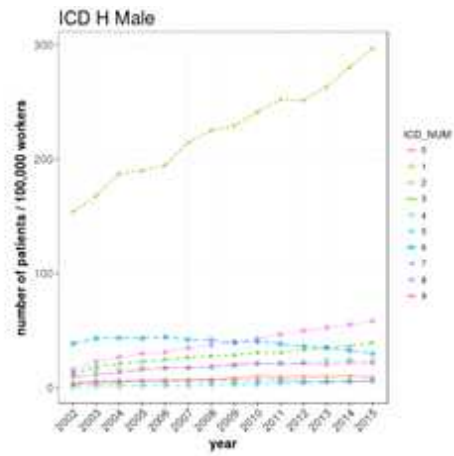
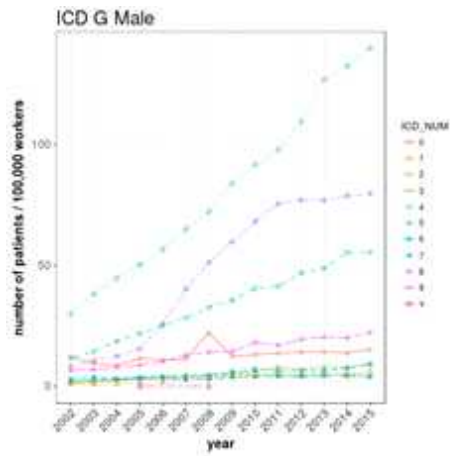
retrospective cohort, occupational cohort, big-data cohort

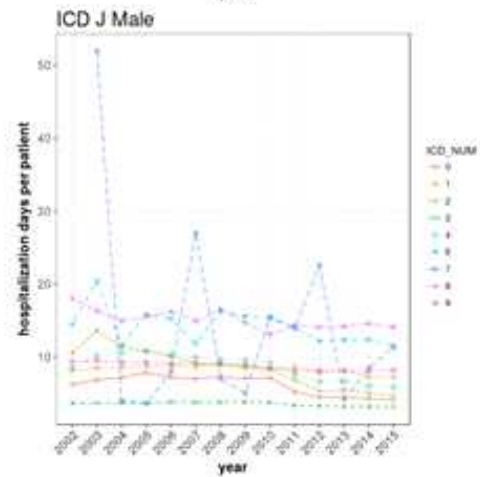
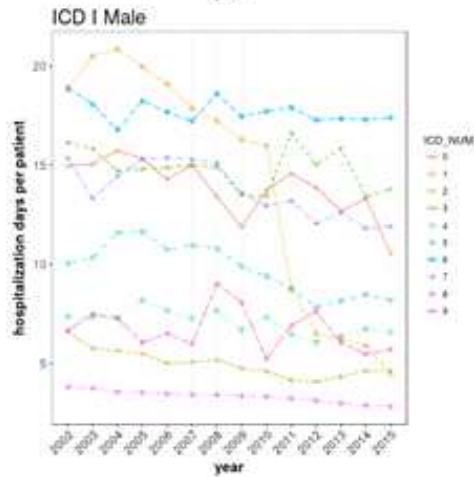
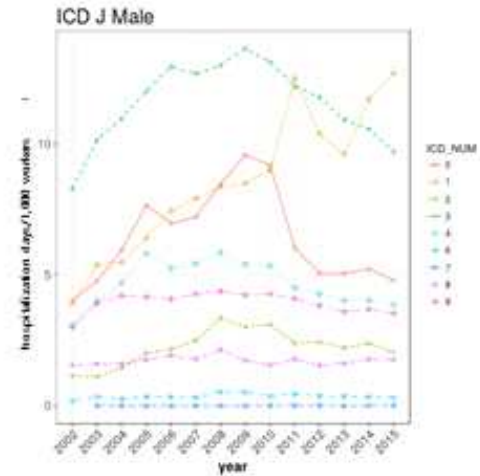
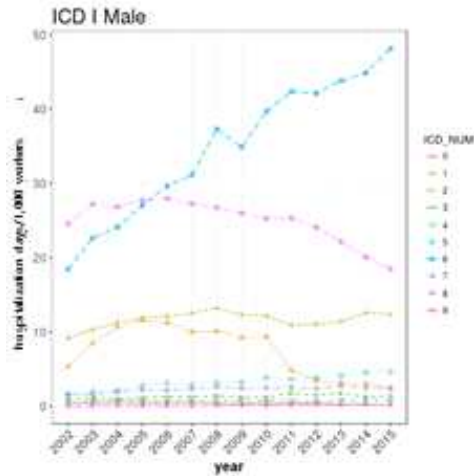
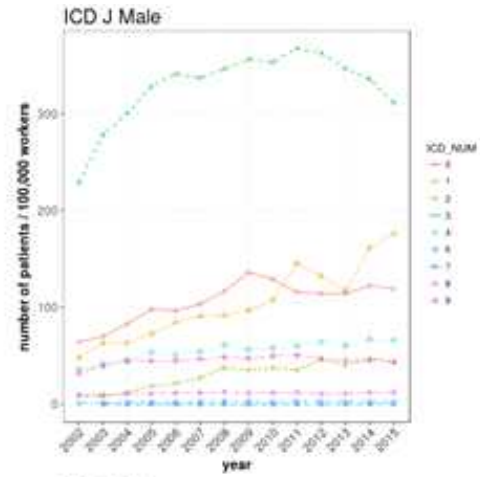
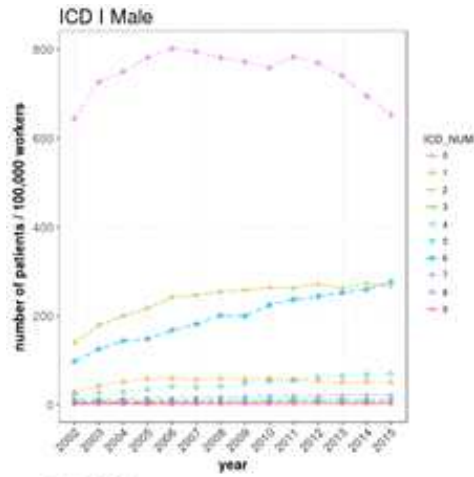
부록 및 부표

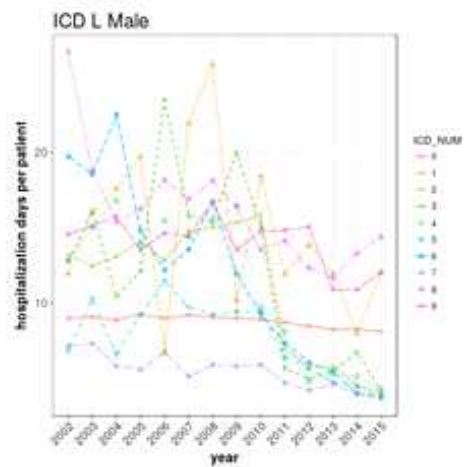
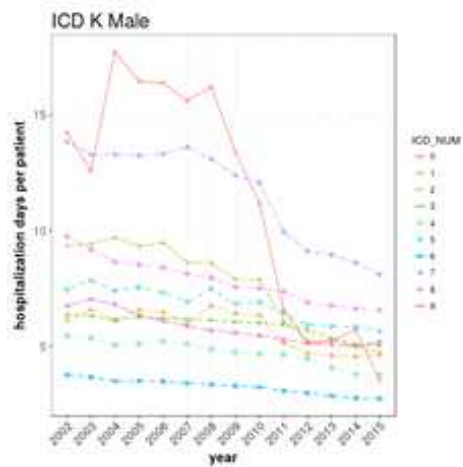
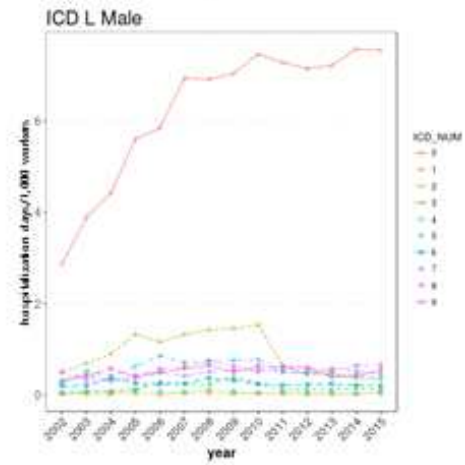
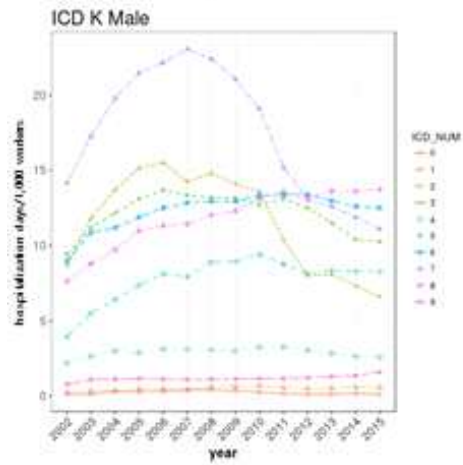
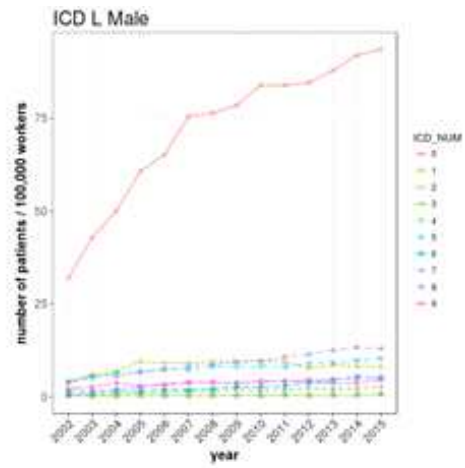
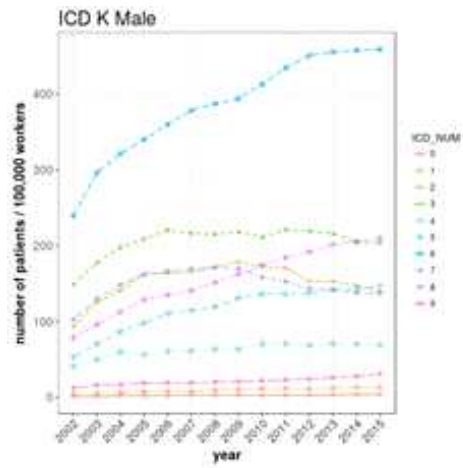


