Technical Report 2023-287

Benchmarking and Crosswalks of Army Data from Multiple Sources

Nathaniel J. Ratcliff, https://orcid.org/0000-0003-4291-1884
Joel Thurston, https://orcid.org/0000-0002-3923-9065
Kathryn Linehan, https://orcid.org/0000-0001-9012-6261
Joshua R. Goldstein, https://orcid.org/0000-0002-1164-1829
Vicki A. Lancaster, https://orcid.org/0000-0002-1517-3807
Aaron D. Schroeder, https://orcid.org/0000-0003-4372-2241
Stephanie S. Shipp, https://orcid.org/0000-0002-2142-2136
Sallie Keller (PI), https://orcid.org/0000-0001-7303-7267

University of Virginia,
Biocomplexity Institute and Initiative,
Social and Decision Analytics Division

March 2022

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE L. ZBYLUT, Ph.D.
Director

Research accomplished under contract
for the Department of the Army by

University of Virginia, Biocomplexity Institute & Initiative, Social and Decision Analytics Division

Technical Review by

Kelly S. Ervin, U.S. Army Research Institute

DISPOSITION

This report has been submitted to the
Defense Information Technical Center (DTIC).

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
March 2022	Interim	April 2021 – December 2021

4. TITLE AND SUBTITLE	5a. CONTRACT OR GRANT NUMBER
Benchmarking and Crosswalks of Army Data from Multiple Sources	W911NF1920164
	5b. PROGRAM ELEMENT NUMBER 0602785A

6. AUTHOR(S)	5c. PROJECT NUMBER
Nathaniel J. Ratcliff, Joel Thurston, Kathryn Linehan, Joshua R. Goldstein, Vicki A. Lancaster, Stephanie S. Shipp, Aaron D. Schroeder, and Sallie Keller	5d. TASK NUMBER
	5e. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT
University of Virginia, Biocomplexity Institute and Initiative, Social and Decision Analytics Division (SDAD) 1100 Wilson Blvd, Suite 2910, Arlington, VA 22209	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. MONITOR ACRONYM
U.S. Army Research Institute for the Behavioral & Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop 5610) Fort Belvoir, VA 22060-5610	ARI
	11. MONITOR REPORT NUMBER

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES ARI Research POC: Dr. Kelly Ervin

14. ABSTRACT Benchmarking and assessing the quality of Army data is a foundational step in analyses. This research aims to benchmark U.S. Army data in the Person-Event Data Environment (PDE) data enclave from 2000 to 2019 against publicly available data that document annual numbers of total Army strength and accessions. We complete three sets of comparisons. In the DOD Demographic Report and SDAD Project Crosswalk, we compare top-level counts of Soldiers derived within the PDE from various Army data sources with those reported within annual DOD demographic report profiles. The RAND Accession Report and SDAD Project Crosswalk, compares top-level accession counts of Soldiers derived within the PDE from various Army data sources with a RAND report on Army accessions. Lastly, in the Examination of Missing Accession Dates for Officers, we examine observed missing data patterns of officers for specific accession years and identify their potential source. These analyses revealed that our top-level sample characteristic numbers and trends over time are generally consistent with those reported annually by the DOD demographic reports for total active-duty Army strength and with reports of annual accessions. Data are available with research approvals through the PDE managed by the Research Facilitation Laboratory of the Army Analytics Group.

15. SUBJECT TERMS
Army; benchmarking; accession; demographics; force strength

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	20	Kelly S. Ervin 703-545-2401

BENCHMARKING CROSSWALKS OF ARMY DATA FROM MULTIPLE SOURCES

EXECUTIVE SUMMARY

Research Requirement:

The research described in this technical report is part of a larger research effort supported by a cooperative agreement (#W911NF1920164) between Biocomplexity Institute's Social and Decision Analytics Division (SDAD) at the University of Virginia and the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). The overarching research aims to leverage Department of Defense (DOD) administrative data to model and optimize individual and team performance. This report benchmarks the unclassified but sensitive DOD data used for the research and accessed through the U.S. Army's Person-Event Data Environment (PDE) against publicly available data from published reports that document annual numbers of total Army strength and accessions for a timeframe from 2000 to 2019. Specifically, given that data assets accessed through the PDE are secondary data sourced from various DOD organizations, we wanted to check the data quality of these data assets and assess the degree to which they match with public reports using similar data.

The PDE data enclave is first described. Then, the benchmarking is divided into three major sections. In the DOD Demographic Reports Crosswalk, we present an analysis of top-level counts of Soldiers derived from data accessed within the PDE for our project with those reported within annual DOD demographic report profiles. In the RAND Accession Report Crosswalk, we present a crosswalk analysis of top-level accession counts of Soldiers for our project with a RAND report on Army accessions. Lastly, in the Examination of Missing Accession Dates for Officers, we examine observed missing data patterns of officers for certain accession years and identify their potential source.

Procedure:

This benchmarking effort used a variety of sensitive person-level data sources, including the Active Duty Personnel Master File, Military Entrance Processing Command (MEPCOM) tables, and Active Duty Military Personnel Transaction File covering a time range from September 1999 to December 2019. Individual-level data were linked between tables using PIDs (person identifiers). For each data table, we profiled the variables for their data quality, including structure (e.g., consistency, uniqueness, and linkability) as well as their completeness of coverage over time. We used filing dates, report dates, and test dates to determine the earliest and latest dates represented by the data. Data were examined by cohort year (indicating a Solider was in the Army for a given year) and by accession year (indicating the year a Soldier accessed to the Army). All

computational work was completed in the PDE using the statistical software R 3.6.1. RFL staff reviewed all requests to export aggregate results out of the PDE to ensure confidentiality is maintained.

Findings:

We find that the linked person-level Army data for our project is aligned with data reported in publicly available reports. Specifically, our top-level sample characteristic numbers and trends over time are consistent with those reported annually by the DOD demographic reports (DOD, 2003, 2010, 2018, 2019) for active-duty Army strength. Similarly, our counts of annual accessions are also consistent with those reported from 2000 to 2016 by other DOD-sponsored researchers utilizing the PDE for analysis (Knapp et al., 2018). Taken together, these findings provide strong support for the high quality of Army data available to projects in the PDE for further modeling and data analysis.

Utlizatiation and Dissemination of Findings:

This evidence contributes to the knowledge and utility of DOD administrative data sources made available within restricted access data enclaves such as the PDE. Moreover, our findings allow for confidence in the available data for further modeling and analyzing individual and team Soldier performance. The information gathered in this report helps identify patterns and trends of Soldier characteristics over time that provide context and will inform the modeling of performance. Beyond this report, our research using these data will also provide empirically sound information to enable Army and DOD leaders to develop human capital management policies and programs that address performance and readiness.

BENCHMARKING CROSSWALKS OF ARMY DATA FROM MULTIPLE SOURCES

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Benchmarking Crosswalks of Army Data from Multiple Sources

Background Overview

The research described in this technical report is part of a larger research effort supported by a cooperative agreement (#W911NF1920164) between the Biocomplexity Institute's Social and Decision Analytics (SDAD) Division at the University of Virginia and the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). The overarching research aims to leverage Department of Defense (DOD) administrative data to model and optimize individual and team performance. This report benchmarks the unclassified but sensitive DOD data used for the research and accessed through the U.S. Army's Person-Event Data Environment (PDE) against publicly available data from published reports that document annual numbers of total Army strength and accessions for a timeframe from 2000 to 2019 (see Figure 1). Specifically, given that that data accessed through the PDE are secondary data sourced from various DOD organizations, we wanted to check the data quality of these secondary data assets and assess the degree to which they match with public reports using similar data. The PDE data enclave is first described. Then, the benchmarking is divided into three major sections. In the *DOD Demographic Report Crosswalk*, we present a crosswalk analysis of top-level counts of Soldiers derived from the sensitive person-level secondary for our project with those reported within annual DOD demographic report profiles. In the *RAND Accession Report Crosswalk*, we present a crosswalk analysis of top-level accession counts of Soldiers derived within the PDE for our project with a RAND report on Army accessions. Lastly, in the *Examination of Missing Accession Dates for Officers*, we examine observed missing data patterns of officers for certain accession years and identify their potential source.

