

사회 정책 계획 또는 평가를 위한 노르웨이 행정 데이터 활용*

Using Norwegian Administrative Data for Social Policy Planning
or Evaluation

로위 닐센(노동 및 사회연구를 위한 Fafo 연구소)
Roy A. Nielsen(Fafo Institute for Labour and Social Research)

이 글에서는 개인 고유식별번호로 연결되는 노르웨이 행정등록부 데이터(Norwegian administrative register data)를 소개하고 이 데이터가 다양한 분야에서 어떻게 활용되는지를 살펴보았다. 다양한 정보로부터 오는 노르웨이 행정등록부 데이터는 전체 인구를 대상으로 하는 데이터로서 개인의 장기간 이력을 추적하는 데 사용될 수 있다. 이 글에서는 사회 정책 연구에 활용된 몇 가지 사례를 제시하고자 한다.

1. 들어가며

노르웨이에서는 사람이 태어나면 개인고유식별번호(PIN: Personal Identification Number)가 부여되고 중앙인구등록부(CPR: Central Population Register)에 추가된다. 성장하면서 개인고유식별번호(PIN)이 국가 교육 데이터베이스(National Education Database)에 입력되며, 직업을 갖게 되어 노동시장에 진입하면 고용주와 함께 피고용인으로 등록된다. 대부분의 국가, 특히 북유럽 국가에서는 통계 및 행정 등 다양한 목적을 위해 광범위한 행정등록 시스템이 구축되었다. 행정등록 시스템은 인구 규모 및 위치, 학력 수준, 노동력 규모, 세금과 급여 수준 등을 계산하는 수단으로 활용된다.

각 등록부(register)에 대해서는 연구자들이 그다지 관심을 두지 않지만, 여러 가지 다른 등록

* 영문으로 작성된 원고로 원문은 한국보건사회연구원 영문홈페이지 참고(<https://www.kihasa.re.kr/english/main.do>)

부의 정보가 서로 연결되면 다양하고 흥미로운 연구들에 이용될 수 있다. 노르웨이와 다른 북유럽 국가에서는 서로 다른 분야를 다루는 등록부 간의 연결이 가능하기 때문에 북유럽의 데이터 금광(NordForsk, 2014; Tupasela, Snell, & Tarkkala, 2020)이라고 불릴 정도다. 지난 수십 년간 컴퓨터 사용 기술 향상과 함께 더욱더 많은 양의 데이터를 활용 수 있게 되었고, 연구 경험과 지식이 축적되었으며, 통계 기법 또한 발전하였다.

이 글에서는 노르웨이 행정등록부의 몇 가지 특성과 자료 활용 방법에 대해 논의하고자 하며, 공공 정책에 대한 조사와 평가에 어떻게 활용되어 왔는지 사례를 통해 살펴보고자 한다. 이러한 논의가 북유럽 국가에 관심이 있는 연구자들에게 유의미한 결과를 제공할 것으로 기대한다.

2. 노르웨이 행정등록 데이터

노르웨이 공식 통계의 대부분은 행정등록부상의 개인정보를 기반으로 한다. 개인을 식별하는 데는 대부분 개인고유식별번호(PIN)이 사용된다. 1964년 노르웨이에 도입된 개인고유식별번호(PIN)는 중앙인구등록부(CPR)에 등록되는데, 이를 통해 개인의 출생, 혼인 상태, 사망¹⁾ 등을 추적할 수 있다. 개인고유식별번호(PIN)를 통해 부모-자녀 관계를 포함한 가족관계와 세대를 식별할 수도 있다. 개인고유식별번호(PIN)는 대부분의 노르웨이 등록부에서 개인을 식별하는 데 사용되며, 서로 다른 등록부의 정보를 이용할 수도 있다. 따라서 그 출처가 다양하더라도 일반적으로 인구통계, 학력, 직업 이력, 소득, 세금 및 혜택에 대한 개별 정보를 함께 이용할 수 있다. 또한 가능한 경우에는 의료, 법 집행 및 군 이력 등과 관련한 다양한 정보를 추가로 이용할 수 있다.