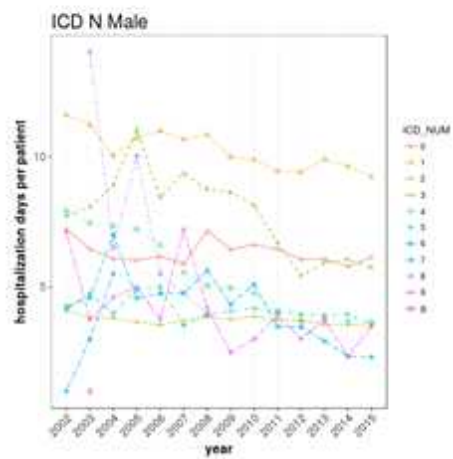
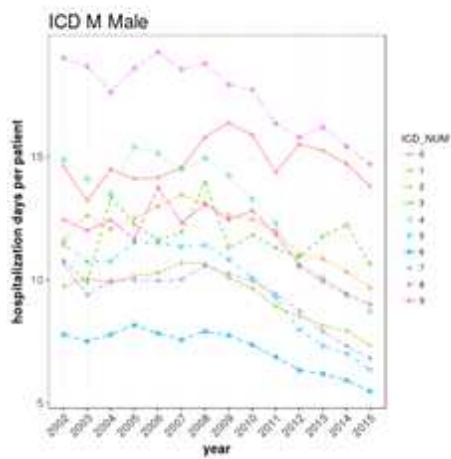
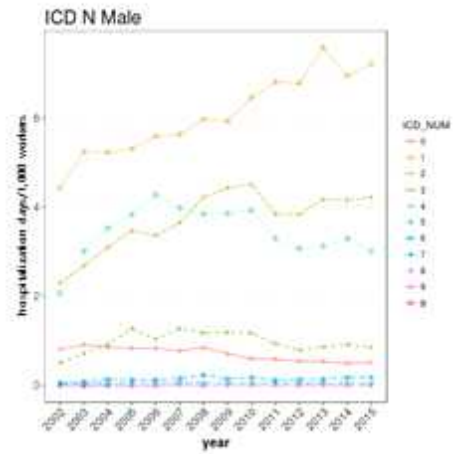
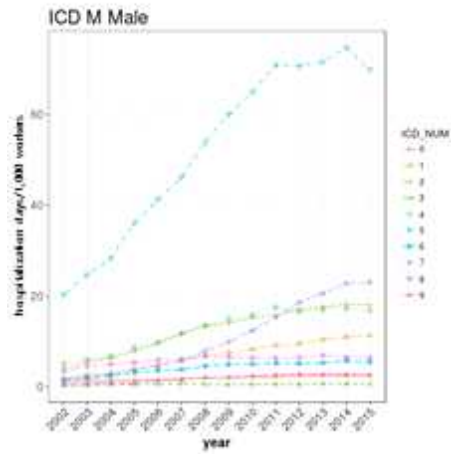
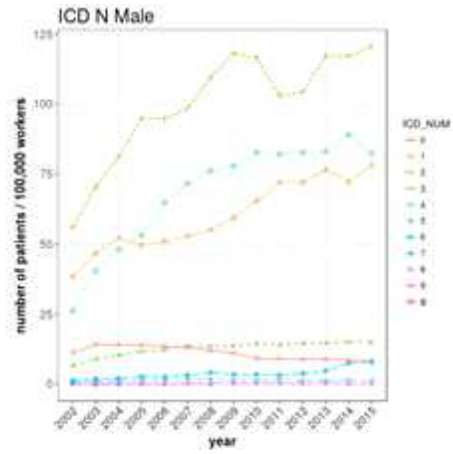
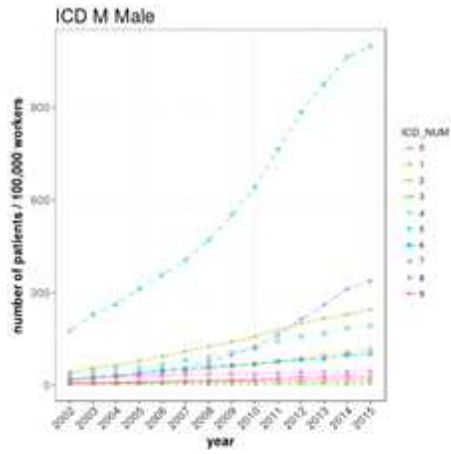


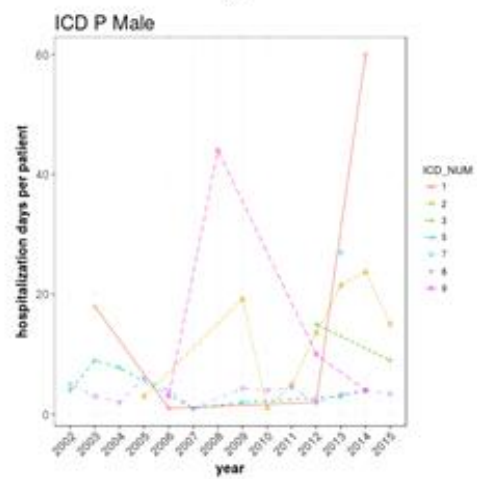
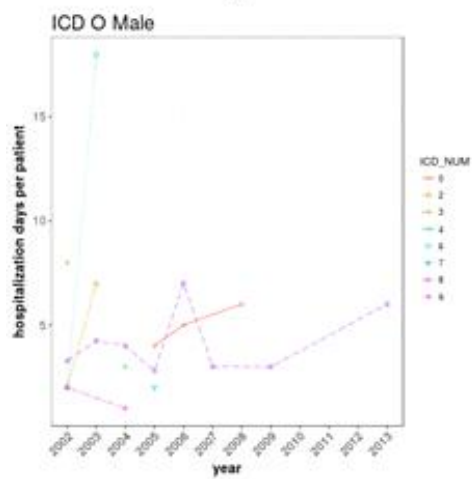
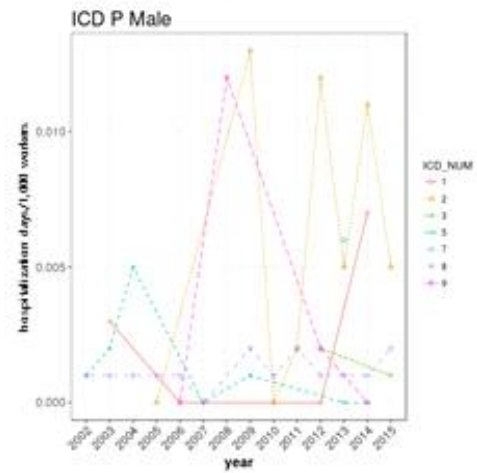
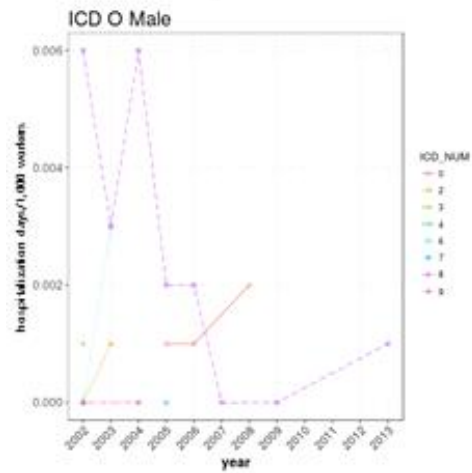
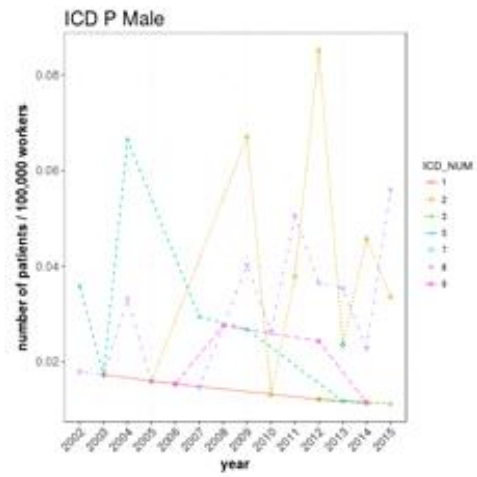
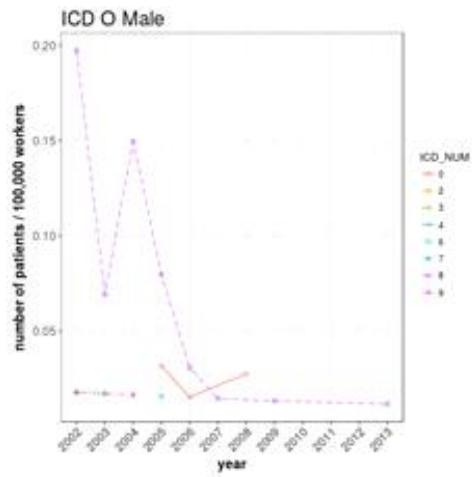


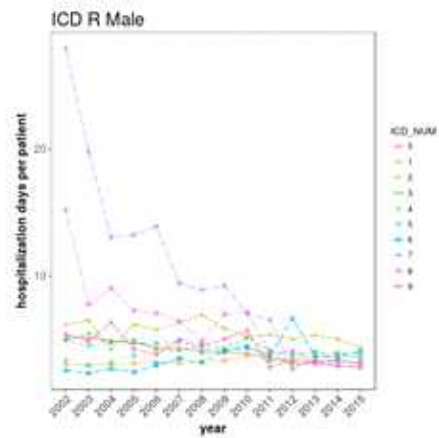
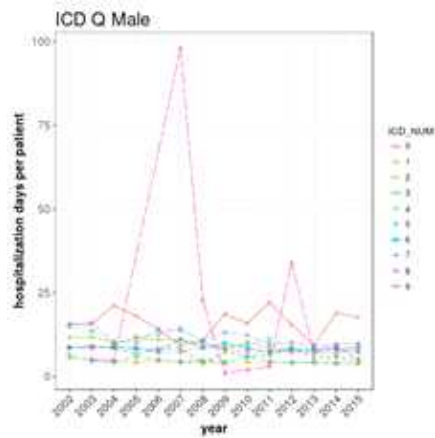
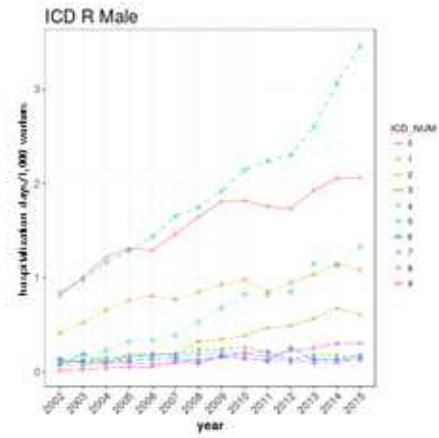
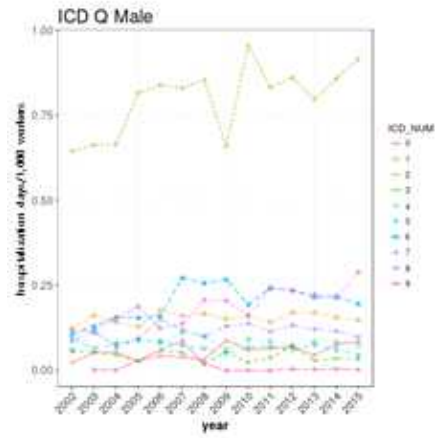
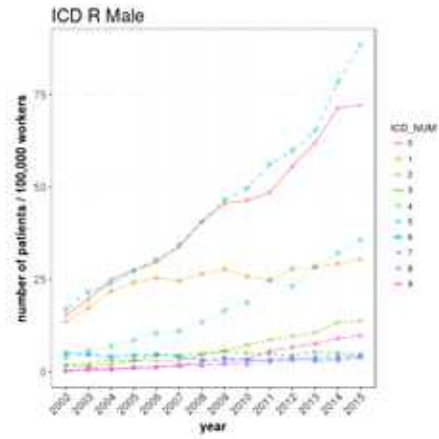


