pISSN 1229-1153 / eISSN 2465-9223 *J. Food Hyg. Saf.*
Vol. 35, No. 6, pp. 607~617 (2020)
https://doi.org/10.13103/JFHS.2020.35.6.607

Journal of Food Hygiene and Safety
Available online at http://www.foodhygiene.or.kr

인체 바이오모니터링 지식 모델링 및 데이터베이스 구축

이장우 · 양세희 · 이헌주*
켐아이넷(주) 통합위해평가연구실

Knowledge Modeling and Database Construction for Human Biomonitoring Data

Jangwoo Lee, Sehee Yang, Hunjoo Lee*
Department of Integrated Risk Assessment Research, Chem.I.Net Ltd., Seoul, Korea

(Received November 16, 2020/Revised November 23, 2020/Accepted November 26, 2020)

ABSTRACT - Human bio-monitoring (HBM) data is a very important resource for tracking total exposure and concentrations of a parent chemical or its metabolites in human biomarkers. However, until now, it was difficult to execute the integration of different types of HBM data due to incompatibility problems caused by gaps in study design, chemical description and coding system between different sources in Korea. In this study, we presented a standardized code system and HBM knowledge model (KM) based on relational database modeling methodology. For this purpose, we used 11 raw datasets collected from the Ministry of Food and Drug Safety (MFDS) between 2006 and 2018. We then constructed the HBM database (DB) using a total of 205,491 concentration-related data points for 18,870 participants and 86 chemicals. In addition, we developed a summary report-type statistical analysis program to verify the inputted HBM datasets. This study will contribute to promoting the sustainable creation and versatile utilization of big-data for HBM results at the MFDS.

Key words : Human biomonitoring, Knowledge model, Database, Statistical analysis

인체 바이오모니터링(Human biomonitoring, HBM)은 인체의 뇨나 혈액과 같은 생체지표에서 환경유해물질의 원물질 또는 그 대사체를 측정하는 것이다[1]. HBM을 통하여 생산된 데이터(이하, HBM데이터)는 인구집단의 특성(예, 연령, 성별 등)에 따른 유해물질의 노출수준을 확인할 수 있으며, 확인된 노출수준은 건강 영향과의 상관성을 확인하는 데에 중요한 소스로써 활용되고 있다[2].

따라서, 평가 대상이 일반인구집단 내지는 취약계층 등 특정인구집단을 대상으로 유해물질의 위해성(risk)을 평가하는 특성을 가지고 있어[2], HBM데이터는 정부기관을 중심으로 정기적인 조사 혹은 특정 목적의 타겟 모니터링 프로그램에 의해 생산되고 있다.

국외의 경우 미국 질병관리본부(Centers for Disease Control and Prevention, CDC)에서는 1999년 이래 국가 바이오모니터링 프로그램(National Biomonitoring Program, NBP)을 통해 매년 정기적으로 조사한 후 HBM 데이터를 수집하고 처리하여, '환경화학물질의 인체노출에 관한 국가보고서'를 매년 발표하고 있다[3]. 한편, 유럽연합은 30개 유럽 국가들이 연합하여 독일 환경청을 코디네이터로 하여 2017-2021까지 5개년 계획에 의해 유럽연합 인체바이오모니터링 프로그램(Human Biomonitoring for EU Program, HBM4EU)을 통하여, 매년 HBM데이터의 생산하고 인구집단별 노출수준에 대하여 평가를 하고 있다[4].

국내의 경우 국립환경과학원에서는 2009년 이래로 3년을 기수로 하여 '국민환경보건 기초조사' 프로그램을 통하여 매년 정기적으로 조사하여 HBM 데이터를 생산하고 그 결과를 통계분석하여 국가환경포털(환경부), 국가통계포털(통계청)을 통하여 발간하고 있으며, 원시자료도 연구목적의 경우 제공하고 있다. 한편, 식품의약품안전처에서는 2009년 이래 용역연구개발사업 프로그램을 통하여 일반·취약인구집단을 대상으로 자체·외부 연구개발과제 형식의 비정기적인 조사를 통하여 HBM데이터를 생산하고 있으며, 그 결과를 연구개발보고서 형태로 해당 기관의 연

*Correspondence to: Hunjoo Lee, Department of Integrated Risk Assessment Research, Chem.I.Net Ltd., Seoul 07964, Korea
Tel: +82-2-2647-4930, Fax: +82-2-2647-4932
E-mail: adstar@cheminet.kr

Copyright © The Korean Society of Food Hygiene and Safety. All rights reserved. The Journal of Food Hygiene and Safety is an Open-Access journal distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution,and reproduction in any medium, provided the original work is properly cited.

구관리시스템을 통하여 제공하고 있다[5].

데이터베이스(Database, DB) 관리 관점에서 살펴보면, 미국 CDC는 HBM데이터를 실험실간 협업적인 입력 네트워크 체계를 갖추어 매년 축적하고[6], 매년 보고서 형식으로 제공하고 있으며, 유럽연합의 HBM4EU 역시도 프로그램 세부과제(Work Package)에 데이터 관리 및 분석 연구를 두어, 레시피 형식의 명시적인 데이터 관리 계획서(Data Management Plan)[7]를 마련하여, 데이터 생산 표준 및 명명법 등을 만들어 체계적으로 DB를 축적하고 있다.

그러나, 국내의 경우, 아직 SAS 파일, 엑셀파일 등 반정형 형태내지는 보고서 등 비정형 데이터 형식으로 관리되고 있어, 서로 다른 부처 자료원간의 통합은 물론 단일 부처에서 생산된 다년간 자료원의 통합에도 호환성 문제 등 난점을 가지고 있는 실정이다.

다년간 생산된 대용량의 HBM 데이터의 비호환성을 개선하고, 지속적으로 일관성 있게 관리하여 통합 분석할 수 있는 환경을 만들기 위해서는 반드시 현행 개별파일시스템을 데이터베이스관리시스템(Database Management System, DBMS)으로 전환하는 것이 필요하다. DBMS 체계로의 전환을 위해서는 HBM 데이터의 개념, 특성, 유형을 컴퓨터가 해석 가능한 모형을 만들어내는 것이 선행되어야 하는데, 그것이 바로 '지식 모델링'이다.

지식 모델링 기법은 온톨로지 기법 등을 비롯하여 다양하게 존재하지만, 데이터 관리를 목적으로 하는 경우 관계형 DB모델이 많이 활용되고 있다[8]. 관계형 데이터베이스 모델은 집합론을 기초로 만들어진 모델로써 엑셀과 같은 파일 시스템이 갖는 데이터 불일치성을 보완할 수 있다. 관계형 DB모델은 질병관리청의 국민건강영양조사 중 다년도 식이조사 자료와 내분비계 장애물질의 생태 독성 자료들에 대하여 제안된 바 있다[9,10].

