

Predicting army reserve unit manning using market demographics

Nathan Parker and Jonathan Alt

TRADOC Analysis Center, U.S. Army, Monterey, California, USA, and

Samuel Buttrey and Jeffrey House

Department of Operations Research, Naval Postgraduate School, Monterey, California, USA

Predicting
army reserve
unit manning

69

Received 28 September 2017
Revised 6 March 2018
18 May 2018
6 July 2018
Accepted 17 July 2018

Abstract

Purpose – This research develops a data-driven statistical model capable of predicting a US Army Reserve (USAR) unit staffing levels based on unit location demographics. This model provides decision makers an assessment of a proposed station location's ability to support a unit's personnel requirements from the local population.

Design/methodology/approach – This research first develops an allocation method to overcome challenges caused by overlapping unit boundaries to prevent over-counting the population. Once populations are accurately allocated to each location, we then develop and compare the performance of statistical models to estimate a location's likelihood of meeting staffing requirements.

Findings – This research finds that local demographic factors prove essential to a location's ability to meet staffing requirements. We recommend that the USAR and US Army Recruiting Command (USAREC) use the logistic regression model developed here to support USAR unit stationing decisions; this should improve the ability of units to achieve required staffing levels.

Originality/value – This research meets a direct request from the USAREC, in conjunction with the USAR, for assistance in developing models to aid decision makers during the unit stationing process.

Keywords Demographics, Data analysis, Logistic regression, Readiness, Manning

Paper type Research paper

Introduction

Location selection for US Army Reserve (USAR) units is a challenging multiple-attribute decision analysis problem. Decision makers consider a location's facility availability, the specialty skills required and most importantly the ability of the geographic area surrounding the location to generate new recruits in the required numbers to fill the unit's authorizations. USAR policies require that soldiers serving in these units reside within either a 50-mile radius or a 90-min drive time of their unit (Department of the Army, 2005). Given that USAR units exist within a fixed recruiting market, and these markets vary considerably in the number of recruits they can reliably produce, it is imperative that stationing decisions consider recruiting market viability as a screening factor (Dertouzos and Garber, 2006). Developing statistical models to estimate an area's ability to generate



© In accordance with section 105 of the US Copyright Act, this work has been produced by a US government employee and shall be considered a public domain work, as copyright protection is not available. Published in *Journal of Defense Analytics and Logistics*. Published by Emerald Publishing Limited.

Journal of Defense Analytics and
Logistics
Vol. 2 No. 2, 2018
pp. 69-79
Emerald Publishing Limited
2399-6439
DOI 10.1108/JDAL-09-2017-0020

recruits in sufficient numbers, while accounting for the pull of adjacent units within this radius, presents a challenging and previously unaddressed problem for Army recruiting.

Background

The Army Reserve is a critical component of the United States' National Defense Strategy. Since the initial deployments to Afghanistan in 2001, the USAR has deployed over 170,000 soldiers in support of the Global War on Terrorism. Today, the USAR supplies 75 per cent of support units and capabilities such as logistics, medical, engineering, military information support and civil affairs. Together, these comprise half of the Army's combat support and combat service support forces [Office of the Chief of Army Reserves (OCAR), 2015]. The Reserve Component allows the Army to maintain a ready and trained force available to meet strategic and operational needs without bearing the cost of maintaining that force in an active duty capacity (Klerman, 2009).

Unlike an active duty Army unit, the geographic location of a USAR Troop Program Unit (TPU) has a direct impact on its ability to meet staffing goals and related personnel readiness requirements. The USAR does not have the flexibility to move part-time soldiers to meet staffing shortfalls, so it must be able to draw a sufficient number of qualified recruits from the local community (Department of the Army, 2005). To achieve required personnel readiness levels, a TPU must be able to meet its staffing requirements across all ranks (Department of the Army, 2010). The wide range of reserve location fill rates (actual staffing as a percentage of authorized requirement) depicted in Figure 1 highlights the USAR struggle to meet staffing goals at the individual TPU level, with some units significantly over- or under-strength.