Person-Event Data Environment (PDE)

The Army's Person-Event Data Environment (PDE) is a restricted data enclave used to access data sources for this study. The PDE is a secure, remote-access, virtual data enclave that houses unclassified, but sensitive, secondary data assets sourced from various DOD systems and organizations including data about psychological measures, performance indicators, medical information, and administrative personnel records across individual service members' careers (National Academies of Sciences, Engineering, and Medicine, 2017; Vie et al., 2015). The Army Analytics Group's (AAG) Research Facilitation Laboratory (RFL) administers the PDE. It provides access to researchers and institutions with support and approval by Army or DOD sponsors (Knapp et al., 2018a). The research project needs determine what data are made available to researchers in conjunction with negotiation between RFL and the data owners with proper study protocol approvals from an Institutional Review Board (IRB) and the Army Human Research Protections Office (AHRPO). Personally Identifiable Information (PII) is protected in the PDE through a Rosetta process which assigns a unique project specific PDE-generated personal

identifier (PID) to each individual in the data. The PID for a given individual remains the same across data sources for a project enabling the linkage of multiple data types but is different across research projects in the PDE to protect privacy. All data results presented in this report were accessed, analyzed, and summarized within the PDE.

Data Source Characteristics

Data Sources

This benchmarking project used a variety of person-level data sources, including the Active Duty Personnel Master File, Military Entrance Processing Command (MEPCOM) tables, and Active Duty Military Personnel Transaction File covering a time range from September 1999 to December 2019 (see Table 1). Individual-level data were linked between tables using PIDs. For each data table, we profiled the variables for their data quality, including structure (e.g., completeness, consistency, uniqueness, linkability) and completeness of coverage over time. We used filing dates, report dates, and test dates to determine the earliest and latest dates represented by the data. All computational work was completed in the PDE using the statistical software R 3.6.1. RFL staff reviewed all requests to export aggregate results out of the PDE to ensure confidentiality is maintained.

Samples

The individual-level project data were examined in two different ways. Data were examined by *cohort year*, indicating if a Soldier was in the Army at the end of a given DOD fiscal year which starts in October and ends in September (e.g., snapshot date of September 30, 2013). The cohort represents the complete set of active-duty Soldiers in the Army at the end of a given fiscal year which is how data are reported in the DOD demographic reports. For example, for the 2003 cohort year, the cohort could include those who accessed to the Army before the year 2003 (i.e., 2002, 2001, and earlier). Data were also examined by *accession year*, indicating if a Soldier accessed to the Army (i.e., joined) at any point in a given calendar year from 2000 to 2019. Across both sampling frames, warrant officers (n = 598) were excluded given their low numbers at accession and their tendency to be excluded from similar reports due to low sample sizes.

Sample By Cohort Year. The cohort year sample consisted of 1,831,356 active-duty Soldiers who either enlisted or commissioned to the Army (see Table 2). As shown in Table 2, the sample includes 43.6% in a combat arms Military Occupational Specialty (MOS), 14.4% in a combat support MOS, and 28.0% in a combat service support MOS. For the distribution by Soldier sex, 83.6% were male and 16.3% were female. The distribution was 70.9% White, 21.1% Black, 3.7% Asian, 0.9% American Indian or Alaskan Native, 0.3% Native Hawaiian or Pacific Islander, and 3.2% Mixed Race/Other. The distribution of the sample by ethnicity was 86.4% not Hispanic

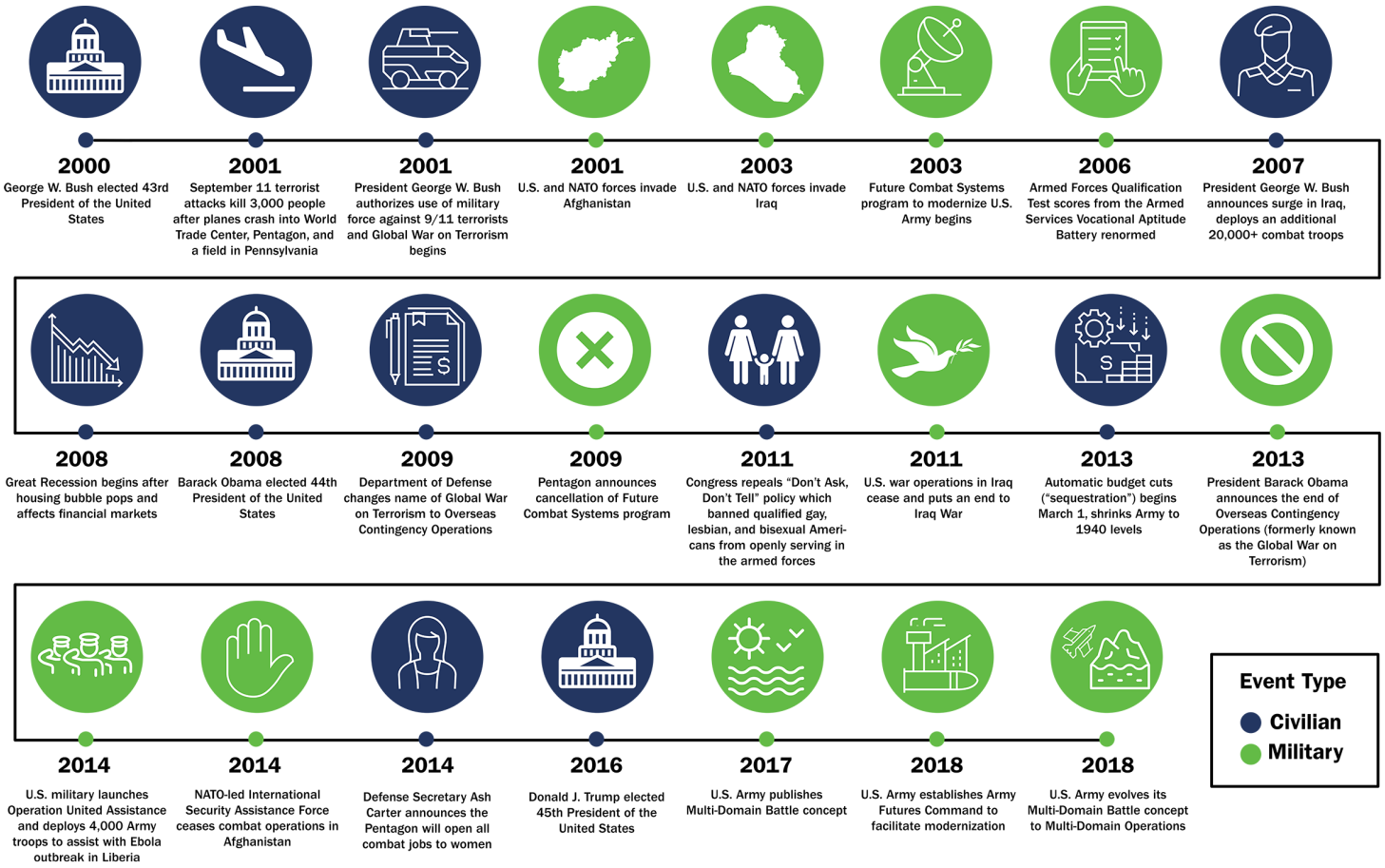
or Latino and 13.6% Hispanic or Latino (Army records of ethnicity started in 2003). The average age upon accession was 22.21 (*SD* = 4.05) years.

Sample By Accession Year. The accession year sample used the same data as the cohort year sample except with filtering criteria applied. The accession year sample consisted of Soldiers selected from the Active Duty Military Personnel Master table who accessed to the U.S. Army sometime between 2000 and 2019 (inclusive). Subjects are active-duty Soldiers who initially joined as enlisted or commissioned as officers. After applying the filtering criteria, the final sample consisted of 1,439,681 Soldiers with 1,356,546 (94.2%) enlistees and 83,135 (5.8%) officers at the time of accession (see Table 2). The sample of officers was lower than expected due to two factors: missing accession dates and a lag in initial record formation. Officers were about ten times more likely to have missing data for accession dates than enlistees (0.9% vs. 0.1%, respectively). Moreover, the rates for missing officer accession dates were especially prevalent from 2000 to 2002 (~3.3%).

Another notable aspect of the data was the lag between the accession date of the Soldier and their first record of entry into the Active Duty Military Personnel Master Table. For example, the first record of the Soldier could appear in 2019, but the actual accession date listed in that first record would be for an earlier date (e.g., 2017-05-24). This lag (in years) was more prominent for officers (*M* = 2.44, *SD* = 2.18) than enlisted (*M* = 0.17, *SD* = 0.82), which explains why the numbers of officers in our sample were much lower for the years 2017 to 2019 as perhaps the Army had not yet entered the records. See the *Examination of Missing Accession Dates for Officers* section below for a detailed breakdown of benchmarking and crosswalk of missing data.

Figure 1

Timeline of Major Events Relevant to the U.S. Army from 2000 to 2019



Note. Event type represents events that were predominately military-driven versus civilian-driven.