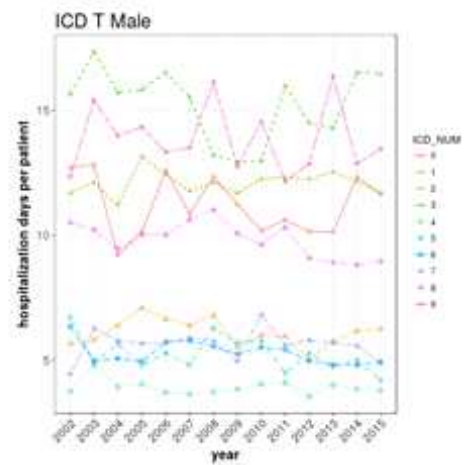
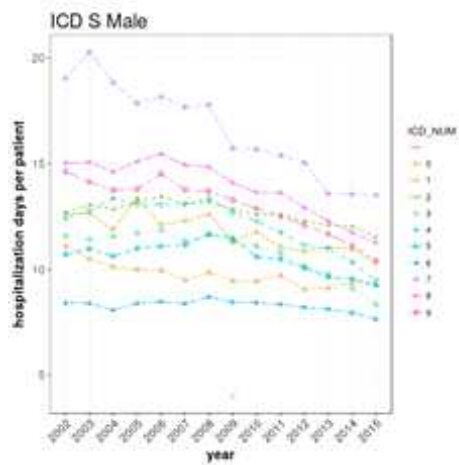
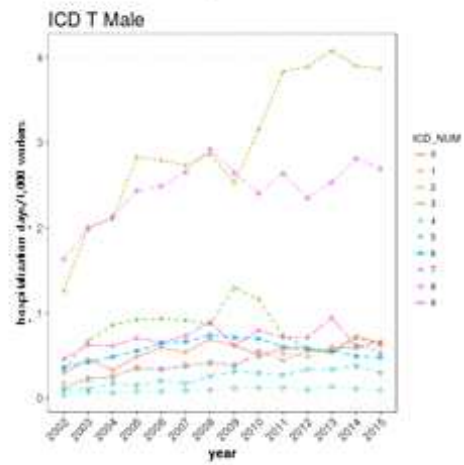
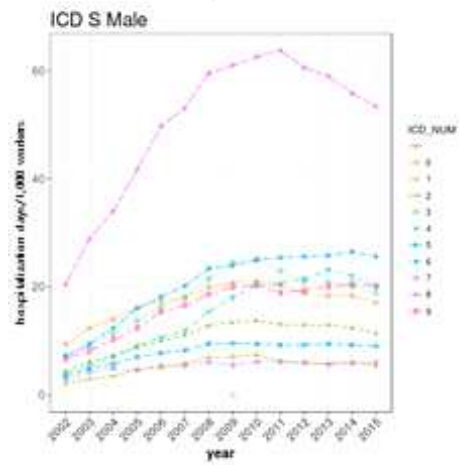
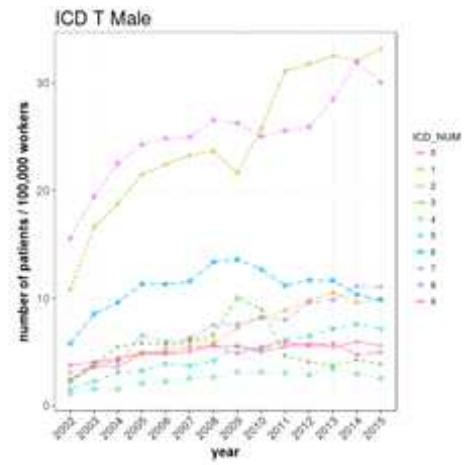
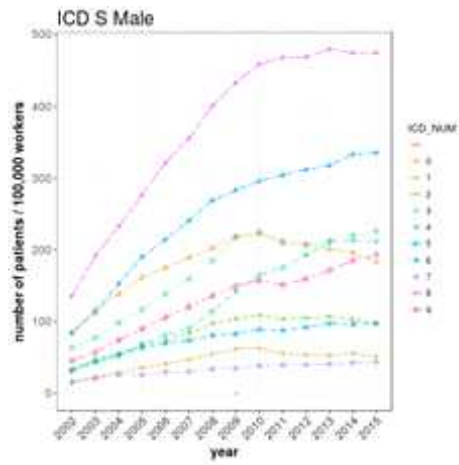


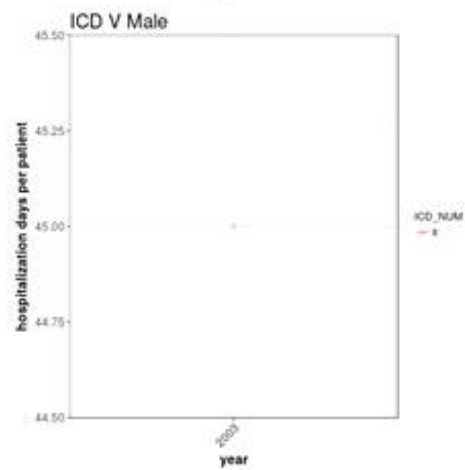
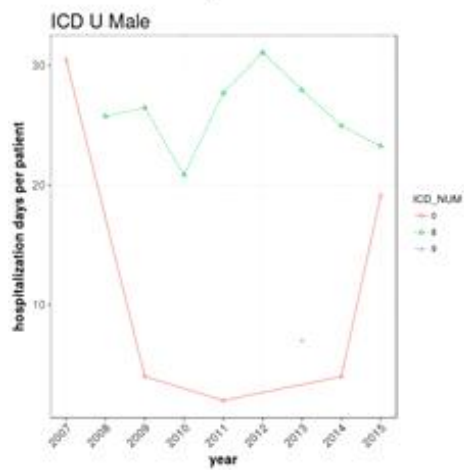
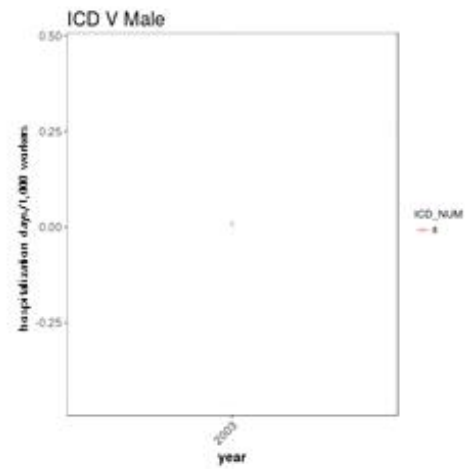
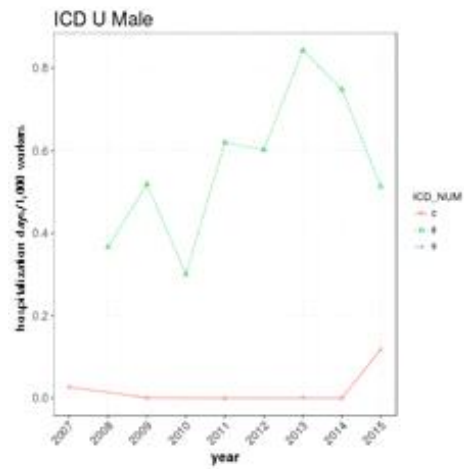
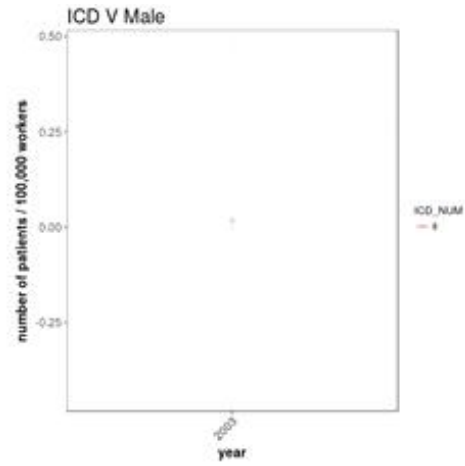
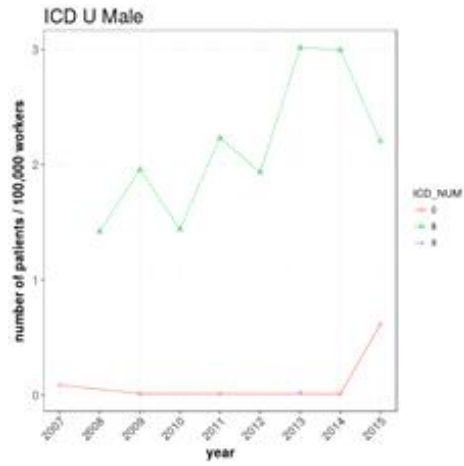


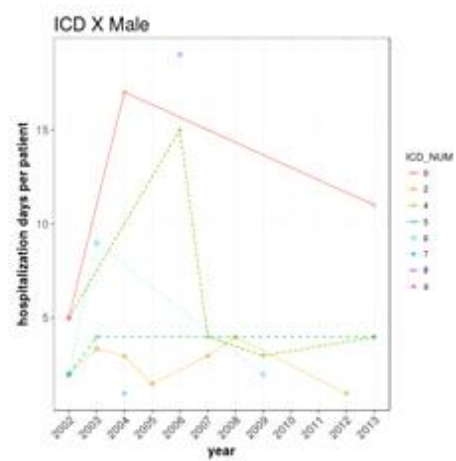
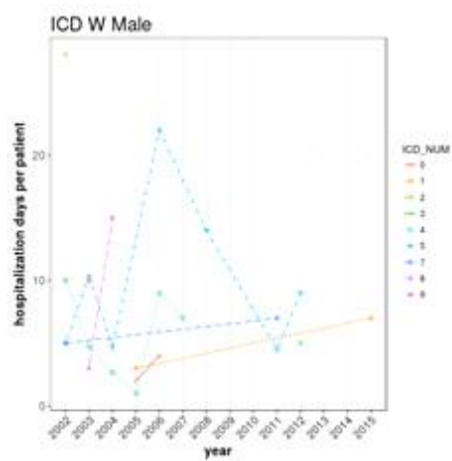
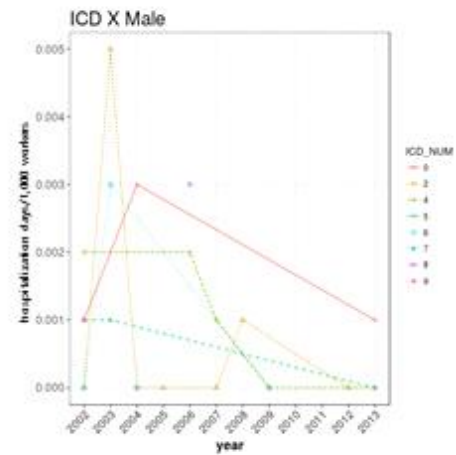
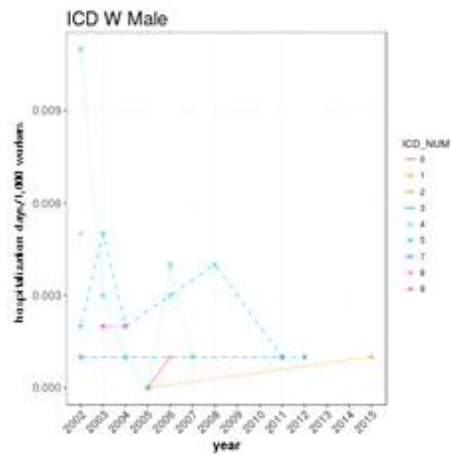
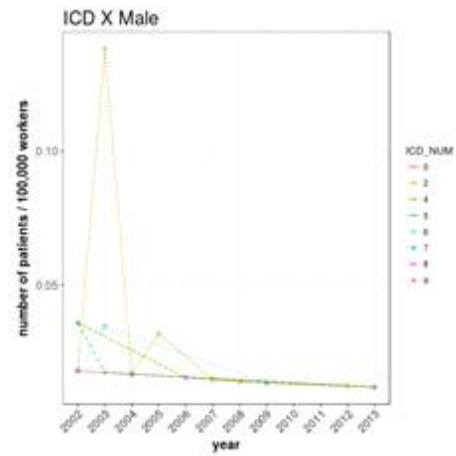
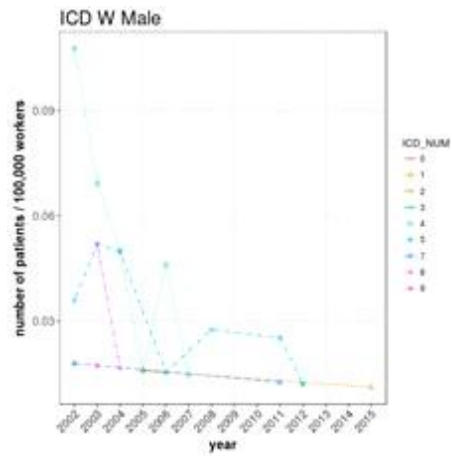


