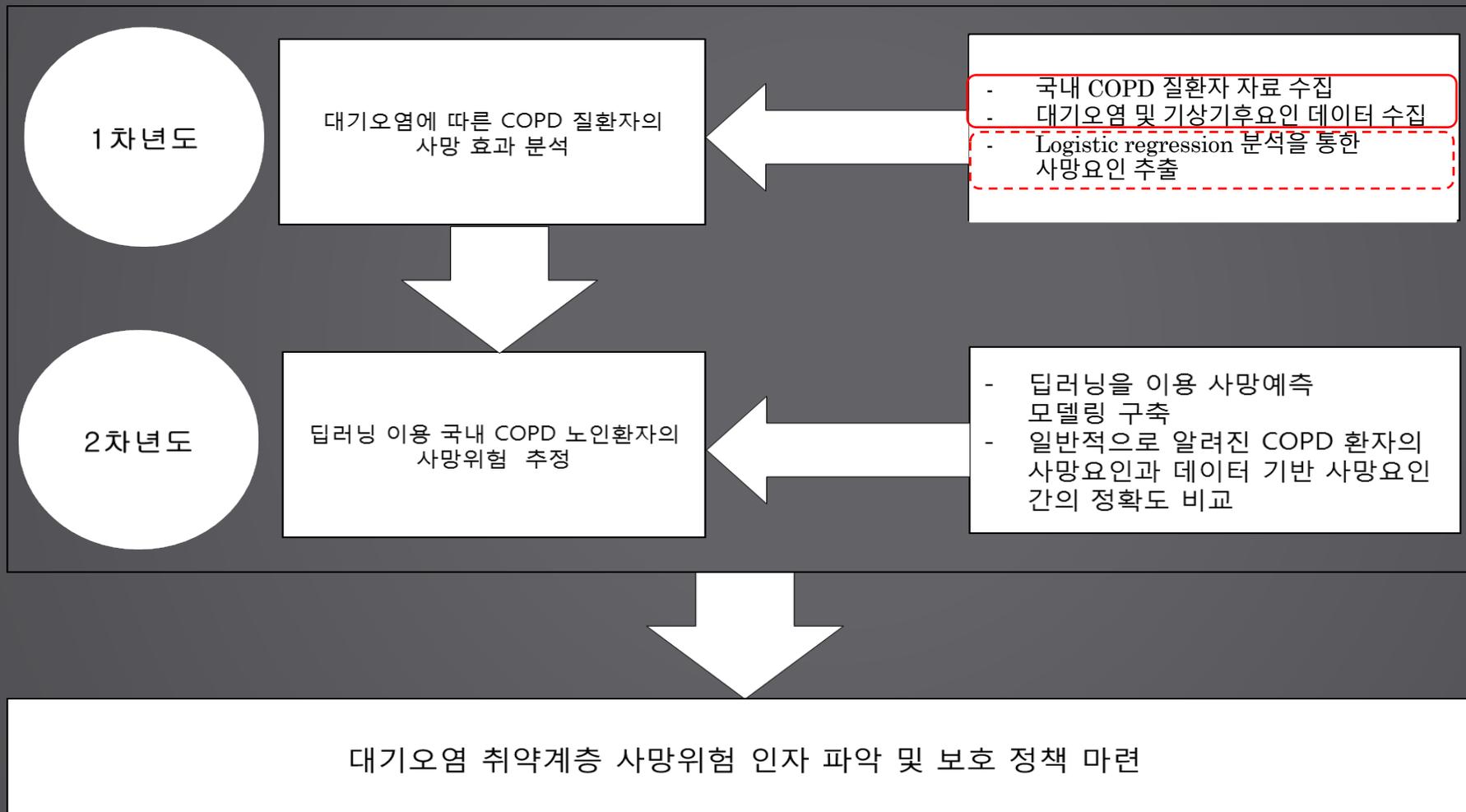


(5) 딥러닝 이용 노인인구 호흡기 질환 사망 위험 추정 [강선아]

