

건설업 유해화학물질 노출 모델의 개발 및 검증: Tier-2 노출 모델

김승원^{1*} · 장지영¹ · 김갑배²

¹계명대학교, ²한국산업안전보건공단

Development and Validation of Exposure Models for Construction Industry: Tier 2 Model

Seung Won Kim^{1*} · Jiyoung Jang¹ · Gab Bae Kim²

¹Keimyung University, ²Korean Occupational Safety and Health Agency

ABSTRACT

Objectives: The major objective of this study was to develop a tier 2 exposure model combining tier 1 exposure model estimates and worker monitoring data and suggesting narrower exposure ranges than tier 1 results.

Methods: Bayesian statistics were used to develop a tier 2 exposure model as was done for the European Union (EU) tier 2 exposure models, for example Advanced REACH Tools (ART) and Stoffenmanager. Bayesian statistics required a prior and data to calculate the posterior results. In this model, tier 1 estimated serving as a prior and worker exposure monitoring data at the worksite of interest were entered as data. The calculation of Bayesian statistics requires integration over a range, which were performed using a Riemann sum algorithm. From the calculated exposure estimates, 95% range was extracted. These algorithm have been realized on Excel spreadsheet for convenience and easy access. Some fail-proof features such as locking the spreadsheet were added in order to prevent errors or miscalculations derived from careless usage of the file.

Results: The tier 2 exposure model was successfully built on a separate Excel spreadsheet in the same file containing tier 1 exposure model. To utilize the model, exposure range needs to be estimated from tier 1 model and worker monitoring data, at least one input are required.

Conclusions: The developed tier 2 exposure model can help industrial hygienists obtain a narrow range of worker exposure level to a chemical by reflecting a certain set of job characteristics.

Key words : Bayesian statistics, construction, exposure model

I. 서 론

전통적인 빈도주의적 접근법(frequentist approach)과 구별되는 베이지안 접근법(Bayesian approach)의 가장 큰 특징은 사전확률분포(prior probability distribution)를 확률의 계산에 결합시킨다는 점이다. 이전의 측정자료, 모델 예측치, 전문가의 판단(professional judgement) 등이 사전확률분포로 활용될 수 있다. 그 이외에도 베이지안 통계에서는 모집단의 모수들에 대하여 일정한

분포를 가정하고 확률에 기초한 방법들(probabilistic techniques)을 사용하여 노출수준을 예측할 수 있다. 이런 확률에 기초한 방법들은 미국과 유럽연합 등지의 환경보건 분야에서는 이미 1990년대부터 개발되어 활용되어 오던 방법들이며 베이지안 통계(Bayesian statistics)는 그 중 하나이다.

작업환경관리를 위해서 측정하는 표본의 수는 사용되는 화학물질의 수에 비해서 적은 편이며, 법적으로 관리가 요구되는 화학물질들에 대한 측정을 하는

*Corresponding author: Seung Won Kim, Tel: 053-580-5197, Email: swkim@kmu.ac.kr, Baekeun-Gwan, #B50, Department of Public Health, Keimyung University. 1095 Dalgubeol-Daero, Dalseo-Gu, Daegu

Received: 2014, Revised: 2014, Accepted: 2014

This is an Open-Access article 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.

경우에도 모집단의 모수들을 추정할 수 있을 만큼 많은 수의 시료를 채취하는 경우는 드물다. 이 점에 착안하여 유럽연합에서는 Registration, Evaluation, Authorization and Restriction of Chemicals(REACH) 제도를 도입하면서 실질적인 측정 없이 과거노출자료를 바탕으로 화학물질의 독성 및 사용방법에 따라 노출정도를 추정하여 위험성 평가(risk assessment)를 수행하거나, 적은 수의 측정자료를 가지고 베이지안 통계를 이용하여 예상되는 노출농도의 범위를 추정할 수 있도록 독려하고 있다.

미국 등지에서도 산업위생 전문가의 판단, 이전 측정자료 등을 일관적이고 과학적으로 작업환경관리에 활용할 수 있는 방법을 찾아왔다. 베이지안 통계는 그 중 가장 주목받고 있는 방법이며 미국 National Institute for Occupational Safety and Health(NIOSH)에서도 Occupational Exposure Assessment Sampling Strategy Manual을 개정하는 과정에서 측정데이터가 적은 경우에 베이지안 통계를 활용하도록 준비 중이다.

본 연구의 주요 목적은 개발된 노출 모델 tier 1 추정치와 실측치를 결합하여 tier 1의 노출예상범위보다 더 좁은 범위를 제시하는 tier 2 노출 모델을 개발하는 것이다. 또한, 엑셀의 매크로와 비주얼 베이직 요소를 결합하여 개발된 도구를 스프레드시트 상에 구현하는 것이다.

II. 연구 내용 및 방법

1. 베이지안 통계 분석 방법 조사

국내외 산업보건분야 및 국외 환경보건분야에서 출판된 문헌 중 베이지안 통계 분석 방법을 사용한 논문들을 분석하였다.

국내 산업보건분야 논문의 경우 한국산업위생학회지를 검색하였다. 국외 산업보건분야의 경우 다음 학술지들을 검색하였다.

- Journal of Occupational and Environmental Hygiene
- Annals of Occupational Hygiene
- Scandinavian Journal of Work, Environment & Health

국외 환경보건분야 및 보건의료분야에서 출판된 문헌은 SCI Journal 중 환경과학(Environmental Science) 분야와 통계 및 확률(Statistics and Probability)분야에서 imaction factor가 높은 보건관련 학술지들을 가

려 다음과 같이 리스트를 만들어 최근 논문들을 검색하였다.

