

# 한국 병적기록표 데이터베이스의 구축과 활용<sup>(1)</sup>

이철희 · 도민영 · 박순영 · 홍두승

이 논문은 병무청 병적기록표 자료의 성격과 최근 이루어진 이 자료의 데이터베이스 구축 내용을 소개함으로써 이 자료의 활용 및 유사한 자료의 수집과 관련된 시사점을 제공하였다. 또한 이 데이터에 포함된 주요 변수를 이용하여 1946년부터 1957년까지 출생한 남성의 개인적인 특성을 개관하고 자료의 활용가능성을 모색하였다. 한국의 병무기록은 장기간에 걸쳐 축적된 행정자료로서 정확성과 대표성이 높으며, 특히 개인별 마이크로 자료가 본격적으로 생산되기 이전 시기의 개인의 사회경제적 특성과 후생상태를 연구하는데 있어서 유용성이 큰 자료로 평가된다. 이번 데이터베이스 구축에서 제외된 1957년 이후 출생자 표본을 추가하고 군복무 기록에 대한 전산입력이 이루어지면 자료의 유용성이 더욱 높아질 것으로 기대된다.

주제어: 병적자료, 데이터베이스, 사회경제적 특성, 후생수준

## 1. 머리말

이 글의 목적은 병무청 병적기록표 자료의 성격을 소개하고, 최근 이루어진 이 자료의 데이터베이스 구축 내용을 설명함으로써 추후 이 자료의 활용과 유사한 자료의 수집에 도움을 주기 위한 것이다. 또한 이 데이터에 포함된 변수를 활용하여 1946년부터 1957년까지 출생한 남성의 개인적인 특성을 개관하고 자료의 활용가능성을 모색하고자 한다.

많은 국가에 있어서 병무자료는 청년기 남성의 개인적인 특성에 대해 장기간에 걸쳐 일관된 정보를 제공해주는 대표적인 자료 가운데 하나이다. 특히 신장이나 체중과 같은 영양상태 내지 후생의 지표의 장기적인 변화를 분석한 선행연구에서 병적자

---

(1) 이 연구는 2014년 서울대학교 기초학문 교육 및 연구지원사업 사회과학분야 융복합 연구과제 지원을 받아 수행되었다. 연구자들은 병적기록부 표본 데이터베이스 구축을 허락해주시고 직접 이미지파일을 생성하여 제공해주신 병무청 관계자들에게 깊은 감사의 뜻을 전한다. 이 과정에는 그 분들의 이해와 협조가 없었다면 불가능했을 것이다. 또한 데이터 입력 작업을 맡아준 서울대학교 대학원의 이용수, 홍석기, 홍혜영 제씨와 서울대학교 학사과정의 권순현, 김혜지, 박종령, 서민교, 안상현, 정다운, 정다정, 정수영, 채병진 제씨에게 고마운 마음을 전한다.

료는 가장 중요한 사료로 이용된 바 있다. 예컨대 19세기 산업화 시기 동안 일인당 소득이나 임금은 상승했지만 평균 신장이 감소했다는 것을 밝힌 연구결과는 영국과 미국의 병무자료를 이용하여 얻은 것이다[Fogel(1986); Floud, Wachter, and Gregory (1990)].

병무자료를 퇴역군인 연금자료나 인구센서스 원사료와 같은 다른 데이터에 연결하여 분석하는 연구도 활발하게 진행되었다. 예컨대 남북전쟁 참전군인의 군복무기록을 퇴역군인 연금기록, 연금신청을 위한 건강검진기록, 여러 연도의 인구센서스 등에 연결하여 한 개인의 일생을 재구성한 종단자료가 구축되었으며[Fogel(1993); Wimmer(2003)], 이 데이터는 건강과 사회경제적 지위 간의 동태적인 상호작용 [Lee(1997, 2005, 2008)], 전쟁 중 군복무 경험이 경제적 이동성에 미친 영향[Costa and Kahn(2008); Lee(2007)], 고령자의 퇴직과 경제적 지위[Costa(1998); Lee(2005, 2015)] 등 다양한 주제에 대한 연구에 활용된 바 있다.

다른 국가의 경우와 마찬가지로 한국의 병무기록도 다음과 같은 점에서 상당한 가치가 있는 자료이다. 첫째, 장기간에 걸쳐 축적된 정부의 중요한 행정자료이며, 명확한 목적 하에 작성되었으므로 조사항목에 일관성이 있다. 즉 장기적으로 일관된 자료의 구축이 가능하다. 둘째, 병적기록표 기재사항에 허위의 사실이 없음을 확인하였다는 문구 아래 관할 시군구청장의 직인을 날인한 데서 알 수 있듯이 거짓보고의 위험성이 낮다. 특히 학력의 경우 위조여부를 별도로 검토하고, 신체이상상황을 기록한 군의관의 직위 및 성명을 명기하기 때문에 다른 서베이 자료에 비해 보고의 정확성이 높다고 사료된다. 셋째, 병적기록표에는 징집대상자 및 가족의 인적사항, 인사분류검사 및 적성검사 결과, 신체검사 결과가 함께 기재되어 있다. 이는 다른 자료로부터는 동시에 얻기 어려운 정보이다.

병적기록자료는 특히 개인별 마이크로 자료가 본격적으로 생산되기 이전 시기의 개인의 사회경제적 특성과 후생상태를 연구하는데 있어서 유용성이 큰 자료이다. 과거 시기의 후생의 지표로 널리 이용되는 신장과 몸무게를 대표성이 있는 표본에 대해 제공하는 자료는 1990년대 이후에 대해 얻을 수 있다. 그 이전 기간에 대해서는 특정한 집단(예컨대 학생, 죄수, 행려사망자 등)에 대한 자료가 있을 뿐이다[최성진(2006); Choi and Schwekendiek(2009); Kim and Park(2011)]. 근래의 데이터를 이용하여 과거 출생코호트의 신장을 분석하는 경우 선택적인 생존으로 인한 편이나 연령에 따른 신장축소의 문제에 직면하게 된다. 또한 근래의 마이크로 자료들도 성장기의 환경을

결정할 수 있는 요인에 대한 정보(예컨대 성장기 부모의 직업)를 충분히 제공하지 못한다. 반면 병적기록자료는 대표성이 있는 청년기 남성의 개인적 특성 및 신장, 체중, 건강검진결과 등 후생지표는 물론 20세 무렵 부모의 특성에 대한 정보를 제공해준다.

## 2. 병적기록표 원자료와 구축과정

### 2.1. 표본추출과 자료 확보

본 연구팀은 병무청 담당자들과의 회의와 조사를 통해 2002년 이후에 징병검사를 받은 장정들부터는 병적기록이 전산화되어 있으며, 그 이전 징병검사자의 경우 병적기록표가 이미지파일이나 마이크로필름 형태로 보관되어 있다는 것을 확인하였다. 그리고 각 출생코호트별 병적기록표 수를 확인한 결과 1960년대 초 이후의 병적기록표 수는 대체로 징병검사 대상이 되는 전체 남성인구의 규모와 유사하다는 것을 확인하였다. 이러한 정보를 기초로 하여 1960년대 중반 이후 징병검사를 받은 1946년생 이후 출생자를 표본추출의 대상으로 정하였다.

애초의 계획은 전산자료에 포함되지 않은 1946년~1981년 출생자 전체를 대상으로 하여 2% 무작위표본을 추출하는 것이었다. 그러나 여러 가지 제약 때문에 우선적으로 1946년~1957년 출생자의 0.5% 무작위 표본을 추출하는 것으로 결정하였다. 해방 직후 출생한 코호트에 대한 자료를 수집하기로 한 것은 통계자료가 상대적으로 부족한 1960년대 이전 시기 출생자에 대해 병무기록이 갖는 가치가 더 클 것으로 판단하였기 때문이다.

표본의 추출은 자료의 형태가 상이한 1971년 이후 전역자(면제자 포함)와 이전 전역자(면제자 포함)에 대해 각기 다른 방법으로 수행하였다. 1971년 이후 전역자들(대체로 1968년 이후 입대한 1948년 이후 출생자들)에 대해서는 병적기록표의 이미지파일이 전산화되어 있다. 이 경우에는 1946년생~1957년생을 대상으로 매월 20일 출생자(주민번호 앞부분 끝자리 수가 20으로 끝나는 전역자)를 선별하고, 이들 가운데 주민등록번호 뒷부분 끝자리 수가 4 혹은 9인 사람들을 추출하여 대략 1/150 표본을 얻었다. 그리고 이들 가운데 기록이 완전하지 않거나 식별하기 어려운 사람들(이미지가 흐리고 일부 지워진 경우, 알아보기 어려운 한자 혹은 흘림체로 기록된 경우)을 제외하였다.

1966년~1970년 전역자(대부분 1946년~1948년 출생자)의 경우 각 연도 마이크로

필름 롤(3,000명 정도 포함)의 맨 앞에 위치한 15명을 추출하여 징병검사 및 군복무 기록을 프린트하였다. 그리고 맨 앞의 15명 가운데 기록이 완전하지 않거나 식별하기 어려운 사람(이미지가 흐리고 일부 지워진 경우, 알아보기 어려운 한자 혹은 흘림체로 기록된 경우)이 있는 경우 이 사람을 건너 뛰어 그 다음 사람의 기록을 추출하였다. 1946년~1950년 출생자의 경우 전산화된 이미지 파일과 마이크로필름 등 두 가지 출처로부터 추출되어 통합되었다.