- ◆ 만성폐쇄성 폐질환 사망 위험을 딥러닝을 이용하여 추정
 - 연구 대상 :65세 이상 만성폐쇄성폐질환(COPD) 환자
- ◆ 연구내용 : 1단계 사망요인 파악 연구 , 2단계 사망확률 추정 연구 2년차로 확대
 - 자료 : 건강보험 맞춤형연구 DB , 인구, 기후, 대기오염도 및 대기오염물질 배출량 자료를 연계
 - 1차년도 (2018) : 대기오염에 따른 COPD 질환자 사망요인 분석(맞춤형 연구 DB)
 - GAM 분석을 통해 사망에 영향을 미치는 주요 변인을 발견하고 사망 효과를 분석
 - 2차년도 (2019) : COPD 노인 질환자 사망위험 추정
 - 딥러닝 이용 사망예측 모델 구축
 - 1차년도 분석 시 발견된 COPD 환자의 사망요인 .vs 데이터 기반 파악 사망요인 : 사망위험 추정치 정확도 비교
- ◆ 연구 진행 상황 : 사망요인 분석 중
 - 환경 및 기후인자, 진단 데이터가 사망에 미치는 요인 분석
 - 사망요인 선정 과정에서 의료 전문가 인터뷰 진행 (서울삼성병원 호흡기내과 박혜윤박사/서울 강남 카톨릭 병원 이진국박사)

연구 프레임워크

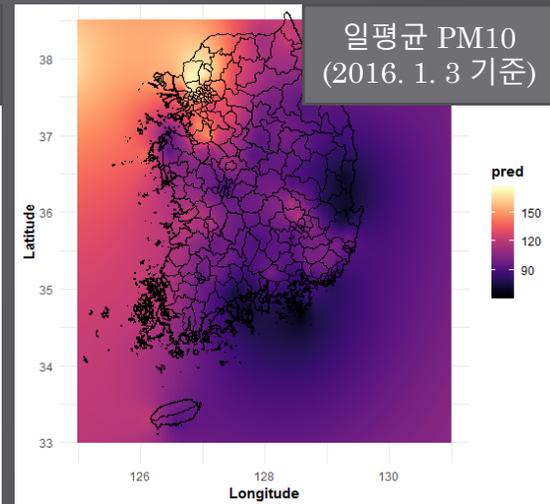
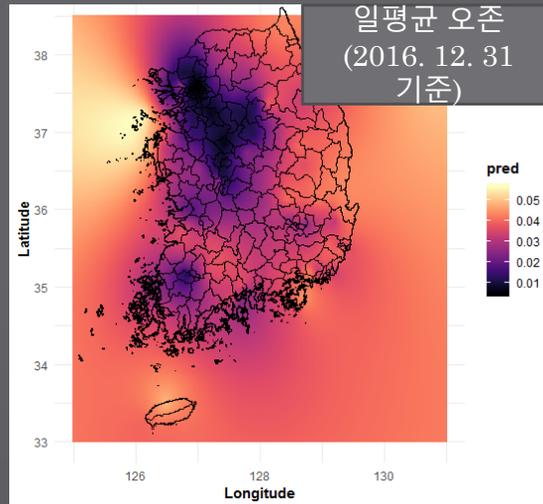
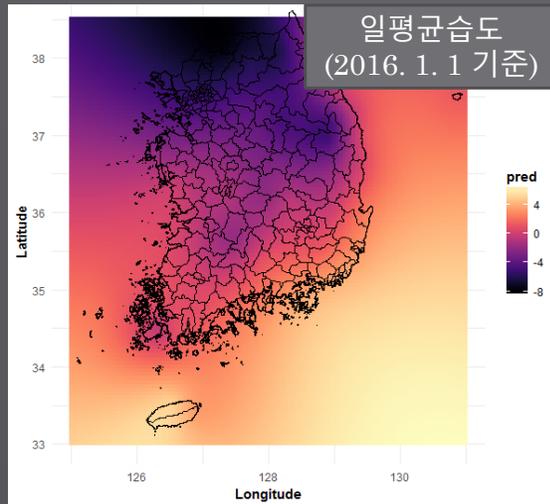
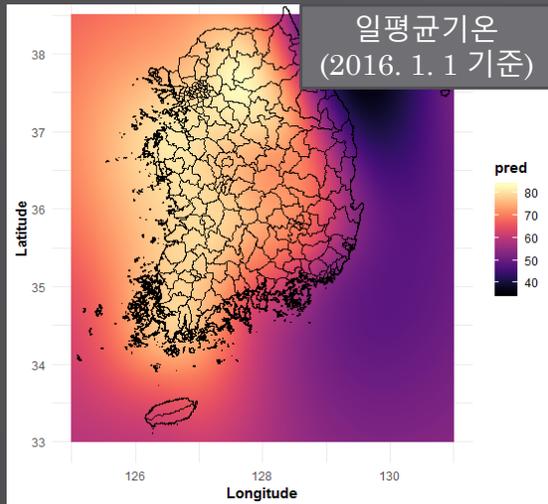