본 연구의 목적은 식품의약품안전처에서 과거 생산한 HBM데이터를 분석하여, 관계형 데이터 모델링 기법을 통해 지식 모델을 도출하고, 그 모델 기반의 HBM DB 구축 사례를 제시하고자 한다.

Materials and Methods

HBM 지식 모델링 및 DB 구축에 사용한 데이터 셋

식품의약품안전평가원에서 자체 또는 용역을 통해 2006-2018년까지 수행한 바이오모니터링 사업 보고서와 원자료 및 코드북을 제공받아 데이터베이스 구축에 사용하였다[11-21]. 해당 기간 중 수집한 원자료는 Table 1에 기술하였다. 데이터 셋은 총 11건의 보고서 및 원자료로 구성되어 있으며, 조사대상자 수가 500명 이하인 연구보고서는 4건이며 그 외 7건은 1,030-4,867명을 대상으로 조사한 보고서였다. 주요 구축 데이터는 모니터링 사업에서 수집한 혈액 또는 뇨 시료에서 화학물질을 분석한 농도 값이며, 일부 산모를 대상으로 수집한 모유 중 화학물질 분석값이 있었다.

Table 1. Informations on biomonitoring studies used as sources

Project title	Institution	Year	Chemicals	Sample	Participants
Human Exposure Asesment Study of Bisphenol A	MFDS	2006	Bisphenol A	Blood	300
Human Exposure Asesment Study of Zearalenone	MFDS	2006	Zearalenone	Blood	401
Biomonitoring of exposure to metals derived from artificial joints	MFDS	2007	Cr, Ni	Blood	205
Survey of human milk for POPs in coperation with WHO	MFDS	2008	DDT, HCH, Mirex etc.	Breast milk	480
Database of biological samples for human biomonitoring of hazard materials	Korean Soiciety of Environmental Health	2009 -2010	phthalates	Blood and urine	3,094
Human Biomonitoring of Phthalates, Heterocyclic amines and Phenols for Korean Children	Eulji University	2011	phthalates and Bisphenol A	Urine	1,030
Biomonitoring of acrylamides in children's urine	MFDS	2011	AA, GA, AAMA	Urine	1,030
Biomonitoring of hazardous materials - 5 chemicals including acrylamide	Kyung Hee university	2009 -2010	PFCs, HCAs, AA, Phenol, POPs	Blood and urine	1,904
A study on the integrated exposure to hazardous materials for safety control	Dankook university	2010 -2012	phthalates 4 and Pb, Cd, Hg, MeHg	Blood and urine	4,867
Integrated Risk Assessment Study of Endocrine Disruptors	Yonsei Universit	2015 -2017	Bisphenol A, phthalates	Serum and urine	1,461
Study on the exposure asesment system based on human biomonitoring (I) (Total exposure asesment of bisphenols, acrylamides, furans	MFDS	2018	Bisphenols, acrylamides, furans	Blood and urine	2,487

연구 목적에 따라 각 수집한 생체시료에서 분석한 화학물질의 종류는 비스페놀 류, 프탈레이트 류, 중금속 류, 농약 류 및 페놀 류 등 이었다.

HBM 대상 물질 기초 코드 표준화

HBM연구 대상 물질의 표준 마련을 위한 코드 표준화 작업을 실시하였다. 코드 표준화 작업 시 화학물질의 명칭(국문 및 영문)은 물론 고유 CAS No.를 비롯하여 향후 바이오모니터링 데이터 분석을 통하여 작성될 표나 그래프에 효과적인 활용을 위하여 화학물질의 약어도 모두 구성하였다. 화학물질의 명칭은 대한화학회에서 제공하는 유기, 무기화합물 명명법[22,23]을 참고하였으며, 물질의 약어는 미국 CDC의 NBP에서 제공하는 물질명과 실용성 제고를 위해 기존 문헌들을 참고하여 부여하였다[3,24-28].

HBM 지식 모델링

지식 모델링을 위해 식품의약품안전처에서 최근 12년간 수행한 11개의 자료원을 단계적으로 활용하여 관계형 DB 모델링 방법론에 입각한 DB 설계 작업을 실시하였다. 첫 단계에서는 가장 광범위한 기간과 변수로 구성된 '유해물질 안전관리 통합노출평가기반 연구사업단'[18] 과제의 원시자료 및 코딩북을 기준으로 기초 모델링을 실시하였다. 이후, 나머지 자료원들의 변수 및 코딩체계를 순서대로 분석하여, 특이성이 있거나 특수목적을 가진 연구와 같이 해당 지식 모델을 정교화하였다.

HBM 데이터의 개념, 생산 방법, 생산 순서 및 지식간 관계성을 분석하여 핵심적인 개체를 확인하여, 개체별 세부변수들을 도출하였다. 또한 세부변수들의 경우 범주형, 연속형 자료는 도메인 분석을 통하여 분류하고, 개체간 혹은 세부 변수간 종속성 및 관계성을 분석하여 최종적으로 개체-관계(Entity-Relationship) 지식 모델을 도출하였다.