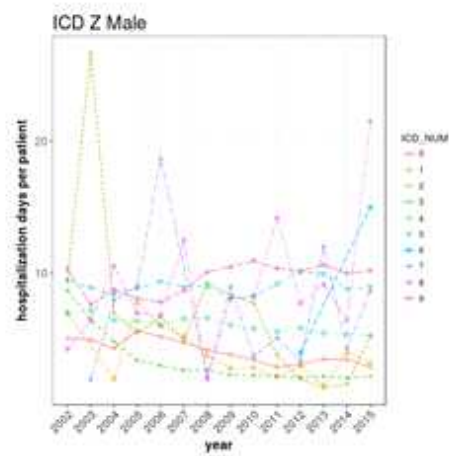
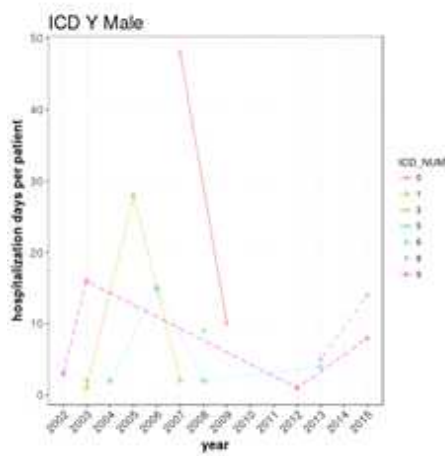
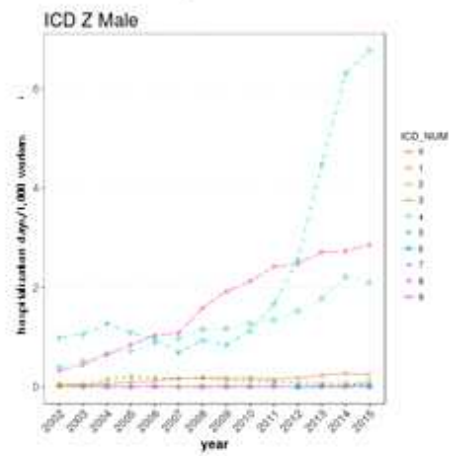
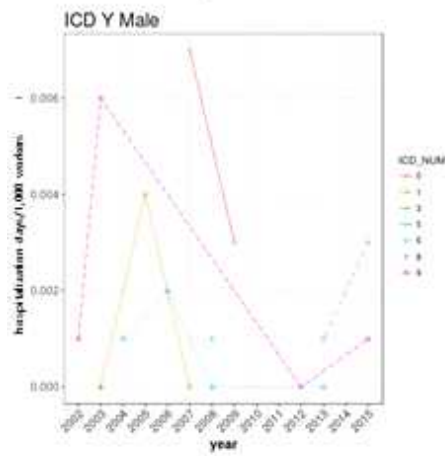
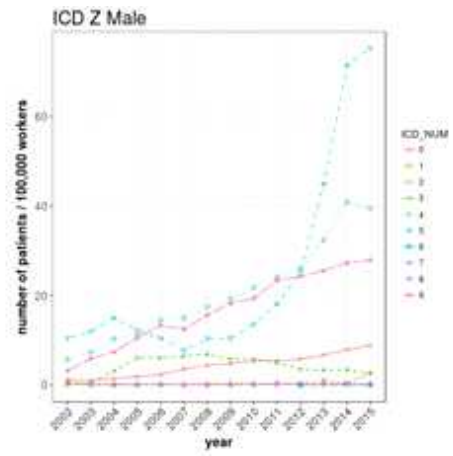
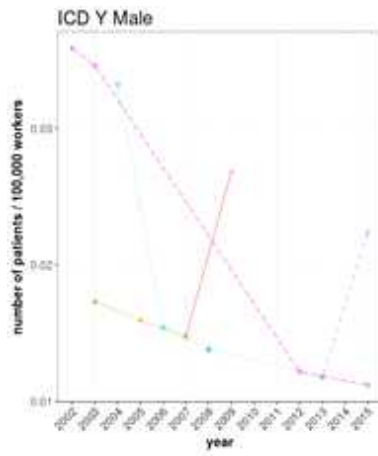


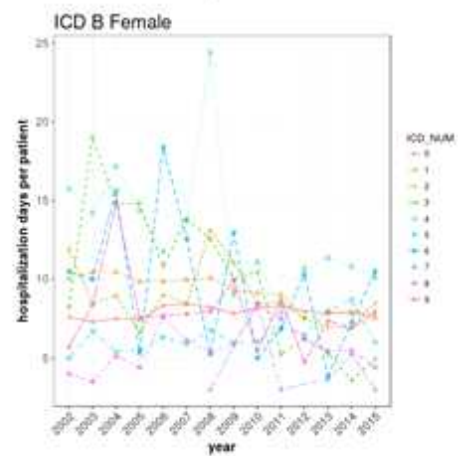
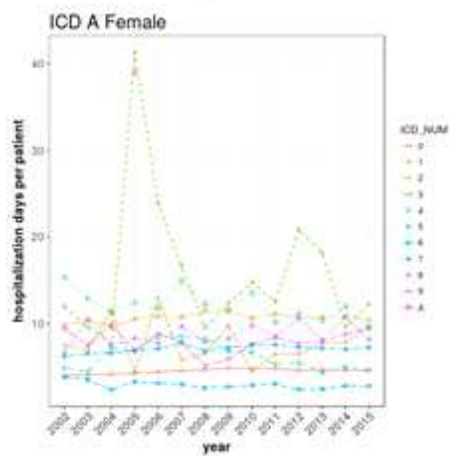
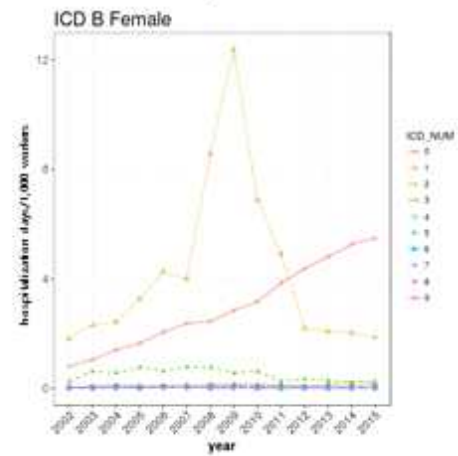
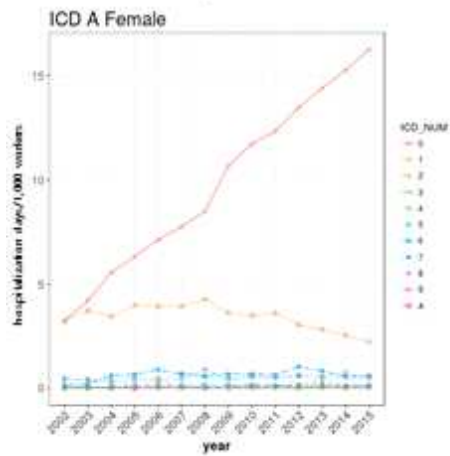
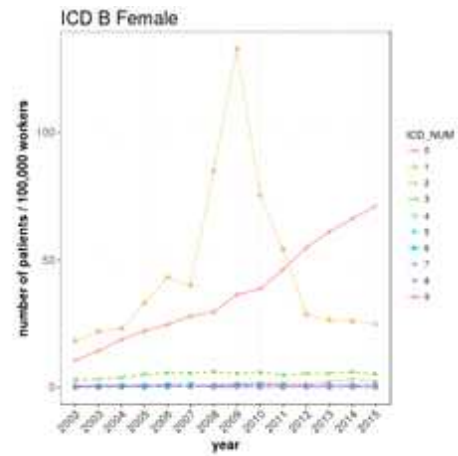
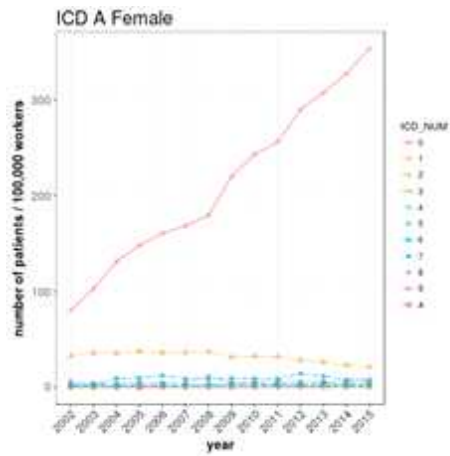