완료

진행 중

분석 자료 및 전처리 과정(대기오염 및 기상자료)

- ◆ 설명변수: 국민건강보험공단 맞춤형 DB 진료 기록, 기상기후 데이터, 대기오염 데이터
- ◆ 대기 및 기상 자료: 건강보험 DB 와 연계 목적 시공간 해상도 조정
 - GIS 기법 중 Kriging을 이용하여 점(point)데이터인 측정소 자료를 면 데이터인 시군구 자료로 변형
 - 시간단위 대기오염 데이터를 일간 평균을 취하여 일 단위 건강보험 DB와 연계

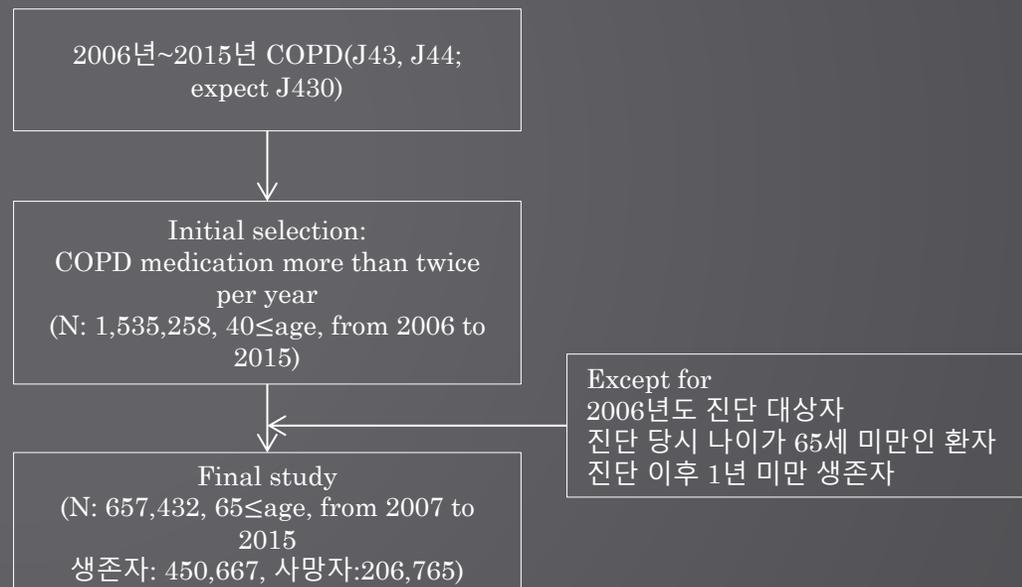


건강보험 맞춤형 DB 신청항목 및 전처리 내용

DB 신청내용

조건	내용
연도	2006년~2015년
상병코드	J43, J44(except J430)
주상병/부상병	주상병 및 모든 부상병
산정특례 특정기호구분	없음
의과/한방/치과/약국	의과
입원/외래	입원, 외래
행위수가코드	전체자료
약제주성분코드	전체자료
기타	없음