이렇게 추출된 표본 병적기록표에는 개인의 식별을 가능하게 해주는 민감한 정보가 포함되어 있다. 개인정보 보호를 위해 병무청과의 협의를 통해 본인 성명, 군번, 부모(세대주) 성명, 주민번호 뒤의 6자리, 주소 중 읍·면·동 이하의 정보 등 다섯 가지 항목을 제거하였다. 최종적으로는 개인식별정보가 제거된 병적기록표가 이미지 파일로 생성되어 연구자들에게 전달되었다.

## 2.2. 원자료의 형식 구분

원자료인 병적기록표의 형식은 시간이 지나면서 변화하였다. 본 연구팀이 얻은 병적기록표 표본에서는 총 9개의 상이한 형식이 발견되었다. 이 중 네 가지 형식의 병적기록표가 대부분이며, 나머지 다섯은 극히 소수이다. 표본의 대부분을 차지하는 네 가지 형식을 등장한지 오래된 순서대로 포맷1, 포맷2, 포맷3, 포맷4 등으로 명명하였으며, 나머지 형태도 등장한지 오래된 순서대로 포맷5부터 포맷9로 명명하였다.<sup>(2)</sup> 포맷1부터 포맷4의 형태는 부록에 제시되어 있다. 자주 등장하는 네 개의 포맷에 한정하여 각 포맷별 수록 정보를 살펴보면 <표 1>에 제시된 바와 같다.

## 2.3. 원자료의 DB화

병무청으로부터 CD로 제공받은 원자료는 병적기록표 작성대상의 생년을 기준으로 분류된 12개의 폴더 안에 그림파일(png 파일) 형태로 저장되어 있었다. 그림파일의 제목은 '0001'부터 시작되는 네 자리 숫자이고, 제목이 홀수인 그림파일과 그 다음의 짝수 제목인 그림파일은 각각 동일한 병적기록표의 앞면과 뒷면에 대한 스캔본인 것이 원칙이었다. 본 연구팀의 DB화 첫 업무는 전술한 원칙에 어긋나는 제목을 수정하는 일이었다. 그림파일의 제목은 향후 개인식별번호(personal identification number,

(2) 병적기록표 우측 하단에 해당 형식 병적기록표 사용을 승인한 연월일이 기재되어 있어, 이것으로 병적기록표들의 등장 순서를 판단할 수 있다.

〈표 1〉 포맷별 수록정보

구분	수록정보	포맷1	포맷2	포맷3	포맷4
병무행정	일련번호	○	○	○	○
인적사항	성명, 생년월일	○	○	○	○
	주민등록번호			○	○
	본적지, 거주지	○	○	○	○
	최기경찰서와 최기우체국		○		
	호주 또는 친권자	○	○	○	○
	명령수령인		○	○	○
	가족사항	○	○	○	○
	재산	○	○		○
	소득수준 또는 재산세등급	○	○		
	종교	○	○	○	
	취미	○	○	○	
	특기	○	○	○	○
	면허 및 자격증		○	○	○
	기호	○	○		
	학력	○	○	○	○
	외국어		○	○	○
	직업	○	○	○	○
	경력		○	○	○
	성품	○			
	사상	○	○		○
상별	○	○		○	
학생군사교육			○	○	
인사분류검사, 적성검사	인사분류검사 및 적성검사 결과	○	○	○	○
	적성분야(가병종)	○		○	○
징병처분사항	징집등급배점				○
	징집등급	○	○	○	○
	징병처분(병종, 징집순서, 종결처분, 징병처분사유)	○	○	○	○

〈표 1〉 포맷별 수록정보(계속)

구분	수록정보	포맷1	포맷2	포맷3	포맷4
신체검사	검사구분, 검사장소, 검사년월일	○	○	○	○
	신장, 흉위, 체중, 혈압, 시력, 색맹, 청력, 흉부X선, 관절, 치아	○	○	○	○
	혈액형	○	○	○	○
	내과, 피부비뇨과, 정신과				○
	신체등급	○	○	○	○
징병생활관련	군번	○	○	○	○
	출신기번		○	○	
	근무특기	○	○	○	○
	군사교육		○	○	○
	상별		○	○	○
	징집년도	○	○		
	계급상황	○	○	○	○
	징집구분, 역종구분	○			
	부대복무기록, 예비군복무기록		○	○	○
	입원기록	○	○	○	○
기타	위의 정보 외의 정보를 자유롭게 기록할 수 있는 공간	가족적요 <sup>(3)</sup> 비고	비고 기타	기타참고사항 비고 기타	비고 기타

이하 PID)를 부여할 때 중요한 기준이 되고, 제목을 통하여 해당 그림파일의 내용을 추측할 수 있기 때문에, 원칙에 어긋나는 제목을 수정하는 작업은 필수적이었다.<sup>(4)</sup>

제목 수정 후에는 원자료를 코딩할 바탕 파일인 마이크로소프트 엑셀 파일을 제작하였다. 우선 코딩 대상 변수로는 ‘병적기록표 작성 대상자가 징병처분을 받기 전 형성된 정보 중 개인정보보호를 위하여 사용이 금지된 것을 제외한 정보와 관련된 변

(3) 가족적요에는 가족 구성원들의 각각의 건강정도 또는 사회적 직위 등을 기재한다(1963년 5월 10일 시행된 「병역법 시행규칙」 [서식 9] 참조).

(4) 포맷1과 5를 제외한 모든 병적기록표의 앞면, 즉 흡수 제목의 그림파일에 개인의 군복무 이전 형성된 정보(인적사항, 인사분류검사 결과, 신체검사 결과, 징병처분 등)가 기록되어 있고, 뒷면, 즉 짝수 제목의 그림파일에 개인의 군복무 정보(계급, 복무기록, 예비군기록 등)가 기록되어 있다.

〈표 2〉 입력 대상 변수 목록

대분류	변수명
PID 및 포맷	PID, format
인적사항	면허 및 자격증, 생년월일, 당시본적(시도), 당시본적(시군구), 현재본적(시도), 현재본적(시군구), 당시본적2(시도), 당시본적2(시군구), 현재본적2(시도), 현재본적2(시군구), 당시주소(시도), 당시주소(시군구), 현재주소(시도), 현재주소(시군구), 학력(학교), 학력(학년), 학력(재/퇴/졸), 문맹, 호주친권자, 직업(직종), 직업(근무월수), 경력(내용), 경력(연수), 가족1(생년월일), 가족1(관계), 가족1(질병사항), 가족2(생년월일), ..., 가족6(질병사항), 친부친모여부, 독자, 이복형제, 외국어(언어), 외국어(기능), 외국어(정도), 상벌, 사상(내용), 사상(사유), 재산(동산), 재산(부동산), 주택(소유), 주택(형태), 재산정도, 재산세등급, 생계부담자, 월수입, 종교(내용), 종교(직위), 취미, 기호, 특기, 취미 및 특기, 성품, 성격, 학생군사교육(이수과정), 학생군사교육(이수연한,단축기간)
인사분류검사 및 적성검사	인사분류검사(구분), 인사분류검사1(종류), 인사분류검사1(점수, 적부), 인사분류검사1(등급,재검점수), 인사분류검사2(종류), ..., 인사분류검사4(등급, 재검점수), 적성검사1(종류), 적성검사1(점수), 적성검사1(등급), 적성검사2(종류), ..., 적성검사4(등급)
징집정보	징집년도, 징집등급배점1(연령), 징집등급배점2(학력), 징집등급배점3(체격), 징집등급배점4(지능), 징집등급배점5(계), 징집등급, 선병등급
신체검사	혈액형, 신체검사(검사구분), 신체검사(연월일), 신체검사(검사장소), 신장, 흉위, 체중, 혈압(고), 혈압(저), 시력(좌), 시력(우), 안과, 이비인후과(좌), 이비인후과(우), 흉부, X선, 외과, 내과, 피부과, 비뇨과, 피부비뇨과, 정신과, 치과, 신체등급, 신체(체격)등위
기타	비고, REMARK

수'로 한정하였고, 모든 포맷이 아닌 일부 포맷에만 있는 정보여도 코딩 대상에 포함시켰다. 그 후 샘플의 식별과 사용자 편의 제공을 위하여 몇 가지 변수를 추가로 생성하였다. 이 과정을 거쳐 확정된 코딩 대상 변수 목록은 〈표 2〉에 제시된 바와 같다.

코딩을 위한 엑셀 파일은 두 개의 시트로 구성되었고, 각각 '본자료'와 '부가자료'로 명명되었다. 징병검사의 대상이 두 번 이상의 검사를 받은 경우 여러 개의 검사결과를 반영하기 위함이다. 본자료에는 PID 및 포맷, 인적사항, 첫 번째 받은 인사분류검사 및 적성검사, 징집년도, 첫 번째 받은 징집등급배점 및 징집등급, 첫 번째 받은 신체검사, 이상의 사항과 관련된 기타사항을 입력하도록 하였다. 부가자료에는 PID 및 포맷, 두 번째 이후 받은 인사분류검사 및 적성검사, 두 번째 이후 받은 징집등급

배점 및 징집등급, 두 번째 이후 받은 신체검사, 이상의 사항과 관련된 기타사항을 입력하도록 하였다.