- Environmental Health Perspective
- Critical Reviews in Environmental Science and Technology
- Environmental Science and Technology
- Atmospheric Environments
- Statistics in Medicine

베이지안 통계 분석이 환경보건 분야에서 주목을 받게 된 이유는 측정자료를 충분히 얻는 것이 불가능한 경우가 많아 자료의 희소성을 보충할 방법을 찾게 되면서부터이다. 실측자료가 부족한 경우 과거의 측정자료, 전문가의 의견, 모델의 예측치 등을 과학적이고 투명한 방법으로 현재의 측정자료와 연결시킬 방법이 필요했고 베이지안 통계는 이런 요구들의 많은 부분을 충족시킬 수 있는 도구였다. 자료가 희소한 경우에 사용될 수 있는 다른 방법과 함께 베이지안 통계 분석이 가지는 장단점에 대해서 조사하였다.

베이지안 통계가 주목받게 된 또 다른 이유는 확률의 형태로 결과를 제시하여 의사결정을 돕는다는 것이다. 이 방법은 허용기준과의 비교 시 초과/미초과 등의 이분법적인 판단이 아닌 허용기준을 초과할 가능성을 확률로서 제시한다. 이러한 확률분포형태의 결과 제시는 기업경영 등에서 사용하던 의사결정방법론을 허용기준 초과 판단여부에 사용한 경우라 할 수 있다. 미국 NIOSH에서 1977년 발행하고 Occupational Safety and Health Administration (OSHA) 이를 채택하여 우리나라를 포함한 여러 나라들의 노출평가의 프레임으로 사용되었던 Occupational Exposure Sampling Strategy Manual이 개정 중에 있다. 이 개정작업에서 베이지안 통계가 어떻게 활용되고 있는지도 조사하였다.

2. 작업환경측정 분야에서 사용되는 통계 방법 및 소프트웨어 개발

1) 베이지안 방법론과 노출모델의 결합

기존에 개발된 노출 모델 tier 1과 베이지안 통계를 결합하여 tier 2 모델을 개발하였다. tier 2 모델은 실측치와 모델의 예측치를 결합하여 더 좁은 범위의 노출예상범위를 제시하는 것이 가능하다. 유효성 검증에 있어서 tier 2 모델은 실측치와 예상치를 비교하

여 만드는 tier1 모델과는 달리 수학적으로 계산하는 모델이기 때문에 실측치를 이용한 검증이 필요 없다.

2) 엑셀을 이용한 소프트웨어 개발

엑셀의 매크로와 비주얼 베이직 요소를 결합하여 앞에서 개발된 도구들을 스프레드시트 상에 구현하였다. Tier 2 모델은 tier 1 계산결과를 필요로 하지만 건설업에 한정되지 않고 측정데이터를 이용하여 예측범위를 좁혀주는 역할을 한다.

III. 연구결과 및 고찰

1. 베이지안 통계 분석 방법 조사

1) 베이지안 통계 분석 방법

베이지안 통계 방법론(Bayesian statistical method)은 알려져 있는 사전 정보(prior information)와 관측한 자료(data)를 결합하여 얻어진 사후 정보(posterior information)를 바탕으로 통계적인 추론(statistical inference)을 하는 방법들을 말한다. 보건 역학 분야에서 베이지안 통계 방법론은 생소한 개념으로 생각될 수 있으나 그 기본적인 개념이 적용된 예는 쉽게 찾을 수 있다.

최근 베이지안 방법론이 많은 분야에서 사용되는 주된 이유는 기존의 전통적인 통계방법(frequentist method)과는 다르게 사전 정보의 확률 분포를 반영하면서 통계적인 추론을 하기 때문이다. 전통적인 통계방법에서는 사전 정보를 사용하지 않고, 오직 실제 관측된 자료로부터 표본이 추출된 모집단의 특성을 통계적으로 추론하는데 비해, 베이지안 방법론은 전문가의 소견, 다른 비슷한 주제의 연구 성과물, 일반적으로 알려져 있는 사실 등을 불확실성(uncertainty)을 반영한 사전 정보의 확률 분포로 표현하고 활용하여 특히 관측된 자료의 수가 적은 경우 보다 의미 있는 추론을 가능하게 한다.

• 베이지안 방법론의 활용

베이지안 방법이 활용되는 대표적인 두 가지는 이산형 자료일 경우와 연속형 자료일 경우인데, 실제 다양한 변수로 구성된 자료 분석이나 변수간의 관련성을 보는 복잡한 모형들 (예를 들면 회귀분석이나 ANOVA)에서도 베이지안 방법론은 쉽게 적용될 수 있다. 이미 WinBUGS와 같은 무료 프로그램을 통해

많이 사용되고 있다. 특히, 자료의 수가 적은 상황에서 이미 알려져 있는 사실을 사전 정보로 활용하고자 할 때, 베이지안 방법론은 매우 유용하게 사용될 수 있다.

2) 문헌 조사: 산업보건분야

산업보건분야에서는 환경보건분야 등과 비교하여 상당히 최근에 들어서 베이지안 통계분석방법을 사용하기 시작하였다. 관련 연구들을 연대기적인 순서로 정리하면 다음과 같다.

• Ramachandran의 Bayesian inference

Ramachandran & Kandlikar(1996)는 베이지안 분석을 이용해 이산적으로 측정된 자료(discrete measurements)를 가지고 연속된 형태의 공기중 입자의 크기분포를 추정하였다. 이러한 역산(inversion)은 베이지안 분석이 많이 활용되는 분야 중 하나이다. 단봉분포(unimodal distribution)를 하는 경우 다른 방법으로도 비교적 쉽게 역산이 가능하지만 쌍봉분포(bimodal distribution)의 경우 계산이 쉽지 않다. Ramachandran & Kandlikar는 베이지안 분석과 몬테카를로 시뮬레이션(Monte Carlo simulation)을 결합하여 쌍봉분포하는 공기중 입자의 크기 분포를 성공적으로 추정하였다. Micro Orifice Uniform Deposition Impactor(MOUDI) 시료채취기 같은 직경분류충돌기(cascade impactor)를 사용하여 공기중 입자의 크기분포를 측정하는 경우 스테이지 수가 많지 않아 각 스테이지의 cutoff 크기별 이산적 자료만 얻을 수 있을 뿐 연속된 형태의 입자크기분포를 알 수 없다. 베이지안 분석은 최대우도추정법(maximum likelihood estimation) 등과 함께 많이 쓰이는 역산의 한 방법이다.