HBM DB 구축 방법론

도출된 관계형 DB 지식 모델 기반으로 원시데이터의 유형에 따라 가공 및 정제하여 구축하였다. 원시데이터의 유형은 총 3가지로 종이문서(Hard copy) 보고서, PDF파일(Soft copy) 보고서, 스프레드시트 파일로 존재하였다. 이중, 선행적으로 종이문서 및 PDF파일 형식의 보고서의 경우 스프레드파일 형태로 전환하기 위한 전처리작업을 실시하였다. 첫째, 종이문서 형태의 보고서는 광학식문자판독장치(Optical Character Recognition, OCR) 프로그램을 통해 변환 후 가공작업을 수행하였는 데, 글자 간격 또는 특수문자에 의한 인식 오류로 오변환 사례가 다수 발생하여 수작업을 통하여 보정 및 정제작업을 실시하였다. 오변환의 예로는 문자의 인식 오류(원자료: 광주광역시 북구, OCR변환: 광주광역시 au 7), 숫자의 소숫점 누락(원자료: 425.98, OCR변환: 42598) 등이 있었다. 둘째, PDF파일 보고서 역시 텍스트로 변환한 후 스프레드시트 파일에 옮긴 후 정제작업을 실시하였다. 이후 상기 전처리작업을 통하여 변환된 스프레드시트 파일과 원본 자체가 스프레드시트로 만들어진 파일은 도출된 관계형 DB지식모델의 개체별 생성순서에 따라 데이터 가공작업을 실시하였다. 특히, 기기분석을 통해 산출된 HBM 농도 원본값을 유지하고, 부가적으로 LOD 이하의 값은 ND 여부를 마킹하는 항목을 추가함은 물론, Hornung 등[29]이 제안한 방법에 따라 LOD 이하의 값을 LOD/$\sqrt{2}$ 값으로 대체하여 계산한 값을 별도로 저장하는 항목을 추가하였다. 또한, '999' 등의 이상치 값은 제거하였다. 그리고 뇨 시료의 농도 값의 경우 화학물질 희석 정도를 고려하기 위하여 크레아티닌 농도를 이용하여 보정한 값도 별도로 DB화하였다. 설문항목은 범주형, 연속형 변수와 자율 텍스트로 입력되어있는 그룹으로 나누어 범주형 변수의 경우 설문 응답 항목 확인 후 범주외에 벗어나는 값의 확인 작업을 실시하고, 연속형 변수는 변수의 최소값 및 최대값 범위 이내의 값이 포함되었는 지를 확인하였다. 이에 따른 정제작업은 보고서의 원문과 비교를 통하여 실시함으로써 데이터의 왜곡이 없도록 작업을 실시하였다. 텍스트의 경우 원본과 함께 오타등에 대한 정제한 값을 별도로 DB화 하였다. 또한, DB화된 결과를 자료원별 혹은 여러자료원들을 병합하여 결과를 확인할 수 있는 통계 분석 프로그램을 개발하였다. 통계분석 프로그램은 대한민국의 공공사업에 표준적으로 적용되는 전자정부프레임워크를 기반으로 구축하였다. 웹서버와 DBMS는 각각 JEUS 5.0, Oracle 10 g 환경에서 구현하였으며, 사용자 환경은 마이크로소프트의 인터넷 익스플로러와 구글의 크롬 브라우저에서 구동되도록 구현하였다.

구축된 HBM데이터의 신뢰성 검증

HBM 데이터의 신뢰성 검증은 참여자들에게 수집한 설문항목에 대한 응답결과와 생체시료 중 화학물질 분석 결과에 대해 수행하였다. 설문항목 응답결과의 경우 원자료와 구축된 DB의 설문항목 응답결과를 각각 매핑하여 결과값의 일치도와 누락된 정보 등의 오류 사항을 확인하였다. 생체시료 중 화학물질 분석 결과의 경우 자료원 명, 참여자 수, 수집 생체시료 수 등의 일반 정보는 원자료 보고서의 내용과 시스템내의 자료원 정보를 비교 확인하였다. 그리고 분석 결과 값은 자료원에서 보고한 통계량(기하평균, 퍼센타일)과 시스템에서 산출한 통계량을 비교하여 DB 구축의 정확성을 확인하였다.

Results

HBM 물질 코드 표준화 결과

물질의 표준화를 위해 코드 분류체계를 수립하였다. 크게 코드는 물질의 그룹코드와 개별물질 코드를 합성한 6자

리 코드체계로 정의하였다. 그룹코드는 아크릴아마이드(Acrylamide), 비스페놀 및 페놀류(Bisphenols and phenols), 중금속(Heavy metals), 헤테로사이클릭 아민류(Heterocyclic amines), 곰팡이 독소(Mycotoxin), 과불화화합물(Perfluorinated Compounds), 잔류유기오염물질(persistent organic pollutants), 프탈레이트(phthalates)에 대하여 2자리코드로 첫 영문자 1자리와 중복시 일련번호 1자리로 총 2자리 코딩 규칙에 따라 부여하였다. 개별물질코드는 그룹코드 2자리를 포함하여 그룹내에 3자리와 대사체 일련번호 1자리로 총 여섯자리로 부여하였다. 예를 들어, 휘발성유기화합물 그룹의 자일렌 및 그 물질의 대사체인 2-메틸히푸르산의 경우 코드 규칙은 휘발성유기화합물은 'V0', 원물질인 자일렌은 'V00010', 2-메틸히푸르산은 'V00011'로 부여된다. 즉, 코드의 3번째자리부터 3자리는 원물질의 일련번호를 나타내며, 맨끝자리가 '0'이면 원물질을, '0'이 아닌 경우 앞의 원물질의 대사체의 일련번호를 부여하는 형태로 코딩을 실시하였다. 이를 통하여 491개 물질에 대한 물질코드를 부여하였다.

HBM 지식 모델링 결과

HBM데이터의 HBM데이터 지식 모델링은 관계형 DB 모델링 기법에 의거하여 개체, 개체별 속성 및 개체간 관계를 모델링하였다. HBM데이터 중 가장 상위개념으로 평가대상 인구집단(Age group)이 최상위 개체로 식별되었다. 평가 대상 인구집단 개체를 중심으로 조사대상자(Participants)를 구분하여, 조사대상자별 설문자료(Questionnaire results), 측정자료(Measurement results), 노출 평가자료(Assessment results) 개체들이 조사대상자 개체의 고유번호로 연계되어 개체간의 관계가 연결되는 형태로 모델링을 하였다. 평가대상 인구집단은 영유아, 아동 및 청소년, 성인, 영유아 부모로 구분하였으며, 각 그룹의 설문자료 개체의 속성은 인구사회학적 조사(Demographic sociological survey), 생활방식(Lifestyle survey), 거주정보(Residence information survey), 환경노출정보(Environmental exposure information), 살충제 사용 및 음용수(Pesticide use/Potable water), 임상증상 및 과거력(Clinical symptoms & Personal history), 알레르기(Allergy), 신경행동발달(Neural behavior development), 식이조사 24시간 회상(Dietary survey by 24 recall)의 자료로 관계를 형성하도록 모델링을 실시하였다. 측정자료 개체는 조사대상자별로 신체 및 생리학적 특성 계측(Body), 생체시료 중 화학물질 농도(Concentration in biomarker), 식품 중 화학물질 농도(Concentration in food) 개체자료가 포함되는 관계를 나타내도록 모델링을 실시하였다. 이중 생체시료 중 화학물질 농도 및 식품 중 화학물질 농도 개체는 물질코드와 연계되어 물질별 결과값을 저장하고 관리하는 구조로 모델링을 실시하였다. 한편, 노출 평가결과 개체의 경우 HBM연구에서 수행한 설문자료의 24시간 회상법에 의한 식품섭취량 정보(Dietary survey by 24 recall)와 기존의 식품 중 화학물질 농도개체 자료간의 연산을 통하여 식이 노출량(Dietary exposure)을 평가한 개체의 DB가 구축되는 형태로 모델링을 실시하였다(Fig. 1). 즉, 설문자료 중 식이 섭취에 대한 조사 결과가 있는 경우 기존의 식품 중 화학물질 농도와 연계하여 식품 섭취를 통한 화학물질 노출량을 계산하여 조사대상자별 개인 노출량이 산정되어 관리될 수 있도록 모델을 도출하였다.