Table 1

Data Tables Made Available within the Person-Event Data Environment (PDE) and their Sources

PDE Table Name	Master	MEPCOM 1	MEPCOM 2	Transaction
Source Name	Active Duty Military Personnel Master	Military Entrance Processing Command	Military Entrance Processing Command	Active Duty Military Personnel Transaction
Data Owner	DMDC	DMDC	DMDC	DMDC
PDE Table Name	MV_MASTER_AD_ARMY_QTR_V3A	MEPCOM_USAREC_RA_ANALYST	MV_DMDC_MEPCOM_700_V2	MV_TRANS_AD_ARMY_30_V3A
Description	Administrative records master table.	Initial entry records.	Initial entry records.	Entry and exit status within the Army.
First Date	2001-09-30	2000-10-01	1999-09-30	2001-09-01
Last Date	2019-12-31	2016-07-19	2019-11-30	2018-12-31

Note. DMDC = Defense Manpower Data Center, MEPCOM = Military Entrance and Processing Command.

Table 2

Sample Characteristics of the Cohort Year and Accession Year Sample for Crosswalk Analyses, 2000–2019

Sample Characteristics		Cohort Year Sample	Accession Year Sample		
			Enlisted and Officer	Enlisted	Officer
Sample Size (*N*)		1,831,356	1,439,681	1,356,546	83,135
Rank Group (%)					
	Enlisted	92.30	94.23	100	0
	Officer	7.70	5.77	0	100
MOS Type (%)					
	Combat Arms	43.62	45.42	45.45	44.79
	Combat Support	14.35	13.52	13.68	10.77
	Combat Service Support	28.03	27.08	27.85	14.40
	Special Service	9.18	8.53	7.29	28.73
	Operations	0.63	0.33	0.33	0.25
	Unknown/NA	4.19	5.13	5.38	1.06
Solder Age (in years) at Accession: *M* (*SD*)		22.21 (4.05)	21.82 (4.16)	21.74 (4.09)	23.03 (4.93)
Soldier Sex (%)					
	Male	83.63	83.16	83.51	77.37
	Female	16.33	16.84	16.49	22.63
Soldier Race (%)					
	White	70.89	73.89	73.69	77.00
	Black	21.08	19.20	19.89	8.00
	Asian	3.65	4.08	3.93	6.49
	American Indian/ Alaskan Native	0.86	0.89	0.92	0.48
	Native Hawaiian/ Pacific Islander	0.26	0.33	0.34	0.23
	Mixed Race/Other	3.20	1.60	1.22	7.80
Soldier Ethnicity (%)††					
	Hispanic or Latino	13.58	13.77	13.77	9.60
	Not Hispanic or Latino	86.42	86.23	85.84	90.40
Soldier Highest Education at Accession (%)†					
	Less than H.S.	0.41	0.40	0.43	0.00
	H.S. or Equivalent	80.09	82.08	86.68	7.12
	Some College	2.22	2.58	2.69	0.75
	Associate Degree	2.43	1.83	1.94	0.05
	Bachelor Degree	9.48	9.13	5.18	73.68
	Master Degree	1.60	0.52	0.41	2.30
	Doctorate Degree	1.03	0.75	0.03	12.58
	Unknown	2.75	2.70	2.65	3.51

Note. † = Highest level of education at first available record for individual Soldier; †† = ethnicity records were 4.7% missing overall (0.3% for enlisted and 75.3% for officers) and the Army did not begin to keep ethnicity records until 2003; NA = not applicable or missing data. Data sources are from the DOD tables in the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transaction tables.

DOD Demographics Reports Crosswalk

The purpose of this examination was to better understand the nature of the demographic data (e.g., top line numbers and breakouts by rank, gender, and race) that we have access to in the PDE for our project and how these compare to what has been reported publicly in DOD demographic reports (e.g., DOD, 2003, 2019) from 2000 to 2019.[1] We first describe the data sources used in the crosswalk comparisons and our findings.

Description of Data Sources

- DOD demographic reports: DOD demographic reports are annual reports published by the Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy (ODASD MC&FP) that provide a profile of the military community across the five military branches (e.g., DOD, 2003). As of the date of this technical report, the annual reports available cover a time span from 2000 to 2019. The DOD reports are based on Defense Manpower Data Center (DMDC) snapshot data from the September file date (end of fiscal year) in the Active Duty Master File and do not apply any filtering. That is, DOD demographic reports do not filter out reservists and national guard Soldiers who may temporarily be on 'active' status.
- SDAD Cohort Sample: Based on DMDC snapshot data deposited in the PDE from the September file date (end of fiscal year) in the Active Duty Master File without applying any filtering. This is a direct replication of the DOD demographic report methodology as we use the same data source and no filtering criteria to take out reservists and national guard Soldiers who may temporarily be on 'active' status.
- SDAD Cohort Active Duty Only Sample (Excluding Reservists and National Guard): Based on DMDC snapshot data deposited in the PDE from the September file date (end of fiscal year) in the Active Duty Master File with filtering out of reservists and national guard on active duty status (i.e., active duty only).

Findings

A graphical comparison of the top-level count of Soldiers between all three data sources can be found in Figure 2. Appendix Table A1 presents the aggregate results pulled from the DOD demographic report data from 17 annual reports (the 2003 report contained top-level information for the years 2000, 2001, and 2002). Appendix Table A2 presents the aggregate results from the SDAD cohort sample. Appendix Table A3 presents the aggregate results from the SDAD cohort active duty only sample.

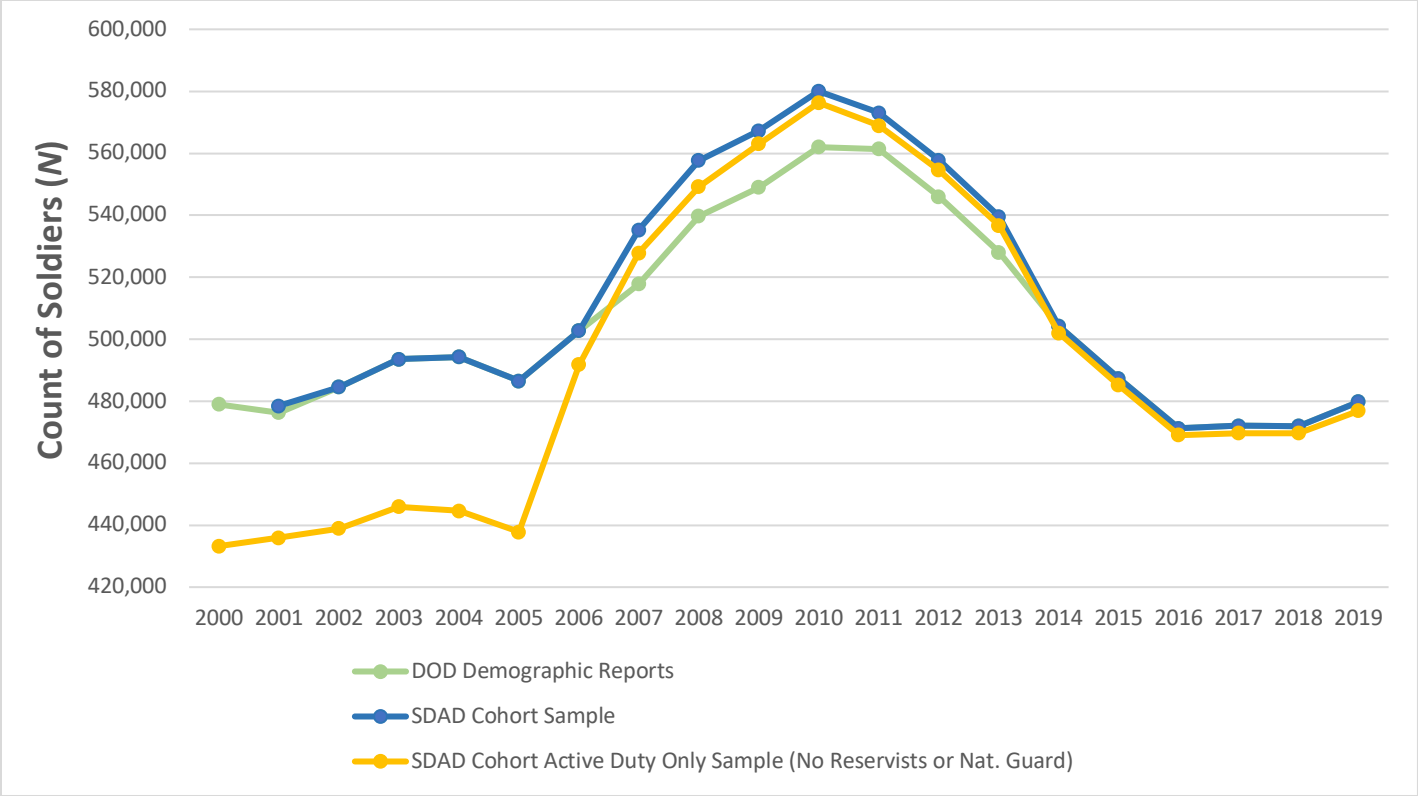
[1] DOD demographics reports for 2003–2015 available from the Office of the Under Secretary for Personnel and Readiness (https://prhome.defense.gov/M-RA/Inside-M-RA/MCFP/Reports/) and Military OneSource (https://www.militaryonesource.mil/) for 2004–2019.

