

# 보건복지 분야 매크로데이터 현황과 매크로데이터 비밀보호방법 제언

*Macro Data on Social Welfare and Health Care:  
Current State and Protection Recommendations*

오미애 한국보건사회연구원 부연구위원  
진재현 한국보건사회연구원 전문연구위원

본 연구에서는 보건복지 분야 매크로데이터 현황을 살펴보고 매크로데이터의 비밀보호방법 중 노출위험을 측정하여 민감한 셀을 결정하는 방법론에 대한 소개와 보건복지 분야 매크로데이터 비밀보호방법 적용 사례를 살펴봄으로써 그 활용가능성을 검토하였다.

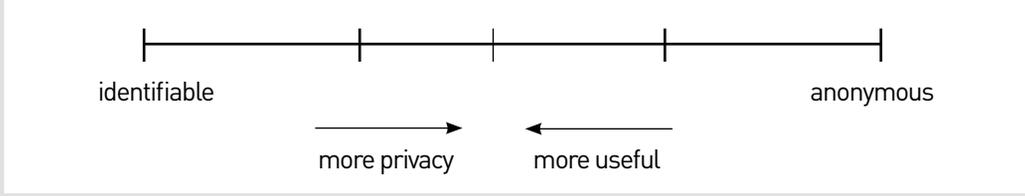
정책 수립 및 평가에서는 데이터의 매우 세분화된 정보까지 필요하므로 매크로데이터에서도 개별정보의 비밀보호라는 제약은 피할 수 없다. 따라서, 매크로데이터 내 개인 및 기업체 등의 정보를 보호하면서 데이터 이용자의 요구에 맞게 데이터가 분석·활용될 수 있도록 매크로데이터의 통계적 비밀보호방법에 대한 관심과 심도 깊은 연구가 필요하다.

## 1. 들어가며

정부는 데이터의 중요성을 인식하고 국정과제로 정부 3.0을 제시하여 공공기관 데이터의 적극적인 개방 및 공유를 통해 국민의 알권리를 충족하고 일자리를 창출하기 위해 노력하고 있다. 이와 함께 「공공데이터의 제공 및 이용 활성화에 관한 법률」을 제정(13.6.27)하여 공공데이터 개방·활용을 위한 제도적 기반을 마련하는 등 공공데이터를 효율적으로 개방하기 위해 노력하고 있다. 이러한 노력으로 공공데이터포털(data.go.kr)로 공공데이터 개방 창구를 일원화하여 2015년 2월 현재 파일데이터 11,898건, Open API 1,623건, 데이터 시각화 251건, 게시물 2,700건을 개방하고 있다.

한편, 데이터를 생산하는 기관에서는 개인 및 사업체의 비밀을 보호하겠다고 서약하고 있지만, 각종 정보를 DB화하고 분석하는 과정에서 또 공개 과정에서 비밀보호 대상 자료가 공개될 위험은 더욱 커지고 있는 실정이다. 특히, 특정한 변수의 값에 특정한 값을 갖는 개인이 한 명만 있는 경우에는, 이 개인의 신상이 노출될 가능성은 높아진다. 반면에, 자료의 비밀보호 조치를 너무 강요하다보면 데이터를 이용하는 사용자에게 제공되는 변수가 너무 제약이 되어서, 자료의 제공으로부터 기대되는 효과를 얻지 못하게 된다. [그림 1]과 같이 식별성에서 개별정보를 보호하고자 하면 원하는 정보를 얻지 못하게 되고, 익명성에서 자료의 활용가치를 높이고자 하면 개별정보의 노출위험

그림 1. 개체 식별성과 익명성과의 관계



자료: Sweeney, L (2001). *Computational Disclosure Control : A Primer on Data Privacy Protection*, Ph. D. thesis, MIT, Dept. of EE and CS. pp.31

이 있기 때문에 이에 대한 적절한 합의점을 찾아야 한다.

이를 위해, 데이터의 개별정보 비밀보호와 관련하여 미국, 캐나다, 호주 등 통계 선진국을 중심으로 노출위험이 있는 자료에 대해 다양한 방법론을 적용하고 있지만, 여전히 마이크로데이터에 대한 비밀보호 방법에 비해 매크로데이터에 관한 비밀 보호 방법에 대한 연구는 상대적으로 미비한 실정이다.