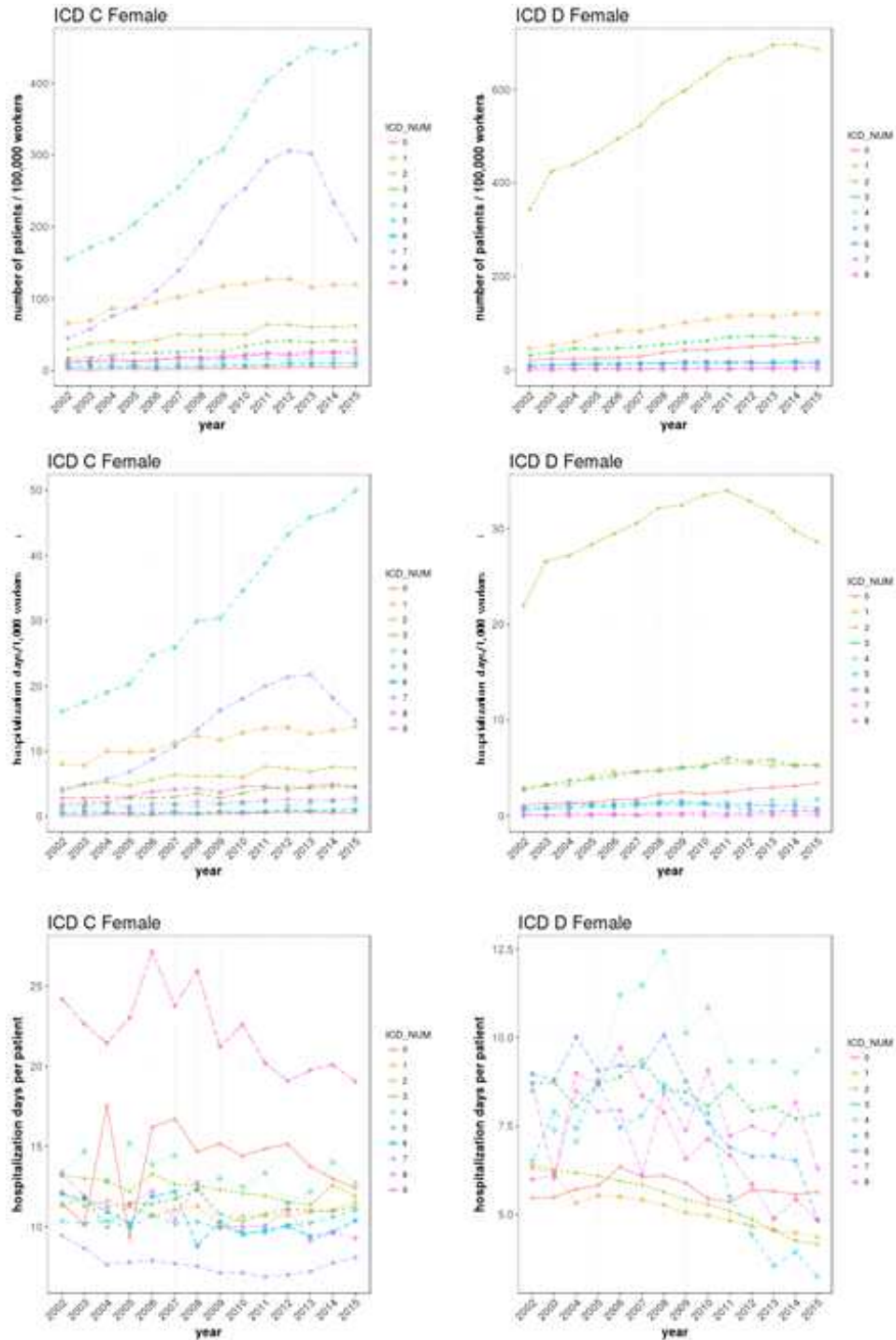


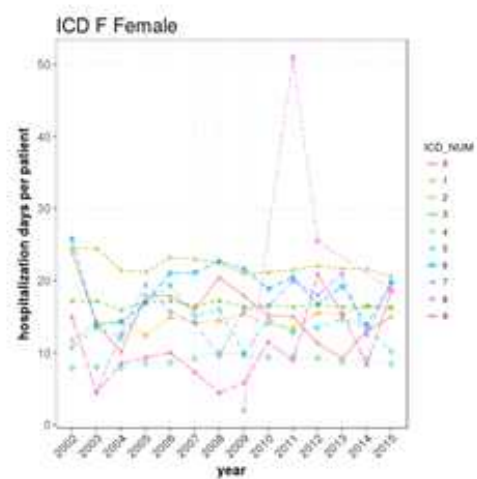
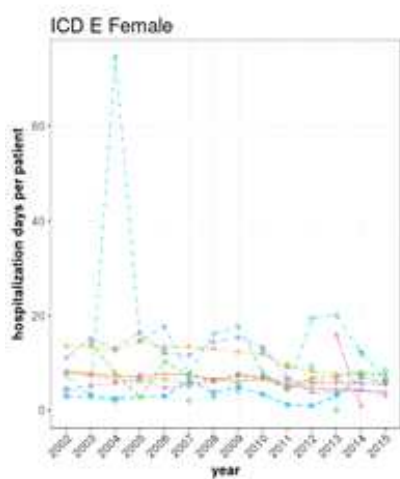
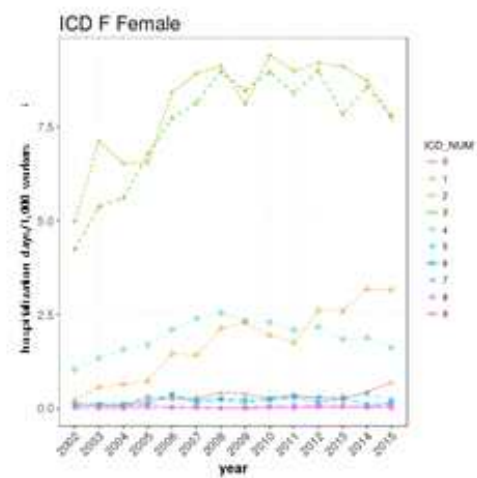
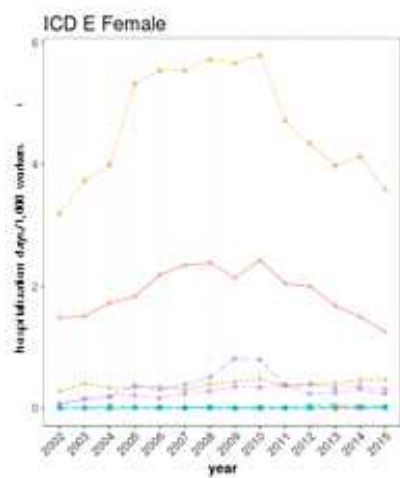
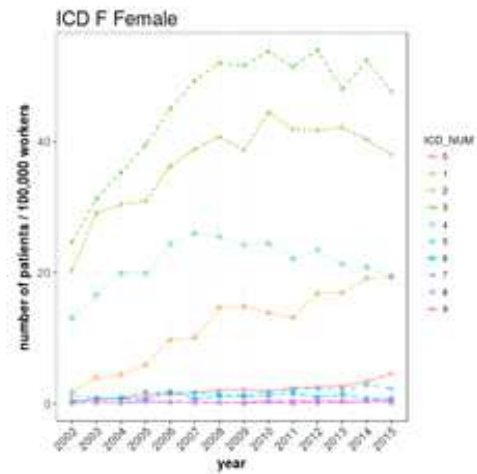
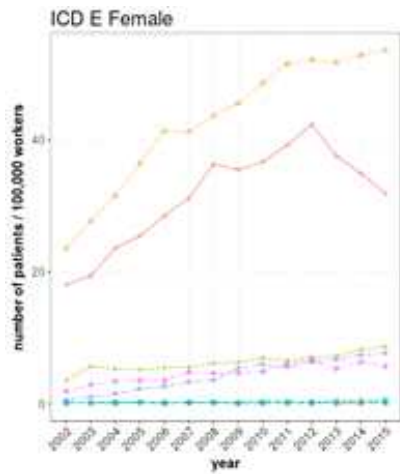


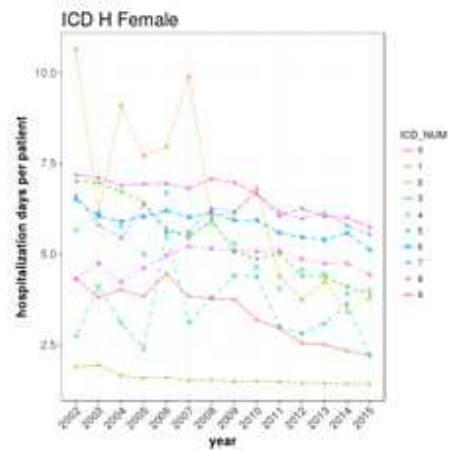
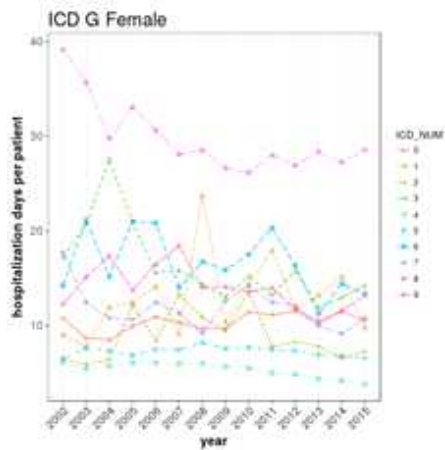
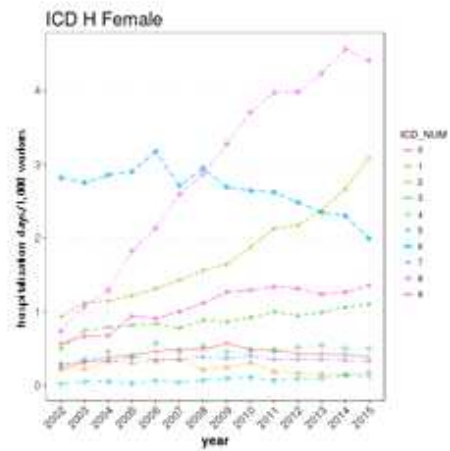
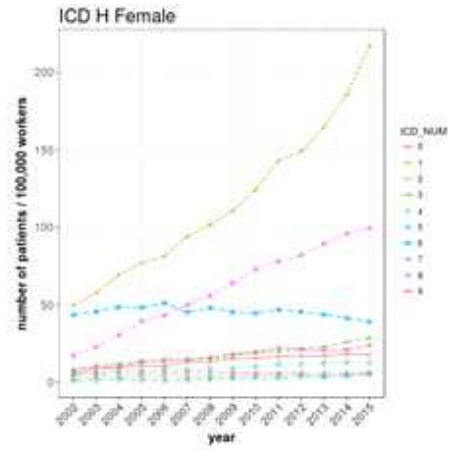
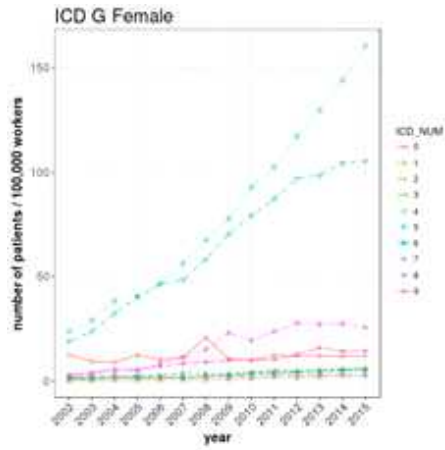


