Leveraging U.S. Army Administrative Data for Individual and Team Performance

Nathaniel J. Ratcliff  
*US Army Research Institute*

Kelly S. Ervin  
*US Army Research Institute*

Joshua Goldstein  
*University of Virginia*

Vicki Lancaster  
*University of Virginia*

Sallie Keller  
*University of Virginia*

See next page for additional authors

Follow this and additional works at: https://repository.upenn.edu/admindata_conferences_presentations_2018

Ratcliff, Nathaniel J.; Ervin, Kelly S.; Goldstein, Joshua; Lancaster, Vicki; Keller, Sallie; Shipp, Stephanie; and Thurston, Joel, "Leveraging U.S. Army Administrative Data for Individual and Team Performance" (2018).  
https://repository.upenn.edu/admindata_conferences_presentations_2018/21

DOI https://doi.org/10.23889/ijpds.v3i5.1086

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/admindata_conferences_presentations_2018/21

For more information, please contact repository@pobox.upenn.edu.
Leveraging U.S. Army Administrative Data for Individual and Team Performance

Abstract
The Army possesses vast amounts of administrative (archival) data about Soldiers. These data sources include screening tests, personnel action codes, training scores, global assessments, physical fitness scores, and more. However, the Army has yet to integrate these data to create a holistic operating picture. Our research focuses on repurposing Army administrative data to (1) operationalize social constructs of interest to the Army (e.g., Army Values, Warrior Ethos) and (2) model the predictive relationship between these constructs and individual (i.e., Soldier) and team (i.e., unit) performance and readiness. The goal of the project is to provide people analytics models to Army leadership for the purposes of optimizing human capital management decisions.

Our talk will describe the theoretical underpinnings of our human performance model, drawing on disciplines such as social and industrial/organizational psychology, as well as our experience gaining access to and working with Army administrative data sources. Access to the archival administrative data is provided through the Army Analytics Group (AAG), Person-event Data Environment (PDE). The PDE is a business intelligence platform that has two central functions: (1) to provide a secure repository for data sources on U.S. military personnel; and (2) to provide a secure collaborative work environment where researchers can access unclassified but sensitive military data.

Comments
DOI https://doi.org/10.23889/ijpds.v3i5.1086

Author(s)
Nathaniel J. Ratcliff, Kelly S. Ervin, Joshua Goldstein, Vicki Lancaster, Sallie Keller, Stephanie Shipp, and Joel Thurston

This presentation is available at ScholarlyCommons: https://repository.upenn.edu/admindata_conferences_presentations_2018/21
Leveraging U.S. Army Administrative Data for Individual and Team Performance

Presentation for:

Administrative Data Research Facilities Network (ADRF) Research Conference: Innovations in Administrative Data

Nathaniel J. Ratcliff, PhD
Kelly S. Ervin, PhD
U.S. Army Research Institute for the Behavioral and Social Sciences (ARI)

Joshua Goldstein, PhD
Vicki Lancaster, PhD
Sallie Keller, PhD
Stephanie Shipp, PhD
University of Virginia, Social and Decision Analytics Division (SDAD)

Joel Thurston, PhD
Center for Army Leadership (CAL)
U.S. Army Combined Arms Center

November 14, 2018
**MISSION:** Drive scientific innovation to enable the Army to acquire, develop, employ, and retain professional Soldiers and enhance personnel readiness.

**Soldier-Oriented S&T**

Develop innovative measures, methods, and models to maximize personnel and unit readiness of the Future Army.

**Basic Research**

Develop fundamental theories and investigate new domain areas in behavioral and social sciences with high potential impact on Army issues.

**Scientific Assessments**

Conduct scientific assessments and provide actionable findings to inform Army leaders and policy.

**ARI Mission**

**ARI: Where Personnel Science Meets Personnel Practice**
The Problem Space

- **Problem**: The Army possesses a trove of administrative data (e.g., personnel records, training scores) but has yet to fully leverage these data.