병적기록표 원자료를 엑셀 파일로 DB화하기 위한 데이터 입력은 2014년 7월부터 2015년 5월까지 총 13명에 의하여 진행되었다. 13인 중 9인은 데이터 1차 입력, 3인은 1차 입력된 데이터의 검토 및 수정, 1인은 2차 검토, 수정 및 진행 총괄업무를 담당하였다. 업무보고는 2주 간격으로 이루어졌으며, 업무보고와 함께 전술한 9인은 2주간 생성한 데이터를, 3인은 1차 수정한 데이터를 상위 보고자에게 제출하였다. 데이터 입력의 효율성을 위한 주요 원칙은 크게 다음의 세 가지로 요약된다.

첫째, 이미 상위 보고자에게 제출된 데이터의 수정권한은 보고받은 자에게만 있다. 단 이미 제출한 데이터에서 본인의 작업오류를 발견하였을 경우 신속히 상위 보고자에게 오류사항을 보고한다. 이 원칙은 수정권한을 일원화하고, 식별이 어려운 대량의 원자료를 입력하는 9인의 조교에게 입력업무에만 전념하도록 하기 위하여 고안되었다.

둘째, 원자료에 보이는 그대로 입력한다. 해당 변수명과 관련이 없어 보이거나 입력하여야 할 값의 의미를 알 수 없어도 우선 보이는 대로 입력하여야 한다. 이는 입력자의 자의적 판단을 최대한 막아 입력의 객관성을 높이기 위하여 고안되었다. 원자료의 × 표시, √ 표시도 데이터 입력 시 생략하지 않고 그대로 입력하도록 하였다.

셋째, 원자료인 병적기록표의 공백에 기재된 다양한 내용도 코딩에 반영한다. 당시 병적기록표 작성 담당 행정직원이 병적기록표에 기록하여야 할 기재사항 외에도 병무행정 에 필요하다 판단되는 정보가 있을 경우 그것을 병적기록표의 공백에 작성하였는데, 이 또한 향후 중요한 연구 소재가 될 가능성이 있으므로 최대한 충실하게 DB화에 반영하고자 하였다. 이를 위하여 엑셀파일에 ‘비고’ 및 ‘REMARK’ 변수를 따로 생성하여 관련 정보를 자유롭게 작성하도록 하였다.

원자료가 지금으로부터 대략 반세기 전 수기로 작성된 자료의 스캔본이고, 애초에 통계분석과 연구를 위하여 생성된 정보가 아니며, 병무청의 가공과 수정을 거친 후 전달받은 자료라는 점에서 데이터 입력 당시 다음과 같은 문제들이 발생하였다.

첫째, 원자료 기재내용이 악필로 작성되었거나, 복잡한 한자(漢字)가 필기체로 작성되었거나, 병적기록표 원본이 오래되어 기록내용의 식별이 곤란한 경우가 발생하였다. 이 경우 식별 불가능한 글자는 ‘?’로 대체하여 입력하였다.

둘째, 당시 병무행정현장에서 통용된 기호화된 변수내용의 뜻을 현재 데이터 입력



자들이 해독할 수 없는 경우가 있다. 이 경우 병적기록표의 반복 관찰을 통해 변수내용의 패턴을 발견하거나 병무행정 관련 법령<sup>(5)</sup>을 살펴봄으로써 해독을 시도하였다. 그 결과 해독이 가능한 변수 값도 발견하였으나, 여전히 해독을 못한 변수 값도 존재한다. 이 경우에는 전술한 데이터 입력 원칙에 따라 기호화된 변수가 기재된 항목에 해당 변수 값을 충실하게 입력하였다. 징집시의 행정절차에 대한 자세한 사료 발견 시 해독에 진전이 있을 것으로 기대된다.

셋째, 당시 징병검사나 신병교육대 검사의 과정을 알 수 있는 사료가 부족하여 일부 데이터 해석이 어려운 문제가 있다. 이 문제는 병적기록표 당시의 상황에 대한 정보 부족에 기인한다는 점에서 위의 문제와 비슷한 성격을 가지고 있다. 특히 인사분류검사와 적성검사에 대한 데이터 입력시 이 문제가 두드러졌다. 인사분류검사와 적성검사 항목의 연도별, 지역별, 검사장소별 이질성이 컸고, 설사 검사항목이 동일하여도 연도별, 지역별 점수나 등급 산정방식의 이질성도 커서, 병적기록표 작성 당시 시행되었던 인사분류검사와 적성검사 관련 자료가 추가로 확보되지 않는 한 이것과 관련된 데이터를 연구에서 사용하는 것에 한계가 분명히 존재할 수밖에 없다.

넷째, 원자료의 개인정보를 지우는 과정에서 발생한 문제가 있다. 먼저 병적기록표 당사자에 관련된 개인정보를 가리기 위한 마커 표시의 범위가 지나치게 넓어 주변 글씨까지 식별되지 않은 경우가 있다. 또한 주소 전체, 최종학력 전체 등 DB화를 위하여 필요한 정보까지 지워진 경우도 있었다. 이 문제는 추후 원자료를 제공한 기관으로부터 관련 정보를 다시 받아야만 해결된다. 따라서 데이터 입력 작업 중 해당 문제가 발견되었을 경우 별도로 이 사실을 기록해 두었다.<sup>(6)</sup>

다섯째, 원자료의 형태 자체로부터 발생한 문제가 있다. 병적기록표 일부분(보통 병적기록표의 하단 끝 부분, 우측 끝 부분)이 상실되어 상실된 병적기록표 부분의 정보

(5) 관련법령으로 『병역법』, 『병역법 시행령』, 『병역법 시행규칙』, 『징병 신체검사 등 검사규칙』 등이 있다.

(6) 반대로 지워야 하는 개인정보임에도 지워지지 않은 경우도 발견되었다. 이러한 경우 데이터 입력의 참고자료로만 사용할 뿐 직접 개인정보를 데이터에 입력하지는 않았다. 예를 들어 1970년 작성된 병적기록표 주소지에 ‘서울 영등포구 봉천동 ○○번지’라고 기재된 경우 엑셀 파일 ‘당시주소(시도)’ 란에 ‘서울’, ‘당시주소(시군구)’ 란에 ‘영등포구’, ‘현재주소(시도)’ 란에 ‘서울’, ‘현재주소(시군구)’ 란에 ‘관악구’로 입력하였다. 만약 봉천동 이하 부분이 원칙대로 마커로 가려졌으면 데이터 입력 시 당시 주소와 현재 주소 모두 ‘서울 영등포구’로 입력을 할 수밖에 없었을 것이다. 이와 같이 미처 가려지지 않은 개인정보를 직접적으로 데이터 입력 대상으로 편입시키지 않고 다른 (입력이 허용된) 데이터의 입력의 참고자료로만 사용하였다.

를 잃는 경우가 있었다. 병적기록표의 앞면 혹은 뒷면만 있는 경우도 존재하였다. 이러한 문제는 가용정보의 개수를 줄이는 결과를 낳는다.

여섯째, 병적기록표 작성 당시의 판정기준의 모호함도 데이터 작업 시 어려움 중의 하나였다. 신체검사 도중 발견된 질환명과 정도, 신체등급, 징집등급 산정 기준은 『장병 신체검사 등 검사 규칙』에 상술되어 있고 시대와 관계없이 비교적 일관적이어서 문제가 덜하였다.<sup>(7)</sup> 그러나 같은 연도, 같은 종류의 인사분류검사 또는 적성검사에서도 동일한 점수를 받은 표본 간 등급 설정에 이질성이 있는 경우는 종종 발견되었다.<sup>(8)</sup>

마지막으로 한 표본이 병적기록표를 반복 사용하면서 추가 기재한 사항 때문에 겪는 어려움이 있었다. 혈액형 재검사로 바뀐 혈액형과 이전 혈액형의 동시 기재, 1차 검사와 2차 검사 간 개인신상 변동 전과 변동 후 사항의 동시 기재 사례가 종종 발견되었다. 따라서 두 개의 기재사항 간 선후관계를 구분할 수 있는 경우 후에 기록된 값을 데이터에 입력하였으나, 구분이 불가능할 경우 불가피하게 두 가지 값 모두를 입력하였다.

### 3. 데이터 소개

여기에서는 포맷1, 포맷2, 포맷3, 포맷4 중 세 개 이상의 포맷에 등장하여 완결도가 높고 변수 성격상 중요성도 큰 변수들을 선별하여 해당 변수의 정의 및 생성방법을 설명하고 해당 변수 값의 예시를 제시하고자 한다. 추가적으로 데이터 입력 작업 중 자유롭게 활용하였던 ‘비고’와 ‘REMARK’ 변수에 대한 간략한 설명도 하고자 한다.<sup>(9)</sup>

(7) 신체질병유무 및 정도, 신체등급, 징집등급 산정 기준에 문제가 없는 것은 아니었다. 예를 들어 그 때 당시 시행된 『장병 신체검사 등 검사 규칙』에 따르면 신장미달 판정을 받아야 하는 신장 153cm의 지원자에게 신장미달 판정을 하지 않은 사례가 발견된다.