피고용인은 회사고유번호(CIN: Corporate Identity Number)를 통해 고용주와 연결되며 직업, 직종, 및 직장 위치까지 식별할 수 있다. 직장 위치는 직장 활동의 지리적 변화를 측정하고 통근자를 파악하는 데도 사용된다. 더불어 거주지와 직장 주소를 연계해 집에서 직장까지의 거리를 측정할 수 있게 된다.

대부분의 등록부는 노르웨이 통계청(Statistics Norway)에 주기적으로 전달되어 국가 통계를 업

1) 사망 원인은 노르웨이 공중보건연구원(Norwegian Institute of Public Health)의 별도 등록부에 기록

데이트하는 데 사용된다. 전달된 등록부는 오류를 줄이기 위해 공개 전에 노르웨이 통계청의 품질보증 시스템을 거치게 된다. 개별 등록부의 신뢰도나 등록부 내 등록된 정보의 신뢰도는 정보의 중요도와 활용도에 따라 달라진다. 세금 기록은 개인이 납부할 세액을 결정하는 근거가 되기 때문에 매우 신뢰할 수 있으며, 암 이력과 같이 연구를 위해 수집된 등록 정보 또한 신뢰할 만하다. 신뢰도가 다소 떨어지는 데이터의 흔한 예는 고용 계약 종료일이다. 연내 다른 날짜 또는 이른 날짜에 계약이 종료됨에도 불구하고 일부 피고용인들은 관례적으로 계약 종료일을 12월 31일로 기입하기 때문이다. 개별적인 고용관계에서 계약 종료일은 크게 중요한 것이 아니기 때문에 엄격한 품질보증을 적용받지 않았으나, 2015년부터 고용 기록을 등록하는 새로운 방식을 도입함으로써 이 문제를 해결하고자 했다. 등록부 데이터의 또 다른 문제는 등록 절차의 변경으로, 데이터에 공백이나 파손을 일으킬 수 있어 데이터를 호환할 수 없는 경우가 발생한다.

이러한 약점에도 불구하고 등록부 데이터는 모든 개인의 정보를 포괄하고 있으며, 없어도 않고 정기적으로 업데이트되기 때문에 매우 신뢰할 수 있는 출처로 간주된다. 특히 비용이나 더 긴 기간 데이터를 사용할 수 있다는 사실을 고려하면, 등록부 데이터는 설문조사 데이터에 비해 장점이 있다. 그러나 등록부 데이터는, 예를 들어 고용, 실업, 비노동인구 등 '상태(status)' 정보만을 날짜 정보와 함께 제공하며 설문조사를 통해 알 수 있는 신념이나 태도, 동기 등의 정보는 포함하지 않는다는 한계도 있다. 등록부 데이터는 인구조사나 특정 하부 모집단 등 설문조사 모집단을 규정하는 데 활용될 수 있다. 등록부 데이터가 표본 추출을 위해 사용된 경우에는 무응답 편향(non-response bias)을 분석하는 데도 활용될 수 있다. 나아가 활용 가능한 등록부 데이터와 응답자를 연결함으로써 설문 문항을 줄일 수 있으며 응답률도 높일 수 있다. 등록부 데이터를 설문조사와 연계한다면, 설문조사 전과 후를 모두 포함하는 데이터를 활용할 수 있을 것이다. 설문조사가 기업과 같은 조직을 대상으로 할 경우에는 회사 고유번호(CIN)를 통해 연계된 등록부 데이터를 활용할 수도 있다.