• Ramachandran의 과거노출 추정

베이지안 역산을 사용하여 연구했던 Ramachandran (2001)는 베이지안 분석을 이용해 니켈 제련소에서 공기중 니켈의 과거노출수준을 추정하였다. 과거노출의 경우 노출정도를 추정하는데 이용 가능한 데이터의 수가 많지 않아 불확실성(uncertainty)이 큰 추정이 되는 경우가 많다. 기록으로 남아있는 자료와 전문가의 의견을 결합하여 사전분포(prior distribution)로 사용하고 베이지안 분석을 통해서 이용가능한 측

정치를 결합하는 경우 얻어지는 사후분포(posterior distribution)의 변이는 작아지게 된다. 베이지안 분석은 이와 같이 과거노출 추정에 많이 사용된다. 기본적으로 베이지안 분석을 이용한 역산과 원리가 같다고 할 수 있다.

• Hewett의 Bayesian decision analysis

Hewett et al.(2006)은 베이지안 의사결정분석(Bayesian decision analysis)을 이용하여 허용기준 초과여부를 확률 형태로 보여주는 방법을 제시하였다. 이들이 사용한 노출 카테고리(exposure category) 분류는 American Industrial Hygiene Association(AIHA)의 노출카테고리 분류방식(exposure category scheme)을 따랐다. AIHA 노출카테고리는 각각 허용기준(occupational exposure limit, OEL)의 0.01, 0.1, 0.5, 1.0배를 기준으로 5개의 그룹으로 나누어진다.

이 연구에서 사용한 방법 중 일부는 본 연구에서도 차용되었으므로 자세하게 설명하고자 한다. 측정데이터가 각각의 노출카테고리에 속할 확률은 다음 식을 이용하여 계산된다.

$$P(\ln G_i, \ln D_i | data) = \frac{P(data | \ln G_i, \ln D_i) \cdot P(\ln G_i, \ln D_i)}{\sum_{i=1}^k [P(data | \ln G_i, \ln D_i) \cdot P(\ln G_i, \ln D_i)]} \quad (1)$$

여기서 G와 D는 각각 기하평균과 기하표준편차, i는 각 노출카테고리, data는 해당 유사노출그룹(similar exposure group, SEG)의 측정값 {y₁, y₂, ..., y_n}을 나타낸다. 여기서 P(data | ln G_i, ln D_i)는 우도분포(likelihood distribution)를 나타내는 함수로 다음의 식에 의해서 계산될 수 있다.

$$P(data | \ln G_i, \ln D_i) = K \cdot \prod_{j=1}^n pdf(y_j | \ln G_i, \ln D_i) \quad (2)$$

여기서 K는 비례상수, pdf는 기하분포에 대한 확률밀도함수(lognormal probability density function)를 나타낸다. 확률밀도함수를 이용한 총 승(누적곱하기)의 계산은 다음과 같이 이루어진다.

$$pdf(y_j | \ln G_i, \ln D_i) = \frac{1}{\ln D_i \sqrt{2\pi}} \cdot \exp\left(\frac{-(y - \ln G_i)^2}{2(\ln D_i)^2}\right) \quad (3)$$

식(2)와 (3)에 설명된 확률밀도함수에 대한 계산을 위해서 앞서 설명한 식(1)을 적분의 형태로 다시 쓰면 측정자료(data)가 각 노출카테고리에 해당될 확률은 식(4)와 같이 계산된다.

$$P(Pop_i | data) = \frac{\int_{\ln G} \int_{\ln D} [P(data | \ln G_i, \ln D_i) \cdot P(Pop_i)] d(\ln G) d(\ln D)}{\int_{\ln G_{min}}^{\ln G_{max}} \int_{\ln D_{min}}^{\ln D_{max}} [P(data | \ln G_i, \ln D_i) \cdot P(Pop_i)] d(\ln G) d(\ln D)} \quad (4)$$

여기서 Pop_i는 각 노출카테고리를 나타낸다.

Vadali et al.(2009)은 Hewett이 개발한 의사결정분석 방법에서 모델 예측치를 사전분포의 자료로 활용하는 법을 제시하였다. 전문가의 의견(expert judgement)을 사전확률로 사용하는 것은 이론적으로 가능하지만 표준화하는 것이 어렵다. 그렇기에 좀 더 재현성이 있는 수학적 노출모델을 사용하여 사전확률분포를 계산하였다. Vadali et al.(2009)이 사용한 모델 중 일반 환기모델(general ventilation model)은 정적인 상태(steady state)에 있는 작업장에서 유해물질의 발생량(G)과 환기량(Q)을 이용하여 공기중 농도(C)를 계산하는 간단한 모델이다.

$$C(mg/m^3) = \frac{G(mg/hr)}{Q(m^3/hr)} \quad (5)$$

이 모델에서 발생량과 환기량은 각각의 최대/최소치를 갖게 되는데 이 값들에 대한 불확실성을 관리하기 위해서 이들 각각의 최대/최소치에 대한 범위를 다시 정한다. Vadali et al.(2009)은 이것을 이차원 몬테카를로 기법(2-dimensional Monte Carlo method)이라고 소개하였다. 식(5)의 모델 이외에도 Nicas & Jayjock(2002)의 near field/far field 모델이나 포화증기압 모델(saturation vapor pressure) 등을 이용한 사전확률을 계산하여 서로 비교하였다.