이 글에서는 매크로데이터의 개념 및 보건복지부 소관 매크로데이터의 현황을 살펴보고, 매크로데이터의 노출위험을 측정하여 민감한 쉘을 결정하는 방법론을 제시 및 적용하여 매크로데이터의 비밀보호를 위한 노력이 필요함을 강조하고자 한다.

## 2. 보건복지 분야 매크로데이터의 현황

통계청에 의하면<sup>1)</sup> 조사를 통해 생산되어 최초

입력된 그대로의 자료를 원시데이터(raw-data)라 하고, 이 원시데이터의 입력오류 등을 제거하여 공표 통계표를 작성하고 분석하기 위해 가공한 자료를 마이크로데이터(micro-data)라고 한다. 매크로데이터(macro-data)는 이 마이크로데이터를 임의의 기준에 따라 집계한 자료로서, 집계의 정도에 따라 세분된 자료에서 통합된 자료까지 공표 정도에 따라 다양하게 제공되고 있다.

보건복지부의 매크로데이터는 먼저 행정자치부에서 관리하는 공공데이터포털<sup>2)</sup>을 통해 살펴볼 수 있는데, 공공데이터포털을 통해 보건복지부 소관 26종, 소속기관 46종, 산하기관 108종을 파일 데이터, 오픈 API, 시각화 등 다양한 방식으로 매크로데이터를 구득할 수 있다.(<표1>, <표2>, <표3> 참조<sup>3)</sup>)

다음으로 보건복지부의 매크로데이터는 보건복지 분야 기본 통계와 국내외 주요 통계를 수록한 보건복지통계연보를 통해 살펴볼 수 있다. 「2014 보건복지통계연보」를 통해 국민건강 영역 68종, 보건의료 인력 및 시설 영역 32종, 보건산업 영역

1) 통계청. 마이크로데이터서비스 시스템(<http://mdss.kostat.go.kr/mdssex/et/Glossary.jsp>)

2) 공공데이터포털은 공공기관이 생성 또는 취득하여 관리하고 있는 공공데이터를 찾을 수 있는 통합 창구임.

3) 공표되어 있는 데이터명 그대로로서, 각 데이터 간 레벨이 다를 수 있음. 매크로데이터에 대한 확정적 정의에 대한 논의가 필요하다고 봄.

**표 1. 공공데이터포털을 통해 구득 가능한 보건복지부 소관 매크로데이터 현황**

기관명	데이터명
보건복지부 (26)	차상위 본인 부담경감 대상자 현황통계, 노숙인 및 긴급복지 지원현황, 응급의료기관현황 및 구급차 현황통계, 장애인 편의시설 설치현황, 사회복지사 자격증 취득 현황통계, 사회복지자원 봉사자 현황, 지역 아동센터 현황, 사회복지관 현황, 장례 지도사 자격증발급 및 화장울현황통계, 보건복지분야관련 사업종사자 현황 통계, 암 발생률 주요통계, 보건복지부소관 법인현황, 의료급여지원현황, 노인장기 요양기관 및 노인복지생활시설현황 등, 장애인 연금 및 장애수당수급자현황통계, 보건의료분야 R&D 사업현황, 입양아동수 및 입양기관현황통계, 공중위생영업소 현황통계, 평가인증어린이집정보현황, 면허의료인 수(의사 수,간호사 수, 한의사 수, 약사 수), 정신건강 관련 기관현황, 영유아보육료, 가정양육수당지원대상자현황통계, 의료기관 이용률(외래, 입원), 기초생활보장수급자 수, 장애인활동지원통계정보

주: 괄호 안은 매크로데이터 중 수입, 매크로데이터가 아닌 것(보건복지부 영문 홈페이지 게시판DB 등)은 제외.  
 자료: 공공데이터포털(www.data.go.kr)을 통해 필자가 정리(2015.2.6.기준)