그러나 각 등록부가 갖는 행정목적이 따로 있기 때문에 분석적 목적이 있다 하더라도 항상 사용 가능한 것은 아니다. 일부 등록부는 매일 업데이트되는 반면, 어떤 등록부는 연간 정보만을 취급한다. 연구 주제나 데이터 유형과 수는 연구 대상의 기간 설정이나 분석 설계를 제한할 수 있어 연구 계획을 세울 때 반드시 고려해야 한다. 어떠한 데이터 출처나 기간 설정이

든 등록부 데이터는 반복적인 횡단면 분석(cross sectional analysis)이나 사건사 분석(event history analysis), 패널 분석(panel analysis)이 가능하다. 또한 적절히 설계하기만 한다면 등록부 데이터를 사용하여 인과 추론도 할 수 있다.

등록부 데이터를 활용하여 분석하기 위해서는 등록부 데이터에 대한 자료 활용 허가를 받아야 한다. 등록부 데이터 자료 활용에 대해서는 법으로 규제하고 있다. 2018년부터 대부분의 유럽 국가들[적어도 유럽연합(EU)이나 유럽경제공동체(EEC) 국가들]은 일반 데이터 보호 규정(GDPR: General Data Protection Regulation)을 도입했다. 노르웨이에서는 인증받은 연구기관(종합대학, 단과대학, 연구소 등)의 연구자들만이 등록부 데이터를 활용할 수 있다. 이들 연구기관은 노르웨이 연구데이터센터(NSD: Norwegian Centre for Research Data)와 협력하여 데이터 보호 영향평가(DPIA: Data Protection Impact Assessment) 방안을 마련했다. 그 내용에 따르면, 연구 제안은 해당 법률과 규정을 따라야 하고 연구 질문을 조사하는 데 필요한 만큼의 데이터만을 사용해야 하며 데이터에 포함된 개인정보를 활용함으로써 생기는 불이익보다 연구로 인해 발생하는 혜택이 더 커야 한다는 것이다. 또한 연구자들은 데이터 저장 방식이 데이터 무단 활용을 충분히 제어할 수 있다는 점을 문서화해야 한다. 매우 민감한 데이터의 경우에는 노르웨이 정보보호국(Norwegian Data Protection Authority)과 지역 윤리위원회(ethics committees)가 데이터 사용 승인에 관여하기도 한다. 또한, 연구자들은 노르웨이 통계청 또는 각 등록부의 소유자, 예를 들어 노르웨이 노동복지청(NAV: Norwegian Labour and Welfare Administration), 노르웨이 건강국(Norwegian Directorate of Health) 등에 데이터 활용 권한 신청을 해야 한다. 지금까지 연구에 사용되는 대부분의 등록부 데이터는 노르웨이 통계청이 관리해 왔으며, 다른 출처의 등록부 데이터가 사용될 경우 일반적으로 노르웨이 통계청 사무소로 연결된다.

일반 데이터 보호 규정(GDPR)의 도입은 데이터 활용에 소요되는 시간을 줄였지만, 여전히 시간 소모적이다. 데이터 활용에 필요한 이용자 수인 등록부 소유자 수는 프로젝트 제안부터 데이터 수령까지의 효율적인 절차에 장애가 되는 것으로 나타났다(Van Der Wel et al., 2019). 더불어 노르웨이에서는 등록부 데이터 활용 비용이 상당히 비싸다. 그러나 이용 시간과 데이터 가격 모두 프로젝트에 포함된 데이터 소스의 수와 변수의 종류에 따라 상이하다.

이상적으로는 데이터는 익명이어야 한다. 즉 자료에서 개인을 식별할 수 없어야 한다. 이를 위해 첫째, 변수의 수, 즉 연구자에게 전달되기 전에 변수들을 분류함으로써 데이터를 최

소화해야 한다. 둘째, 일반적으로 연구자들은 개인의 고유식별번호(PIN)에 접근할 수 없는데, 노르웨이 통계청으로부터 연구자들에게 전달되는 연계 데이터에서 개인고유식별번호(PIN)은 개인의 익명성을 보장하기 위해 무작위 아이디(ID) 번호로 대체된다. 마지막으로, 데이터 전달에 앞서 연구자들은 비공개 서약에 서명해야 한다.