Lee et al.(2013)은 Cherrie & Schneider(1999)가 개발한 structured subjective assessment(SSA) 모델과 control banding 모델 등의 개념적인 모델이나 전문가의 판단을 사전확률분포로 사용하려는 시도를 하였다. Cherrie & Schneider(1999)의 SSA 모델은 이후 유럽연합에서 REACH에 따라 화학물질을 등록할 때 요구되는 노출수준 추정에 사용되는 노출모델들의 근간이 되었다. 이 연구에 사용된 개념적인 모델들도

만족스러운 수준의 사전확률분포를 제공하였으나, 여전히 할당되는 변수의 범위에 대한 주관적인 선택을 어떻게 반복성 있게 만들 수 있는가 등의 문제를 안고 있다.

• **NIOSH의 Occupational Exposure Sampling Strategy Manual 개정 작업**

미국 NIOSH는 1977년 Leidel, Busch, Lynch 3인의 저자에 의해서 작성된 Occupational Exposure Sampling Strategy Manual(OESSM)에 대한 개정작업을 진행하고 있다. OESSM은 발간 이후 미국 OSHA가 허용기준을 적용하는 전략의 근간이 되었으며 한국을 비롯하여 많은 나라에서 이를 채택하여 사용하였다. 현재의 OESSM(1977)에서 허용기준 초과여부의 판단은 빈도주의 통계(frequentist statistics)에서 사용하는 신뢰구간을 계산하여 허용기준과 비교하도록 되어있다. 신뢰구간은 충분히 많은 수의 자료수를 모집단에서 추출하는 경우 모집단의 분포와 상관없이 정규분포하고 이를 이용하여 모집단의 모수를 추정할 수 있다는 중심극한정리(central tendency theorem)에 의존한다. 좀 더 자세히 설명하면 95% 신뢰구간에서 upper confidence limit(UCL)을 계산하여 이 값이 허용기준 값보다 크면 허용기준을 초과할 확률이 있다고 보는 방법이다. 이 방법의 단점 중 하나는 중심극한정리를 성립시킬 만큼 많은 수의 시료가 채취되어야 한다는 점인데 실제 작업환경측정 시에 여러 가지 이유로 한 개 내지 두 개 정도의 시료가 채취되는 것이 대부분이라는 점이다. 이 상황은 국내에만 국한되지 않고 OESSM을 만든 미국에서도 마찬가지이다.

이에 NIOSH는 개정을 준비하면서 적은 수의 측정이 이루어지는 경우에 대한 별도의 내용을 준비 중이다. 이 부분에서 허용기준과 비교하게 되는 이론적 근간은 Hewett et al.(2006)이 제시한 베이지안 통계를 이용하는 것이다. 베이지안 통계기법은 중심극한정리에 의존하지 않기에 이를 이용하면 한 개의 측정값을 가지고서도 반복 가능한 방법으로 노출수준 및 허용기준을 초과할 확률을 계산할 수 있다.

• **유럽연합의 노출모델들**

유럽연합에서는 Registration, Evaluation, Authorisation and Restriction of Chemicals(REACH) 법령을 실시하면서 등록하려는 화학물질에 대해 예상되는 노출수

준을 시나리오별로 보고하도록 하고 있다. 시나리오 중에는 작업장에서 근로자들이 노출되는 상황을 포함하여 소비자가 제품을 사용했을 때 노출될 수 있는 상황 등을 포함하고 있다. 모든 화학물질에 대하여 실험적으로 노출수준을 평가하는 것은 실제로 불가능하므로 1차적으로 노출모델을 이용한 평가를 통해 노출수준이 낮을 것이라고 알려진 상황들은 실측하지 않도록 걸러내는 것(screening out)을 허용하고 있다. 이 목적으로 개발된 노출모델을 tier 1 모델이라고 하며 tier 1 모델에서는 노출농도에 대한 실제 평가 없이 노출 상황에 대한 자료만을 가지고 노출수준을 예측한다. 일부 실측 데이터를 사용하여 tier 1 모델에서 예측된 노출수준을 더 좁혀가는 것을 tier 2 모델이라고 한다. 이러한 노출모델의 개발을 위하여 유럽은 몇몇 기관들이 지난 수년간 노력해 왔고 여러 가지의 모델이 개발되고 업데이트 되어왔다.

본 연구에서는 이들 노출 모델 중 작업환경에 대한 시나리오에 초점을 맞추어 개발된 노출 모델에 대해서 간략하게 기술하였다.

작업장 환경에 사용할 수 있도록 개발된 노출모델 중 가장 많이 알려지고 활발히 연구 및 사용되고 있는 모델은 Advanced REACH Tools(ART)와 Stoffenmanager를 들 수 있다. 이 두 모델은 모두 앞서 설명한 Cherrie & Schneider(1999)의 structured subjective assessment (SSA) 모델에 기반을 두고 있다. SSA 모델이 처음 학계에 소개된 것은 Cherrie et al.(1996)에 의해서였는데 초기의 모델은 조금 더 다듬어져 1999년 논문에서 현재의 모습을 갖추게 되었다. 그렇기에 유럽연합의 노출모델에 관한 연구들은 Cherrie & Schneider (1999) 논문을 인용하고 있으며 본 보고서에서 지칭하는 SSA 모델은 이 논문을 가리킨다.