이와 같이 수립한 관계형 DB 모델 기반으로 도출된 지식 모델은 특정 조사대상자의 모든 정보를 연결하여 조회 및 추출할 수 있을 뿐만 아니라 인구집단(영유아, 청소년, 성인 등)별로 그룹핑하여 다양한 통계량을 산출할 수 있도록 설계하였다.

HBM DB 구축 결과

모니터링 DB를 구축한 자료원에 대한 정보는 Table 1에 기술하였다. 각 자료원 별 구축 결과에서 아크릴아마이드 5종, 과불화합물 10종, 헤테로싸이클린 아민 8종, 페놀류 15종, 농약류 18종을 분석한 2009-2010년 유해물질 인체모니터링 연구 사업단 자료가 102,816건(50.0%)으로 가장 많았다. 2010-2012 유해물질 안전관리 통합노출평가

Table 2. DB construction status for biomonitoring data

Groups		Records number
Sources		11
Chemicals		86
Subject		18,870
Total records		205,491
Age	Adult	159,376
	Child and Student	29,688
	Infant	16,427
Sex	Male	94,031
	Female	111,460
Sample	Blood and serum	72,167
	Urine	132,774
	Breast milk	550
Year	2006	1803
	2007	442
	2008	550
	2009	121,556
	2010	38,630
	2011	20,600
	2015	9,475
	2018	12,435

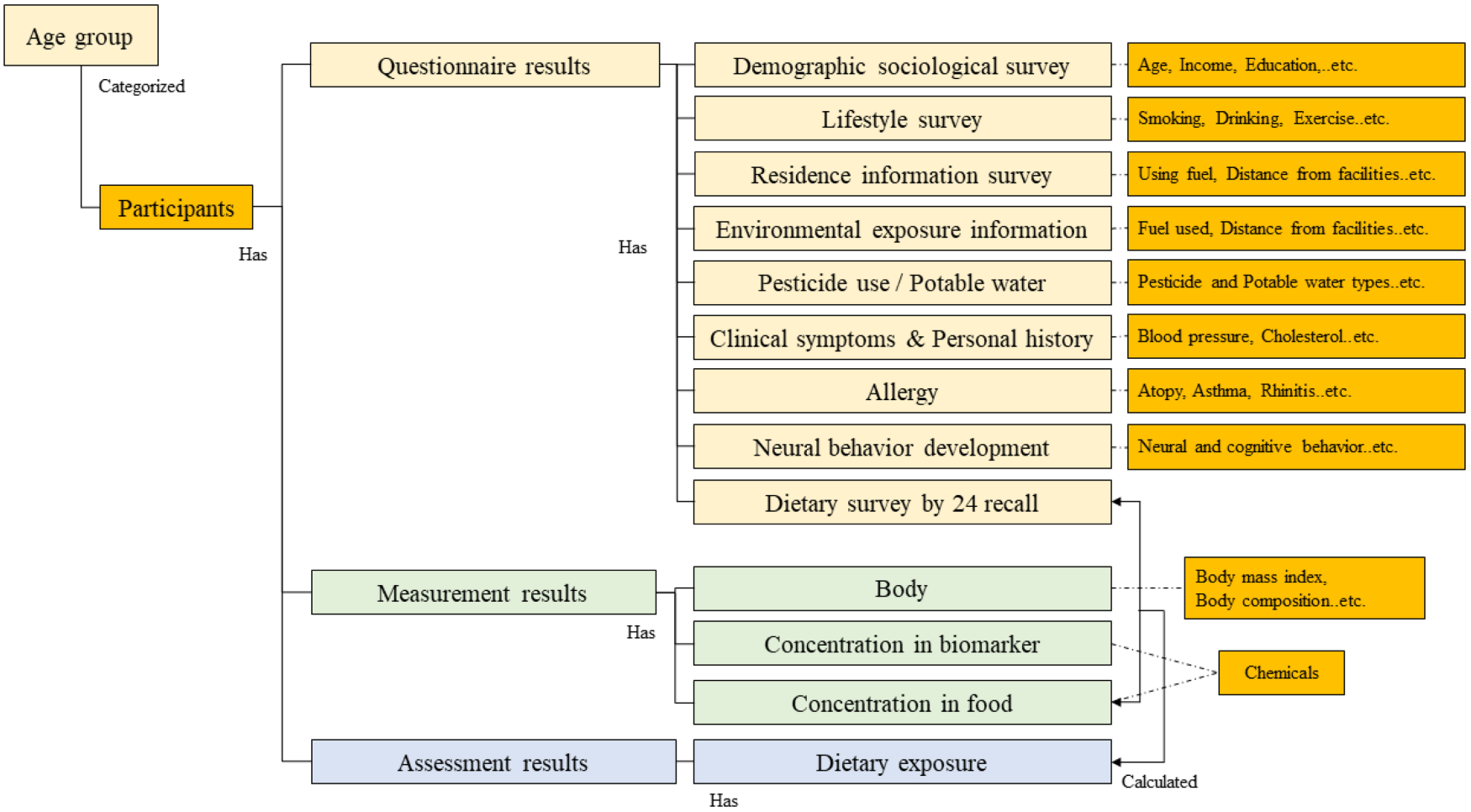

Fig. 1. Conceptual Knowledge Modeling Result for HBM Data.

기반 연구사업단(중금속 4종, 프탈레이트 8종, 비스페놀류 1종)의 자료가 38,630건(18.8%)으로 두 번째로 많았으며 그 외 9건의 연구보고서에서는 총 64,045건 (31.2%)이 구축되었다.