지난 몇 년간 노르웨이 통계청은 노르웨이 연구데이터센터(NSD)와 협력하여 연구자와 학생이 사용할 수 있는 익명의 등록부 데이터에 대한 웹 기반 액세스를 개발했다. 지금까지는 이용 가능한 데이터와 통계 기법에 한계가 있었지만 이를 확대하고자 노력하고 있다("About microdata.no," n.d.). 활용 가능한 등록부 데이터에 대한 개요는 노르웨이 통계청 홈페이지에서 확인할 수 있다("Data for research," 2018).

아래에서는 등록부 데이터가 활용된 사회 정책 연구 사례 몇 가지를 제시하고자 한다. 경제 연구에서 노르웨이 등록부 데이터에 관심이 더 있는 연구자들은 Røed와 Raaum(2003)의 연구를, 범죄학에 관심 있는 연구자들은 Lyngstad와 Skardhamar(2011)의 연구를 살펴볼 수 있다. 노르웨이 보건등록부에 대한 개요는 노르웨이 공중보건연구원에서 찾아볼 수 있다("Overview of the national health registries," 2020).

3. 일용직 일자리와 이민자 통합 간 연관성 연구

Kavli와 Nielsen(2019)은 고용 기록을 활용한 분석을 통해 불안정한 직종으로 여겨지는 일용직 일자리가 이민자들 사이에서 정규직이 되는 디딤돌이었는지, 아니면 종착역이었는지를 조사하고 있다. 시작일과 종료일, 근로계약 시간 변경일 등의 고용 기록을 포함한 등록부 데이터가 사용되었기 때문에 분석에는 당시 노르웨이의 거의 모든 일용직 근로자가 포함되었으며, 근무시간과 퇴직 날짜의 변화를 파악할 수 있었다. 고용 기록은 노르웨이 노동복지청 고용주 및 고용자 국가 등록부(NAV State Register of Employers and Employees)와 급여 및 세금 증명서 등록부(Register of Certificates of Pay and Tax)의 연계를 통해 확인할 수 있었다. 또한, 데이터에는 국가 교육 데이터베이스의 인구통계(중앙인구등록부(CPR)에서의 이민자의 출신국 포함)와 교육 수준에 대한 등록 정보도 포함되었다. 5년간의 추적을 통한 경쟁위험 이력분석(risk event history analyses)을 통해 그들이 일용직 근무에서 정규직 또는 노동시장 이탈로 전환하는 것을 비교 분석하였다. 분석 결과, 일반적으로 이민자가 비이민자에 비해 정규직으로 전환될 가능성이 같

거나 더 높은 것으로 나타났다. 그러나 근무시간이 짧으면, 즉 근무시간이 주당 15시간 미만인 경우, 노동시장 이탈 위험이 증가한다는 것 또한 발견했다. 결론적으로, 이 연구는 이민자들에게 근무시간이 짧은 일용직 일자리의 경우 종착역이 되기도 하지만, 장시간 근무하는 일자리는 정규직을 위한 디딤돌이 된다는 것을 밝혔다.