Because of these wide variations in individual TPU fill rates, the USAR must temporarily augment deploying units with soldiers from other units to meet required staffing thresholds. Of 22 units included in a 2009 Government Accountability Office study, 21 required augmentation from non-deploying units to meet staffing requirements for deployment (Pickup, 2009). This "cross-leveling" of personnel induces stress in individual soldiers as well as both the gaining and losing units (Laurent, 2005). Many current campaign plans require the deployment of a large number of USAR units within the first 30-45 days of

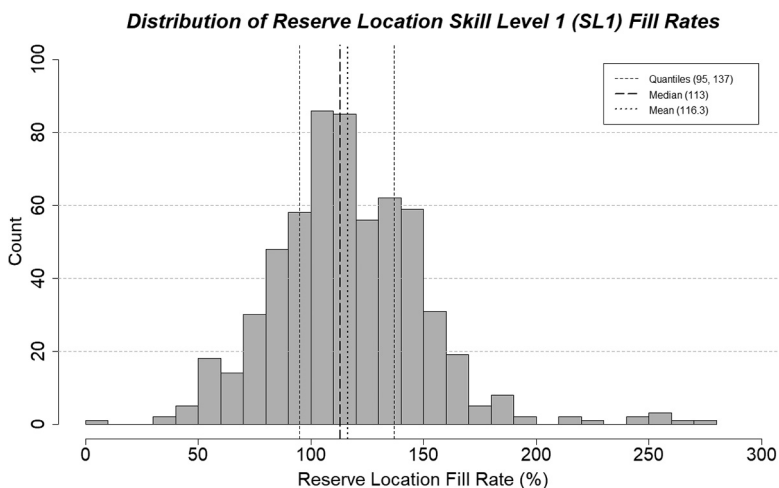


Figure 1. Distribution of reserve location skill level (SL1) fill rates (As of February 2015)

operations. This timeline does not allow time for a major cross-leveling of personnel (Department of Defense, 2011). To ensure the USAR can meet the personnel readiness requirements of the United States' national defense strategy, units must be located in areas where the market is able to meet and sustain their staffing requirements.

Since 2008, the Stationing Tool Army Reserve (STAR) has been the USAR's official decision-support tool used in the stationing process. STAR relies on weights elicited from subject matter experts to generate an overall utility score based on a location's ability to meet staffing, force structure and facilities requirements (Bradford and Hughes, 2007). Beginning in 2006, the USAR experienced a major expansion, generating the need for a tool to support basing decisions. Since then the tool has supported periodic, but infrequent, repositioning of USAR units. In anticipation of potential force re-structuring, a better screening methodology, one that takes into account a location's potential to meet staffing levels, was required. USAR staffing data as of February 2015 shows that almost 20 per cent of USAR locations selected using the current methodology are unable to support the staffing requirements of their TPUs.

We consider a data-driven approach to develop a statistical model focused solely on identifying those locations with the ability to support the increased staffing requirements from a potential stationing action. Specifically, we focus on the identification of potential stationing locations with a high probability of supporting the TPU's staffing requirements in the Skill Level 1 (SL1) ranks, defined as E-1 through E-4. Such a model would augment the STAR model currently used by USAR in its stationing decision process to screen out locations with low probabilities of supporting the TPU's staffing requirements. In the following sections, we review related work, discuss our data sources and review our development of a method for allocating populations and their demographic attributes to competing reserve centers. Finally, we explore the use of multiple methods including linear regression, classification trees and logistic regression to estimate a location's ability to meet a unit's SL1 requirements.

Related work

In the early 1990s, USAREC developed the *Market Supportability Study* (MSS) to meet requirements from DOD Directive 1225.7, *Reserve Component Facilities Programs and Unit Stationing*, directing services to review the manpower potential of an area to determine its adequacy for meeting and maintaining authorized officer and enlisted strengths (Deputy Secretary of Defense, 1996).

Fair (2004) developed the Unit Positioning and Quality Assessment Model to improve the USAR stationing process. Fair constructed a database capturing demographic statistics at the ZIP-code level including factors related to the population quality and vocation inclination. He then fit a linear regression model in which the vocational groups, lifestyle segments, military available population, quality of accessions and unemployment rate were the predictor variables and annual USAR production the response variable. Fair's work does not address the distribution of a population between multiple reserve centers so ZIP codes within the range of multiple reserve centers are, at times, counted more than once. This work provides our starting point for identifying the data necessary to develop the predictor variable our models will consider.