**표 2. 공공데이터포털을 통해 구득 가능한 보건복지부 소속기관 소관 매크로데이터 현황**

기관명	데이터명
질병관리본부(18)	장기이식통계(분기별,연보), 장기기증희망현황, 장기이식대기자현황, 장기기증현황, 장기이식현황, 지역별일별·주별·월별감염병발생통계, 성별·연령별연간감염병발생통계, 지역별·월별역학적연관성별감염병발생통계, 지역별전년대비주별·월별감염병발생통계, 연도별·월별연간감염병발생통계, 지역별연간감염병발생통계, 질병별일별·주별·월별감염병발생통계, 질병별성별·연령별감염병발생통계, 질병별전년대비주별·월별감염병발생통계, 직업별·질병별감염병발생통계, 지역별성별·연령별감염병발생통계, 질병별·지역별·감염병발생통계, 국민건강통계
국립서울병원(2)	정신보건시설정보DB, 국립서울병원연보DB
국립부곡병원(3)	국립부곡병원진료통계, 국립부곡병원약물중독자치료보호실적, 국립부곡병원고객만족도
국립춘천병원(4)	입원재원자수현황, 외래환자수현황, 환자진료정보, 외래/입원수입금현황
국립공주병원(3)	진료통계DB, 원무통계DB, 병동정보
국립재활원(8)	입원환자상병별/장애유형별/유형별진료비및재원일수, 병상별입원환자통계, 의료진(진료과)별내원환자통계정보, 장애인운전교육현황, 상병별환자통계정보, 유형별(의료종별)내원환자통계정보, 장애인인별통계정보, 국립재활원연보
국립마산병원(4)	진료비청구현황, 외래/입원수입금현황, 외래환자수현황, 입원재원자수현황
국립목포병원(4)	입원재원자수현황, 외래내원환자수현황, 진료비청구현황, 외래/입원수입금현황

주: 괄호 안은 매크로데이터 중 수입, 매크로데이터가 아닌 것(국내외 의과학 전문학술 서지정보 등)은 제외.  
 자료: 공공데이터포털(www.data.go.kr)을 통해 필자가 정리(2015.2.6.기준)

**표 3. 공공데이터포털을 통해 구득 가능한 보건복지부 산하기관 소관 매크로데이터 현황**

기관명	데이터명
국민건강보험공단 (36)	직장부과자료관리, 직장소득월액보험료정보, 노인장기요양보험급여비지급현황, 지역부과자료관리, 사업장관리현황, 웨디가입사업장현황, 의료급여통계연보, 복지용품목별이용현황, 노인장기요양보험운영센터별등급판정결과현황, 지역별이용통계, 주요수술통계, 노인장기요양보험인정현황, 만성질환자건강지원서비스 교육 프로그램일정, 가상계좌수납현황, 건강N제공정보, 독촉고지(4대보험료), 정기고지, 사회공헌, 현금급여비(광역시별)지급현황, 현금급여(출산비)지급현황, 현금급여비(도단위별)지급현황, 현금급여비(가정산소치료)지급현황, 현금급여(자가노노소모성재료)지급현황, 현금급여비(제1형당뇨환자소모성재료)지급현황, 현금급여비(장애구보장구)지급현황, 장기요양기관월별현황, 마이헬스뱅크(MyHealthBank)이용현황, 현금급여(요양비)지급현황자료, 임신출산진료비지원내역현황, 급여수준및복지후생, 건강검진통계연보, 건강보험통계연보, 보험료부담대비급여비현황, 건강보험주요통계, 노인장기요양보험통계연보, 노인장기요양보험주요통계
건강보험심사평가원 (18)	상병코드별수진자자료, 요양기관현황정보, 코호트자료, 질환별약제사용정보, 질병별환자정보, 의약품생산/수입/소비량, 약제별사용량정보, 종별과목별진료비현황, 원외처방약제통계자료, 의료각지역별/진료과목별분류, 질병통계, 질병소분류(3단상병)통계자료, 요양기관자원정보, 건강보험통계연보, 행위통계, 심사기준정보, 진료경향모니터링정보, 건강보험통계자료
국민연금공단(38)	급여_원금수취내역, 자격EDI규모별현황, 급여_노후연금자급대부심사결재의뢰내역, 급여_일시금내역, 해외채권투자정보, 급여_수급권기분, 급여_기본연금액, 급여_장애연금내역, 급여_분할연금내역, 급여_연금최종산정, 해외대체투자정보, 급여_유족연금, 급여_노령연금내역, 국내대체투자정보, 해외주식투자현황, 자격지역가입자평준신고금액, 급여_연금내역, 급여_급여심사, 징수건보일일수납현황, 관할지사별유족연금지급현황, 관할지사별장애연금지급현황, 자격지역가입자변동자료일별현황, 자격_일일지사별변동사업장및가입자현황, 자격_자격변동확인통지현황, 자격일일가입자읍면동별현황, 자격지역가입자정기자료현황, 일자별일시금이체현황, 장애연금지급현황, 노령종류별연금지급현황, 유족연금지급현황, 징수수납기관별입금전문현황, 징수종별기금입금통계, 자격일일지사별사업장및가입자현황, 자격외국인체류자격별가입현황, 관할지사별노령연금지급현황, 부당이득환수기금입금현황, 자격외국인국적별현황, 자격시구신고평균소득월액
한국보건산업진흥원(3)	외국인환자유치실적통계, 의료서비스산업통계, 의약품산업통계
한국사회복지협의회 (8)	사회복지자원봉사지역사회봉사단현황통계, 사회복지자원봉사자원봉사관리센터현황통계, 사회복지자원봉사통계연보(2013년), 사회복지자원봉사인증관리요원통계, 전국푸드뱅크일반통계, 자원봉사단체정보, 자원봉사인증센터정보, 사회복지자원봉사활동현황통계
국립중앙의료원(1)	전국공공병원데이터
국립암센터(1)	암등록통계자료
대한적십자사(2)	대한적십자사혈액통계정보, 대한적십자사기부관리정보
한국보육진흥원(1)	어린이집평가인증현황