- **Purpose**: Use state-of-the-art analytics to develop models that integrate existing DOD data and make predictions about Soldier behavior and performance. This research will provide statistical models that forecast and visualize individual and unit performance.

- **Payoff**: Knowledge about how best to utilize data from disparate sources to form a holistic picture of Soldier and unit performance.
  
  - Streamline/Reduce training
  - Identify informative and efficient metrics of performance
  - Optimize talent management decisions across Soldier lifecycle
Leveraging U.S. Army Administrative Data 
for Individual and Team Performance

Joshua Goldstein, PhD

Biocomplexity Institute & Initiative
Data Science Framework

Problem Identification: Relevant Theories & Working Hypotheses

Data Discovery

ADMINISTRATIVE DATA, DESIGNED DATA, OPPORTUNITY DATA, & PROCEDURAL DATA

Inventory, Screening & Acquisition

DATA INGESTION & GOVERNANCE

DATA PROFILING
Data Structure, Data Quality, Provenance & Meta Data

DATA PREPARATION
Cleaning, Transformation, Restructuring

DATA LINKAGE
Ontology, Selection & Alignment, Entity Resolution

DATA EXPLORATION
Characterization, Summarization, Visualization

FITNESS-FOR-USE-ASSESSMENT

STATISTICAL MODELING & DATA ANALYSES
Data Science Framework

Problem Identification: Relevant Theories & Working Hypotheses

Data Discovery

ADMINISTRATIVE DATA,
DESIGNED DATA, OPPORTUNITY DATA,
& PROCEDURAL DATA

Inventory, Screening & Acquisition

DATA INGESTION & GOVERNANCE

DATA PROFILING
Data Structure, Data Quality, Provenance & Meta Data

DATA PREPARATION
Cleaning, Transformation, Restructuring

DATA LINKAGE
Ontology, Selection & Alignment, Entity Resolution

DATA EXPLORATION
Characterization, Summarization, Visualization

FITNESS-FOR-USE-ASSESSMENT

STATISTICAL MODELING & DATA ANALYSES
Literature – Relevant Theories

**Integrity**

**Definition 1**
Integrity: Honesty, Genuineness, Trustful

**Definition 2**
Integrity [authenticity, honesty]

**Definition 3**
Speaking the truth but. More broadly presenting oneself in a genuine way and action in a sincere way; being without pretense; taking responsibility for one’s feelings and actions.
Data Science Framework

Problem Identification: Relevant Theories & Working Hypotheses

Data Discovery

DATA INGESTION & GOVERNANCE

ADMINISTRATIVE DATA, DESIGNED DATA, OPPORTUNITY DATA, & PROCEDURAL DATA

Inventory, Screening & Acquisition

DATA PROFILING
Data Structure, Data Quality, Provenance & Meta Data

DATA PREPARATION
Cleaning, Transformation, Restructuring

DATA LINKAGE
Ontology, Selection & Alignment, Entity Resolution

DATA EXPLORATION
Characterization, Summarization, Visualization

FITNESS-FOR-USE-ASSESSMENT

STATISTICAL MODELING & DATA ANALYSES
Administrative Data Sources

Person-Event Data Environment (PDE)

- DOD maintains numerous datasets about military personnel including deployments, demographics, accessions/attrition, pay, promotions, awards, and training records.

- Army Analytics Group Research Facilitation Lab (AAG-RFL) provides access to many of these datasets through the PDE in a secure, cloud-based enclave.