11개의 자료원에 대해 조사대상자 수는 18,870명, 화학물질의 종류는 86종이다. 전체 DB구축 건수는 205,491건으로 성인 159,376건, 아동 및 청소년 29,688건, 영유아 16,427건이며, 남녀 성별로 구분하였을 때는 각각 94,031건과 111,460건으로 조사되었다. 수집된 시료별로 데이터 구축 수를 보면 전혈과 혈청은 72,167건, 뇨는 132,774건, 그리고 모유는 550건으로 조사되었다. 또한 조사대상자에 따른 DB 구축 건수에서 연령별 구축 건 수는 성인이 159,376명(77.6%)으로 가장 많았으며, 성별에서는 여성이 111,460건(54.2%)로 약간 높았지만 큰 차이는 없었다. 생체 시료별 DB 구축건 수는 뇨 시료에서의 화학물질 분석이 132,774건(64.6%)으로 가장 많은 DB를 구축하였다 (Table 2).

86종의 화학물질 별 구축 결과를 보면 프탈레이트 류가 63,226건(30.8%)으로 가장 많았으며, 페놀류 41,611건 (20.2%), 농약류 34,772건(16.9%), 과불화합물 19,040건 (9.3%), 중금속 16,581건(8.1%) 등의 순서로 화학물질에 따라 구축 건 수의 차이가 있었다(Table 3).

또한, Fig. 2와 같이 해당 DB를 활용하여 평가대상 인구집단별 통계량을 분석할 수 있는 프로그램을 개발하였다. 통계분석 절차는 먼저 11개의 자료원을 대상으로 통계분석을 하고자 하는 단일 혹은 복수의 자료원을 선택한

Table 3. List of chemicals in the established biomonitoring DB

Acrylamide (13,776)	Bisphenols and phenols (41,611)	Heavy metals (16,581)	Heterocyclic amines (15,282)
Acrylamide	Bisphenol A	Cadmium	2-amino-3-methyl-imdazo[4,5-f]quinoline
Glycidamide	Bisphenol S	Chromium	2-amino-3,4-dimethylimidazo[4,5-f]quinoline
N-Acetyl-S-(2-carbamoylethyl)-L-cysteine	Bisphenol F	Lead	2-amino-3,8-dimethylimidazo[4,5-f]quinoxaline
	Benzophenone-3	Mercury	2-amino-1-methyl-6-phenylimidazo [4,5-b]pyridine
	Triclosan	Methyl mercury	2-amino-6-methyldipyrido[1,2-a:3’,2’-d]imidazole
	2,4-Dichlorophenol	Nickel	2-aminodipyrodo[1,2-a:3’,2’-d]imidazole
	2,4,5-Trichlorophenol		2-amino-9H-pyrodo[2,3-b]indole
	2,4,6-Trichlorophenol		2-Amino-3-methyl-9H-pyrodo[2,3-b]indole
	2,5-Dichlorophenol		
	4-butylphenol		
	4-Tert-Butylphenol		
	4-Pentylphenol		
	4-Hexylphenol		
	4-heptylphenol		
	4-Tert-octylphenol		
	4-Octylphenol		
	4-Nonylphenol		

Table 3. (Continued)

Mycotoxin (1,203)	Perfluorinated Compounds (19,040)	Persistent organic pollutants (34,772)	Phthalates (63,226)
α-Zearalanol	Perfluoropentanoic acid	Aldrin	Mono-methyl phthalate
β-Zearalanol	Perfluorohexanoic acid	Dieldrin	Mono-ethyl phthalate
Zearalenone	Perfluoroheptanoic acid	Endrin	Mono-n-butyl phthalate
	Perfluorooctanoic acid	Beta-Hexachlorocyclohexane	Mono-isobutyl phthalate
	Perfluorononanoic acid	Toxaphene	Mono-benzyl phthalate
	Perfluorodecanoic acid	Chlordane	Mono-(2-ethylhexyl) phthalate
	Perfluorohexane sulfonic acid	p,p'-Dichloro-Diphenyl-Trichloroethane	Mono-(2-ethyl-5-carboxypentyl) phthalate

Table 3. (Continued)

Mycotoxin (1,203)	Perfluorinated Compounds (19,040)	Persistent organic pollutants (34,772)	Phthalates (63,226)
	Perfluoroheptanesulfonic acid	o,p'-Dichloro-Diphenyl-Trichloroethane	Mono-(2-ethyl-5-hydroxyhexyl) phthalate
	Perfluorooctanesulfonic acid	p,p'-Dichloro-Diphenyl-Dichloroethane	Mono-(2-ethyl-5-oxohexyl) phthalate
	perfluorononane sulfonic acid	o,p'-Dichloro-Diphenyl-Dichloroethane	Mono-(2-carboxymethylhexyl) phthalate
		p,p'-Dichloro-diphenyl-dichloroethylene	Mono-5-carboxy(2-ethylpropyl) phthalate
		o,p'-Dichloro-diphenyl-dichloroethylene	Mono-cyclohexyl phthalate
		Heptachlor	Mono-(3-carboxypropyl) phthalate
		Chlordane-cis	Mono_n_pentylphthalate
		Chlordane-trans	Mono-n-octyl phthalate
		Heptachlor epoxide	Mono-isononyl phthalate
		Nonachlorcis	Mono-iso-decyl phthalate
		trans-Nonachlor	
		Hexachlorobenzene	
		Oxychlordane	
		Mirex	
		2,2',3,4,4',5' and 2,3,3',4,4', 6-Hexachlorobiphenyl	

다(1). 검색버튼을 클릭 후(2), 선택한 자료원에 대한 화학물질 목록이 나타나고 아래 검색창에서 원하는 물질을 검색한다(3). 선택한 물질의 분석된 생체시료에 따라 농도값을 확인할 수 있으며(4), 뇨의 경우 크레아티닌을 보정한 값도 선택하여 확인할 수 있다(5). 모든 조건을 선택하면 관련된 통계분석 결과가 나타나며, 통계분석 결과는 연령그룹에 따라 조사대상자 수, 검출률, 기하평균 그리고 퍼센타일(최소, 25분위수, 중위수, 75분위수, 95분위수, 최대값)을 제공한다(6). 최종결과는 엑셀파일 형식으로 다운로드 할 수 있도록 구현하였다(7).

Discussion

국내의 HBM데이터는 목적, 시기, 연구자 및 측정장비의 차이에 의하여 서로다른 형식에 따라 생산되다 보니, 데이터의 호환성의 문제로 인하여 특정 HBM데이터를 신속하게 조회해야 하거나 인구집단별 시간적 추이분석내지는 다른 국가의 자료와 비교에 난점을 가지고 있는 것이 현실이다. 따라서, 본 연구를 통하여 데이터 호환성 문제를 개선하고 통합적이고 관리가능한 HBM데이터 관리체계 구축을 위해 일관성 있는 표준화된 코딩 체계 및 HBM의 개념에 적합한 지식 모델을 관계형 DB모델 기반으로 도출하였다. 또한, 도출된 관계형 DB 모델의 기반하에 모니터링 자료의 자료원, 조사대상자 정보, 화학물질측정 및 임상검사 자료, 노출평가 결과자료를 구축하였다. 그 결과 11개 자료원의 조사대상자 수 18,870명을 대상으로 86종의 화학물질에 대한 HBM-DB 총 205,491건을 구축하였다.