SSA 모델은 작업자와 오염원과의 거리에 따라 근거리공간(near field)과 원거리공간(far field)으로 나누고 별도로 계산을 수행한다. 이 두 공간 간의 구분은 정교하게 정의된 것이 아니라 오염원을 중심에 두고 있는 지면에 붙어있는 2 m의 정육면체 공간(체적 = 8 m³) 정도로 정의하고 있다. ART와 Stoffenmanager에서도 오염원으로부터 1미터 떨어진 공간 정도로 조약하게 정의되어 있다. 이 두 공간을 각각 NF와 FF로 표기하고 각 공간 안에서는 화학물질이 완벽하게 공기와 혼합되어 어디서든 농도가 같다고 가정한다. 작업자는 두 개의 가상공간 사이를 오가며 작업

할 수 있으므로 T시간동안 어느 작업자가 노출될 수 있는 농도는 다음과 같이 정의할 수 있다.

$$C_T = \sum_{j=1}^n (C_{NF,j} + C_{FF,j}) \cdot \Delta_j \quad (6)$$

여기서 j는 각 작업(task)을 나타내며, Δ_j 는 각 작업별 할애시간(fraction of time each task is performed)을 나타낸다. 각 공간별로 오염물질의 농도에 크게 영향을 미치는 인자는 비슷하다. 공통적인 요소는 4가지로 오염물질 방출률(emission of pollutant, ϵ), 오염물질을 다루는 방식(handling, h), 사용 중인 오염물질 관리방법의 효율(the efficiency of local controls, η_{lc}), 착용하고 있는 개인보호구의 효율(the efficiency of personal protective equipment, η_{ppe})이다. 이 네 가지 인자들은 오염원에서 방출된 화학물질에 비례해서 농도를 증가시키거나 감소시키므로 식 (6)을 이용한 공기 중 농도 계산에서는 다음과 같이 승수(multiplier) 형태로 사용된다.

$$C_{NF} = (\epsilon_{i,NF} \cdot h_{NF} \cdot (1 - \eta_{lc,NF}) \cdot t_{a,NF} + \epsilon_p) \cdot (1 - \eta_{ppe}) \quad (7)$$

$$C_{FF} = (\epsilon_{i,FF} \cdot h_{FF} \cdot (1 - \eta_{lc,FF}) \cdot t_{a,FF} + \epsilon_p) \cdot (1 - \eta_{ppe}) \cdot d_{gv} \quad (8)$$

여기서 t_a 는 오염원에서 화학물질이 능동적으로 배출되는 시간(time that the source is actively emitting)이며, d_{gv} 는 희석 혹은 전체 환기(dilution or general ventilation)에 대한 승수로 사용되지 않는 경우 1이 되며 원거리공간에서만 사용된다. 식 (7)과 (8)에서 앞에서 설명한 오염물질 방출(ϵ)이 내재적 방출(intrinsic emission, ϵ_i)과 수동적 방출(passive emission, ϵ_p)의 두 가지로 구분됨을 알 수 있다. 내재적 방출은 화학물질의 고유한 성질에 의해서 일어나는 방출이며, 유기용제의 증기 방출이나 입자상 물질의 비산 등을 예로 들 수 있다. 수동적 방출은 누출(leak)과 같은 예측할 수 없는 상황에서 의해서 일어나는 방출을 의미한다. 오염물질 관리방법의 효율과 개인보호구의 효율은 농도와 역의 상관관계를 보이므로 1에 대한 보수(1-n)의 형태로 사용되고 최댓값은 1이다.

각각의 승수들은 1차적으로 전문가들의 합의에 의해서 결정되며, 그러한 이유로 모델 이름에 ‘subjective’라는 단어가 들어가게 된다. 2차적으로 측정자료와의 비교를 통해 보정하여 사용하기도 하지만 직관적인

이해를 위해 언제나 유효숫자를 한 자리수로 유지한다. 또한 할애시간(Δ_j)과 배출시간(t_a)을 제외한 이 승수들은 연속변수가 아닌 이산변수로 사용된다. 예를 들면 오염물질 관리방법의 효율의 경우 관리하지 않는 경우 1, 국소배기나 다른 관리방법이 사용되고 있는 경우 0.3, 잘 설계되고 유지·보수되는 국소배기나 다른 고효율의 관리방법을 사용하는 경우 0.1을 사용한다. 이들에 대한 좀 더 자세한 설명은 별도의 논문인 tier 1 노출모델 부분에서 설명하였다. SSA 모델을 실측치와 비교하여 본 결과 오염물질의 성질에 따라 예측력에 차이를 보였다. Cherrie & Schneider(1999)는 석면, 톨루엔, MMMF 호흡성분진, 스티렌 측정자료를 가지고 분석하였으며 각 오염물질별로 양 또는 음의 오차를 보였고 다른 정도의 상관성을 보였다.

3) 문헌 조사: 환경보건분야 및 보건의료 분야

환경보건분야 및 보건의료분야에서 베이지안 통계를 연구에 활용한 사례들이 있었다. 이들 연구 중에는 산업보건분야에서 사용되었던 용도인 모집단 모수 혹은 모델 변수들에 대한 역산에 이용하거나 역산된 모델 변수들을 이용하여 새로운 상황에 대한 예측을 하는 경우가 많았다. 산업보건에서 이제껏 사용되지 않은 두 가지 방법인 베이지안 계층적 모형과 성향 점수 분석에 대해서 정리하면 다음과 같다.

• 베이지안 계층적 모형

계층적 모형은 복잡한 구조의 자료를 단순한 확률 모형들로 표현되는 몇 개의 계층적 구조로 분해하여 알고자 하는 모수들(parameter) 이나 관측되지 않는 잠재 변수들(latent variables)에 대한 추론을 가능하게 하는 방법이다. 이러한 계층적 구조를 전통적 통계 방법에서는 변량모형(random effect model)이나 혼합 모형(mixed effect model)에 적용하였으나, 전통적 통계의 개념상 여러 층으로 구성된 계층 모형을 적용하는데 구조적 어려움이 있다. 하지만, 베이지안 관점에서는 베이즈 정리를 확장하여, 복잡한 자료의 구조를 여러 단순한 층들로 각각 단순화하고 그에 맞는 확률 분포 모형(probability distribution model)을 지정할 수 있다는 장점이 있다. 하지만, 이러한 베이지안 계층적 모형은 전통적인 통계 방법에 비해 많은 연산 시간(computing time)을 요구하기 때문에 과거에는 많이 사용되지 않았으나, 최근 컴퓨터의 비약

적인 발전을 통해 이러한 단점을 극복함으로써 보건 분야뿐 아니라 의학이나 환경 연구에서도 많이 사용되고 있다(Gilks et al., 1996; Su et al., 2001; Gronewold & Borsuk, 2010).