- Accessing the data requires:
  - Obtain a Common Access Card (CAC)
  - Human Research Protections/IRB approvals
  - Approval to access the PDE
Data Discovery – DOD Data Sources

Accessible in the PDE:

- Digital Training Management System
- Army Training and Requirements Resource System
- Individual Training History / DOD Formal Course File
- Army Workforce Transaction File: Military Awards
- Unit Risk Inventory
- Defense Equal Opportunity Management Institute (DEOMI) Organization Climate Survey
- Omaha 5 Behavioral Survey
- Active Duty Military Personnel Master
- Active Duty Military Personnel Transaction
- United States Military Entrance Processing Command (MEPCOM) Regular Army Analyst
- Interactive Personnel Elective Records Management System (IPERMS)
- Army Career and Alumni Program (ACAP)
- Pre-Deployment, Post-Deployment, Periodic Health Assessments
- Global Assessment Tool (GAT)
Data Discovery – Non-DOD Data Sources

Take a systematic approach in selecting from a huge inventory of public data sources and archives. Documented accessibility of data, geographies of interest, and relevant variables.

**American Community Survey**
- Factfinder, Integrated Public Use Microdata Sample (IPUMS)
- Aunt Bertha’s
- Bureau of Economic Analysis

**Bureau of Labor Statistics**
- Quarterly Census of Employment and Wages (QCEW), Consumer Price Index (CPI), Current Employment Statistics (CES), etc.

**Community Commons**

**County Health Rankings**
- Gapminder World
- Global Health Data Exchange
- Headwater Economics

**International Household Survey Network**

**Integrated Postsecondary Education Data System (IPEDS)**
- Postsecondary Education
- IRS Tax Records
- Penn State Simple Online Data Archive
- MIT Living Wage Calculator

**National Center for Education Statistics**
- Population Reference Bureau

**Simply Analytics**
- Statistical Atlas
- U.S. Government Open Data
- World Bank Open Data
# Steps in the Data Quality Analysis Process

<table>
<thead>
<tr>
<th>Completeness</th>
<th>Proportion of Elements Properly Populated for a Given Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Issue types include: Record fields containing no data; records not containing necessary fields; datasets not containing the requisite records (e.g., testing for NULLs and empty strings existing where not appropriate)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Validity</th>
<th>Proportion of Elements whose Attributes Possess Proper Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Checking for value validity generally comes in the form of straight-forward domain constraint rules (e.g. where gender is not one of (male, female), or where age is not between ([0, 110]))</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Uniqueness</th>
<th>Count of Unique Values Taken by an Attribute or Combination of Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Frequency distribution of an element. (Note: The more homogeneous the data values of an element, the less useful the element is for analysis)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Duplication</th>
<th>Degree of Replication of Distinct Observations Per Observation Unit Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• For example, greater than one registration per student per official reporting period.</td>
</tr>
<tr>
<td></td>
<td>• Note: Duplication occurs as a result of choice of level of aggregation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consistency (Record)</th>
<th>Degree to which two or More Data Attributes Satisfy a Dependency Constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Relationship validation</td>
</tr>
<tr>
<td></td>
<td>• For example, zip-code – state consistency; gender – pregnancy consistency</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Consistency (Longitudinal)</th>
<th>Degree an Attribute, Combination of Attributes, Remains Consistent Over Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• For example, an individual's gender changing and then changing back (clerical error) or an individual's race classification changing from one to two races (change in number of options)</td>
</tr>
</tbody>
</table>
Demographic Log Analysis & Reduction

Basic demographics between data files need to be reconciled and reusable data products created.

**Demographics Table**
- Information about the enlistedee that typically remains static over time (e.g., gender, race, ethnicity, entry test scores)
- Simple rules are applied to resolve duplicates and invalid entries
- Contains one row per personal identifier (PID)

**Transaction Table**
- Events or enlistedee information that can change periodically (e.g., duty station, rank, pay grade)
- Contains multiple rows per PID