The results reveal several interesting findings:

1. DOD demographic reports have a higher number and percentage of officers than the SDAD Cohort Active Duty Only Sample given that reservists and national guard Soldiers on temporary active duty status are not filtered out and are disproportionally in the officer category [composition of guard and reservists on active status from 2000 to 2019: Overall (N = 100,610), Enlisted (N = 6,541, 6.5%), Officer (N = 63,734, 63.3%), Warrant Officer (N = 30,327, 30.1%), NA (N = 8, 0.0%). This is true for the years from 2000 to 2006 during the start of the Global War on Terrorism in which reservists were likely called upon to serve in active status before recruiting could catch up with the need for Soldiers. See Figure 2 and Appendix Tables A1, A2, and A3.
2. Comparing the DOD demographics report with project cohort sample analysis, numbers were close or identical. See Figure 2 and Appendix Tables A1 and A2.
3. Comparing the DOD demographics report with the project cohort active duty only sample, the project cohort sample data consistently shows fewer active-duty Soldiers from 2000 to 2006, given the filtering of reservists and national guard Soldiers. This aligns with our first finding and the need for reservists to serve in active status until the Army could recruit sufficient numbers of replacements after the start of the Global War on Terror. Across both SDAD cohort samples, we observed a slight increase of counts from 2007 to 2013 (driven solely by enlisted numbers) compared to the DOD demographics report counts and cannot speculate the reason for this slight increase beyond a relation to the surge in Iraq that started in 2007. See Figure 2 and Appendix Tables A1 and A3.

Figure 2

Comparison of Top-Level Soldier Counts between DOD Demographic Reports with SDAD Cohort Sample and SDAD Cohort Active Duty Only Sample (Excluding Reservists and National Guard on 'Active' Status)



Note. DOD = Department of Defense, SDAD = Social and Decision Analytics Division. SDAD samples are drawn from the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transaction tables and depicted in this figure. The data are highlighted in Appendix Tables A1, A2, and A3.

RAND Accession Report Crosswalk

To better understand the nature of the accession data (e.g., top line numbers and breakouts by rank, gender, race), we compared the data our project had available within the PDE to publicly reported data in a 2018 RAND report (see Knapp et al., 2018b) from 2000 to 2016.

Description of Data Sources

- RAND Accession Report: The data in this report are based on aggregated DMDC data on accessions for each fiscal year from 2000 to 2016 (see Knapp et al., 2018b, Table 2.1, p. 11). Though not explicitly stated, numbers appear to represent counts of enlisted Soldiers as the report focuses on this group. Further information on the methodology used in the report was unavailable.
- SDAD Project Accession Year Sample: Based on available Army data available to our project in the PDE, we used accession dates for active-duty Soldiers and filtered out accession dates for reservists and the national guard. We present aggregated accession counts of officers as well as enlisted Soldiers. Accession year was determined via a derived accession date variable coalesced from three different variables and data tables (DATE_ACC (MEPCOM 1 table), AFMS_BASE_DT (Transaction table), and USVC_INT_ENT_DT (Master table).

Findings

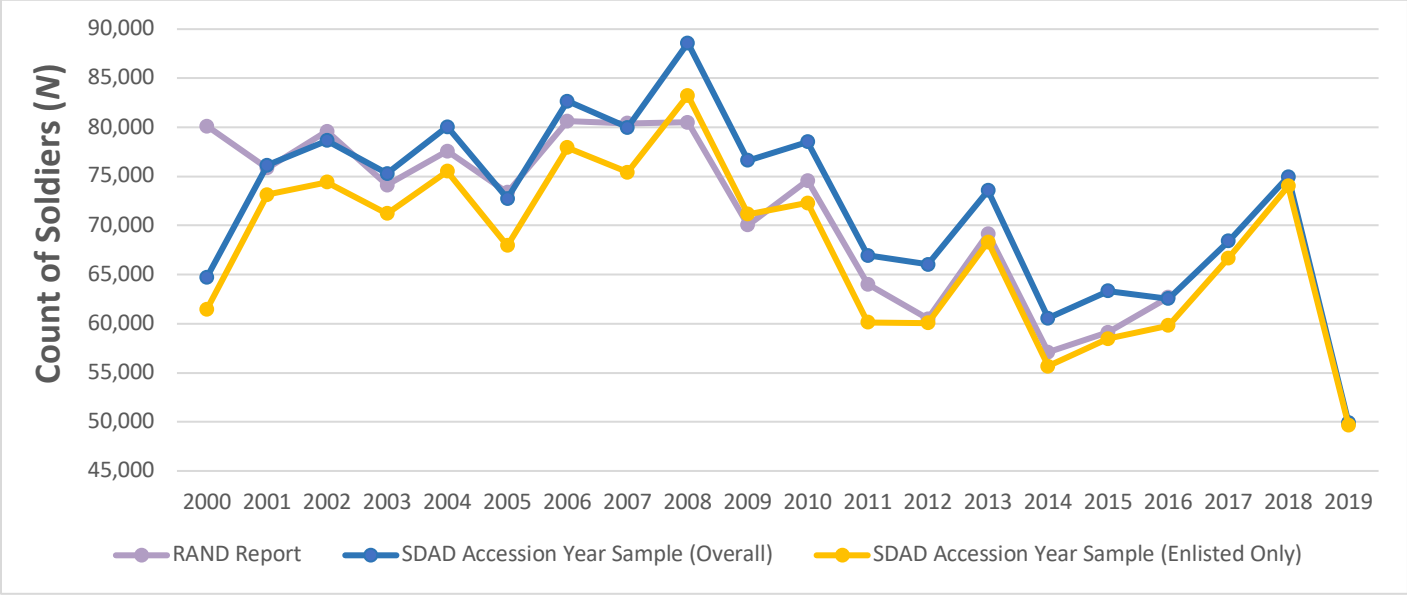
A graphical comparison of the top-level count of accessions between the two data sources can be found in Figure 3. Appendix Table A4 presents the data pulled from the RAND accession report (Knapp et al., 2018b) and Appendix Table A5 presents the SDAD accession year sample counts overall and for officers and enlisted separately.

Results reveal several interesting findings:

1. The RAND accession numbers for enlisted and SDAD accession year sample numbers are close, typically falling within 3,500 Soldiers of one another. See Figure 3 and Appendix Tables A4 and A5.
2. The year 2000 marks a notable exception between the two counts, with SDAD numbers for accessions in that year almost 19,000 Soldiers lower than the RAND report. This may be due to a prevalence of missing data for that year in the SDAD sample. See Figure 3 and Appendix Tables A4 and A5.
3. The SDAD accession numbers for the years 2017 to 2018 for officers are low in comparison to previous years. These two years see fewer than 1,000 officers access each year, in comparison to an average of nearly 5,000 officers per year through the earlier time period. See Appendix Table A5.

Figure 3

Comparison of Top-Level Enlisted Soldier Accession Counts between RAND Accession Report and SDAD Accession Year Sample



Note. SDAD = Social and Decision Analytics Division, Overall = accession counts for officers and enlisted combined, Enlisted only = accession counts only for enlisted Soldiers. The RAND report numbers are assumed to be only for enlisted Soldiers. (Knapp et al., 2018b). SDAD samples are drawn from the DOD tables in the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transactions and depicted in this figure. The data are highlighted in Appendix Tables A4 and A5.

Examination of Missing Accession Dates for Officers

The purpose of this examination was to further understand the nature of the lower number of officers observed to have accessed to the Army in the years 2017 to 2019 in the SDAD samples. We examine data from the individual data sources (i.e., Master table and Transaction table) as well as data from the SDAD sample using derived estimates of accession dates (i.e., SDAD Accession Year sample).

Description of Data Sources

- Master Table Accession Variables: Based on DMDC Active Duty Personnel Master File table using all available Soldier records (see Table 1).
- Transaction Table Accession Variables: Based on DMDC Active Military Personnel Duty Transaction File table using all available Soldier records (see Table 1).
- SDAD Accession Year Sample: Based on accession date of Soldiers for a given calendar year (filters out reservists and national guard Soldiers). Accession year was determined via a derived accession date variable coalesced from three different variables and data tables (DATE_ACC (MEPCOM 1 table), AFMS_BASE_DT (Transaction Military Personnel Duty Transaction File table), and USVC_INT_ENT_DT (Active Duty Personnel Master File table).