• **성향 점수(propensity score) 분석**

어떤 보건 연구에서 치료(treatment)에 의한 피실험자에 대한 효과를 살펴보기 위해 가장 확실한 연구 방법론은 치료군(treatment group)과 대조군(control group)에 피실험자를 무작위(randomization)로 배정할 수 있는 실험적(experimental) 연구이다. 하지만, 보건 연구는 인간을 대상으로 하기 때문에 실험적 방법으로 연구를 수행할 수 없고, 따라서 관찰(observational) 연구를 통해 인과관계를 파악하고자 하는 경우가 많다. 이런 관찰 연구에서 올바른 인과관계를 파악하기 위해서는 치료군과 대조군에 있는 피실험자들의 치료 유무를 제외한 다른 중요한 변수들에 있어서는 서로 비슷해야 한다는 조건이 반드시 만족되어야 한다. 하지만, 치료 유무 외의 다른 관찰된 변수들이 많은 경우, 이러한 변수들 각각에 대해서 치료를 받은 사람과 치료를 받지 않은 사람이 비슷한 지 여부를 비교하는 것은 상당히 어려운 일이다. 이러한 문제를 다루기 위해 제시된 방법이 성향 점수이다. 성향 점수는 처리 유무를 제외한 다른 중요한 변수들의 합수로서, 만약 치료를 받은 사람과 치료를 받지 않은 사람의 성향 점수가 같다면 그 두 사람은 원래의 변수들의 분포가 같다는 균형 성질(balancing property)을 만족하게 되고, 따라서 이 성향 점수를 적절히 반영한 모형은 비록 관찰 연구라 하더라도 인과 관계에서의 치료의 효과를 볼 수 있다.

전통적 통계 방법에서 성향 점수를 반영한 치료 효과의 분석은 우선 성향 점수를 추정한 다음, 몇 개(일반적으로 5개)의 계층(stratification)으로 나누어 모형에 포함시킨 후 치료의 효과를 보았다(Cochran, 1968; Rosenbaum & Rubin, 1983). 하지만, 최근 연구에서 추정된 성향 점수를 사용하는 방법은 성향 점수의 불확실성(uncertainty)을 제대로 반영하지 못하기 때문에 치료 효과를 제대로 추정하지 못 하는 단점이 있음을 보여 주었고, 그 대안으로서 베이저안 방법을 통해 성향 점수의 불확실성을 반영하면서 동시에 처리 효과를 추정하는 방법이 제시되었다 (McCandless et al., 2009). 또한, 베이저안 방법의 장점인 선행 연

구의 결과나 전문가의 의견을 반영하여 여러 개의 성향 점수 모형들 간의 비교가 가능하고, 특히 표본의 수가 적은 경우에 베이저안 성향 점수 분석 방법이 전통적인 분석 방법보다 조금 더 좋은 장점을 가지고 있음을 모의실험과 실제 자료를 통해 보여줬다 (Kaplan & Chen, 2012).

2. **작업환경측정 분야에서 사용되는 통계 방법 및 소프트웨어 개발**

1) **베이저안 방법론과 노출모델의 결합**

개발된 노출 모델과 베이저안 통계를 결합하여 tier 2 모델을 개발하였다. tier 2 모델은 실측치와 모델의 예측치를 결합하여 더 좁은 구간의 신뢰구간을 제시하는 것이 가능하다.

유럽연합에서 개발되어 사용 중인 노출모델 소프트웨어는 크게는 tier 1 모델과 tier 2 모델로 나눌 수 있다. 가장 큰 차이는 측정데이터를 이용하는가의 여부에 달렸다. Tier 1 모델의 경우 사용되는 화학물질과 사용량, 작업환경 등에 기반에서 기존에 측정된 데이터베이스에서 동일한 화학물질이 유사작업에서 사용된 경우에 대한 측정값들을 검색하여 예상되는 노출범위를 제시해준다. Tier 2 모델의 경우 tier 1 모델에서 계산된 값을 사전자료(prior)로 사용하고 해당 공장에서 실제로 측정된 값을 베이저안 업데이트(Bayesian update)를 통해 예측치를 계산한다.

베이저안 업데이트는 우선 주어진 데이터에 대한 우도분포(likelihood distribution)을 계산하고, 이를 tier 1 모델에서 계산된 95% 신뢰구간을 사전확률(prior distribution)로 하여 식 (2)에 따라 product를 계산하였다.

우도분포의 계산에서 기하평균 범위는 데이터의 최대값(y_{max})으로 표준화(y_{max}=1)한 후에 0.001-10의 범위에서 계산하였다. 기하표준편차의 경우 1.05-4의 범위에서 계산하였으며 이는 Hewett et al.(2006)이 사용한 범위와 같다. 이것을 식으로 표현하면 다음과 같다.

$$P(y|\ln G, \ln D) = \int_{G=0.001}^{G=10} \int_{D=1.05}^{D=4} \prod_{j=1}^n pdf(y_j|\ln G, \ln D) d(\ln G) d(\ln D) \quad (9)$$

• **노출모델의 형식**

노출모델을 어떤 플랫폼(platform)으로 만들 것인가에 대한 논의를 위하여 유럽모델 3종과 미국의 베

이지안 통계프로그램을 분석하였다.

- Advanced REACH Tools (ART)

웹을 기반으로 만들어져 인터넷 접속이 가능해야 이용할 수 있다. 웹 기반 모델의 문제는 인터넷 접속이 불가능한 경우 이용할 수 없다는 것과 인터넷 보안이 취약한 경우 회사의 내부정보가 외부로 유출될 수 있다는 점 등을 들 수 있다. 이런 이유로 웹 기반의 도구 사용을 꺼리는 회사가 많다. ART에서는 tier 1 모델과 tier 2 모델을 모두 이용할 수 있다.