4. 40세 이상 정규교육 이수자의 노동시장 참여 연구

전 세계 인구고령화를 배경으로 하는 Midtsundstad와 Nielsen(2019)의 논문에서는 40세 이후 정규교육을 이수하는 것이 이후 노동시장 지위에 영향을 미치는지를 조사하고 있다. 분석 자료에는 2003~2013년 40~54세 노르웨이 거주자 전원이 포함되어 있다. 이전보다 더 높은 교육 수준의 정규교육을 받은 40세 이상의 개인을 식별하기 위해 노르웨이 통계청의 국가 교육 데이터베이스를 활용하였다. 노동시장 참여는 근로 계약에 따른 연간 일수로 측정하고 병가 일수를 차감하여 이른바 ‘실제’ 또는 적극적인 노동시장 참여도를 측정하였다. 고용 기록 및 근로시간은 노르웨이 노동복지청 고용주 및 고용자 국가 등록부와 급여 및 세금 증명서 등록부의 연계를 통해 확인할 수 있었다. 병가 일수는 노르웨이 노동복지청(NAV)의 질병수당 등록부에서 확인할 수 있었다. 정규교육이 노동시장 참여 변화의 원인인지 알아보기 위해 고정 효과 회귀 분석(fixed-effects regression analyses)을 사용하였다. 분석 결과는 정규교육의 수준을 높이는 것이 이후 노동시장 참여에 상당한 영향을 미친다는 것을 보여 준다. 중등교육 이수자 남성들에게 강력하고 장기적인 영향을 미치는 것으로 나타난 반면, 낮은 수준의 3차 학위 취득은 남성과 여성 모두에게 가장 큰 영향을 미치는 것으로 나타났다. 이 논문은 노년층 근로자의 정규교육에 대한 접근을 용이하게 하는 것이 노동 생활 확대, 즉 일하는 삶을 연장하는 데 기여할 수 있다고 결론짓고 있다.

5. 사업장의 건강 문제 개입과 직장 생활 연장 간 연관성 연구

Nielsen과 Midtsundstad(2020)의 최근 연구는 ‘건강에 문제가 있고 업무 능력이 저하된 직원들의 근무 기간 연장을 촉진하기 위해 특별 조치’를 제공하는 사업장에 취업하는 것이 전반적인 병가 및 장애 위험에 영향을 미쳤는지 조사하고 있다. 연구 모집단은 2000년부터 2010년까지 등록부 데이터의 모든 피고용인과 연계된 2010년 설문조사 참여 사업장으로 구

성되었다. 등록부 데이터에는 질병 결근, 근로 장애, 성별, 교육 및 연령에 대한 정보가 포함되어 있었다. 기업 등록부(Register of Business Enterprise)에 있는 회사 아이디(ID)는 제도와 피고용인 연계를 위해 사용되었다. 질병 및 장애 휴가는 노르웨이 노동복지청(NAV) 등록부 데이터를 통해 확인하였다. 추가로 인구통계적 변수는 중앙인구등록부(CPR)에서, 교육 변수는 노르웨이 통계청의 국가 교육 데이터베이스에서 추출하였다. 개별 고정 효과와 함께 선형확률 회귀분석(linear probability regression)을 이용하여 사업장의 개입이 병가나 장애의 변화를 유발하는지를 조사하였다. 분석 결과, 유사한 데이터로 진행했던 초기 연구(Midtsundstad & Nielsen, 2016)를 입증하듯 화이트칼라 근로자, 특히 여성 근로자 사이에서 장애 위험이 어느 정도 감소한 것으로 나타났다. 그러나 장애 감소는 질병 결근 위험의 증가와 함께 나타나 일부 근로자들의 장시간 근무가 무단결근자를 증가시킬 수 있음을 시사한다. 이 논문은 사업장에서의 건강 증진이 일부 근로자에게는 근무 기간을 연장시킬 수 있지만, 노르웨이 사업장들이 이러한 개입을 계속하기 위해서는 재정적 인센티브가 필요할 수 있다고 결론짓고 있다.

6. 장기적인 사회보장, 건강 및 소득 변화 궤적 연구

사회보장 혜택을 장기간 받는 것은 사회에서 다소 취약한 위치에 있음을 나타낸다. 최근 논문에서 Løyland, Hermansen, Dahl과 Wahl(2020)은 장기 사회보장 수급자 집단의 개인에게 고통과 심리적 스트레스가 각각 그리고 둘이 결합하여 9년에 걸친 소득 변화 궤적과 어떻게 연관되는지를 조사하였다. 연구에서는 개인고유식별번호(PIN)와 노르웨이 통계청의 등록부 데이터에 연계된 2005년 서베이 데이터, 노르웨이 노동복지청(NAV), 사망 사유 등록부를 결합하였다. 이 연구에는 장기 사회보장 수급자 456명만이 포함되지만, 취약계층에 대한 설문조사와 등록부 데이터를 결합하여 소득 변화 궤적을 그룹화하기 위한 순차 분석(sequence analysis)과 다항로지스틱 회귀분석(multinomial logistic regression analysis)을 함께 이용한다는 점에서 흥미롭다. 분석 결과, 장기 사회보장 수급자들은 건강 문제로 더 큰 부담을 느끼고 있으며, 이는 그들의 삶의 기회와 소득 변화 궤적에도 영향을 미치는 것으로 나타났다.