본 HBM-DB는 관계형 DB모델로 구축되어 지속적인 대용량 DB 축적이나 HBM 데이터 해석을 위한 활용 측면에서 효율적으로 수행할 수 있다. 첫 번째, 조사대상자 또는 화학물질을 중심으로 HBM 관계 데이터 추출이 용이하다. HBM-DB 사업은 목적에 따라 그 사용이 다른데 현 노출 수준을 파악할 목적으로 작은 규모의 대상자를 모집하여 연구를 수행하거나 노출평가를 위한 목적으로 대규모 인구집단을 대상으로 노출된 화학물질의 현 노출 수준과 함께 화학물질이 노출될 수 있는 원인을 설문 형태로 정보를 수집하여 수행하는 연구 두 가지로 나눌 수 있다. 전자의 경우 연령 성별에 대한 기본 정보만 포함되어 있지만, 후자의 경우 연령, 성별 뿐만아니라 인구학적 특성, 거주정보, 생활습관, 신체계측 및 임상정보 등의 다양한 자료가 포함되어 있다. 따라서 기본적으로 연령과 성별을 비롯하여 다양한 변수에 따라 화학물질 수준을 확인할 수 있으며, 화학물질 노출 수준에 따라 나타나는 연령, 거주정보, 직업, 생활습관, 임상증상 등 관련 특징에 대한 정보들을 쉽게 추출할 수 있다. 또한, 코드화된 화학물질 정보에 원물질과 대사체 정보를 모두 DB화하여 바이오마커로써 대사체를 사용한 분석 결과를 원물질로 선택하여 추출할 수 있는 기반을 마련하였다.

두 번째, 이러한 추출 자료는 통계 패기지 처리가 가능한 형태이므로 탐색적 분석의 기반을 마련할 수 있다.

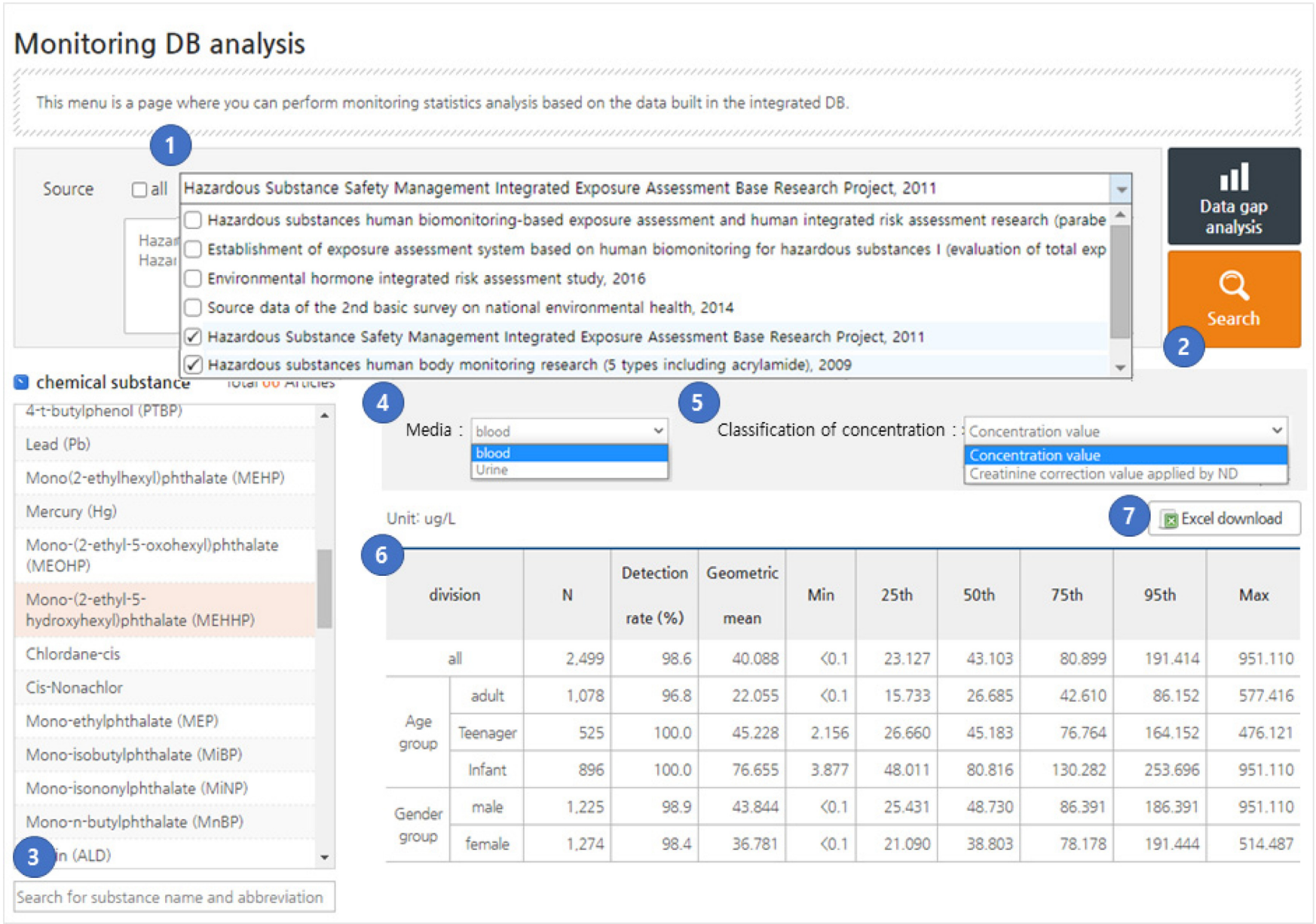

Fig. 2. Statistical monitoring analysis system for HBM database.