- Stoffenmanager

네덜란드에서 주관하여 개발하고 있는 Stoffenmanager도 웹을 기반으로 하고 있어 ART와 문제점을 공유한다. Tier 1 모델과 tier 2 모델 모두 이용가능하다. 지속적인 업데이트로 사용자층을 많이 확보하고 있다.

- ECETOC TRA

엑셀 매크로로 만들어졌으며 현재 Tier 1 모델만 제공한다. 한 동안 업데이트를 하지 않았으나 근래 들어 규칙적으로 업데이트를 하고 있다. 국내에서 Lee et al.(2012)이 ECETOC TRA 일부 변수를 조절하는 방법으로 현지화를 시도하였으나 노출 모델 자체의 개발은 아니었다.

- IH Data Analyst(IHDA)

Hewett et al.(2006)이 개발한 프로그램으로 베이직 안 통계에 기반을 둔 데이터분석을 제공한다. 노출모델은 제공하고 있지 않으며 AIHA Exposure Category에 따라 입력한 데이터가 각 카테고리에 속할 확률을 계산하여 보여준다. 독립적인 프로그램으로 PC에 설치해서 사용해야한다.

모든 프로그램에서 사용되는 알고리즘은 기본적으로 같다고 할 수 있으므로 한 가지 플랫폼으로 개발하는 경우 다른 플랫폼으로의 포팅이 용이하다. 개발상의 편의성을 생각하고 유지 및 배포비용을 고려하여 엑셀에서 매크로(macro)를 이용한 개발을 선택하였다.

2) 엑셀을 이용한 소프트웨어 개발

엑셀의 매크로와 비주얼 베이직 요소를 결합하여 앞에서 개발된 도구들을 스프레드시트 상에 구현하였다. 식 (9)의 구현에 있어서 분석적인 계산이 불가

능하므로 리만 합(Riemann sum)을 이용하여 적분값을 추산하였다. 리만 합(S)은 함수 f(x)에 대하여 적분 구간을 n개로 분할하고 각 지점(i)에서의 함수값과 다음 지점(i-1)까지의 거리를 둘 사이의 한 점($x_{i-1} \leq y_i \leq x_i$)에서의 함수값 f(y_i)을 곱한 것을 모두 합하여 계산한다.

$$S = \sum_{i=1}^n f(y_i)(x_i - x_{i-1}) \quad (10)$$

적분 구간은 로그변환을 한 상태에서 200개의 작은 구간으로 나뉘어져 계산되었다. 200개로 나눈 이유는 엑셀이 필요한 계산을 수행하는데 걸리는 시간이 3초 이내로 걸리도록 경험적으로 결정한 임의의 값이다. pdf 함수는 엑셀에서 제공하는 NORMDIST(x, mu, sigma, cumulative) 함수를 이용하였다.

노출모델에서 자료의 입력은 이산변수의 경우 드롭다운박스(drop-down box)를 이용하여 입력 시 오류로 인한 에러를 최대한 줄이도록 디자인 하였으며 컬러 코딩(color coding)을 통하여 설명문의 배경에 사용된 색을 해당 메뉴에도 사용해 사용법을 쉽게 인지하도록 하였다. 노란색은 연속변수에 파란색은 이산변수에 사용하였다.

3) 엑셀 프로그램 이용 시나리오

• 측정대상 작업 선정

작업환경을 측정할 수 있는 자원이 한정되는 경우 가장 노출농도가 높을 것으로 예상되는 작업을 개발된 프로그램을 통해 선정할 수 있다. 이것이 유럽에서 노출모델을 개발한 본래 목적이기도 하다.

• 보호구 선정

작업을 시작하기 전에 보호구를 미리 준비하기를 원하는 경우 예상되는 농도의 범위를 개발된 엑셀 프로그램을 통해서 예측할 수 있고, 그 범위에 맞는 호흡보호구를 준비할 수 있다. 개발된 노출 모델에서 제시하는 예측범위의 최대값은 신뢰구간 95%의 값이므로 이 수준에 맞추어 보호구를 선정하면 작업자에게 충분한 보호를 제공할 수 있다.

• 새로운 관리방법 적용 시 효과 예측

일부 건설업 작업은 국소배기장치를 설치하는 등의 관리방법을 사용하는 것이 가능하다. 예를 들면

견출 작업 시에 그라인더에 진공집진기를 연결하는 것을 들 수 있다. 이런 방법을 도입하는 경우 어느 정도의 관리효과를 얻을 수 있는지 노출모델을 통해 예측하는 것이 가능하다.

4) 엑셀 스프레드시트

사용자가 사용하는 엑셀 스프레드시트는 tier 1과 tier 2의 두 개의 탭(tab)으로 나누어진다. 하나는 노출모델을 이용하여 측정자료 없이 노출수준을 예측하는 ‘노출모델-Tier 1’ 탭이고, 다른 하나는 측정자료와 노출모델의 결과를 결합하여 예측의 정확도를 높이는 ‘노출모델-Tier 2’ 탭이다.

현재 버전의 엑셀 스프레드시트의 특징 중 하나는 색을 이용한 안내(color coding)이다. 노란색 셀은 사용자가 직접 숫자를 입력해야하는 경우에 사용하고 제일 위에 있는 안내 페이지도 같은 배경색을 사용하였다. 파란색 셀은 사용자가 주어진 목록 중에 선택(drop-down menu)해야 하는 경우이다. 입력해야 하는 각 셀 옆에는 단위가 있고 도움말이 노트 형태로 삽입되어 있다. 인쇄를 하는 경우 한 페이지에 나올 수 있도록 크기를 제한하였다.