From 2006 to 2008, the Office of the Chief of the Army Reserve (OCAR) commissioned a series of studies centered on the USAR stationing process. The Center for Army Analysis (CAA) completed the year-long *Army Reserve Stationing Study* (ARSS) in 2007 to:

[. . .]develop a unit stationing methodology and tool that considers important factors including: capacity of a local area to recruit and maintain unit personnel, the ability to provide career

progression opportunities for USAR soldiers, and the location and capacity of existing Reserve facilities. (Bradford and Hughes, 2007)

The ARSS team focused on multiple-objective decision analysis as the core of their analysis identifying 18 separate measures and a value function for each measure. The development of the value functions and comparative weights drew primarily from subject matter experts from the stationing teams within the regional commands.

In 2008, the OCAR also used the CAA expertise and methodologies developed during the ARSS series to assist in the developing the STAR. This web-based tool automates the process developed by CAA, providing USAR analysts the ability to quickly conduct the initial analysis to evaluate a new station location. Because STAR relies on a multiple objective utility function to rank locations, in some cases, the high-value contributions from other attributes, such as facilities and career advancement opportunities, may mask a location's poor potential to meet staffing requirements.

In addition to previous research related to stationing USAR units, we also draw on the research regarding the economics of recruiting to confirm our selection of demographic factors. In reviewing research by Warner (1990), Brockett *et al.* (1997) and Knowles *et al.* (2002), we are confident that we have not left out any of the demographic factors identified by previous work as significant.

This research will provide a data-driven, objective and repeatable method to screen out candidate locations with poor potential to meet SL1 staffing requirements as a means of augmenting the STAR model currently in use.

Data and methods

Sources of data

USAREC and USAR provided the bulk of the data for this analysis with the exception of publicly available data from the Center for Disease Control and Prevention (CDC) and the US Census Bureau. Table I provides details regarding the data used in this research including the source and contribution of each data set. The USAR data sets provide the majority of the information we need to determine the staffing levels for each reserve center whereas the USAREC data sets provide the majority of demographic information about the recruitable populations. We use CDC and Census Bureau data to fill a couple of identified demographic data gaps (Center for Disease Control and Prevention, 2010; USA Census Bureau, 2014a, 2014b). The majority of our data is at the ZIP code level though the finest granularity available of a couple of them is the county level. For these data sets, we distribute county-level data across the ZIP codes within a county using a county-to-ZIP code crosswalk based on 2010 Census data (USA Census Bureau, 2014a, 2014b). Our method parallels that used in recent research into population demographic impacts on regular Army recruiting (McDonald *et al.*, 2017).

Prior to this study, the USAREC G2 prepared a table of distances and drive times from the centroid of each ZIP code containing a USAR reserve center to the centroid of each population ZIP code within the center's recruitable market range. The recruitable market range is defined by the USAR as either a 50-mile radius or a 90-min drive time (Department of the Army, 2005). A ZIP code's population can be within the recruitable market range of multiple reserve centers – in some cases, up to 25. Simply counting the population within range of a reserve center, without accounting for reserve center overlap, would overestimate the population available to reserve centers in high-concentration areas. To allocate each ZIP code's population to its nearby reserve locations, we expand on the approach of Mehay (1989).

Data set (Source)	Contribution
USAR unit and personnel (USAR G1)	Contains each TPU's total authorizations and current fill statuses which allows us to calculate the current fill rate for each unit by skill level
USAR cohort (USAR G1)	
USAREC production (USAREC G2)	Contains the current status of each USAR accession which allows us to calculate the attrition rate for each unit. The characterization of service code for each accession allows us to determine if they separated under adverse or non-adverse conditions
DoD production (USAREC G2)	Contains demographic information for every USAREC enlisted accession (USAR and RA). This allows us to calculate the quality and type of Soldier that each ZIP code is producing
USAREC recruiter laydown (USAREC G2)	Contains count data, by ZIP code, for enlisted accessions into all DoD services which allows us to account for market competition beyond the USAR and RA
USAREC recruiter laydown (USAREC G2)	Contains the number of recruiters assigned to each recruiting station and the ZIP codes they are responsible for servicing. This provides information to account for the impact of additional recruiters on overall production
Unemployment rate (USAREC G2)	Contains the Bureau of Labor Statistics U-3 unemployment rate at the county level. This allows us to explore the impact of local labor market health on a location's ability to produce USAR accessions
Obesity (CDC)	Contains obesity rates at the county level. This allows us to explore the impact of local population health/fitness on a location's ability to produce USAR accessions
Qualified military available (USAREC G2)	Contains the estimated number of 17-24-year-olds within each ZIP code who are eligible for enlistment in the armed services. This allows us to control for variations in the populations age distribution across locations
Post-secondary enrolled (Census Bureau)	Contains the number of individuals currently enrolled in resident, post-secondary institutions at the ZIP code level. This allows us to correct for the fact that the QMA data screen these individuals out, as they are not seen as available for full-time military service
Regional location (USAREC G2)	Contains the areas of responsibility for each of the five Recruiting Brigades. This allows us to segment the United States into five regional identities