Findings

Our results provide several key insights about the accession data for officers available in the PDE (see Tables 3 and 4, Appendix Tables A6 to A11, and Figure 4). Results are first presented by data source with an examination of accession variables within the Master table (Table A6 and A7) and the Transaction table (Table A8 and A9).

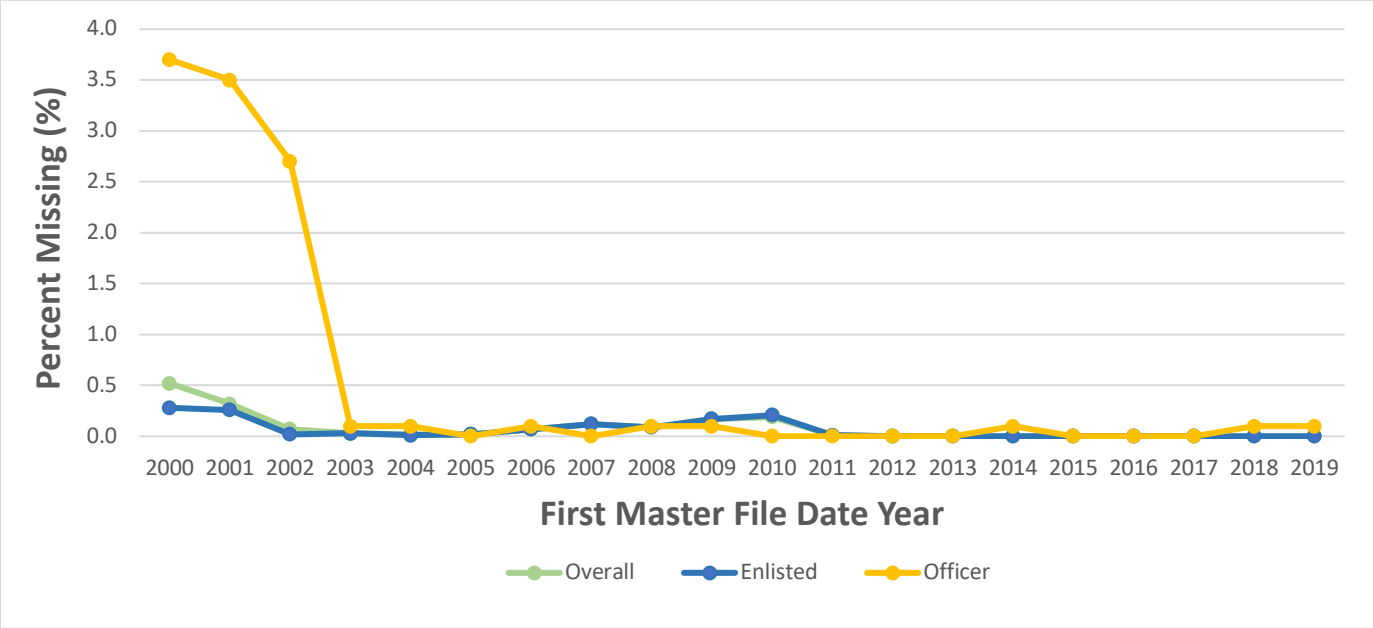
1. Officers have greater number of missing data for accession dates overall (see Appendix Table A10).
2. Missing data on accession dates for officers is more prevalent from 2000 to 2003 (see Figure 4 and Appendix Table A11).
3. There appears to be a lag between when the first file date for a Soldier is made in the master file and their date of accession (e.g., Accession Date: 2015-02-16, First File Date: 2019-12-30; see Table 3). This is supported by a comparison of the first file date in the Master table and Transaction table, with the Transaction table showing a shorter lag between date of the first

recorded Transaction file and our Combined Accession Date (see Table 4). The lag is greater for officers, which explains the low numbers of officers for 2018 and 2019 observed in the accession crosswalk. These numbers are likely not dramatically lower than in previous years, rather the full set of officer records were likely not yet available in the PDE at the time of analysis.

4. The lag for officer data might reflect instances in which officers commission to the Army (i.e., access) but are not truly 'in' the Army to receive a record in the Master table. For example, officers joining special programs like scholarships, ROTC, and delayed entry to complete school might commission at a given accession date but not officially join a unit in the Army until after this time period (Research Facilitation Laboratory, personal communication, June 17, 2021)—which is typically 2–4 years, aligning with the lag of time we observed.
5. Lags for enlisted Soldiers may reflect the time of initial training (i.e., basic training), which is a period of time of about 180 days (Research Facilitation Laboratory, personal communication, June 17, 2021)—this aligns with the lag of time we observed that was around 1–3 months.

Figure 4

Percentage Missing on Combined Accession Variable (DATE_ACC.CB) by Year of First Master Table Filing Date



Note. Overall = data with officer and enlisted Soldiers. Combined accession date variable represents coalescing of three different variables and data tables in the Person-Event Data Environment (PDE): DATE_ACC (MEPCOM 1 table), AFMS_BASE_DT (Transaction table), and USVC_INT_ENT_DT (Active Duty Military Personnel Master File table). Data are highlighted and presented in Table A11.

Table 3

Lag Between First Master Table Record Date and Combined Accession Date

Group	File Date Variable	Accession Date Variable	*M* (Years)	*SD* (Years)
Overall	FILE_DT	DATE_ACC.CB	0.30	1.09
Enlisted	FILE_DT	DATE_ACC.CB	0.17	0.82
Officers	FILE_DT	DATE_ACC.CB	2.44	2.18

Note. Data are from the FILE_DT in the Active Duty Personnel Master File in the Person-Event Data Environment (PDE). DATE_ACC.CB is a derived variable coalesced from accession variables (see Appendix Table A10) in the MEPCOM table, Transaction table, and Master table.

Table 4

Lag Between First Transaction Table Record Date and Combined Accession

Group	File Date Variable	Accession Date Variable	*M* (Years)	*SD* (Years)
Overall	FILE_DT	DATE_ACC.CB	0.47	1.45
Enlisted	FILE_DT	DATE_ACC.CB	0.37	1.27
Officers	FILE_DT	DATE_ACC.CB	2.11	2.62

Note. Data are from the FILE_DT in the Transaction table in the Person-Event Data Environment (PDE). DATE_ACC.CB is a derived variable coalesced from accession variables (see Appendix Table A10) in the MEPCOM table, Transaction table, and Master table.

Conclusion

We find that the Army data that our project has access to in the PDE is aligned with data reported in publicly available reports. Specifically, our top-level sample characteristic numbers and trends over time are consistent with those reported annually by the DOD demographic reports (DOD, 2003, 2010, 2018, 2019) for active-duty Army strength. Similarly, our counts of annual accessions are also consistent with those reported from 2000 to 2016 by other non-DOD researchers utilizing the PDE for analysis (Knapp et al., 2018a). Taken together, these findings provide strong support for the high quality of Army data available to projects in the PDE for further modeling and data analysis.

References

Department of Defense. (2003). *2003 Demographics: Profile of the military community*. Washington, DC. https://prhome.defense.gov/M-RA/Inside-M-RA/MCFP/Reports/

Department of Defense. (2010). *2010 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2010-Demographics-Report.pdf

Department of Defense. (2018). *2018 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2018−demographics−report.pdf

Department of Defense. (2019). *2019 demographics report: Profile of the military community*. Washington, DC. https://download.militaryonesource.mil/12038/MOS/Reports/2019-demographics-report.pdf

Knapp, D., Asch, B. J., DeMartini, C., Ruder, T., & Hanley, J. M. (2018a). *Using the Person-Event Data Environment for military personnel research in the department of defense: An evaluation of capability and potential uses*. RAND Corporation: Santa Monica, CA., https://www.rand.org/pubs/research_reports/RR2302.html

Knapp, D., Orvis, B. R., Maerzluft, C. E., & Berglund, T. (2018b). *Resources required to meet the U.S. Army's enlisted recruiting requirements under alternative recruiting goals, conditions, and eligibility policies*. RAND Corporation: Santa Monica, CA.

National Academies of Sciences, Engineering, and Medicine. (2017). *Strengthening data science methods for department of defense personnel and readiness missions*. National Academies Press.