(8) 징병검사 대상자에 대한 정확한 검사와 엄정한 병역처분을 위한 『징병검사 규정』(병무청 훈령)은 2009년 8월 제정되어 그 이전 징병검사 사항에 대한 공개적인 문서는 병무청에 요청하기 전까지는 알 수 없다.

(9) 참고로 전술하였듯 변수명과 관련성이 없어 보이나 그 의미를 알 수 없는 기호화, 숫자화된 변수 값도 해당 변수 내용으로 입력을 하였다. 또한 식별이 불가능한 수기나 도장의 경우 ‘?’로 대체하였다. 이 장에서는 지면의 한계 상 이러한 사례에 대한 자세한 설명은 생략하도록 하겠다.

### 3.1. 개인식별번호 및 포맷

개인식별번호(변수명: PID)는 특정한 개인을 구분하여 한 개인에 대해 복수로 존재하는 파일들을 연결하거나 병적기록표 원자료에서 해당 개인을 찾을 수 있도록 생성된 변수이다. 여덟 자리 숫자로서 원자료 제공기관에서 폴더로 구분한 생년 네 자리와 파일번호 네 자리를 붙여 만들었다.

포맷(변수명: format)은 원자료의 형태에 따라 분류한 아홉 개의 포맷 중 해당 표본의 병적기록표가 속한 포맷을 나타내는 변수이다. 1부터 9까지의 숫자 중 하나의 값을 가진다. 이 정보는 특정한 변수의 누락이 애초 정보의 부재에서 비롯된 것인지(해당 변수를 포함하지 않은 병적기록표 형식) 아니면 다른 원인 때문에 발생한 것인지를 결정할 때 이용될 수 있다.

### 3.2. 인적정보

〈표 3〉는 인적정보에 관한 변수 및 생성방법을 정리해준다. 이를 간략하게 소개하면 다음과 같다. 우선 면허 및 자격증, 직업(직종), 경력(내용) 등에는 관련 내용을 병적기록표에 기재된 그대로 입력하였다. 생년월일은 해당 표본의 생년월일을 여덟 자리 숫자(YYYYMMDD) 형태로 나타낸 변수이다.

주소 및 본적지에 관한 변수는 당시본적(시도), 당시본적(시군구), 현재본적(시도), 현재본적(시군구), 당시본적2(시도), 당시본적2(시군구), 현재본적2(시도), 현재본적2(시군구), 당시주소(시도), 당시주소(시군구), 현재주소(시도), 현재주소(시군구) 등 총 12개이다. 주소 및 본적지 구분은 세 개의 차원으로 이루어져 있다. 먼저 변수명에 ‘당시’가 붙은 변수는 병적기록표가 작성될 당시 기준의 행정구역으로 표시한 지역명이고, ‘현재’가 붙은 변수는 데이터 입력 업무가 시작된 2014년 기준의 행정구역으로 표시한 지역명이다. ‘시도’가 붙은 변수는 시도 단위의 주소지이고, ‘시군구’가 붙은 변수는 시군구 단위의 주소지이다. 마지막으로 포맷 1, 2, 5의 경우 본적지를 두 개 기재할 수 있었는데, 병적기록표 상 (2)본적지, 신본적의 경우 ‘본적’ 변수에 입력하였고, (1)본적지 혹은 본적의 경우 ‘본적2’ 변수로 입력하였다. 나머지 포맷의 경우 ‘본적’ 변수에만 본적지를 입력하였다.<sup>(10)</sup>

(10) 두 개의 본적 변수 중 ‘본적’ 변수가 주된 변수이고 ‘본적2’ 변수가 보충적인 변수이다. 본적 변수는 해당 표본이 영유아기를 보낸 지역에 대한 대리변수(proxy variable)의 성격을 가지고 있다. 본적지를 두 개 기재할 수 있을 경우 일반 본적지 혹은 처음 작성한 본적지는 해당

〈표 3〉 데이터 소개: 인적정보

변수명	변수 존재 포맷	정의	입력방법	변수값의 예
면허 및 자격증	2, 3, 4	해당 표본이 취득한 면허 및 자격증	기록된 내용을 보이는 대로 입력	운전 2종 면허
생년월일	모든 포맷	해당 표본의 생년월일	여덟 자리 숫자로 입력	19500501
당시본적(시도)	모든 포맷	해당 표본의 본적지/주소지 -‘당시’: 병적기록표 작성 당시 기준 -‘현재’: 2014년 기준 -‘시도’: 시도 단위 -‘시군구’: 시군구 단위 -‘본적2’: 본적지 입력란이 두 개인 포맷인 포맷1, 2, 5에서의 ‘(1)본적지’ 혹은 ‘본적’	-시도 단위는 두 글자 입력 -시군구 단위는 ‘시’, ‘군’, ‘구’ 글자를 제외하고 입력 -‘주소불명’ 기재시 ‘주소불명’으로 입력	-시도: 서울, 경기 -시군구: 수원, 연천, 종로
당시본적(시군구)	모든 포맷			
현재본적(시도)	모든 포맷			
현재본적(시군구)	모든 포맷			
당시본적2(시도)	1, 2, 5			
당시본적2(시군구)	1, 2, 5			
현재본적2(시도)	1, 2, 5			
현재본적2(시군구)	1, 2, 5			
당시주소(시도)	모든 포맷			
당시주소(시군구)	모든 포맷			
현재주소(시도)	모든 포맷			
현재주소(시군구)	모든 포맷			
학력(학교)	모든 포맷	해당 표본이 마지막으로 다닌 학교급	학교급 입력	고등학교
학력(학년)	모든 포맷	해당 표본이 마지막으로 다닌 학교급에서의 최종학년	학년을 숫자로 입력	3
학력(재/퇴/졸)	모든 포맷	해당 표본이 마지막으로 다닌 학교의 재학, 퇴학, 졸업, 수료 여부	재, 퇴, 졸, 수료 중 입력	졸
호주친권자	모든 포맷	해당 표본의 호주 또는 친권자	해당 표본 기준에서의 호주 또는 친권자의 호칭을 한자어로 입력	부
직업(직종)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 직업 내용	기록된 내용을 보이는 대로 입력	농업

표본의 조부모의 생활의 근거지, 신본적지 혹은 나중에 작성된 본적지는 해당 표본의 부모의 생활의 근거지일 확률이 높다. 따라서 본적 변수로 알고자 하는 표본의 성격에 더욱 가까운 신본적지 혹은 나중에 작성된 본적지를 주된 변수인 ‘본적’ 변수에 입력하였다.

〈표 3〉 데이터 소개: 인적정보(계속)

변수명	변수 존재 포맷	정의	입력방법	변수값의 예
경력(내용)	2, 3, 4, 6, 7, 8, 9	해당 표본의 경력 내용	기록된 내용을 보이는 대로 입력	건설업
가족n(생년월일) (n=1,2,3,4,5,6)	모든 포맷	해당 병적기록표 가족관계 작성란에 n번째로 기록된 가족의 생년월일	-여덟자리 숫자로 입력 -연령이 기재된 경우 연령으로 입력	19100301
가족n(관계) (n=1,2,3,4,5,6)	모든 포맷	해당 병적기록표 가족관계 작성란에 n번째로 기록된 가족의 표본과의 관계	한자어로 입력하는 것 이 원칙이나 ‘손윗’, ‘누 이’, ‘동생’은 순우리말 로 입력	부
가족n(직업) (n=1,2,3,4,5,6)	모든 포맷	해당 병적기록표 가족관계 작성란에 n번째로 기록된 가족의 직업	기록된 내용을 보이는 대로 입력	농업
재산(동산)	1, 2, 4, 6, 7, 8, 9	해당 표본의 보유 동산의 가치	기록된 내용을 보이는 대로 입력	300,000
재산(부동산)	1, 2, 4, 6, 7, 8, 9	해당 표본의 보유 부동산의 내역 또는 가치	-포맷 2, 4: 기록된 내용 을 보이는 대로 입력 -나머지 포맷: 기록된 내용에 부동산 종류 및 단위도 함께 입력	전 400평

학력에 관한 변수는 학력(학교), 학력(학년), 학력(재/퇴/졸) 등으로 구성되어 있다. 해당 표본의 최종학력을 입력하되 각각 학교, 학년, 재학/퇴학/수료/졸업 중 해당사항을 입력하였다.

호주친권자 변수에는 해당 표본의 호주 또는 친권자를 입력하였다. 원자료에는 표본 자신의 기준에서 호주 또는 친권자의 위치를 작성하는 경우(부, 조부 등)와 호주 또는 친권자의 기준에서 자신의 위치를 작성하는 경우(자, 손자) 모두 존재하였는데 데이터 입력 시에는 이 중 전자의 기준을 적용하였다.<sup>(11)</sup>

(11) 후자의 기준을 전자로 변환하는 데에는 ‘가족관계에 등장→남성→위사람’의 우선순위로 호칭을 부여하였다. 예를 들어 원자료에 호주 또는 친권자로 ‘자(子)’를 기재한 경우, i) 가족관계에 부친과 모친 모두 존재하거나 모두 존재하지 않는 경우 ‘부(父)’로 바꾸었다. ii) 가족관계에 모친만 존재할 경우 ‘모(母)’로 바꾸었다. 또한 원자료에 호주 또는 친권자로 ‘조카’가

병적기록표에는 최대 6명의 가족을 적을 수 있는 칸이 마련되어 있다. 이에 따라 각각의 가족에 대해 가족n(생년월일), 가족n(관계), 가족n(직업) ( $n=1, \dots, 6$ ) 등의 변수를 생성하였다. 각각 가족구성원의 생년월일과 해당 표본과의 관계, 직업을 순차적으로 입력하였다.