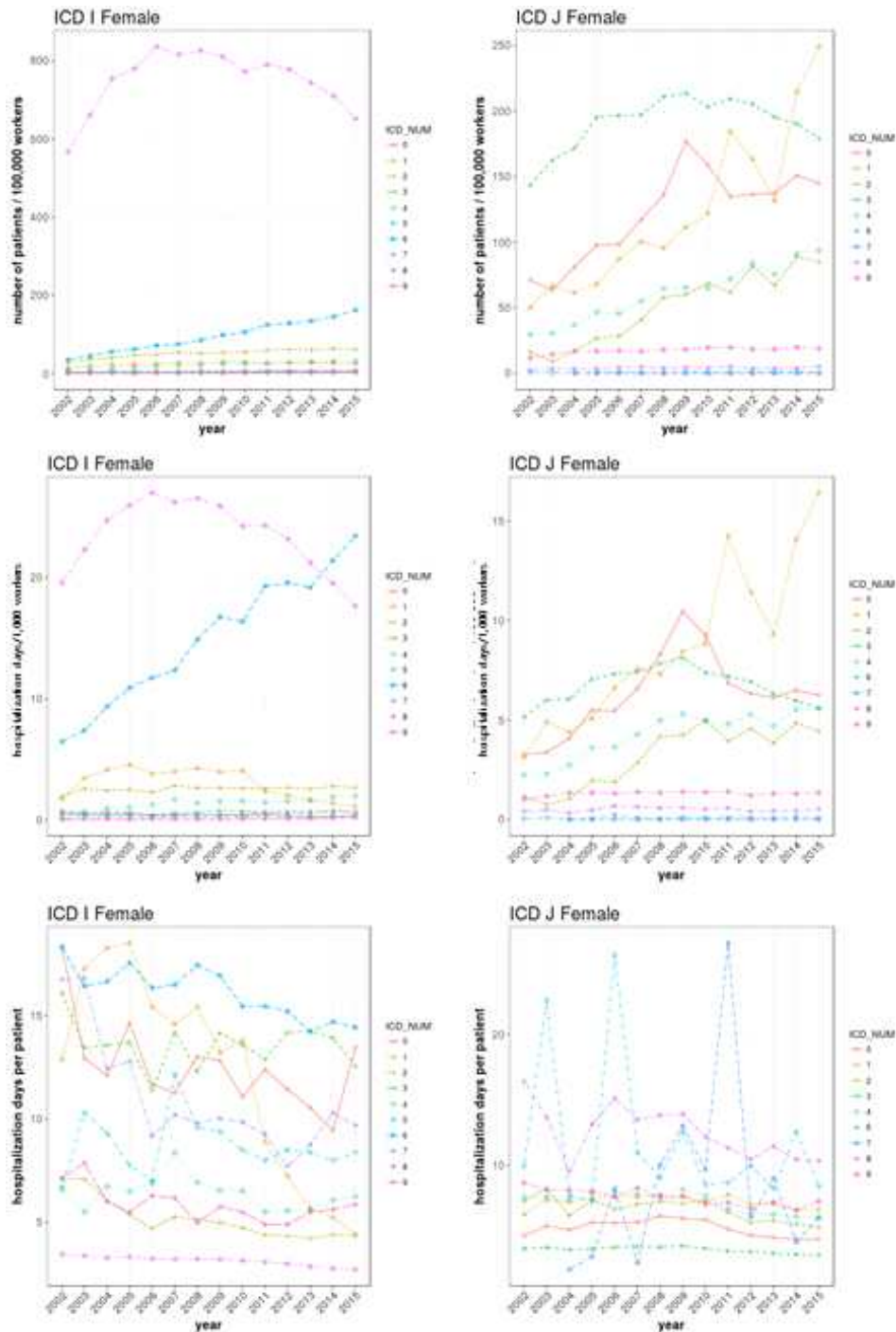


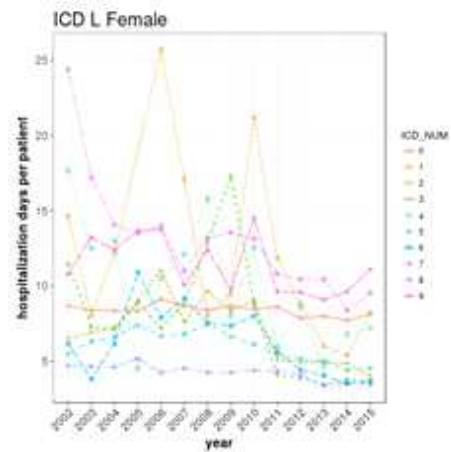
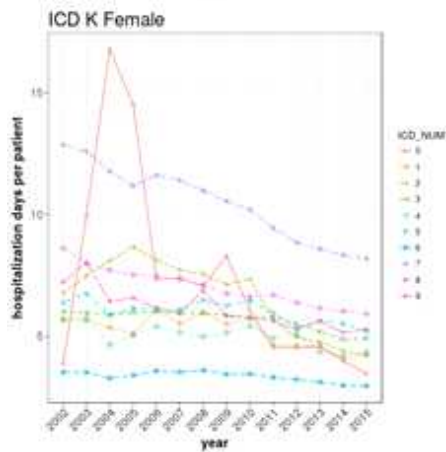
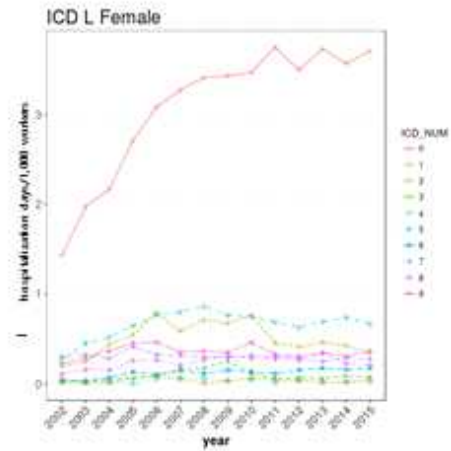
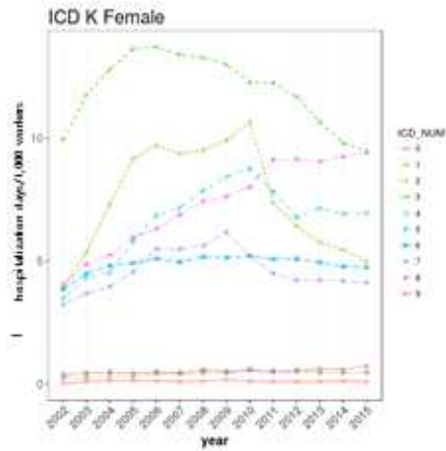
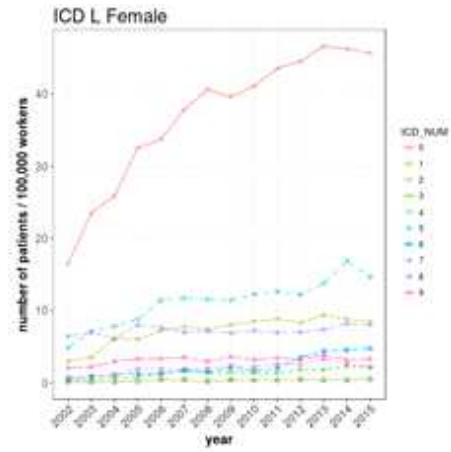
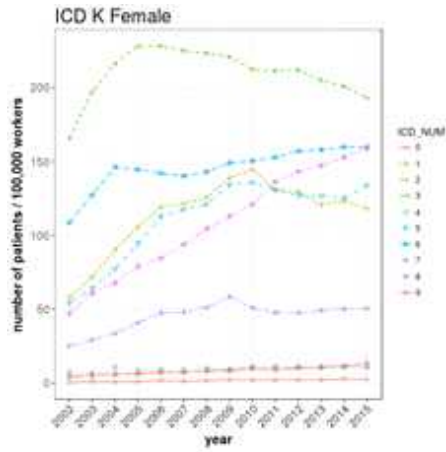