HBM 연구의 목적은 다양한 노출원으로부터 다양한 노출경로를 통해 노출되는 화학물질의 수준와 추이를 파악하고 그 조사 결과를 해석함으로서 노출 원인을 설명할 수 있어야 한다[30-32]. 따라서 조사대상 물질 별 노출 수준 분포 특성 및 영향요인과의 상관성 분석이 필요하며, 생체 내 유해물질 농도의 영향요인을 분석하기 위해서는 흡연과 음주, 운동 등 생활습관과 식이습관, 거주환경 등의 정보가 필요하다. 또한, 외적노출량으로부터 체내 흡수된 내적노출량 일부는 DNA나 효소 또는 알부민과 같은 거대분자(macromolecule)와 공유 결합을 하여 부가체(adduct)를 형성하여 건강영향을 일으킬 수 있다. 그러므로 노출영역(exposure- domian)과 영향영역(effect-domain)을 작용기전(mode of action)으로 연결시켜 노출평가 및 기전 연구가 대두되고 있다[33]. 이와 같이 연구결과 해석을 위해 상용화된 통계프로그램을 활용하거나 전용 프로그램을 개발한다. 본 HBM-DB는 이러한 기존 통계프로그램과 연계되어 쉽게 사용할 수 있을 뿐만 아니라 새로운 HBM 분석 프로그램 개발에도 활용이 가능하므로 유해물질에 대한 노출 특성 및 건강영향 기전을 확인을 위한 연구에 기여할 수 있을 것으로 생각된다.

세 번째, 역산 노출량을 이용하여 전통적인 노출량 산출방법을 검증할 수 있다. 유해화학물질에 대한 노출량 추정은 환경, 식품, 제품 등 다양한 노출매체(노출원)와 경구, 흡입, 경피 등의 노출경로를 고려한 노출 시나리오 모형에 노출계수를 이용하여 추정하는 전통적인 방법을 많이 사용하고 있다[34,35]. 그러나 노출 시나리오 모형에 따라 과소평가 또는 과대평가 될 수 있으므로 이를 보완하기 위해 Monte-Carlo 시뮬레이션을 이용한 확률론적 접근방법을 활용하고 있다. HBM데이터 역시 전통적인 방법으로 추정된 노출량을 검증할 수 있는 방법으로 사용될 수 있다. HBM데이터를 이용하여 phamacokinetics 모델을 통해 외적 노출량을 역추정(reverse dosimetry) 할 수 있으며 이는 통합노출량을 반영한다[36]. 그러나, 반감기가 짧은 물질은 반복적인 노출을 통해 일관적으로 노출된다는 가정이 필요하고 노출원이나 노출경로를 파악할 수 없는 제한점이 있다. 그럼에도 불구하고 전통적인 방법을 이용한 노출량과 비교하여 노출 시나리오에서 누락된 부분이나 보완할 점을 확인할 수 있다.

마지막으로, 정부의 정책입안자들은 기존 생산된 자료를 통합된 자료를 이용하여 다년간의 물질별 모니터링 추세분석 등 노출 수준의 경향을 파악할 수 있을 뿐만 아니라, 연령 및 성별에 따른 화학물질의 대푯값 선정 및 노출 저감 등의 정책 결정 사항에 대한 자료로도 활용될 수 있다. 다만, 다량의 DB를 확보하였음에도 불구하고, 서로 다른 HBM의 연구설계에 따른 물질별 데이터 비대칭성 문제로 인하여 대푯값 산출에는 한계점을 가지고 있다. 또한, 본 연구에서는 일관성 있는 HBM 빅데이터 구축의 기반을 마련하였으나, 활용성 증진을 목적으로 HBM 분야에서 필요한 다양한 통계분석 및 시각화 기능을 가진 프로그램 개발은 이루어지지 못하였다. 따라서, 향후 해당 지식 모델 기반의 지속적인 DB축적 및 HBM데이터의 해석에 필요한 프로그램의 추가적인 개선 연구가 이루어져야 할 것으로 생각된다.

Acknowledgements

본 연구는 2019년도 식품의약품안전처의 연구개발비(19162위해기096)로 수행되었으며 이에 감사드립니다.

국문 요약

인체 바이오모니터링(Human biomonitoring, HBM) 데이터는 뇨와 혈액 등 생체지표으로부터 환경유해물질의 측정을 통해 획득되며, 다양한 노출원과 경로로부터 노출되는 유해물질의 인체노출수준 및 건강영향과의 상관성을 파악하기 위해 매우 중요하다. 국내의 경우 식품의약품안전처를 비롯한 국가기관의 다양한 HBM 프로그램을 통하여 HBM 데이터가 생산되고 있다. 그러나, 목적, 시기, 연구자 및 측정 장비의 차이에 의하여 서로 다른 형식에 따라 생산되다 보니, 데이터의 호환성의 문제로 인하여 특정 HBM 데이터를 신속하게 조회해야 하거나 인구집단별 시간적 추이분석 내지는 다른 국가의 자료와 비교에 난점을 가지고 있다. 따라서, 본 연구에서는 HBM 데이터를 체계적으로 데이터베이스(Database, DB)화하고 활용성을 증진하게 시킬 목적으로 지식 모델링을 실시하였다. 지식 모델링은 HBM 데이터의 생산되는 변수들을 그룹화하고 관계를 분석하여 2차원 구조의 개체 및 집합론에 기초한 방법론인 관계형 데이터 모델링 기법을 활용하여 실시하였다. 지식 모델은 조사대상자를 인구집단으로 중심으로 설문자료, 측정자료, 노출 평가자료 개체로 구성하고 그 안에 속성들을 정의하고, 개체간에 관계를 설정하는 방식으로 구성하였다. 또한, 도출된 지식 모델을 기반으로 식품의약품안전처에서 2006년-2018년까지 수행한 HBM의 원시데이터를 수집, 정제 및 정규화하여 통합 DB를 구축하였다. 이와 같이 통합된 HBM-DB는 개별 자료원 내지는 특정 자료원들을 선택하여 기간별 농도 수준에 대한 통계분석은 물론, 다양한 검색조건을 통하여 데이터 추출을 할 수 있는 구조로 구축하였다. 본 HBM-DB는 관계형 DB모델로 구축되어 지속적인 대용량 DB 축적이나 HBM 데이터 해석을 위한 도구로써 효율적으로 수행할 수 있을 것으로 생각된다.

Conflict of interests

The authors declare no potential conflict of interest.