재산(동산), 재산(부동산) 변수에는 각각 해당 표본의 동산과 부동산 가치를 입력하였다. 포맷 1, 6, 7, 8, 9의 원자료에는 재산의 종류(동산, 대지, 전답 등) 및 단위(원, 평, 제곱미터 등)가 기록되어 있어 병적기록표 작성자가 아라비아 숫자만 기재할 수 있다. 이 포맷의 원자료를 입력할 때에는 ‘원’ 단위만 생략하였다. 반면 포맷 2, 4의 원자료에는 재산의 종류 및 단위가 인쇄되어 있지 않았다. 따라서 병적기록표 작성자가 아라비아 숫자만 기재하는 경우와 아라비아 숫자에 재산의 종류 또는 단위를 함께 기재하는 경우 모두 존재하였다. 이 포맷의 원자료를 엑셀 파일에 입력할 때에는 원자료에 기재된 그대로 입력하였다.

### 3.3. 인사분류검사 및 적성검사 기록

〈표 4〉는 인사분류검사 및 적성검사 기록에 관련된 변수의 내용을 정리한 결과이다. 인사분류검사 및 적성검사는 징병대상자가 징병검사 혹은 신병교육대 검사를 받을 때 실시하며, 검사 담당자가 징병대상자의 지능, 인성, 적성 등을 측정 후 그 결과를 점수 및 등급으로 기록한다. 전술하였듯 병적기록표 작성 당시의 인사분류검사 및 적성검사에 관한 자료가 부족하여 데이터 입력에 어려움이 있었으나, 병적기록표의 반복 검토를 통하여 최대한 일관적인 값을 입력하고자 하였다.

인사분류검사 및 적성검사는 각각 네 종류까지 입력이 가능하다. 인사분류검사n(종류), 인사분류검사n(점수, 적부), 인사분류검사n(등급, 재검점수) ( $n=1, 2, 3, 4$ ) 등에는 각각 병적기록표의 인사분류검사란에 n번째로 기록되어 있는 검사의 종류, 점수, 등급을 입력하였다. 적성검사n(종류), 적성검사n(점수), 적성검사n(등급) ( $n=1, 2, 3, 4$ ) 등도 마찬가지로 병적기록표의 적성검사란에 n번째로 기록되어 있거나 인사분류검사란의 일부분에 적성검사로 명시되어 있는 검사 중 n번째로 기록되어 있는 검사의 종류, 점수, 등급을 각각 입력하였다.

특정 검사종류를 인사분류검사, 적성검사 중 어느 곳으로 분류할지 여부는 검사명

---

기재되었고 가족관계에 백부모와 숙부모 모두 존재하지 않는 경우 ‘백부(伯父)’로 바꾸어 입력하였다.

〈표 4〉 데이터 소개: 인사분류검사 및 적성검사

변수명	변수 존재 포맷	정의	입력방법	변수값의 예
인사분류검사n (종류) (n=1,2,3,4)	1, 2, 3, 4, 6, 7, 8, 9	-포맷 1, 2: 해당 표본이 받은 인사 분류검사에서 n번째로 실시한 인 사분류검사의 종류 -나머지: 해당 병적기록표에 인쇄 된 인사분류검사 중 n번째 종류	기록된 내용을 보 이는 대로 입력	일반능력검사
인사분류검사n (점수, 적부) (n=1,2,3,4)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 ‘인사분류검사n(종 류)’에 입력된 인사분류검사의 점 수	기록된 내용을 보 이는 대로 입력	120
인사분류검사n (등급,재검점수) (n=1,2,3,4)	1, 2, 3, 4, 6	해당 표본의 ‘인사분류검사n(종 류)’에 입력된 인사분류검사의 등 급	기록된 내용을 보 이는 대로 입력	ㄱ
적성검사n(종류) (n=1,2,3,4)	1, 2, 3, 4, 6, 7, 8, 9	-포맷 1, 2: 해당 표본이 받은 적성 검사에서 n번째로 실시한 적성검 사의 종류 -나머지: 해당 병적기록표에 인쇄 된 적성검사 중 n번째 종류	기록된 내용을 보 이는 대로 입력	전자
적성검사n(점수) (n=1,2,3,4)	1, 2, 3, 4, 6, 7	해당 표본의 ‘적성검사n(종류)’에 입력된 적성검사의 점수	기록된 내용을 보 이는 대로 입력	55
적성검사n(등급) (n=1,2,3,4)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 ‘적성검사n(종류)’에 입력된 적성검사의 등급	기록된 내용을 보 이는 대로 입력	ㄴ

이 아닌 그 검사종류가 기록되어 있는 항목을 기준으로 결정하였다. 따라서 어느 표본에는 적성검사로 분류된 검사종류가 다른 표본에는 인사분류검사로 분류되는 경우가 있을 수 있다. 이는 자의적 판단을 배제하고 병적기록표의 기록에 충실하게 데이터 입력을 하여야 한다는 원칙에 따른 것이다.

또한 포맷 1, 2의 경우 검사의 종류, 점수, 등급 모두 기록관이 직접 기록해야 하고, 나머지 포맷의 경우 검사의 종류는 병적기록표에 이미 인쇄되어 있어 이질성이 존재한다. 포맷 1, 2의 병적기록표 중에서 인사분류검사와 적성검사 기재란에 검사명만 기록되어 있고 점수와 등급이 적혀있지 않은 경우 그 검사가 해당 표본에 대해서는 실시되지 않았다고 판단하여 데이터 입력 대상에서 제외하였다. 그러나 나머지 포맷의 경우 병적기록표에 인쇄되어 있는 검사명은 점수 및 등급이 적혀있지 않아도 데이

터에 검사명은 입력하였다.

### 3.4. 신체검사 기록

〈표 5〉는 신체검사 기록의 입력내용을 정리해 놓은 것이다. 혈액형은 해당 표본의 혈액형을 기록하는 변수이다. 대개 A, B, O, AB 중 한 개의 값을 가지나 간혹 두 개의 혈액형이 기록된 경우가 있다. 처음의 신체검사 결과 이미 판정된 혈액형이 다음 신체검사 때 반복된 사례로 추측되나, 처음 판정된 혈액형과 뒤에 판정된 혈액형을 구분할 수 없어 이 경우 두 개의 혈액형 값을 모두 입력하였다. 신체검사(검사구분), 신체검사(연월일), 신체검사(검사장소) 등은 해당 표본이 받은 신체검사의 구분, 일시, 장소 등을 기록한 변수이다.

신장, 흉위, 체중, 혈압(고), 혈압(저), 시력(좌), 시력(우) 등은 각각 신장, 흉위, 체중, 최고혈압과 최저혈압, 왼쪽시력과 오른쪽시력을 측정하여 기록한 변수이다. 이 변수들은 대개의 경우 통상적인 입력 방식을 따르고 있다. 즉 신장과 흉위는 cm 단위, 체중은 kg 단위, 혈압은 세 자리 혹은 두 자리 숫자, 시력은 2.0 이하의 소수점 한 자리 숫자로 입력한다. 그러나 통상적인 방법이 아닌 숫자로 기록되거나, 측정 군의관의 확인 도장만 찍혀있거나, 정상 여부만 기록된 표본도 있다. 이런 경우에도 병적 기록의 내용을 그대로 입력하였다. 나안시력과 교정시력 모두 기록된 경우에는 ‘나안시력(교정시력)’ 형태로, 교정시력과 안경도수가 모두 기록된 경우에는 ‘교정시력(안경도수)’ 형태로 입력하였다.<sup>(12)</sup>

안과, 이비인후과(좌), 이비인후과(우), 흉부, X선, 외과, 치과 등은 각각 눈, 왼쪽 귀, 오른쪽 귀, 흉부, 관절 등 외과 소관 기관, 치아가 정상인지 여부를 기록한 변수이다. 대부분 정상 여부가 기록되어 있으며(정, 정상, 이상무, 혹은 이상, △), 징병군의관의 확인이 기록된 경우도 있다. 징병군의관의 확인은 보통 검사 결과가 정상일 때에는 ‘진료과를 나타내는 대문자+1+군의관 성’의 형태로 기록되어 있으며,<sup>(13)</sup> 정상이 아닐 경우에는 ‘진료과를 나타내는 대문자+2부터 5까지의 숫자+병종의 심각도를

(12) ‘나안시력(교정시력)’ 형태와 ‘교정시력(안경도수)’ 형태는 그 이질성이 쉽게 구분되어 혼란의 여지는 적다.

(13) 보통 내과는 ‘P’, 외과는 ‘P’, ‘U’ 또는 ‘L’, 안과는 ‘E’, 이비인후과는 ‘H’, 정신과는 ‘S’, 치과는 ‘D’이다. ‘1’은 정상임을 나타내는 숫자이다. 해당 사항의 예로 ‘P1장’, ‘S1조’ 등이 있다.