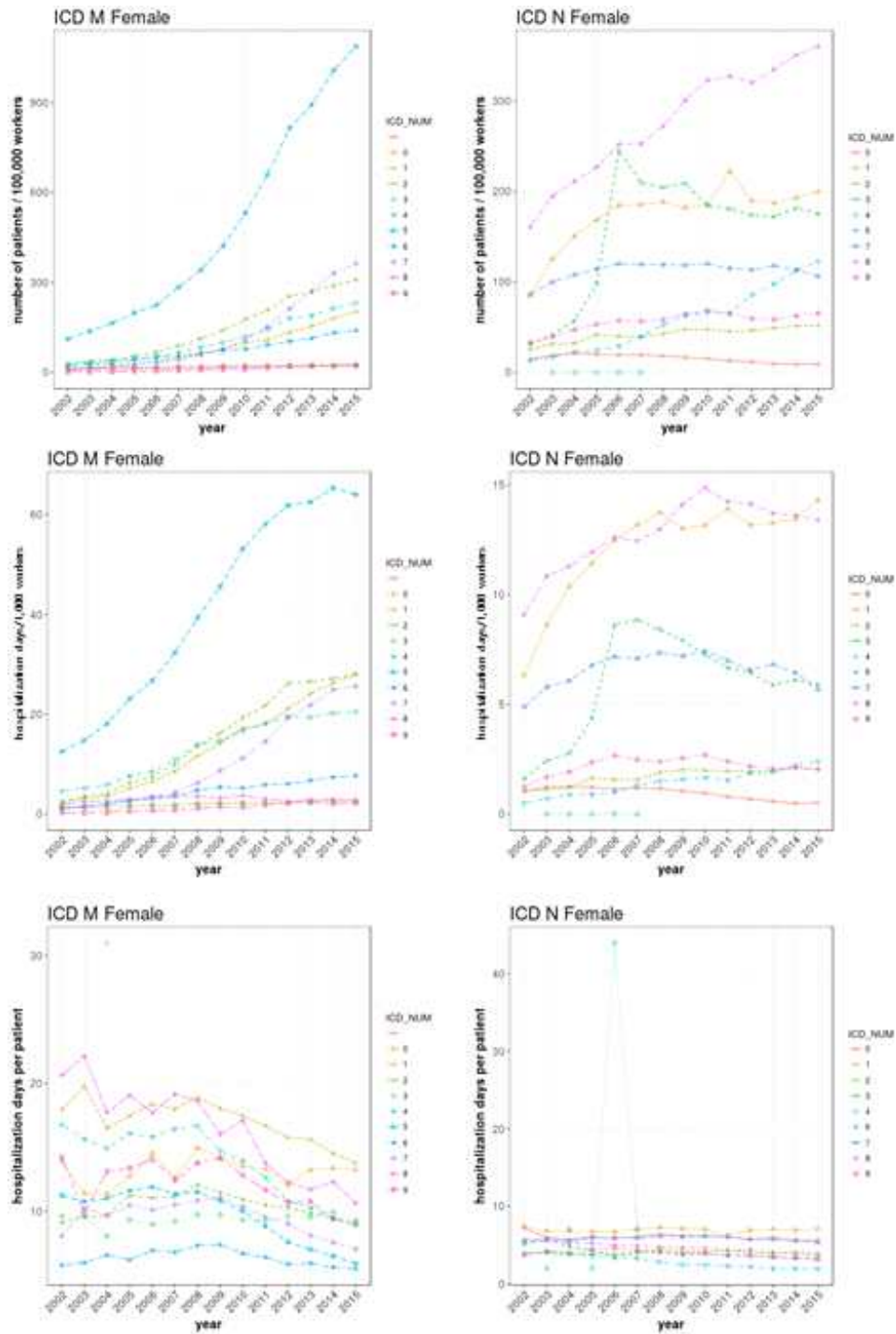


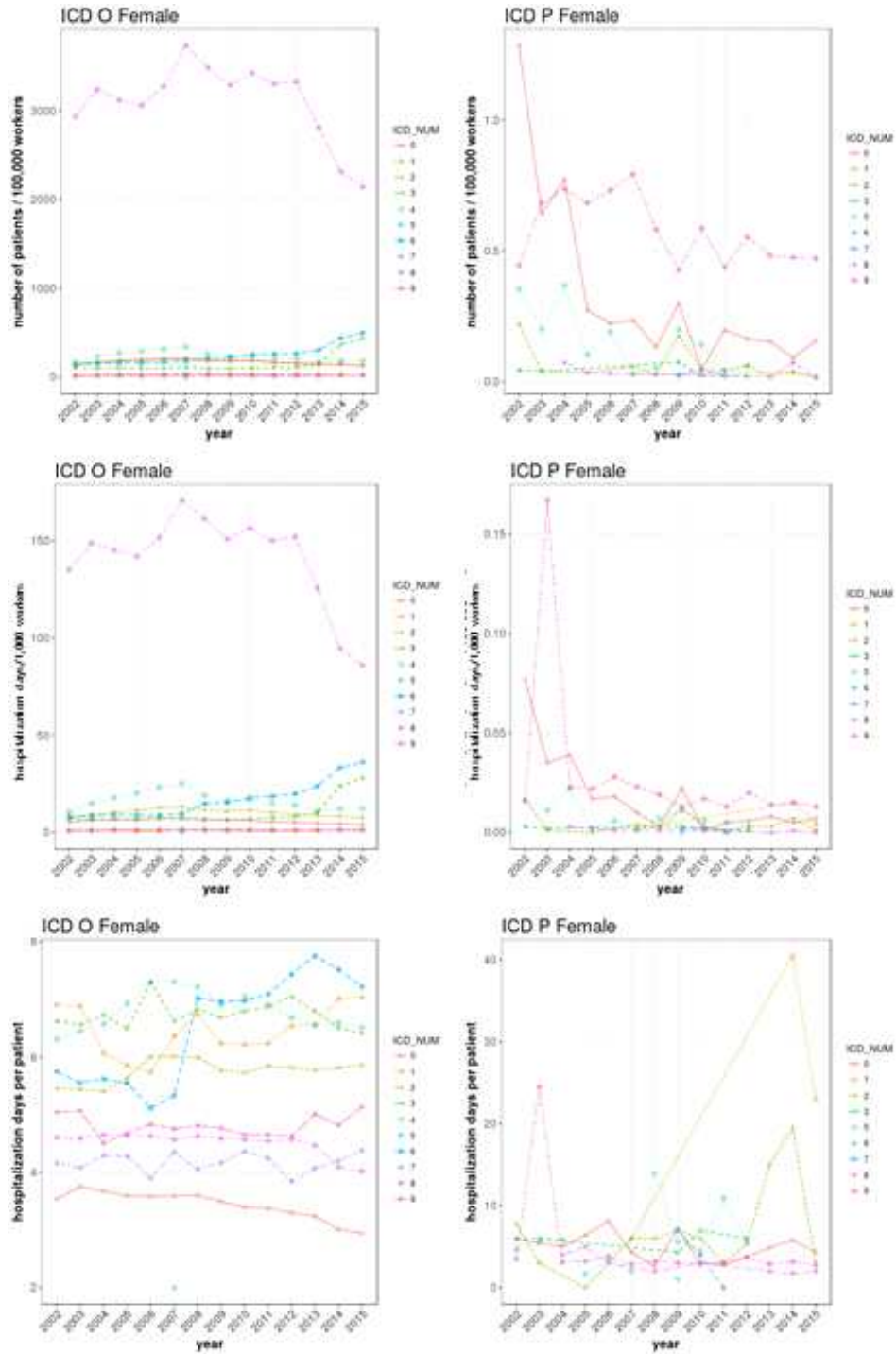


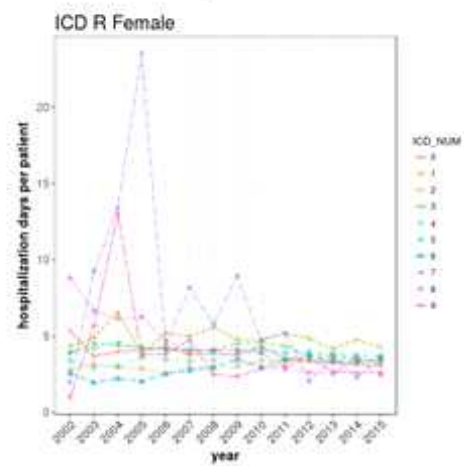
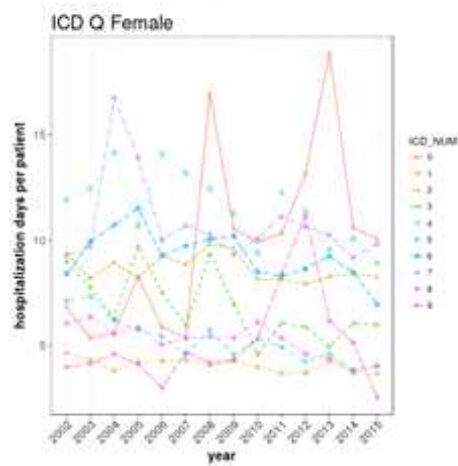
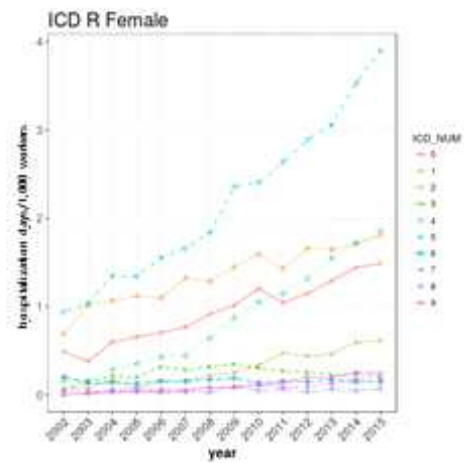
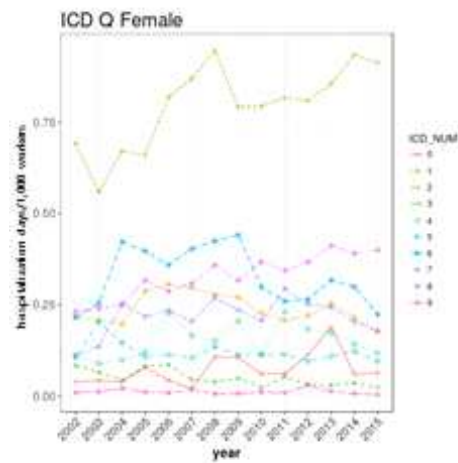
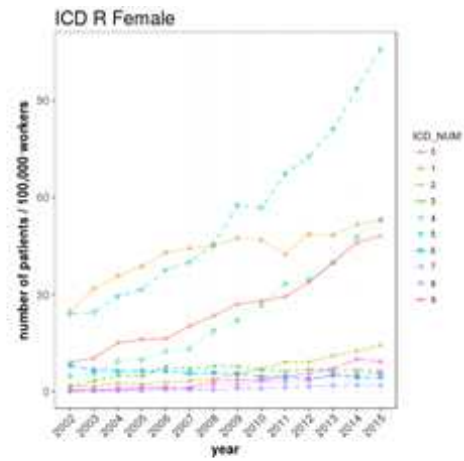
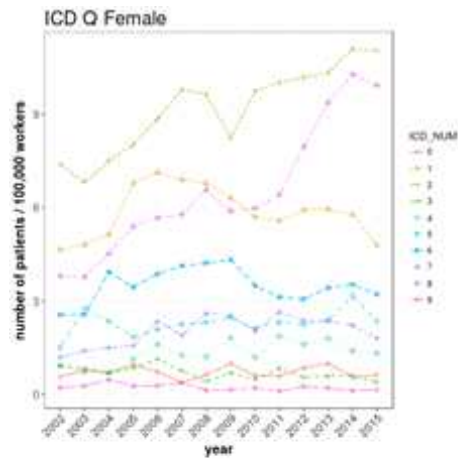