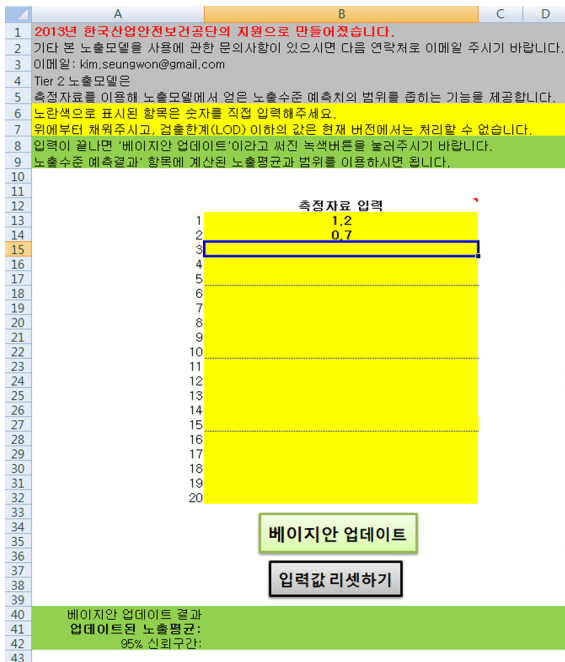


Figure 1. Formation of 'exposure model tier 2' excel spreadsheets tap '노출모델-Tier 1' 탭은 노출모델의 입력값 만으로 예

상 노출평균과 노출범위를 계산한다. 이 탭에서는 매크로가 사용되지 않았다. Figure 1와 같이 ‘노출모델-Tier 2’ 탭에서는 20개까지 자료를 입력할 수 있도록 구성하였으며 데이터 입력 후 초록색의 베이지안 업데이트 버튼을 누르면 ‘노출모델’에서 계산된 노출범위와 측정 자료를 결합하여 더 정교한 예측치를 제시한다. 자료 입력 후 ‘베이지안 업데이트’라는 녹색 버튼을 누르면 업데이트된 노출평균과 95% 신뢰구간이 제시된다. 자료 입력 시에는 비워두는 칸 없이 1번부터 차례대로 채워야한다. 0보다 작거나 같은 값은 입력할 수 없다.

‘노출모델-Tier 2’는 ‘노출모델-Tier 1’을 먼저 입력해야만 계산결과를 얻을 수 있다. 이것은 ‘노출모델-Tier 1’의 결과를 가지고 측정 자료를 이용해서 예측범위를 줄여주기 때문이다. ‘노출모델-Tier 2’ 기능은 매크로 사용을 허용해야 사용할 수 있다. 엑셀 스프레드시트에서 행이나 열을 지우는 경우 제대로 작동하지 않을 수 있다.

IV. 결 론

베이지안 통계기법은 다양한 분야에서 활용되고 있으며 측정치의 수가 적을 때 활용하기 적합한 것으로 조사되었다. Tier 2 모델은 베이지안 업데이트를 주기능으로 구현되었다. Tier 2 모델은 tier 1 모델과 별도의 탭에 구현되었으며 계산을 위하여 측정치와 tier 1 모델의 예측범위를 필요로 한다.

모든 모델은 제한점을 내재하고 있지만 그 한계범위 내에서 활용하는 경우 유용한 도구가 될 수 있다. Tier 2 모델은 tier 1 모델과 함께 측정대상사업장 우선순위 결정, 보호구 선정, 새로운 관리방법 적용시 효과 예측 등에 활용할 수 있을 것으로 기대된다.

감사의 말씀

이 논문은 2013년도 산업안전보건연구원의 위탁연구 용역사업 지원을 받아 수행된 것(2013-연구원-1196)으로 이에 감사를 드립니다.

References

Cherrie JW, Schneider T, Spankie S, Quinn M. A new method for structured, subjective assessments of past

- concentrations. *Occup Hyg* 1996;3:75-83
- Cherrie JW, Schneider T. Validation of a new method for structured subjective assessment of past concentrations. *Ann Occup Hyg* 1999;43(4):235-245
- Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968;24:295-313
- Gilks WR, Richardson S. Analysis of disease risks using ancillary risk factors, with application to job - exposure matrices. *Statist Med* 1992;11(11):1443-1463
- Gronewold AD, Borsuk ME. Improving Water Quality Assessments through a Hierarchical Bayesian Analysis of Variability. *Environ Sci & Tech* 2010;44(20):7858-7864
- Hewett P, Logan P, Mulhausen J, Ramachandran G, Banerjee S. Rating exposure control using bayesian decision analysis. *J Occup Environ Hyg* 2006;3(10):568-581
- Kaplan D, Chen J. A Two-step Bayesian approach for propensity score analysis: simulations and case study. *Psychometrika* 2012;77:581-609
- Lee EG, Kim SW, Feigley CE, Harper M. Exposure models for the prior distribution in Bayesian decision analysis for occupational hygiene decision-making. *J Occup Environ Hyg* 2013;10(2):97-108
- Lee JH, Lee KS, Hong MK. Evaluation of the Application of a European Chemical Risk Assessment Tool in Korea. *J Korean Soc Occup Environ Hyg* 2012;22(3):191-199
- McCandless LC, Gustafson P, Austin PC. Bayesian propensity score analysis for observational data. *Statistics in Medicine* 2009;28:94-112
- Nicas M, Jayjock M. Uncertainty in exposure estimates made by modeling versus monitoring. *AIHA J* 2002;63(3):275-283
- Ramachandran G, Kandlikar M. Bayesian analysis for inversion of aerosol size distribution data. *J Aerosol Sci* 1996; 27(7):1099-1112
- Ramachandran G. Restrospective expoosure assessment using Bayesian methods. *Ann Occup Hyg* 2001;45(8):651-667
- Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 1983;70:41-55
- Su ZM, Adkison MD, Van Alen BW. A hierarchical Bayesian model for estimating historical salmon escapement and escapement timing. *Can J Fisher and Aquat Sci* 2001;58(8):1648-1662
- Vadali M, Ramachandran G, Mulhausen J. Exposure modeling in occupational hygiene decision making. *J Occup Environ Hyg* 2009;6(6):353-362