〈표 5〉 데이터 소개: 신체검사 기록

변수명	변수 존재 포맷	정의	입력방법	변수값의 예
혈액형	모든 포맷	해당 표본의 혈액형	-A, B, O, AB 중 하나로 입력 -서로 다른 혈액형 두 개가 동시 기재된 경우 두 개 모두 입력	A
신체검사 (검사구분)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본이 받은 신체검 사가 실시된 검사	기록된 내용을 보이는 대로 입 력	68징병검사
신체검사 (연월일)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본이 받은 신체검 사의 일시	여덟자리 숫자로 입력	19680401
신체검사 (검사장소)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본이 받은 신체검 사장소	기록된 내용을 보이는 대로 입 력	수도통합병원
신장	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 신장	기록된 내용을 보이는 대로 입 력(보통 cm 단위)	170
흉위	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 흉위	기록된 내용을 보이는 대로 입 력(보통 cm 단위)	88
체중	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 체중	기록된 내용을 보이는 대로 입 력(보통 kg 단위)	64
혈압(고)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 최고혈압	기록된 내용을 보이는 대로 입 력	120
혈압(저)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 최저혈압	기록된 내용을 보이는 대로 입 력	80
시력(좌)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 왼쪽 시력	기록된 내용을 보이는 대로 입 력	1.0
시력(우)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 오른쪽 시력	기록된 내용을 보이는 대로 입 력	1.0
안과	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 눈 이상 여부	기록된 내용을 보이는 대로 입 력	정상
이비인후과 (좌)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 왼쪽 귀 이상 여 부	기록된 내용을 보이는 대로 입 력	정상
이비인후과 (우)	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 오른쪽 귀 이 상 여부	기록된 내용을 보이는 대로 입 력	정상
흉부, X선	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 X선 촬영결 과 나타난 흉부 이상 여부	기록된 내용을 보이는 대로 입 력	정상

〈표 5〉 데이터 소개: 신체검사 기록(계속)

변수명	변수 존재 포맷	정의	입력방법	변수값의 예
외과	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 외과 소관 기관 이상 여부	기록된 내용을 보이는 대로 입력	정상
치과	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 치아 이상 여부	기록된 내용을 보이는 대로 입력	정상
신체등급, 신체(체격) 등위	1, 2, 3, 4, 6, 7, 8, 9	해당 표본의 신체등급	갑, 일을, 이을, 삼을, 병, 정, 무중 택일	일을

나타내는 대문자(A, B, C 중 하나)’의 형태<sup>(14)</sup>로 기록되어 있다. 징병군의관의 확인은 도장 혹은 수기로 기록되어 있다.

신체등급, 신체(체격)등위 변수는 신체검사 결과를 종합한 해당 표본의 신체등급을 나타낸다. 갑중, 일을중, 이을중, 삼을중, 병중, 정중, 무중 중 한 개의 값을 갖는다.

### 3.5. 징집등급

징집등급/선병등급 변수에는 해당 표본이 받은 징집등급이 입력되어 있다. 징집등급의 판정시 징병대상자의 신체조건, 인사분류검사와 적성검사 결과, 연령, 학력, 개인신상 등을 종합적으로 고려하여 1급, 2급, 3급, 4급 중 한 개의 처분을 내리거나, 처분연기 결정을 한다. 따라서 해당 변수의 값 또한 1, 2, 3, 4, 처분연기 중 한 개의 값을 가진다.

### 3.6. 비정형 변수

병적기록표 원자료에는 의무적으로 기재할 사항이 아니지만 중요한 정보를 담고 있는 사항이 공란 곳곳에 기재되어 있는 경우가 많다. 또한 의미를 알 수 없거나 확실하지 않은 정보를 담고 있는 경우도 적지 않다. 따라서 공통의 형식에 맞지 않는 정보들을 따로 입력할 공간의 필요성이 제기되었고, 이를 해결하기 위하여 비고 및 REMARK 변수를 데이터에 추가하였다.

비고란은 원자료에 의무적으로 기재할 사항은 아니지만 중요한 정보를 입력하는

(14) 해당 사항의 예로 ‘P2A’, ‘E5’ 등이 있다.

공간이다. 여기에 입력되는 내용은 주로 다음과 같다.

- 병명, 병명기호, 병명진단일, 병명진단기관: (예) 신장미달 P2A 19700401 수도통합병원
- 해당 포맷에는 의무적으로 기재할 사항이 아니지만 병적기록표 공란에 기재되어 있고, 다른 포맷에는 같은 내용이 의무적으로 기재할 사항일 경우: (예) 포맷 2에 기재된 학생군사교육을 받은 정보, 포맷 3에 기재된 재산사항
- 7번째 이상의 가족에 관한 정보, 5번째 이상의 인사분류검사 및 적성검사 정보, 두 번째 이상의 가능외국어 정보 등
- 기타 잠재적으로 유용하다고 판단되는 정보: (예) 배급받은 군복과 군화 사이즈, 받아쓰기검사결과, 포맷 1 병적기록표의 ‘가족적요’에 기재된 사항

REMARK 란은 병적기록표 내용을 엑셀 파일로 옮기는 과정에서 생기는 의문점 혹은 특이한 사항을 입력하거나, 자신이 해당 변수를 특정한 값으로 입력하게 된 경위를 형식에 구애받지 않고 기재하는 공간이다. 여기에 기재된 내용은 주로 다음과 같다.

- 그릇된 정보로 추정되는 경우: (예) 가족관계에서 부모의 생년월일이 지나치게 늦은 경우
- 변수 값이 극단적인 경우: (예) 재산이 지나치게 많은 경우, 체중이 지나치게 가벼운 경우
- 동일한 병적기록표가 서로 다른 파일명으로 두 개 이상 존재하는 경우
- 가리지 않아도 되는 정보가 마커로 가려져 있어 코딩을 못 한 경우
- 병적기록표의 스캔이 잘못 되어 정보들의 다수를 잃은 경우
- 특정 변수에 두 개 이상의 변수 값이 있었고, 그 중 한 개의 변수 값을 선택하여 입력하였을 때 데이터 입력자가 그러한 선택을 한 이유
- 병적기록표에는 주소지가 불완전하게 기록되어 있으나 데이터 입력 시 시도 구분과 시군구 구분의 주소지를 모두 기재하였을 경우 그 이유 설명: (예) 병적기록표에 시도구분과 읍면동 구분만 기재되어 있을 때, 읍면동 구분을 통하여 시군구 구분을 추측하였을 경우

#### 4. 병적기록을 통해 살펴본 출생코호트별 특성

이 절에서는 병적기록표 원자료를 입력하여 구축한 데이터베이스를 이용하여 각 출생코호트의 특성을 살펴보고자 한다.<sup>(15)</sup> 특히 학력, 신장, 체중과 같은 인적자본 및 후생의 지표가 1946년생부터 1957년생까지 어떻게 변화하였는지를 개관할 것이다. 이는 데이터를 좀 더 자세하게 소개하고 추후의 활용도를 보여주기 위한 것으로, 분석은 현재까지 클리닝(cleaning)이 충분히 이루어졌고, 사회경제적인 중요성이 높은 변수에 제한되었다. 징집남성의 인적자본 및 후생의 지표가 어떻게, 왜 변화했는지에 대해서는 추후에 보다 체계적인 분석이 진행될 예정이다.

##### 4.1. 출생코호트별 학력<sup>(16)</sup>

먼저 최종학력에 관한 정보를 이용하여 징집대상자들의 교육연수를 추정하였다. 데이터에서는 최종학력만 식별되기 때문에 각 급 학교에 진학하기 위해 필요한 통상적인 교육연수에 자료에 기록된 최종학년을 더하여 교육연수를 추정하였다. 예를 들어 어떠한 표본이 고등학교를 3학년으로 졸업한 경우, 국민학교 6년, 중학교 3년, 고등학교 3년을 더하여 12년의 교육을 받았다고 가정하였다. 최종학력의 학교 급은 알 수 있으나 학년이 식별되지 않는 경우, 재학이나 퇴학 상태이면 졸업에 필요한 연수의 중간 값으로 교육연수를 추정하였고, 졸업 상태이면 졸업에 필요한 통상적인 연수로 교육연수를 추정하였다. 최종학력 학교 급이 누락되었을 경우 교육연수를 결측값으로 처리하였다.

〈표 6〉은 각 출생코호트의 징집 시 교육연수 추이를 보여준다. 1946년~1957년 출생코호트 전체의 평균 교육연수는 약 10년으로 추정되었다. 연도에 따른 등락을 보이지만 전반적으로 보아 1951년 이후에 출생한 코호트의 교육연수가 약 반년 가량 긴 것으로 나타난다. 1951년 출생자들은 가장 낮은 교육연수를 기록하였는데 이는 태아

(15) 여기에서의 ‘출생코호트’는 데이터에서의 ‘생년월일’ 변수값에 기초하여 분류한 결과이다. 대부분의 표본이 PID의 첫 네 자리 숫자와 생년월일의 첫 네 자리 숫자가 일치하지만, 207개(1.14%)의 표본에서는 양자가 불일치하거나(79개) 생년월일이 공란이다(128개). 1946년 이전 혹은 1957년 이후 출생한 표본도 11개 존재하지만 이들은 이 절의 분석대상에서 제외시켰다.