7. 나가며

이 글에서는 개인고유식별번호(PIN)로 연결되는 노르웨이 행정등록부 데이터를 소개하고 이

데이터가 다양한 분야에서 어떻게 활용되는지를 살펴보았다. 등록부 데이터는 전체 인구를 대상으로 하는 데이터로서 축적되어 개인의 장기간 이력을 추적하는 데 사용될 수 있다. 이러한 데이터는 활용이 제한되어 있기는 하지만, 권한 승인을 통해 가능한 한 많은 분석과 연구가 수행될 수 있다.

참고문헌

- About microdata.no. (n.d.). Retrieved from <https://microdata.no/en/about/#background> on 2020, September 2
- Data for research. (2018, September 13). Retrieved from <https://www.ssb.no/en/omssb/tjenester-og-verktoy/data-til-forskning> on 2020-09-02
- Kavli, H. C., & Nielsen, R. A. (2019). Stepping in, stepping out or staying put? Part-time work and immigrant integration in Norway. In H. Nicolaisen, H. C. Kavli, & R. S. Jensen (Eds.), *Dualisation of part time work. The development of labour market insiders and outsiders*. Bristol: Policy Press.
- Løyland, B., Hermansen, Å., Dahl, E., & Wahl, A. K. (2020). Differences in income trajectories according to psychological distress and pain: A longitudinal study among Norwegian social assistance recipients. *Scandinavian Journal of Public Health*, 1403494820939812. doi:10.1177/1403494820939812
- Lyngstad, T. H., & Skardhamar, T. (2011). Nordic Register Data and Their Untapped Potential for Criminological Knowledge. *Crime and Justice*, 40, 613–645. doi:10.1086/658881
- Midtsundstad, T., & Nielsen, R. A. (2016). Do workplace interventions reduce disability rates? *Occupational Medicine*, 66(9), 691–697. doi:<https://doi.org/10.1093/occmed/kqw169>
- Midtsundstad, T., & Nielsen, R. A. (2019). Lifelong learning and the continued participation of older Norwegian adults in employment. *European Journal of Education*, 54(1), 48–59. doi:10.1111/ejed.12322
- Nielsen, R. A., & Midtsundstad, T. I. (2020). Do workplace health-promotion interventions targeting employees with poor health reduce sick-leave probability and disability rates? *Scandinavian Journal of Public Health*, 1403494820946543. doi:10.1177/1403494820946543
- NordForsk. (2014). Joint Nordic Registers and Biobanks – A goldmine for health and welfare research. In *NordForsk Policy Paper 5 -2014*. Oslo: Norden.
- Overview of the national health registries. (2020, September 22). Retrieved from <https://www.fhi.no/en/more/access-to-data/about-the-national-health-registries2/> on 2020, September 2
- Røed, K., & Raam, O. (2003). Administrative registers – Unexplored reservoirs of Scientific Knowledge? *The Economic Journal*, 113(488), F258–F281. doi:10.1111/1468-0297.00134
- Tupasela, A., Snell, K., & Tarkkala, H. (2020). The Nordic data imaginary. *Big Data & Society*, 7(1), 2053951720907107. doi:10.1177/2053951720907107
- Van Der Wel, K. A., Östergren, O., Lundberg, O., Korhonen, K., Martikainen, P., Andersen, A.-M. N., & Urhoj, S. K. (2019). A gold mine, but still no Klondike: Nordic register data in health inequalities research. *Scandinavian Journal of Public Health*, 47(6), 618–630. doi:10.1177/1403494819858046