Vie, L. L., Scheier, L. M., Lester, P. B., Ho, T. E., Labarthe, D. R., Seligman, M. E. P. (2015). The U.S. Army Person-Event Data Environment: A military-civilian big data enterprise. *Big Data*, *3*, 1–13. https://doi.org/10.1089/big.2014.0055

Appendix

Table A1

DOD Demographic Report Demographic Characteristics of the active-duty Army, from 2000 to 2019

Fiscal Year	*N* Overall	*N* Enlisted	*N* Officer	% Enlisted	% Officer	% Female	% Female Enlisted	% Female Officer	% White	% Black	% Asian	% AI/AN	% NH/PI	% MR/O
2000	479,026	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2001	476,289	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2002	484,551	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2003	493,563	413,697	67,953	83.8%	13.8%	15.2%	15.2%	15.0%	60.3%	24.0%	3.5%	0.9%	NA	2.3%
2004	494,291	413,515	68,634	83.7%	13.9%	14.7%	14.6%	15.3%	60.1%	22.7%	3.5%	0.9%	NA	2.4%
2005	486,483	405,275	68,932	83.3%	14.2%	14.3%	14.1%	15.3%	60.8%	21.7%	3.7%	0.9%	NA	2.3%
2006	502,790	420,165	69,616	83.6%	13.8%	14.0%	13.8%	15.3%	61.4%	20.7%	3.8%	1.0%	NA	2.2%
2007	517,783	433,101	70,839	83.6%	13.7%	13.7%	13.4%	15.3%	62.5%	20.1%	3.4%	0.9%	NA	2.4%
2008	539,675	452,065	72,928	83.8%	13.5%	13.6%	13.2%	15.5%	62.7%	19.8%	3.4%	0.9%	NA	2.3%
2009	549,015	458,220	75,619	83.5%	13.8%	13.4%	13.0%	15.8%	69.6%	20.1%	3.6%	0.9%	NA	5.9%
2010	561,979	467,537	78,893	83.2%	14.0%	13.4%	12.9%	16.0%	69.5%	20.2%	3.8%	0.8%	NA	5.7%
2011	561,437	463,886	81,698	82.6%	14.6%	13.5%	13.0%	16.2%	69.3%	20.4%	4.0%	0.8%	NA	5.6%
2012	546,057	447,308	82,856	81.9%	15.2%	13.5%	12.8%	16.2%	69.1%	20.6%	3.7%	0.8%	1.0%	4.9%
2013	528,070	429,103	83,233	81.3%	15.8%	13.6%	13.0%	16.4%	68.5%	21.0%	3.9%	0.8%	1.0%	4.9%
2014	504,330	406,669	82,144	80.6%	16.3%	13.9%	13.2%	16.6%	67.7%	21.6%	4.1%	0.7%	1.0%	4.9%
2015	487,366	392,434	79,878	80.5%	16.4%	14.2%	13.6%	16.8%	67.1%	21.9%	4.4%	0.7%	1.1%	4.8%
2016	471,271	378,835	77,861	80.4%	16.5%	14.6%	14.1%	17.0%	66.7%	22.0%	4.7%	0.7%	1.1%	4.7%
2017	472,047	379,937	77,755	80.5%	16.5%	14.9%	14.3%	17.3%	67.0%	21.8%	4.8%	0.7%	1.1%	4.5%
2018	471,990	379,675	78,046	80.4%	16.5%	15.0%	14.4%	17.5%	67.6%	21.4%	4.9%	0.7%	1.1%	4.3%
2019	479,785	387,075	78,423	80.7%	16.3%	15.3%	14.7%	17.9%	67.6%	21.4%	5.0%	0.7%	1.2%	4.1%

Note. DOD = Department of Defense, AI/AN = American Indian/Alaskan Native, NH/PI = Native Hawaiian/Pacific Islander, MR/O = mixed race/other. Data from DOD demographic reports (2003–2019) each using data from the end of the fourth quarter each fiscal year (i.e., September) and include reservists and guard that are on 'active' status. Highlighted column represents data depicted in Figure 2.

Table A2

SDAD Cohort Sample of DOD Demographic Characteristics of the active-duty Army, from 2000 to 2019

Fiscal Year	*N* Overall	*N* Enlisted	*N* Officer	% Enlisted	% Officer	% Female	% Female Enlisted	% Female Officer	% White	% Black	% Asian	% AI/AN	% NH/PI	% MR/O
2000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2001	478,491	401,849	65,191	84.0%	13.6%	15.5%	15.7%	15.8%	61.7%	26.5%	2.3%	0.8%	NA	8.7%
2002	484,551	406,170	66,578	83.8%	13.7%	15.5%	15.5%	16.0%	62.8%	25.4%	2.5%	0.8%	NA	8.5%
2003	493,563	413,696	67,953	83.8%	13.8%	14.8%	15.2%	16.4%	63.9%	24.4%	2.7%	0.8%	NA	8.1%
2004	494,281	413,501	69,633	83.7%	14.1%	14.7%	14.6%	16.4%	65.7%	23.1%	3.0%	0.9%	NA	7.3%
2005	486,485	405,270	68,932	83.3%	14.2%	14.3%	14.1%	16.7%	67.0%	22.1%	3.2%	0.9%	NA	6.9%
2006	502,790	420,159	69,616	83.6%	13.8%	14.0%	13.8%	16.6%	67.9%	21.1%	3.3%	0.9%	NA	6.3%
2007	535,226	450,544	70,839	84.2%	13.2%	13.6%	13.2%	16.7%	68.8%	20.7%	3.3%	0.9%	NA	6.2%
2008	557,722	470,112	72,928	84.3%	13.1%	13.4%	13.0%	16.9%	69.3%	20.4%	3.4%	0.9%	NA	6.0%
2009	567,308	476,513	75,619	84.0%	13.3%	13.2%	12.8%	17.2%	69.4%	20.3%	3.5%	0.9%	NA	5.9%
2010	580,010	485,568	78,893	83.7%	13.6%	13.2%	12.6%	17.3%	69.3%	20.5%	3.7%	0.9%	NA	5.7%
2011	573,099	475,548	81,698	83.0%	14.3%	13.4%	12.9%	17.5%	69.2%	20.5%	3.9%	0.8%	NA	5.6%
2012	557,815	459,066	82,856	82.3%	14.9%	13.3%	12.7%	17.6%	69.0%	20.7%	3.6%	0.8%	1.0%	4.9%
2013	539,615	440,648	83,233	81.7%	15.4%	13.5%	12.8%	17.7%	68.4%	21.1%	3.8%	0.8%	1.0%	4.9%
2014	504,330	406,699	82,144	80.6%	16.3%	13.9%	13.2%	17.9%	67.7%	21.6%	4.1%	0.7%	1.0%	4.9%
2015	487,366	392,434	79,878	80.5%	16.4%	14.2%	13.6%	18.2%	67.1%	22.0%	4.4%	0.7%	1.1%	4.8%
2016	471,271	378,835	77,861	80.4%	16.5%	14.6%	14.1%	18.4%	66.7%	21.8%	4.7%	0.7%	1.1%	4.7%
2017	472,045	379,935	77,755	80.5%	16.5%	14.9%	14.3%	18.7%	67.0%	21.8%	4.8%	0.7%	1.1%	4.5%
2018	471,989	379,675	78,045	80.4%	16.5%	15.0%	14.4%	19.0%	67.6%	21.4%	4.9%	0.7%	1.1%	NA
2019	479,785	387,075	78,423	80.7%	16.3%	15.3%	14.7%	19.3%	67.6%	21.4%	5.0%	0.7%	1.2%	NA

Note. DOD = Department of Defense, AI/AN = American Indian/Alaskan Native, NH/PI = Native Hawaiian/Pacific Islander, MR/O = mixed race/other. Army data is taken from the end of the fourth quarter of the fiscal year (i.e., September) and does not filter reservists or national guard who may be on 'active' status. SDAD cohort sample data are accessed from the from the DOD tables in the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transaction tables and depicted in this figure. Highlighted column represents data depicted in Figure 2.