Table I.
Summary of data sets

Methodology

We will first discuss the methodology used to allocate the ZIP code based population data to each reserve center to avoid potential over-counting, then briefly introduce the statistical modeling approaches used in this research.

Our allocation method uses two factors: each reserve center's number of SL1 authorizations, and the drive time from (the centroid of) each population ZIP code to each reserve center. Each ZIP code's proportional allocation to a nearby reserve center is a weighted sum of two proportions: that center's authorization as a proportion of all nearby centers' authorizations, and that center's "ease of drive" (i.e. 90 minus the number of driving minutes from the ZIP code's centroid) as a proportion of all nearby centers' ease of drive. [Figure 2](#) shows an example of the results of the computations for a ZIP code with four reserve locations within range, using weights of (0.50, 0.50) for the two components. Recruiting center A's ease of drive of (90 - 50) represents 23.5 per cent of the total ease of drive across the four centers, and its SL1 authorization represents 28.6 per cent of the total. Therefore, it is allocated $(0.5)(23.5) + (0.5)(28.6) = 26.1$ per cent of the population of the ZIP code. [Parker \(2015\)](#) provides the full development of this population allocation method.

We apply this allocation method globally to produce a $22,680 \times 598$ matrix, M , containing allocation weightings for all combinations of population ZIP codes and current reserve locations. A second matrix, R , holds the $22,680 \times 17$ matrix of population

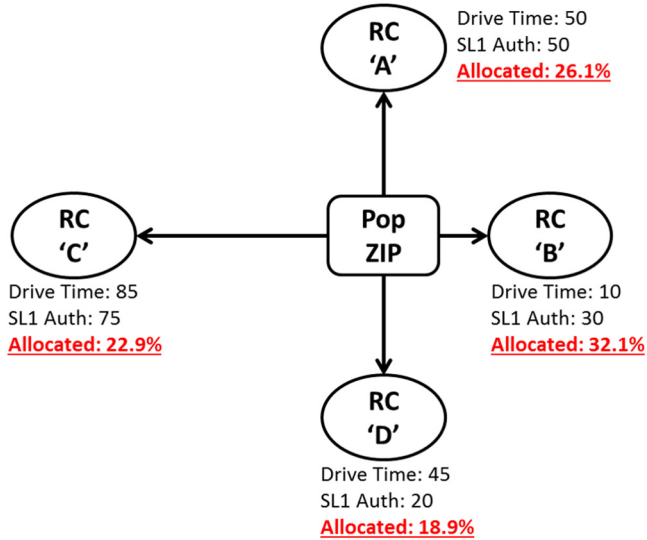


Figure 2.
Distribution of a
single population ZIP
code between four
reserve locations

demographics for each ZIP code, so the product $M^T R$ produces the 598×17 matrix of population demographics by current reserve location. For the normalized demographic factors (unemployment, obesity, Armed Forces Qualification Test scores and attrition rates), we divide the results by the sum of the associated allocation weighting factors to account for the fact that these factors cannot be combined in an additive manner. We then join our reserve center population demographic data set with our SL1 staffing data for each center to produce our master reserve location data set. Next, we remove those locations with missing values or other data anomalies (the vast majority of these are located outside the continental USA) to arrive at our reserve location data set. [Parker \(2015\)](#) provides the full development of our data processing method.

This final data set contains our response variable (SL1 staffing level) and our 17 initial independent variables as shown in [Table II](#). We divide the data set into a training set containing 399 observations and a test set containing 199 observations before model construction. From this data set, we develop a model to predict a location's ability to meet the SL1 staffing requirements of the USAR TPUs in a particular location.