ORCID

Jangwoo Lee https://orcid.org/0000-0002-9172-7090
Sehee Yang https://orcid.org/0000-0001-6055-3000
Hunjoo Lee https://orcid.org/0000-0001-8864-3358

References

1. Kim, S.K., Application of biomonitoring to activities on environmental health and recommendations for korean national environmental health survey. *KJPH*., **52(1)**, 59-74 (2015).
2. World Health Organization, 2015. Human biomonitoring: facts and figures. Copenhagen: WHO Regional Office for Europe, Geneva, Swizerland.
3. Centers for Disease Control and Prevention, 2019. Fourth National Report on Human Exposure to Environmental Chemicals. Atlanta, GA:Centers for Disease Control and Prevention, National Center for Environmental Health.
4. European Human Biomonitoring Initiative (HBM4EU), (2020, November 3). Retrieved from https://www.hbm4eu.eu/
5. National Institute of Environmental Reaearch, 2016. Korean National Environmental Health Survey (KoNEHS): Annual Report on third stage, 2nd year, Incheon, Korea.
6. Association of Public Health Laboratories (APHL), (2020, November 3). Retrieved from https://www.aphl.org/programs/environmental_health/nbn/Pages/application.aspx
7. European Human Biomonitoring Initiative (HBM4EU), (2020, November 3). Data management. Retrieved from https://www.hbm4eu.eu/data-management/
8. Codd, E.F., Derivability, Redundancy, and Consistency of Relations Stored in Large Data Banks, Research Report, IBM, 1969.
9. Kwon, N.J., Suh, J.H., Lee, H.J., Data Cleaning and Integration of Multi-year Dietary Survey in the Korea National Health and Nutrition Examination Survey (KNHANES) using Database Normalization Theory, *J. Environ. Health Sci.,* **43(4)**, 298-306 (2017).
10. Kwon, B.R., Lee, H.J., Developing a Multi-purpose Ecotoxicity Database Model and Web-based Searching System for Ecological Risk Assessment of EDCs in Korea, *J. Environ. Health Sci*., **43(5)**, 412-421 (2017).
11. Ministry of Food and Drug Safety, 2006. Human Exposure Asesment Study of Bisphenol A, Cheongju, Korea.
12. Ministry of Food and Drug Safety, 2006. Human Exposure Asesment Study of Zearalenone, Cheongju, Korea.
13. Ministry of Food and Drug Safety, 2007. Biomonitoring of exposure to metals derived from artificial joints, Cheongju, Korea.
14. Ministry of Food and Drug Safety, 2008. Survey of human milk for POPs in coperation with WHO, Cheongju, Korea.
15. Ministry of Food and Drug Safety, 2010. Database of biological samples for human bio monitoring of hazard materials, Cheongju, Korea.
16. Ministry of Food and Drug Safety, 2011. Human Biomonitoring of Phthalates, Heterocyclic amines and Phenols for Korean Children, Cheongju, Korea.
17. Ministry of Food and Drug Safety, 2011. Biomonitoring of acrylamides in children's urinen, Cheongju, Korea.
18. Ministry of Food and Drug Safety, 2010. Biomonitoring of hazardous materials - 5 chemicals including acrylamide, Cheongju, Korea.
19. Ministry of Food and Drug Safety, 2012. A study on the integrated exposure to hazardous materials for safety control, Cheongju, Korea.
20. Ministry of Food and Drug Safety, 2017. Integrated Risk Assessment Study of Endocrine Disruptors, Cheongju, Korea.
21. Ministry of Food and Drug Safety, 2018. Study on the exposure asesment system based on human biomonitoring(I)(Total exposure asesment of bisphenols, acrylamides, furans, Cheongju, Korea.
22. The Korean chemical society, 2013. Organic Compound Nomenclature, Seoul, Korea.
23. The Korean chemical society, 2017. Inorganic Compound Nomenclature, Seoul, Korea.
24. Assens, M., Frederiksen, H., Petersen, J.H., Larsen, T., Skakkebæk, N.E., Juul, A., Andersson, A., Main, K.M., Variations in repeated serum concentrations of UV filters, phthalates, phenols and parabens during pregnancy. *Environ. Int*. **123**, 318-324 (2019).
25. Attaullah, M., Yousuf, M.J., Shaukat, S., Anjum, S.I., Ansari, M.J., Buneri, I.D., Tahir, M., Amin, M., Ahmad, N., Khan, S.U., Serum organochlorine pesticides residues and risk of cancer: A case-control study. Saudi. *J Biol. Sci*. **25**, 1284-1290 (2018).
26. Goerke, K., Ruenz, M., Lampen, A., Klaus, A., Bakuradze, T., Eisenbrand, G., Richling, E., Biomonitoring of nutritional acrylamide intake by consumers without dietary preferences as compared to vegans. *Arch Toxicol,* **93(4)**, 987-996 (2019).
27. Husoy, T., Andreassen, M., Hjertholm, H., Carlsen, M.H., Norberg, N., Sprong, C., Papadopoulou, E., Sakhi, A.K., Sabaredzovic, A., Dirven, H.A.A.M., The Norwegian biomonitoring study from the EU project EuroMix: Levels of phenols and phthalates in 24-hour urine samples and exposure sources from food and personal care products. *Environment International,* **132**, 105103 (2019).
28. Wu, X.M., Bennett, D.H., Calafat, A.M., Kato, K., Strynar, M., Andersen, E., Moran, R.E., Tancredi, D.J., Tulve, N.S., Hertz-Picciotto, I., Serum concentrations of perfluorinated compounds (PFC) among selected populations of children and adults in California. *Environ. Res*. **136**, 264-273 (2015).
29. Hornung, R.W., Reed, L.D., Estimation of average concentration in the presence of nondetectable values. *Applied Occupational and Environmental Hygiene*, **5**, 46-51 (1990).
30. Larry, L.N., Antinia, M.C., Dana, B.B., Uses and issues of biomonitoring. *Int. J. Hyg. Environ. Health,* **210**, 229-238 (2007).
31. National Research Council (U.S.), 2006. Committee on

Human Biomonitoring for Environmental Toxicants. Human biomonitoring for environmental chemicals. Washington, DC: National Academies Press, pp. 291.
32. National Institute of Environmental Reaearch, 2015. Research development and planing for Korean National Environmental Health Survey (KoNEHS) (III), Incheon, Korea.
33. Ministry of Food and Drug Safety, 2010. Study on human biomonitoring of hazardous substances, MFDS, Cheongju, Korea.
34. Ministry of Environment, 2007. Korean exposure factors handbook. Seoul, Korea.
35. Ministry of Food and Drug Safety, 2010. Integrated risk assessment on Bisphenols, MFDS, Cheongju, Korea.
36. Tan, Y.M., Liao, K.H., Clewell, H.J., 3rd. Reverse dosimetry: interpreting trihalomethanes biomonitoring data using physiologically based pharmacokinetic modeling. *J. Expo Sci. Environ. Epidemiol.*, **17(7)**, 591-603 (2007).