주: 괄호 안은 매크로데이터 중 수입, 매크로데이터가 아닌 것(건강검진기관 안내 등)은 제외.  
 자료: 공공데이터포털(www.data.go.kr)을 통해 필자가 정리(2015.2.6.기준)

표 4. 「2014 보건복지통계연보」를 통해 구득 가능한 보건복지 분야 매크로데이터 현황

영역명	세부영역명
국민건강	국민영양(6), 질병 및 사망원인(26), 보건 의식행태(16), 보건사업(20), 공중위생(3)
보건 의료인력 및 시설	보건 의료인력(22), 보건 의료시설(10)
보건산업	의약품 등 제조 및 판매업(10), 보건 의료분야 R&D 사업(6)
사회복지서비스	복지인력 및 시설(7), 장애인 복지(14), 보육(9), 아동복지(16), 청소년(11), 여성 및 가족(7), 노인복지 및 기타(16)
공공부조	기초생활보장(9), 의료급여(7)
사회보험	공공연금(3), 국민연금(9), 기초노령연금(1), 건강보험(5), 노인장기요양보험(3), 산재보험(5), 고용보험(3)
재정, 경제	보건복지예산(6), 국민경제(7), 경제활동상태(3)

주: 괄호 안이 매크로데이터 수.  
 자료: 보건복지부. 2014 보건복지통계연보를 통해 필자가 정리

표 5. 국가통계포털을 통해 구득 가능한 보건복지부 및 산하기관 소관 조사데이터 내 매크로데이터 현황

기관명	조사명
보건복지부	가정위탁국내입양소년소녀가정현황(5), 결핵현황(4), 공중위생관계업소실태보고(3), 국민건강영양조사(94), 국민구강건강실태조사(116), 국민기초생활보장수급자현황(5), 국민보건 의료실태조사(117), 국민 의료비추계및국민보건계정(3), 노인복지시설현황(3), 노인실태조사(79), 노인학대현황(57), 노후준비실태조사(16), 법정감염병발생보고(6), 보건복지관련산업일자리조사(7), 보건소및보건지소운영현황(5), 보육실태조사(103), 사회복지자원봉사현황(43), 사회복지서비스수요 공급실태조사(86), 아동복지시설보호아동및중사자현황보고(1), 아동종합실태조사(271), 암등록통계(11), 어린이집및이용자통계(11), 요보호아동현황보고(1), 의료기기화장품제조유통실태조사(31), 의약품소비량및 판매액통계조사(2), 자살실태조사(45), 장애인실태조사(56), 장애인편의시설설치현황조사(5), 장애인현황(6), 전국민장내기생충감염실태조사(8), 정신질환실태조사(104), 지역사회건강조사(35), 청소년건강행태온라인조사(31), 최저생계비계측조사(16), 퇴원 손상심층조사(33), 학대피해아동보호현황(30), 한국의사회복지지출(4), 한방의료이용 및한약소비실태조사(30), 환자조사(11), HIV/AIDS신고현황(16)
국민건강보험공단	건강검진통계(140), 건강보험주요수술통계(31), 건강보험통계(229), 건강보험환자진료비실태조사(파일1), 노인장기요양보험통계(34), 지역별의료이용통계(279)
건강보험심사평가원	건강보험통계(73), 의료기관별급여적정성평가현황(9), 완제의약품유통정보통계(39)