<table>
<thead>
<tr>
<th>Column Name</th>
<th>Description</th>
<th>Original Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>PID_PDE</td>
<td>Enlistee's Unique ID</td>
<td>Master</td>
</tr>
<tr>
<td>PN_SEX_CD</td>
<td>Gender</td>
<td>Master</td>
</tr>
<tr>
<td>RACE_CD</td>
<td>Race Code</td>
<td>Master</td>
</tr>
<tr>
<td>INIT_ENT_TRN_END_DT</td>
<td>Initial Entry Training End Date</td>
<td>Master</td>
</tr>
<tr>
<td>DATE_BIRTH_PDE</td>
<td>Person Birth Date</td>
<td>Master</td>
</tr>
<tr>
<td>PN_BIRTH_PLCT_CTRY_CD</td>
<td>Person Birth Place Country Code</td>
<td>Master</td>
</tr>
<tr>
<td>HOR_ZIP_CODE_PDE</td>
<td>Home of Record Zip Code</td>
<td>Analyst</td>
</tr>
<tr>
<td>ACT_SCORE</td>
<td>ACT Score</td>
<td>Analyst</td>
</tr>
<tr>
<td>SAT_SCORE</td>
<td>SAT Score</td>
<td>Analyst</td>
</tr>
<tr>
<td>AP</td>
<td>ASVAB: Auditory Perception Score</td>
<td>Analyst</td>
</tr>
<tr>
<td>CO</td>
<td>ASVAB: Combat Score</td>
<td>Analyst</td>
</tr>
</tbody>
</table>
Conceptual Model of the Social Characteristics of Performance

ANTECEDENTS OF PERFORMANCE: Army Values, Warrior Ethos, Big 5 Personality Traits, Creativity, Motivation, Job & Community Embeddedness, Vocational Interests

DIRECT DETERMINANTS

| Cognitive Component | Social Component | Physical Component |

BEHAVIORS OF PERFORMANCE

Productive Behaviors
Actions/behaviors that contribute to the organization’s goals anchored at the floor by achieving the minimum standard needed to meet the contractual obligations of the job.

Impacts
Organization, Colleagues, & Self

Counterproductive Behavior
Actions/behaviors that impede the progress towards an organization’s goals and can also harm self and colleagues.

CRITERIA OF PERFORMANCE (Outcomes)
Speed of Promotion, Training as a Reward, Evaluation Reports
Linking Data to the Performance Framework

**ANTECEDENTS OF PERFORMANCE:** Army Values, Warrior Ethos, Big 5 Personality Traits, Creativity, Motivation, Job & Community Embeddedness, Vocational Interests

**DIRECT DETERMINANTS**
- Cognitive Component
- Social Component
- Physical Component

**BEHAVIORS OF PERFORMANCE**
- **Productive Behaviors**
  Actions/behaviors that contribute to the organization’s goals anchored at the floor by achieving the minimum standard needed to meet the contractual obligations of the job.
- **Counterproductive Behavior**
  Actions/behaviors that impede the progress towards an organization’s goals and can also harm self and colleagues.

**CRITERIA OF PERFORMANCE (Outcomes)**
- Speed of Promotion
- Training as a Reward
- Evaluation Reports

**Global Assessment Tool**
- External Community Data (e.g. Army Community Service (ACS), BLS)
- Armed Services Vocational Aptitude Battery (ASVAB) scores
- Global Assessment Tool (GAT)
- DEOMI Command Climate Survey Unit Risk Inventory
- Digital Training Management System (DTMS) Height and Weight
- DTMS Army Physical Fitness Test

**IPERMS Courts Martial**
- Article 15 Proceedings
- Letters of Reprimand

**Army Transaction File**
- Army Training and Resource Requirement System
- Individual Training History
Questions

Contacts:
Kelly S. Ervin, PhD
U.S. Army Research Institute for the Behavioral and Social Sciences
Kelly.s.ervin.civ@mail.mil

Joshua Goldstein, PhD
University of Virginia, Social and Decision Analytics Division (SDAD)
joshua.goldstein@virginia.edu

We acknowledge the significant contributions of all co-authors on this project:

Nathaniel Ratcliff, PhD
U.S. Army Research Institute for the Behavioral and Social Sciences

Sallie Keller, PhD
Vicki Lancaster, PhD
Stephanie Shipp, PhD
University of Virginia, Social and Decision Analytics Division (SDAD)

Joel Thurston, PhD
Center for Army Leadership (CAL) U.S. Army Combined Arms Center