(16) 병적기록부에 기록된 학력은 모든 경우 최종학력인 것은 아니다. 병적기록표 작성 당시 재학생인 경우 교육을 완료한 것이 아니기 때문이다. 재학 중인 표본은 총 2,390개이며 이 중 1,906명은 대학생, 390명은 고등학생이다.

〈표 6〉 출생코호트별 교육연수 추이

(단위: 년)

	평균	표준편차	최소값	중간값	최대값	표본수
1946	9.55	3.18	0	9	19	1,136
1947	10.39	2.87	1	12	18	1,010
1948	9.61	3.12	0	9	18	1,412
1949	10.00	3.11	0	9	20	1,104
1950	9.70	2.92	2	9	20	1,283
1951	9.51	3.04	1	9	20	1,243
1952	10.09	2.95	3	9	20	1,550
1953	10.39	2.84	4	11	20	1,348
1954	10.07	2.97	3	10	19	1,688
1955	10.08	2.89	4	10	20	1,834
1956	10.12	2.89	3	10	19	2,112
1957	10.33	2.86	5	11	19	2,165
계	10.01	2.97	0	9	20	17,885

기에 한국전쟁을 경험한 세대의 학력이 낮았다는 Lee(2014)의 결과와 부합한다.

#### 4.2. 각 징집등급의 비율 추이

〈표 7〉은 징병검사에서 각 등급(1~4급)을 받은 개인들의 비율이 어떻게 변화했는지를 보여준다. 결과가 보여주는 특이점은 1950년대 초 이후 3급 혹은 4급 판정을 받은 사람의 비율이 높아진다는 것이다. 1급과 2급의 비율은 연도별로 상당한 등락을 보이지만 두 등급 중 하나를 받은 사람들의 비율은 장기적으로 하락하는 추이를 보인다. 현재로서는 이러한 추이가 나타난 이유를 정확하게 알기 어렵다. 다만 1950년대 초 이후 출생자가 증가하면서 군 입대 자원이 충분해짐에 따라 등급판정의 기준이 높아졌을 가능성을 고려해 볼 수 있다.

〈표 7〉 출생코호트별 징집등급 비율 추이

(단위: %)

	1급	2급	3급	4급	징병처분연기
1946	38.03	50.70	9.86	1.41	0
1947	48.34	38.41	10.60	1.99	0.66
1948	55.34	32.82	9.91	1.54	0.39
1949	50.56	43.21	4.90	1.02	0.31
1950	33.71	62.60	2.58	0.60	0.52
1951	59.01	32.90	4.09	3.31	0.70
1952	64.56	23.82	7.10	3.59	0.94
1953	40.70	34.32	12.58	11.56	0.83
1954	32.41	42.74	12.90	11.88	0.07
1955	22.38	50.72	11.10	15.44	0.04
1956	23.81	50.40	13.44	12.25	0.10
1957	24.24	37.22	20.69	17.55	0.30
계	38.07	41.63	10.88	9.01	0.42

### 4.3. 신장과 체중<sup>(17)</sup>

〈표 8〉은 각 출생코호트의 평균 신장 추이를 보여준다. 1946년~1957년 출생자의 평균키는 166.4센티미터로 추정된다. 장기적으로 볼 때 1946년생부터 1951년생까지는 165.5센티미터 수준에서 정체하다가 1952년생부터 상당히 빠르게 커지는 추이를 보인다. 1957년생은 167.5센티미터로 1950년생에 비해 2센티미터 이상 큰 것으로 나타났다. 한국전쟁으로 인한 민간인 피해가 가장 심각했던 시기에 태아기 혹은 유아기를 보낸 출생코호트(1950년생과 1951년생)의 신장이 가장 작았던 것도 관찰된다.

〈표 9〉는 각 출생코호트의 평균체중 추이를 보여준다. 신장의 경우와 달리 체중은 분석기간을 통해 출생코호트에 따른 장기적인 추세를 보이지 않는다. 반면 한국전쟁의 영향은 매우 강하게 관찰된다. 한국전쟁으로 인한 민간인의 피해가 심각했던 시기에 태아기나 유아기를 보낸 1950년생과 1951년생의 체중은 그 전후에 출생한 코호트의 체중에 비해 훨씬 낮았던 것으로 나타났다. 이상의 결과는 전쟁으로 인해 태아기

(17) 신장과 체중, 흉위는 소수점 이하 숫자가 식별 불가능할 경우에는 정수 부분만 사용하였고, 일의 자리 이상에서 식별 불가능한 경우에는 결측 값으로 처리하였다.

〈표 8〉 출생코호트별 신장 추이

(단위: cm)

	평균	표준편차	최소값	중간값	최대값	표본수
1946	165.8	5.32	151.0	166.0	188.0	965
1947	165.7	5.04	150.0	166.0	190.0	837
1948	165.6	5.34	149.0	166.0	193.0	1,204
1949	165.6	5.46	148.0	165.6	183.0	1,020
1950	165.4	5.26	148.6	165.3	183.0	1,228
1951	165.5	5.04	151.0	165.0	183.5	1,195
1952	166.4	5.21	148.0	166.0	186.0	1,481
1953	166.7	5.26	151.0	167.0	188.0	1,303
1954	166.6	5.15	150.0	167.0	191.0	1,650
1955	167.1	5.06	154.0	167.0	186.0	1,805
1956	167.1	5.16	154.0	167.0	183.0	2,104
1957	167.5	5.33	150.0	167.0	184.0	2,145
계	166.4	5.26	148.0	166.0	193.0	16,937

〈표 9〉 출생코호트별 체중 추이

(단위: kg)

	평균	표준편차	최소값	중간값	최대값	표본수
1946	58.31	5.65	44.2	58.0	81.0	974
1947	58.33	5.97	43.0	58.0	105.0	815
1948	58.33	5.85	40.0	58.0	87.0	1,229
1949	58.10	5.74	42.5	58.0	85.0	1,006
1950	57.57	5.44	43.5	57.0	84.0	1,196
1951	57.80	5.33	43.5	57.0	87.0	1,172
1952	58.32	5.55	41.0	57.9	115.0	1,499
1953	58.68	5.73	43.5	59.0	88.0	1,316
1954	58.39	5.76	42.0	58.0	87.5	1,673
1955	58.56	5.66	45.0	58.0	87.0	1,811
1956	58.21	5.48	45.0	58.0	87.0	2,099
1957	58.65	5.80	45.0	58.0	84.0	2,151
계	58.31	5.66	40.0	58.0	115.0	16,941

혹은 유아기에 굶주림을 겪은 세대의 성인기 이후 건강이 나빠졌다는 기존연구의 결과와 부합되는 것으로 보인다.

## 5. 맺음말

이상에서 소개한 병적기록표 표본 데이터베이스가 자료로서의 가치를 높이기 위해서는 다음의 몇 가지 작업이 추가적으로 이루어질 필요가 있다. 첫째, 입력상의 오류를 바로잡고 정보의 누락을 채우기 위한 데이터 클리닝(cleaning) 작업이 필요하다. 현재까지 기본적인 클리닝 작업이 완료되었지만 데이터를 이용하는 과정에서 새롭게 드러나는 여러 가지 오류를 수정하는 작업은 앞으로도 지속적으로 이루어질 필요가 있다. 그리고 일부 표본의 정보가 누락된 변수들 가운데는 원자료를 재검토하거나 다른 변수들과의 대조를 통해 부분적으로 채워 넣을 수 있는 것들이 있다. 예컨대 현 주소의 경우, 병적기록표에 가까운 우체국이나 경찰서가 기록되어 있는 경우가 있기 때문에 그 명칭으로부터 주소지를 파악하는 것이 가능하다.

둘째, 입력된 자료를 효과적으로 분석에 이용할 수 있도록 추가적인 변수들을 생성할 필요가 있다. 본적지 혹은 주소지의 경우 광역시도와 시군구의 현재 및 병적기록표 작성 당시의 센서스 코드를 생성한 바 있다. 이 외에 추가적으로 필요한 작업들의 예를 살펴보면 다음과 같다. 우선 부모와 본인의 직업을 적절한 방법으로 분류하여 코딩하면 사회경제적 지위와 가정환경을 이해하는 데 도움이 될 수 있다. 가족 구성원을 관찰하여 가구의 유형을 파악한 후 이에 대한 코드를 부여한 변수도 유용하게 이용될 수 있을 것이다.

셋째, 보다 장기적으로는 추가적인 자료의 수집과 전산입력 작업을 수행하여 병적기록표 데이터베이스를 완성할 필요가 있다. 이번 과제에서는 연구비의 부족으로 병적기록표의 뒷면에 있는 군복무 기록을 전산화하지 못하였다. 이 정보가 추가적으로 입력되면 과거 출생코호트의 군 생활을 이해하는 데 유용하게 이용될 수 있을 것이다. 그리고 앞에서 설명했듯이 전체 징병자들에 대한 전산화가 이루어지지 않은 2002년 이전 기록 전체에 대한 표본의 추출과 전산입력 작업을 수행할 필요가 있다. 현재의 데이터를 가지고는 1950년대 말부터 1980년대 초에 출생한 남성들의 징집당시 특성을 분석할 수 없다는 한계가 있다.