국민연금공단	국민노후보장패널조사(22), 국민연금통계(4)
한국보건산업진흥원	병영경영실태조사(339), 의약품·의료기기연구개발실태조사(14)
국립중앙의료원	응급의료현황통계(20)

주: 괄호 안이 매크로데이터 수.  
 자료: 통계청. 국가통계포털을 통해 필자가 정리(2015.2.11.기준)

16종, 사회복지서비스 영역 80종, 공공부조 영역 16종, 사회보험 영역 29종, 재정 및 경제 영역 16종의 매크로데이터를 구득할 수 있다.

통계청에서 제공하는 국가통계포털은 240여 개 기관이 작성하는 경제, 사회, 환경에 관한 750여 종의 국가승인통계를 수록하고 있다. 국가통계포털을 통해 보건복지부 40종, 산하기관 14종의 조사에 관한 매크로데이터 각각 1,510종, 1,234종을 구득할 수 있다.

앞에서 살펴보았듯이 보건복지분야에서 많은 매크로데이터가 공개되어 활용되고 있지만 매크로데이터에 대한 비밀보호방법을 적용하여 공표하고 있지는 않은 실정이다. 따라서, 활용 가능한 매크로데이터의 비밀보호 기법으로 민감한 셀 결정 방법론을 살펴보고자 한다.

### 3. 매크로데이터의 민감한 셀을 결정하는 방법론<sup>4)</sup>

매크로데이터(macro-data)는 그것이 빈도데이터(count data)인지 크기데이터(magnitude data)인지에 따라 민감한 셀을 결정하는 방법 또한 서로 다르다<sup>5)</sup>. 본 장에서는 매크로데이터의 민감한 셀 결정 방법론에 대한 연구를 정리하고자 한다.

#### 1) 빈도데이터의 민감한 셀 결정 방법론

빈도데이터는 기준(base)이 되는 B값을 정의

하고 임의의 값이 B 이하인 경우 PDC(Primary disclosure cell)로 결정한다. 국가통계의 사업체수 관련 데이터에서 기준 B의 경우 미국의 센서스국(Census Bureau)은 5, 미국의 국세청(IRS)는 3, 우리나라 통계청은 2를 기준으로 하고 있다.

#### 2) 크기데이터의 민감한 셀 결정 방법론

크기데이터의 경우 상위의 소수의 데이터가 전체 셀의 크기에서 차지하는 비중이 큰 것을 노출위험이 크다고 결정하게 된다. 대표적인 방법론인 ① (n,k)-dominance rule, ② p-percent rule, ③ p/q ambiguity rule을 살펴보고자 한다.

##### ① (n,k)-dominance rule

$$S_{(n,k)}(X) = \sum_{i=1}^n x_i - \frac{k}{100-k} \sum_{i=n+1}^N x_i$$

(n,k)-dominance rule은 과거 미국의 센서스국에서 주로 사용했던 방법으로 N개의 데이터의 합으로 구성된 셀 값이 노출에 민감한지 여부를 결정하는 방법이다. N개의 데이터를 크기 순으로 정렬하여,  $x_1, x_2, \dots, x_{N-1}, x_N$  이라고 할 때, 상위 n개의 데이터가 전체 셀 값의 k% 이상을 차지하는 경우에 민감한 셀로 정의한다.