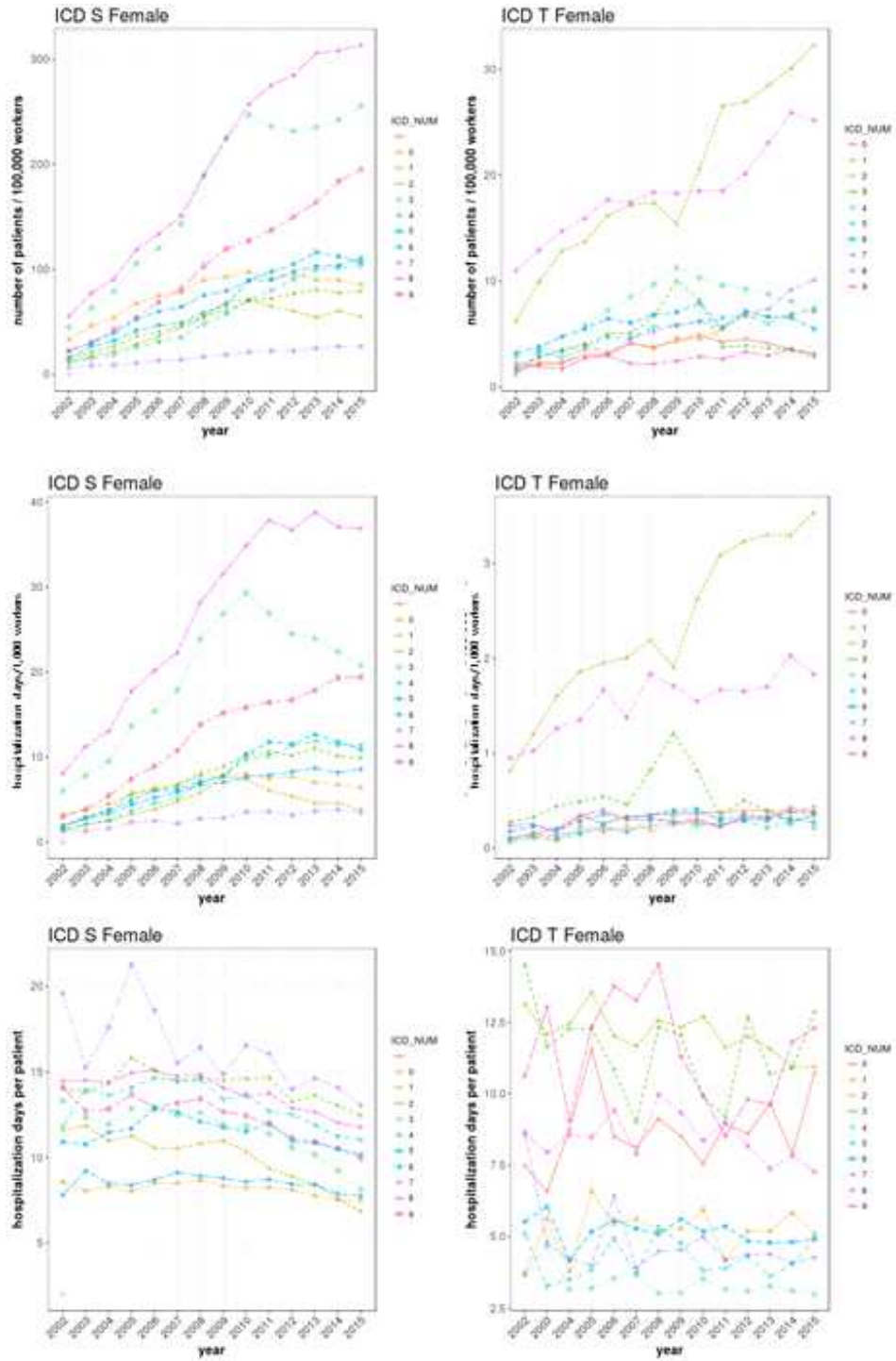


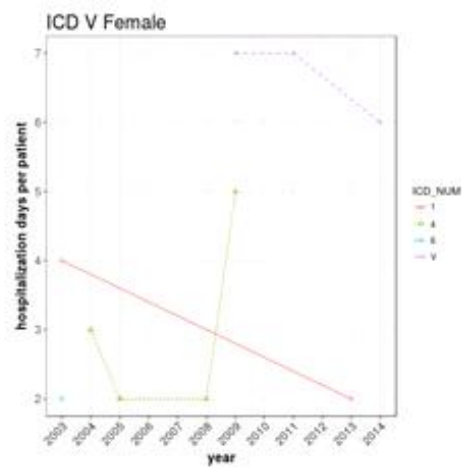
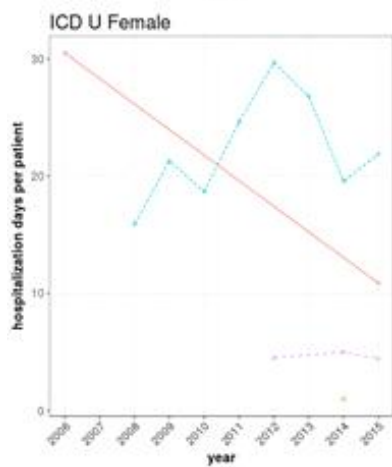
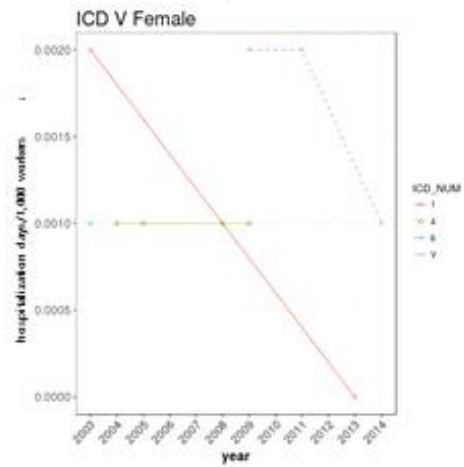
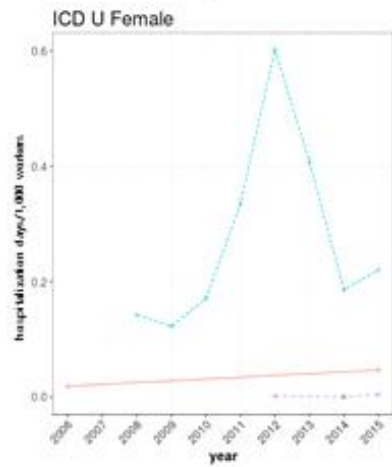
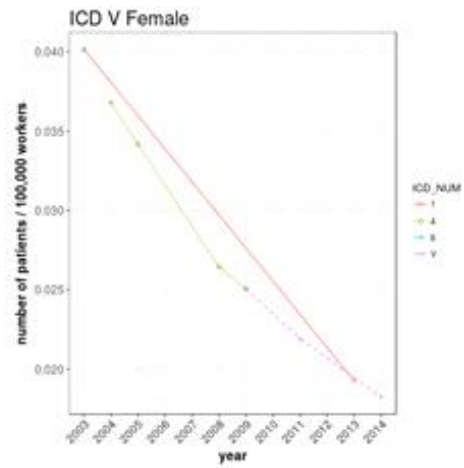
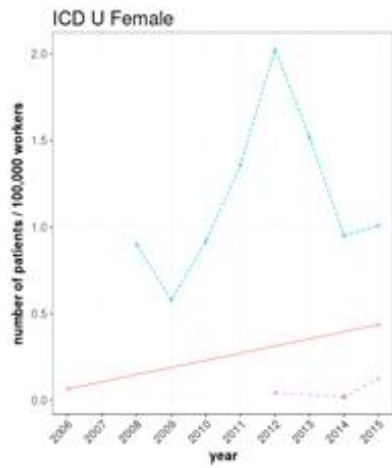


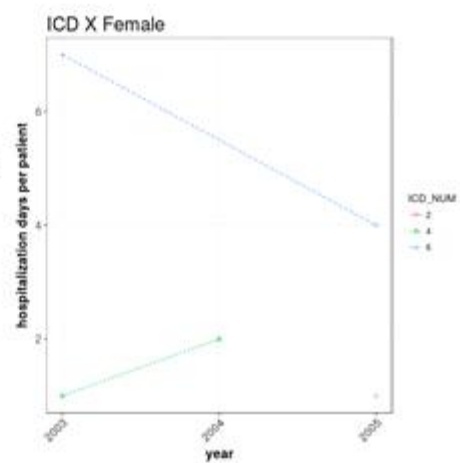
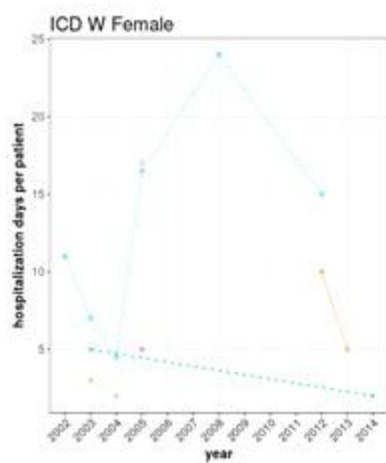
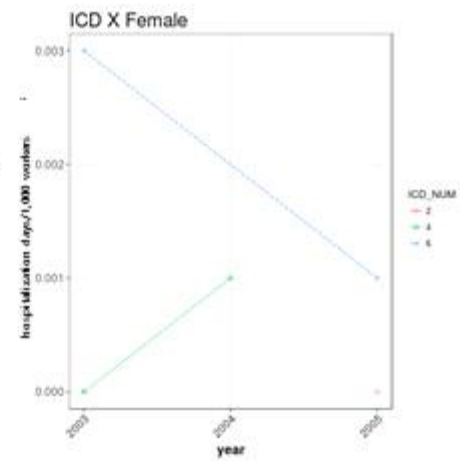
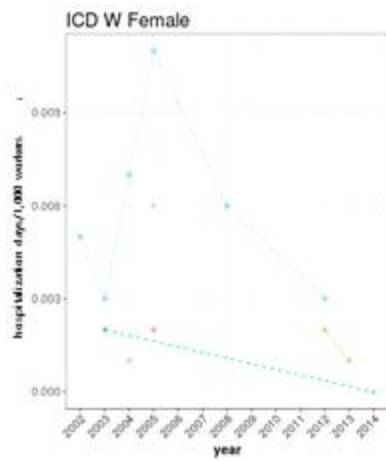
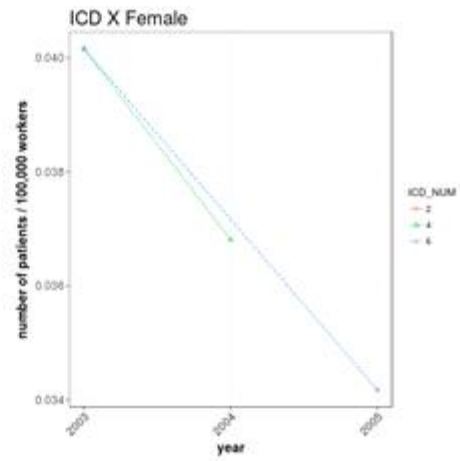
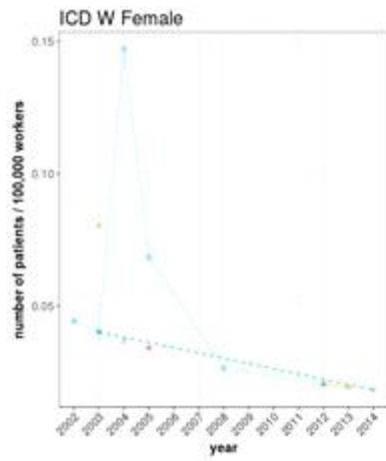












〈〈연 구 진〉〉

연구 기관 : 연세대학교 산학협력단

연구책임자 : 윤 진 하 (교수, 연세대학교 의과대학)

연구원 : 고 상 백 (교수, 연세대학교 원주의과대학)
안 연 순 (교수, 동국대학교 일산병원)
김 환 철 (교수, 인하대학교 의과대학)
강 모 열 (교수, 서울성모병원)
박 철 용 (교수, 삼성창원병원)

연구보조원 : 김 지 현 (전공의, 연세대학교 보건대학원)
임 성 실 (전공의, 연세대학교 보건대학원)
김 민 석 (전공의, 서울대학교 의과대학)
육 지 후 (전공의, 서울대학교 의과대학)
주 진 한 (대학원생, 연세대학교(원주))
김 은 미 (대학원생, 한양대학교 보건대학원)

〈〈연 구 기 간〉〉

2017.04.12 ~ 2017.11.30.

본 연구는 산업안전보건연구원의 2017년도 위탁연구 용역사업에 의한 것임

본 연구보고서의 내용은 연구책임자의 개인적 견해이며, 우리 연구원의 공식견해와 다를 수도 있음을 알려드립니다.

산업안전보건연구원장

**빅데이터 기반 직업코호트 구축을 통한
질병 발생 연구**
(2017-연구원-1057)

발 행 일 : 2017년 11월 30일
발 행 인 : 산업안전보건연구원 원장 김 장 호
연구책임자 : 연세대학교 윤진하
발 행 처 : 안전보건공단 산업안전보건연구원
주 소 : (44429) 울산광역시 중구 종가로 400
전 화 : (052) 7030-878
F A X : (052) 7030-336
Homepage : <http://oshiri.kosha.or.kr>

[비매품]