Table A3

SDAD Cohort Active Duty Only Sample (Excluding Reservists and Nat. Guard) Demographic Characteristics of the active-duty Army, from 2000 to 2019

Fiscal Year	*N* Overall	*N* Enlisted	*N* Officer	% Enlisted	% Officer	% Female	% Female Enlisted	% Female Officer	% White	% Black	% Asian	% AI/AN	% NH/PI	% MR/O
2000	433,240	396,995	31,907	91.6%	7.4%	15.2%	15.5%	12.0%	60.1%	28.0%	2.1%	0.8%	0.0%	9.0%
2001	435,970	400,744	30,596	91.9%	7.0%	15.3%	15.7%	12.3%	60.7%	27.8%	2.2%	0.8%	0.0%	8.5%
2002	438,921	404,954	29,380	92.3%	6.7%	15.2%	15.5%	12.3%	62.1%	26.7%	2.4%	0.8%	0.0%	8.0%
2003	445,958	413,343	27,876	92.7%	6.3%	14.9%	15.2%	12.6%	63.4%	25.6%	2.4%	0.8%	0.0%	7.8%
2004	444,618	412,770	27,055	92.8%	6.1%	14.4%	14.6%	12.6%	65.6%	24.2%	2.7%	0.8%	0.0%	6.6%
2005	437,832	404,640	28,522	92.4%	6.5%	13.9%	14.1%	12.9%	67.4%	23.1%	3.0%	0.9%	0.0%	5.8%
2006	491,756	426,510	58,840	86.7%	12.0%	14.0%	13.8%	16.3%	69.5%	21.2%	3.2%	0.9%	0.1%	5.1%
2007	527,801	461,127	60,845	87.4%	11.5%	13.6%	13.3%	16.5%	70.3%	20.8%	3.3%	0.9%	0.1%	4.6%
2008	549,238	482,604	61,385	87.9%	11.2%	13.4%	13.1%	16.7%	70.9%	20.6%	3.4%	0.9%	0.1%	4.2%
2009	563,020	495,938	62,332	88.1%	11.1%	13.3%	12.8%	17.0%	71.2%	20.5%	3.5%	0.8%	0.1%	3.9%
2010	576,328	508,703	63,356	88.3%	11.0%	13.2%	12.8%	17.2%	71.2%	20.6%	3.7%	0.8%	0.0%	3.6%
2011	568,932	500,661	64,503	88.0%	11.3%	13.4%	13.0%	17.5%	71.2%	20.7%	4.0%	0.8%	0.0%	3.4%
2012	554,601	486,257	65,256	87.7%	11.8%	13.3%	12.8%	17.6%	70.9%	20.8%	4.1%	0.8%	0.1%	3.3%
2013	536,768	468,712	65,450	87.3%	12.2%	13.5%	12.9%	17.7%	70.4%	21.3%	4.3%	0.7%	0.2%	3.1%
2014	502,006	435,252	64,571	86.7%	12.9%	13.9%	13.3%	18.0%	69.8%	21.8%	4.5%	0.7%	0.3%	3.0%
2015	485,208	420,472	62,984	86.7%	13.0%	14.3%	13.7%	18.3%	69.3%	22.1%	4.8%	0.7%	0.4%	2.8%
2016	469,124	406,395	61,365	86.6%	13.1%	14.6%	14.1%	18.5%	69.0%	22.2%	5.0%	0.7%	0.5%	2.6%
2017	469,663	406,971	61,589	86.7%	13.1%	14.9%	14.4%	18.8%	69.3%	22.0%	5.1%	0.7%	0.6%	2.4%
2018	469,663	407,006	61,517	86.7%	13.1%	15.0%	14.4%	19.2%	69.8%	21.6%	5.1%	0.7%	0.6%	2.2%
2019	476,914	414,955	61,209	87.0%	12.8%	15.3%	14.7%	19.6%	69.9%	21.6%	5.2%	0.7%	0.7%	1.9%

Note. DOD = Department of Defense, AI/AN = American Indian/Alaskan Native, NH/PI = Native Hawaiian/Pacific Islander, MR/O = mixed race/other. Army data is taken from the end of the fourth quarter of the fiscal year (i.e., September) and filters out reservists or national guard who may be on 'active' status. SDAD active duty sample data are from the DOD tables in the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transaction tables and depicted in this figure. Highlighted column represents data depicted in Figure 2.

Table A4

RAND Accession Report from 2000 to 2016

Fiscal Year	*N* Overall End Strength	*N* Enlisted End Strength	*N* Officer End Strength	% Enlisted End Strength	% Officer End Strength	*N* Actual Enlisted Accessions	% HQ Recruits
2000	482,170	401,414	80,756	83.3%	16.7%	80,113	53.6%
2001	480,801	400,461	80,340	83.3%	16.7%	75,855	52.5%
2002	486,543	404,305	82,238	83.1%	16.9%	79,585	57.3%
2003	499,301	414,769	84,532	83.1%	16.9%	74,132	60.0%
2004	499,543	412,324	87,219	82.5%	17.5%	77,586	60.9%
2005	492,728	406,923	85,805	82.6%	17.4%	73,373	56.2%
2006	505,402	419,353	86,049	83.0%	17.0%	80,635	45.2%
2007	522,017	433,109	88,908	83.0%	17.0%	80,407	44.2%
2008	543,645	451,846	91,799	83.1%	16.9%	80,517	45.6%
2009	553,044	457,980	95,064	82.8%	17.2%	70,045	54.1%
2010	566,045	467,248	98,797	82.5%	17.5%	74,577	61.4%
2011	565,463	463,605	101,858	82.0%	18.0%	64,019	60.6%
2012	550,064	447,075	102,989	81.3%	18.7%	60,490	60.0%
2013	532,426	428,923	103,503	80.6%	19.4%	69,154	60.5%
2014	508,210	406,519	101,691	80.0%	20.0%	57,101	57.6%
2015	491,365	392,327	99,038	79.8%	20.2%	59,117	57.9%
2016	475,400	378,778	96,622	79.7%	20.3%	62,682	55.8%
2017	NA	NA	NA	NA	NA	NA	NA
2018	NA	NA	NA	NA	NA	NA	NA
2019	NA	NA	NA	NA	NA	NA	NA

Note. HQ = high-quality. Officers = both commissioned officers and warrant officers, NA = data not available from RAND report. The RAND report did not explicitly specify if their top-level accession numbers were only representing enlisted accessions but given that the report was related to enlisted reporting requirements, an enlisted focus is assumed here *(Knapp et al., 2018b).* Highlighted column represents data depicted in Figure 3.

Table A5

SDAD Accession Year Sample Demographic Characteristics of the active-duty Army, from 2000 to 2019

Fiscal Year	*N* Overall	*N* Enlisted	*N* Officer	% Enlisted	% Officer	% Female	% Female Enlisted	% Female Officer	% White	% Black	% Asian	% AI/AN	% NH/PI	% MR/O	% HQ Recruits
2000	64,723	61,478	3,173	95.0%	4.9%	19.9%	19.9%	20.6%	65.5%	23.6%	2.7%	1.1%	0.0%	7.1%	60.7%
2001	76,145	73,127	2,968	96.0%	3.9%	19.0%	19.0%	19.7%	69.8%	21.0%	2.8%	1.0%	0.0%	5.5%	62.0%
2002	78,652	74,402	4,192	94.6%	5.3%	17.7%	17.5%	21.4%	73.6%	17.8%	3.2%	1.0%	0.0%	4.4%	65.8%
2003	75,305	71,202	4,049	94.6%	5.4%	17.9%	17.6%	23.1%	76.0%	15.4%	2.9%	1.0%	0.0%	4.7%	66.2%
2004	80,044	75,532	4,483	94.4%	5.6%	17.0%	16.7%	22.5%	78.2%	15.8%	4.2%	1.0%	0.0%	0.8%	67.1%
2005	72,751	67,956	4,756	93.4%	6.5%	16.3%	16.0%	21.0%	79.8%	14.4%	3.8%	1.1%	0.0%	0.9%	60.1%
2006	82,649	77,928	4,689	94.3%	5.7%	16.3%	16.0%	21.2%	78.7%	15.6%	3.6%	1.1%	0.0%	1.0%	54.8%
2007	79,990	75,435	4,480	94.3%	5.6%	15.9%	15.6%	20.9%	79.2%	15.5%	3.4%	1.0%	0.0%	0.9%	58.3%
2008	88,590	83,253	5,271	94.0%	5.9%	15.8%	15.5%	21.7%	77.8%	17.0%	3.5%	0.9%	0.0%	0.8%	57.6%
2009	76,621	71,136	5,456	92.8%	7.1%	15.7%	15.1%	23.2%	77.2%	17.1%	4.2%	0.8%	0.0%	0.7%	60.7%
2010	78,507	72,284	6,207	92.1%	7.9%	16.0%	15.5%	22.2%	75.4%	18.4%	4.9%	0.8%	0.0%	0.6%	59.1%
2011	66,958	60,127	6,816	89.8%	10.2%	15.9%	15.2%	21.9%	74.7%	18.8%	4.8%	0.8%	0.5%	0.8%	60.0%
2012	66,026	60,074	5,924	91.0%	9.0%	15.1%	14.4%	22.3%	73.7%	20.0%	4.2%	0.7%	0.8%	0.7%	60.6%
2013	73,578	68,316	5,256	92.8%	7.1%	15.5%	15.1%	21.2%	70.4%	22.6%	4.8%	0.7%	0.9%	0.5%	59.3%
2014	60,566	55,675	4,874	91.9%	8.0%	16.0%	15.4%	22.9%	69.8%	22.8%	5.1%	0.8%	1.0%	0.4%	59.2%
2015	63,327	58,457	4,864	92.3%	7.7%	17.6%	16.9%	26.4%	68.8%	23.1%	5.6%	0.8%	1.0%	0.2%	57.2%
2016	62,562	59,808	2,747	95.6%	4.4%	17.3%	16.9%	26.8%	69.8%	23.2%	4.8%	0.8%	1.0%	0.1%	57.1%
2017	68,422	66,690	1,730	97.5%	2.5%	16.8%	16.4%	29.6%	71.2%	22.1%	4.6%	0.8%	1.0%	0.0%	58.1%
2018	74,967	74,006	961	98.7%	1.3%	17.5%	17.3%	35.8%	72.3%	21.0%	4.7%	0.8%	0.9%	0.0%	61.2%
2019	49,899	49,660	239	99.5%	0.5%	18.2%	18.1%	41.8%	68.8%	24.6%	5.0%	0.8%	0.8%	0.0%	57.3%