##### ② p-percent rule

$$S_{p\%}(X) = x_1 - \frac{100}{p} \sum_{i=c+2}^N x_i$$

4) Federal Committee on Statistical Methodology Statistical Policy WP 22, *Report on statistical Disclosure Limitation Methodology*, Chapter IV Methods for Tabular Data pp.42~60. 논문을 참고하여 정리함.

5) Edwards G., et al(2001), *Alcohol Policy and the Public Good*, 3rd ed, Oxford University Press.

이 방법은 하나의 셀을 구성하는 N개의 데이터를 마찬가지로 크기 순으로 정리하였을 때, 어느 하나의 값을 알면 다른 상위의 값을 얼마나 추정할 수 있는가 하는 정도로 노출위험 정도를 결정한다. 이 때 c는 가장 큰 데이터의 값을 추정하기 위해 자료를 결합하는 그룹의 크기이다. 즉, c값을 1로 정하면 두 번째로 큰 데이터 값을 이용해 제일 큰 값에 해당하는 데이터를 얼마나 추정할 수 있는가 하는 정도로 노출위험 여부를 결정한다.

$$S_{p/q}(X) = x_1 - \frac{q}{p} \sum_{i=3}^N x_i$$

p-percent rule을 좀 더 강화한 방법으로서 값이 감소할수록 p-percent rule보다 노출위험 정도를 더 민감하게 판단하는 방법이 된다.

#### 4. 보건복지 분야 매크로데이터 비밀보호방법 적용 예시

##### ③ p/q ambiguity rule

<표 6>은 보건복지통계연보를 통해 구득 가능

표 6. 의료기관종별 의원 및 조산원 종사 의료인력 수: 2013년

(단위: 명)

	의 원	치과의원	한의원	부속의원	조산원
의사	35,399	0	0	157	0
치과의사	1	18,595	0	0	0
한 의사	0	0	14,393	0	0
조산사	193	0	0	0	43
간호사	13,522	472	1,134	227	3
간호주무사	58,911	14,826	11,762	51	7
임상병리사	6,225	8	1	43	0
방사선사	7,055	18	2	39	0
물리치료사	13,539	1	1	58	0
작업치료사	105	0	1	0	0
치과기공사	2	1,937	0	0	0
치과위생사	12	22,061	0	6	0
의무기록사	434	115	38	0	0
약사	32	0	9	5	0

자료: 보건복지부, 2014 보건복지통계연보, 2014

표 7. 퇴원환자 수 각 셀을 구성하는 데이터: 2008~2013년

(단위: 명)

	2008년	2009년	2010년	2011년	2012년	2013년
의료기관수	667,419	655,568	702,907	759,046	789,740	831,764
기관종류1	338,356	339,034	368,098	387,159	400,007	416,276
기관종류2	188,523	186,627	211,918	243,713	255,402	277,845
기관종류3	131,636	120,926	111,021	114,303	118,927	120,573
기관종류4	7,443	7,890	9,929	11,383	13,308	15,178
기관종류5	613	492	833	1,548	1,010	680
기관종류6	440	306	648	561	559	606
기관종류7	261	251	359	318	426	354
기관종류8	147	42	101	61	101	252

주: 각 연도별 내림차순으로 정렬. 각 년도의 기관종류1~8은 서로 관계 없음.  
 자료: 보건복지부·한국보건사회연구원. 환자조사, 각 년도

표 8. 퇴원환자 수의 선형민감도: 2008~2013년

(단위: 명)

	2008년	2009년	2010년	2011년	2012년	2013년
$S_{(2,80)}(X)$	-35,281	6,033	88,452	118,176	118,085	143,549
$S_{20\%}(X)$	-364,344	-310,501	-246,357	-253,711	-271,648	-271,939
$S_{20/50}(X)$	-12,994	14,267	60,871	66,724	64,180	72,169

한 의료기관종별 의원 및 조산원 종사 의료인력 수이다. 빈도데이터의 민감한 셀 결정 방법론에 의해 음영으로 표시된 의원의 치과외과 수 및 치과기공사 수, 치과위원의 물리치료사 수, 한의원의 임상병리사, 방사선사, 물리치료사, 작업치료사 수는

노출위험이 있다고 판단된다.