새로운 데이터의 발굴, 수집, 입력, 분석은 학문적인 진보가 이루어질 수 있는 중요



한 근간이다. 비록 일부시기에 대해 비교적 작은 표본이 수집·입력되기는 했지만 이 글이 소개한 병적기록표 표본 데이터베이스의 구축은 그 동안 연구에 이용된 적이 없는 공공기록이 정부기관과 민간연구자들의 협력에 의해 세상의 빛을 보게 된 사례라는 점에서 의미가 크다고 하겠다. 이 데이터는 추후 해방 이후 우리나라의 후생수준 변화 추이와 그 결정요인을 비롯한 여러 중요한 학문적, 정책적 문제를 연구하는데 유용하게 이용될 것이다. 데이터를 제공한 병무청도 완성된 데이터베이스를 병무행정 개선에 활용할 수 있을 것으로 기대한다. 추후에도 이러한 협력이 지속되어 유용한 데이터가 많이 생산되고 이것이 학문적 발전과 더 나은 정책의 수립에 이용되기를 희망한다.

이철희

서울대학교 사회과학대학 경제학부 교수

151-742 서울 관악구 관악로 1번지

전화: 02-880-6396

팩스: 02-886-4231

E-mail: chullee@snu.ac.kr

도민영

서울대학교 사회과학대학 경제학부 박사과정

151-742 서울 관악구 관악로 1번지

전화: 010-7175-6936

팩스: 02-886-4231

E-mail: knkn73@snu.ac.kr

박순영

서울대학교 사회과학대학 인류학과 교수

151-742 서울 관악구 관악로 1번지

전화: 02-880-9020

팩스: 02-878-8261

E-mail: suny@snu.ac.kr

홍두승

서울대학교 사회과학대학 사회학과 명예교수

151-742 서울 관악구 관악로 1번지

전화: 02-880-6402

팩스: 02-873-3799

E-mail: dshong@snu.ac.kr

### 참고문헌

- 최성진(2006): “식민지기 신장변화와 생활수준,” 『경제사학』, **40**, 57-82.
- Costa D. L.(1993): “Health, Weight, Wartime Stress, and Older Age Mortality: Evidence from Union Army Records,” *Explorations in Economic History*, **30**, **4**, 424-449.
- \_\_\_\_\_(1998): *Evolution of Retirement*, Chicago, University of Chicago Press.
- Costa, Dora L., and Matthew E. Kahn(2008): *Heroes and Cowards*. Princeton, NJ, Princeton University Press.
- Choi, Seong-Jin, and Daniel Schwegendiek(2009): “The biological standard of living in colonial Korea, 1910-1945,” *Economics and Human Biology*, **7**, **2**, 259-264.
- Fogel, R. W.(1986): “Nutrition and the Decline in Mortality since 1700: Some Additional Preliminary Findings,” in S. L. Engerman, and R. E. Gallman (eds.), *In Long-Term Factors in American Economic Growth*, Chicago, University of Chicago Press, 439-527.
- Fogel R. W.(1993): “New Sources and New Techniques for the Study of Secular Trends in Nutritional Status, Health, and Mortality, and the Process of Aging,” *Historical Methods*, **26**, **1**, 5-44.
- Kim, Duol, and Heejin Park(2011): “Measuring Living Standards from the Lowest: Height of the Male Hangryu Deceased in Colonial Korea,” *Explorations in Economic History*, **48**, **4**, 590-599.
- Lee, Chulhee(1997): “Socioeconomic Background, Disease, and Mortality among Union Army Recruits: Implications for Economic and Demographic History,”

- Explorations in Economic History*, **34**, **1**, 27-55.
- Lee, Chulhee(2005): “Wealth Accumulation and the Health of Union Army Veterans, 1860-1870,” *Journal of Economic History*, **65**, **2**, 352-385.
- Lee, Chulhee(2007): “Military Positions and Post-Service Occupational Mobility of Union Army Veterans, 1861-1880,” *Explorations in Economic History*, **44**, **4**, 680-698.
- Lee, Chulhee(2008): “Health, Information, and Migration: Post-Service Geographic Mobility of Union Army Veterans, 1860-1880,” *Journal of Economic History*, **65**, 352-385.
- Lee, Chulhee(2015): “Industrial Characteristics and Employment of Older Manufacturing Workers in the Early-Twentieth-Century United States,” *Social Science History*, **39**, **4**, 551-579.
- Wimmer, L.(2003): “Reflections on the ‘Early Indicator’s Project’: A Partial History,” in D. Costa, (ed.), *Health and Labor Force Participation over the Life Cycle*, Chicago, University of Chicago Press, 1-10.



(1) 2,3,4,6,8,9,10,11,13,14,15,16,17,18,19,20. 좌측은 구.시.읍.면장이 기재하고 확인일자(날짜)한다  
 ※ (2) 8항은직과 (2) (8)과 9항은 연령과 기재한다  
 (3) 이 병적기록표에 관하여는 기재하거나 단조하였을 때는 최종 처리한다

24	검사구분	검사장소	신장	중위	체중	혈압	시력	좌안	시력	우안	색명	청력	좌	우	음부X선	관절	치아	폐격동위	검사년월일	판정판
			164-80								X	X	6401						66-2-7	

62	신장	중위	체중	혈압	시력	좌안	시력	우안	색명	청력	좌	우	음부X선	관절	치아	폐격동위	검사년월일	판정판	
	164	80							X	X	6401							66-2-7	

25	신장	중위	체중	혈압	시력	좌안	시력	우안	색명	청력	좌	우	음부X선	관절	치아	폐격동위	검사년월일	판정판	
									X	X	6401							66-2-7	

27 비 고

<그림 2> 포맷1 형태-뒷면(신체검사란), PID 19460045

12-7229 / 1632

병적기록표

1. 성명	2. 생년월일	3. 혈액형	4. 병력사항	5. 출신지	6. 입원번호
	67				214

7. 직위	8. 직종	9. 직급	10. 직책	11. 직위	12. 직종	13. 직급	14. 직책

15. 신장	16. 중위	17. 체중	18. 혈압	19. 시력	20. 색명	21. 청력	22. 음부X선	23. 관절	24. 치아	25. 폐격동위	26. 검사년월일	27. 판정판
164	80				X	X	6401				66-2-7	

27 비 고

<그림 3> 포맷2 형태, PID 19470055

12-7643 0560

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>① 인번번호</td> <td>② 성명</td> <td>③ 생년월일</td> <td>④ 주민등록번호</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> </table>				① 인번번호	② 성명	③ 생년월일	④ 주민등록번호					<b>병적기록표</b> ⑤ 내역 ⑥ 진단 ⑦ 처치 ⑧ 경과	⑨ 내역서 ⑩ 진단서 ⑪ 처치서 ⑫ 경과서 ⑬ 기타					
① 인번번호	② 성명	③ 생년월일	④ 주민등록번호															
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th>성명</th> <th>성년월일</th> <th>성명</th> <th>성년월일</th> <th>성명</th> <th>성년월일</th> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>				성명	성년월일	성명	성년월일	성명	성년월일							⑭ 진단서 ⑮ 진단서 ⑯ 처치서 ⑰ 경과서 ⑱ 기타		
성명	성년월일	성명	성년월일	성명	성년월일													

⑳ 1973. 8. 21 324신입 146.89.52.5  
 A명 (남) 비정규직 134.14부 3644(병)이신

<그림 4> 포맷3 형태, PID 19520019

12 8006 2067

<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td>① 인번번호</td> <td>② 성명</td> <td>③ 생년월일</td> <td>④ 주민등록번호</td> </tr> <tr> <td></td> <td></td> <td></td> <td></td> </tr> </table>				① 인번번호	② 성명	③ 생년월일	④ 주민등록번호					<b>병적기록표</b> ⑤ 내역 ⑥ 진단 ⑦ 처치 ⑧ 경과	⑨ 내역서 ⑩ 진단서 ⑪ 처치서 ⑫ 경과서 ⑬ 기타					
① 인번번호	② 성명	③ 생년월일	④ 주민등록번호															
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <th>성명</th> <th>성년월일</th> <th>성명</th> <th>성년월일</th> <th>성명</th> <th>성년월일</th> </tr> <tr> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </table>				성명	성년월일	성명	성년월일	성명	성년월일							⑭ 진단서 ⑮ 진단서 ⑯ 처치서 ⑰ 경과서 ⑱ 기타		
성명	성년월일	성명	성년월일	성명	성년월일													

⑳ 1975. 1. 11 71.19.20.18 (남)

<그림 5> 포맷4 형태, PID 19550027

**Abstract**

## Construction and Utilization of a Database on Korean Military Records

Chulhee Lee, Minyoung Do, Sunyoung Pak,  
and Doo-Seung Hong

This paper introduces the primary features of Korean military records kept in the Military Manpower Administration and the database recently constructed from the records, which provides information related to the utilization of the data and collections of similar sources. This article also investigates secular changes in personal characteristics of male cohorts born between 1946 and 1957 using a number of key variables contained in the data, which demonstrates how the data can be utilized. As government administrative data accumulated over a prolonged period, the Korean military records are relatively accurate and representative. These sources are particularly useful for studying socioeconomic characteristics and wellbeing of individuals in the periods for which individual-level micro data are available. Collecting samples of the cohorts born after 1957 and inputting service records into machine-readable form are expected to enhance the value of the data.

**Keywords:** Military Records, Database, Socioeconomic Characteristics, Standards of Wellbeing