Note. AI/AN = American Indian/Alaskan Native, NH/PI = Native Hawaiian/Pacific Islander, MR/O = mixed race/other. Army data uses data by the calendar year in which a Soldier accessed to the Army and filters out reservists or national guard who may be on 'active' status. SDAD accession year sample data are from the DOD tables in the Person-Event Data Environment (PDE): Active Duty Military Personnel Master, Military Entrance Processing Command, and Active Duty Military Personnel Transaction tables. Highlighted columns represent data depicted in Figure 3.

Table A6

Master Table Accession Variables and Percentage Missing

Variable Name	PDE Name	PDE Data Catalogue Description	% NA Overall	% NA Enlisted	% NA Officer
Active Federal Military Service Base Date	AFMS_BASE_DT	The date for which DoD Military Service member's creditable Active Military Service begins. This constructed date functions to indicate a date on which a DoD Military Service member's creditable active military service begins for calculating time.	0.1	0.2	8.9
Uniformed Service Initial Entry Date	USVC_INT_ENT_DT	The date for which the member was first appointed, enlisted, or conscripted into a Uniformed Service of the US. Also referred to as Date of Initial Entry to a Uniformed Service. The data is received daily and monthly from data feeds for employment reporting.	2.2	3.0	6.8
Active Federal Military Service Base Date	AFM_OSVC_BASE_DT	The date for which DoD Military Service member's creditable Active Military Service begins. This constructed date functions to indicate a date on which a DoD Military Service member's creditable active military service begins for calculating time.	86.2	100.0	35.5
File Date	FILE_DT	The data is created daily and monthly by DMDC for reporting.	0.0	0.0	0.0

Note. % NA = percent missing. All variables sourced from the Master table in the Person-Event Data Environment (PDE) with percent missing calculated by SDAD using all available Soldier records. Data catalogue descriptions taken verbatim accessed from the PDE.

Table A7

Master Table Accession Variables and Percentage Missing for Officers by Master Table File Date Year

Variable Name	PDE Name	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Active Federal Military Service Base Date	AFMS_BASE_DT	0.9	2.4	4.6	0.6	0.2	1.6	3.5	8.1	6.5	4.0	6.4	14.8	10.6	19.1	16.7	20.7	19.0	28.7	31.8	43.1
Uniformed Service Initial Entry Date	USVC_INT_ENT_DT	28.6	13.9	9.6	0.9	0.7	0.0	0.2	0.1	0.4	0.1	0.1	0.6	0.7	0.4	0.2	0.7	0.1	0.2	0.1	0.1
Active Federal Military Service Base Date	AFM_OSVC_BASE_DT	89.6	25.6	2.12	0.42	0.31	0.0	13.9	0.3	0.8	0.8	0.7	0.5	0.6	0.4	0.32	31.9	92.3	87.5	93.5	91.5

Note. All variables sourced from the Master table in the Person-Event Data Environment (PDE) with percent missing calculated by SDAD using all available Soldier records.

Table A8

Transaction Table Accession Variables and Percentage Missing

Variable Name	PDE Name	PDE Data Catalogue Description	% NA Overall	% NA Enlisted	% NA Officer
Active Federal Military Service Base Date	AFMS_BASE_DT	The date for which DoD Military Service member's creditable Active Military Service begins. This constructed date functions to indicate a date on which a DoD Military Service member's creditable active military service begins for calculating time.	2.3	0.2	45.8
Active Federal Military Officer Service Base Date	OFCR_APNT_DT	The date for which DoD Military Service member's creditable Active Military Service as an officer begins. This constructed date indicates a date on which a DoD Military Service officer's creditable active military service begins for calculating time. Also referred to as Basic Officer Service Date. The data is the actual or adjusted date from which the amount of active military service performed is calculated. The data is received daily and monthly from personnel data feeds. The data is used for human resources actions and updating DEERS.	97.8	100.0	66.7

Note. % NA = percent missing. All variables are sourced from the Transaction table in the Person-Event Data Environment (PDE) with percent missing calculated by SDAD using all available Soldier records. Data catalogue descriptions taken verbatim from PDE.

Table A9

Transaction Table Accession Variables and Percentage Missing for Officers by Transaction Table File Date Year

Variable Name	PDE Name	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Active Federal Military Service Base Date	AFMS_BASE_DT	100.0	30.5	30.2	17.5	37.6	32.8	44.6	44.1	44.1	57.6	49.7	12.8	49.8	64.2	60.2	56.5	60.4	69.0	69.8	100.0
Active Federal Military Officer Service Base Date	OFCR_APNT_DT	100.0	100.0	100.0	82.6	60.8	63.8	97.0	98.5	98.2	95.2	94.4	94.1	87.0	93.2	17.8	0.8	0.4	0.6	0.3	100.0

Note. All variables are sourced from the Transaction table in the Person-Event Data Environment (PDE) with percent missing calculated by SDAD using all available Soldier records.

Table A10

SDAD Accession Year Sample Accession Variables and Percentage Missing

Variable Name	Table Name	PDE Name	PDE Data Catalogue Description	% NA Overall	% NA Enlisted	% NA Officer
Actual Accession Date	MEPCOM 1	DATE_ACC	Actual accession date.	43.27	35.5	99.9
Active Federal Military Service Base Date	Transaction	AFMS_BASE_DT	The date for which DoD Military Service member's creditable Active Military Service begins. This constructed date functions to indicate a date on which a DoD Military Service member's creditable active military service begins for calculating time.	11.19	8.6	42.3
Uniformed Service Initial Entry Date	Master	USVC_INT_ENT_DT	The date for which the member was first appointed, enlisted, or conscripted into a Uniformed Service of the US. Also referred to as Date of Initial Entry to a Uniformed Service. The data is received daily and monthly from data feeds for employment reporting.	3.33	3.0	6.8
Combined Accession Date	Derived	DATE_ACC.CB	Coalesced variable with DATE_ACC, AFMS_BASE_DT, and USVC_INT_ENT_DT	0.17	0.11	0.9

Note. % NA = percent missing. Variables sourced from the from the DOD tables in the Person-Event Data Environment (PDE): MEPCOM 1, Transaction, and Master tables with percent missing calculated by SDAD using accession year sample.

Table A11

SDAD Accession Year Sample Accession Variables and Percentage Missing for Officers by Master Table File Date Year

Variable Name	PDE Name	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Combined Accession Date (Overall)	DATE_ACC.CB	0.5	0.3	0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Combined Accession Date (Enlisted)	DATE_ACC.CB	0.3	0.3	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Combined Accession Date (Officer)	DATE_ACC.CB	3.7	3.5	2.7	0.1	0.1	0.0	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1

Note. Data are from the DOD tables in the Person-Event Data Environment (PDE): DATE_ACC.CB sourced from the MEPCOM 1, Transaction, and Master tables using a combined estimate of accession date. Percent missing of the combined DATE_ACC.CB variable is calculated by SDAD using accession year sample with a breakdown by a Soldier's first file date year in the Master table. Highlighted rows are depicted in Figure 4.

About the University of Virginia's Social and Decision Analytics Division

The **Social and Decision Analytics Division (SDAD)** is a leading Division in the Biocomplexity Institute at the University of Virginia. The Biocomplexity Institute is at the forefront of a scientific evolution, applying a deeply contextual approach to answering some of the most pressing challenges to human health and well-being within our changing environment. SDAD was created in the fall of 2013 to extend the Biocomplexity Institute's capabilities in social informatics, policy analytics, and program evaluation. The researchers at SDAD form a multidisciplinary team, with expertise in statistics, policy and program evaluation, economics, political science, psychology, computational social science, and data governance and information architecture. SDAD's mission is to embrace today's data revolution, developing evidence-based research and quantitative methods to inform policy decision-making and evaluation.