전처리 내용

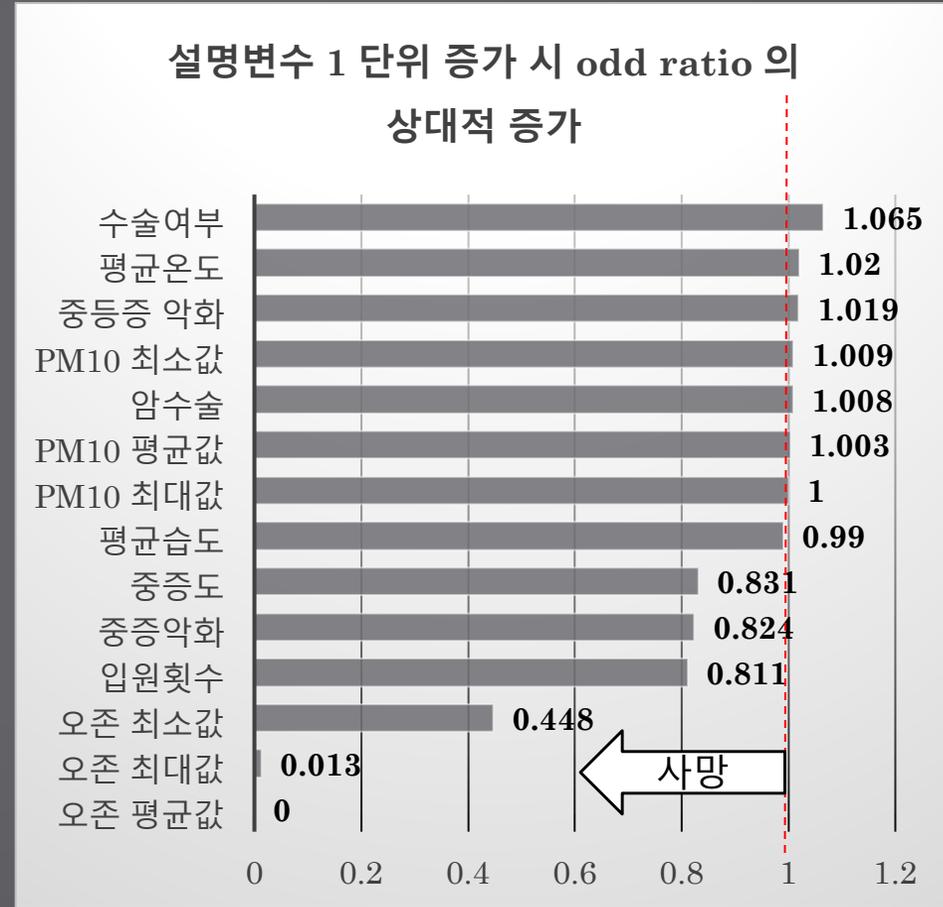


설명변수

구분		변수명	설명
설명 변수	개인정보 데이터	나이	연속형 변수
		성별	연속형 변수
		거주지	범주형 변수
		보험료 20분위	범주형 변수
		직종	범주형 변수
		장애등급	범주형 변수
	진단 데이터	Charlson 동반질환지수	범주형 변수
		입원횟수	연속형 변수
		수술여부(암/비암)	범주형 변수
		COPD 중증도	범주형 변수
		외래 경증 악화-중등증 악화	범주형 변수
		외래 경증 악화-중증 악화	범주형 변수
	기후 데이터	산소 처방전	범주형 변수
		평균 기온	연속형 변수
	대기오염 데이터	평균 습도	연속형 변수
		일평균 오존, 일최대 오존, 일최소 오존	연속형 변수
		일평균 PM10, 일최대 PM10, 일최소 PM10	연속형 변수
		일평균 PM25, 일최대 PM25, 일최소 PM25	연속형 변수

COPD 사망요인 분석 : Logistic Regression

- ◆ 시군구별, 성별, 연령별 층화추출(sample size = 131,478명, 20% sampling)
- ◆ 조정변수 : 성별, 연령, 소득(보험료 20 구간), 직종, 거주지(대도시, 중소도시)
- ◆ 미세먼지 농도는 사망확률에 거의 영향을 주지 않음
 - 사망일자 11월~2월 집중 : 미세먼지 오염도가 낮은 기간
- ◆ 오존 농도는 사망확률을 대폭 증가시키는 경향
- ◆ 향후 계획
 - 통계청 사망원인통계자료 연계
 - 계절요인 통제 등 fine tuning 작업
 - 딥러닝을 이용한 사망위험 예측 작업



$$\exp(b_i) = \frac{\frac{\Pr[Y=1]}{\Pr[Y=0]}_{x_i=x_{i,0}}}{\frac{\Pr[Y=1]}{\Pr[Y=0]}_{x_i=x_{i,0+1}}}$$