크기데이터의 민감한 셀 결정 사례는 보건복지부의 환자조사를 통해 살펴보고자 한다<sup>6)</sup>. 환자조사에 의하면 2008년부터 2013년까지의 퇴원환자 수는 8개 종류의 의료기관 종류에서 나타났는데(<

6) 환자수에 대한 데이터를 통해 1~2개의 의료기관 종이 전체 셀에서 차지하는 크기가 노출 위험이 있지 않는지 판단하고자 함.  
 7) 의료기관의 장 또는 의료기관의 종사자(의무기록실장 등)가 환자조사관리시스템에 직접 입력하는 방식으로 조사된 환자조사에서 퇴원환자 수는 의료기관별 지정된 1개월(31일)간 퇴원한 환자에 대한 조사결과임.

표 7> 참조), 그 선형민감도를 <표 8>에서 살펴 보면 음영으로 표시된 2009년부터 2013년의 퇴원 환자 수가 (n,k)-dominance rule와 p/q ambiguity rule에 의해서 노출위험 정보로 결정된다.

## 5. 나가며

본 연구에서는 보건복지 분야 매크로데이터 현황을 살펴보고 매크로데이터의 비밀보호방법 중 노출위험을 측정하여 민감한 셀을 결정하는 방법론에 대한 소개와 보건복지 분야 매크로데이터 비밀보호방법 적용 사례를 살펴봄으로써 그 활용가능성을 검토하였다. 그 결과 빈도데이터 적용 사례로 살펴본 보건복지통계연보의 의료인력 수 통계와 환자조사의 퇴원환자 수 통계에서 노출 위험이 있는 셀이 나타났다. 이 연구에서는 조사대상자의 입장에서 비밀이 반드시 보호되어야 하는 민감한 통계가 아닌 것을 예시로 들었지만, 통계를 작성하는 기관에서는 수집된 통계를 공표하는데 있어서 비밀보호를 위한 노력이 필요하다고 인지 할 수 있을 것이다.

통계청 통계자료 제공규정을 살펴보면<sup>8)</sup>, 각 통

계작성기관이 생산·보유하고 있는 통계자료를 외부에 제공함에 있어, 필요한 제공의 범위, 방법 및 절차, 비용 등을 정함으로써 개인, 가구, 사업체, 법인 또는 단체 등 통계조사응답자의 비밀을 보호하고 통계자료 이용자의 편의를 도모하는 한편, 통계자료 제공업무의 효율성을 향상시켜야 한다<sup>9)</sup>. 이와 같이, 개별정보의 비밀보호는 개인 및 기업체 정보보호라는 측면과 자료의 정보를 이용한 사회의 기여도를 동시에 고려하여 신중하게 이루어져야 한다.

보건복지분야의 매크로데이터 현황에서 살펴본 것처럼 많은 매크로데이터가 정부3.0 실천을 위해 공개되어 활용되고 있다. 하지만, 마이크로데이터에 대한 비밀보호 기법의 중요성과 관심에 비해 매크로데이터의 비밀보호 기법에 대한 연구는 여전히 관심도가 낮은 실정이다. 정책 수립 및 평가에서는 데이터의 매우 세분화된 정보까지 필요하므로 매크로데이터에서도 개별정보의 비밀보호라는 제약을 피할 수 없다. 따라서, 매크로데이터 내 개인 및 기업체 등의 정보를 보호하면서 데이터 이용자의 요구에 맞게 데이터가 분석·활용될 수 있도록 매크로데이터의 통계적 비밀보호방법에 대한 관심과 심도 깊은 연구가 필요하다. ■

8) 「통계법」 제27조, 제30조, 제31조, 제33조, 제37조와 「통계법 시행령」 제42조, 제46조부터 제50조까지, 제52조에 근거함.

9) 2014년도 개정함.