Initial exploratory analysis revealed high correlation among several predictors including recruiters, qualified military available (QMA) populations, regular army (RA) accessions, army reserve (AR) accessions, Department of Defense (DOD) accessions and post-secondary enrollment. This is understandable, as two subsets of them (QMA I-IV and Post-secondary Enrollment) and (Recruiters) are inputs to the third subset (RA, AR and DOD accessions). We made the decision to remove the first two subsets and allow models to consider only RA, AR and DOD accessions.

Our early modeling efforts used the numeric fill rate of a unit as the dependent variable in a series of linear regression models. These models provided insights into the interactions among predictors, but lacked explanatory power. Our best performing linear regression model, capable of meeting standard regression diagnostic tests, produced an adjusted R-squared of 0.292. We determined that this model did not provide the explanatory power necessary to assist decision makers involved in TPU stationing process. One interesting

Table II.
Independent
variables considered
in model building

Independent variable name	Explanation
Required	Total number of SL1 billets authorized to the TPUs
Adverse attrition	3-year weighted average of the annual count of adverse attrition from TPUs
Non-Adverse attrition	3-year weighted average of the annual count of non-adverse attrition from the TPUs
Recruiters	Count of army recruiters
Army Reserve (AR) accessions	3-year weighted average of the annual count of Army Reserve accessions
Regular Army (RA) accessions	3-year weighted average of the annual count of Regular Army accessions
DoD accessions	3-year weighted average of the annual count of DoD (non-Army) accessions
AFQT	3-year weighted average of the annual average Armed Force Qualification Test score of all Army accessions
Unemployment	Unemployment rate for the population
Obesity	Percentage of the population identified as obese by the CDC data
QMA I	Count of the Qualified Military Available (Category I) population
QMA II	Count of the Qualified Military Available (Category II) population
QMA IIIA	Count of the Qualified Military Available (Category IIIA) population
QMA IIIB	Count of the Qualified Military Available (Category IIIB) population
QMA IV	Count of the Qualified Military Available (Category IV) population
Post-secondary enrollment	Count of the individuals enrolled in post-secondary institutions
Region	Regional location by US Army Recruiting Command Brigade Boundary

insight was that *Unemployment* did not prove to be a statistically significant predictor of fill rate. This suggests that reserve location fill rates are less sensitive to the unemployment rate within their local communities than might be expected.

We then recoded the fill rate as a binary variable, with fill rates less than 100 per cent as the zero value and rates above 100 per cent as the one (Table III). We constructed logistic regression models using this binary variable as the response.

Results

Logistic regression

The logistic regression model, fit using the generalized linear model (glm()) function included in R (R Core Team, 2013), started with a saturated, main-effects model including all 17 independent variables listed in Table II. Variables with a *p*-value greater than 0.05 were systematically removed from the model starting with those identified in the linear regression model as being highly collinear. The logistic regression model initially retained both the *Non-Adverse* and *Adverse* attrition variables independently. Further analysis

Table III.
Response variable for
reserve location fill
rate

Binary value	Classification criterion	No. (%) of reserve locations
0	Fill rate < 100%	176 (29.4%)
1	Fill rate ≥ 100%	422 (70.6%)

indicated that the model performance improved when we additively combined these two variables to create a single *Attrition* variable. We initially use a five-level categorical variable for location, derived from the five USAREC Brigade boundaries. The only region retained by the model is the Southeast (SE) region. We verified the final logistic regression model structure using the stepwise (step()) function included in R to consider adding any previously removed variable if it improved the model’s AIC. Table IV displays the final logistic regression model, including coefficients and associated *p*-values. We see that those locations with higher Required (*RQD*) values are expected to have lower fill rates. There was no evidence of unusual over-dispersion in this model.

We evaluated the observations from the test set using the logistic regression model from Table IV. Table V displays these results in a standard confusion matrix style using a cutoff for the predicted probability of 0.5.

Table V shows the logistic regression model’s overall misclassification rate of 26.1 per cent. The misclassification rate is 32.7 per cent on those locations below the 100 per cent fill level and 23.4 per cent for those locations above that level. The area under the ROC curve is 0.765, suggesting that the model provides acceptable discrimination and can help inform decision makers (Hosmer *et al.*, 2013).

Classification trees

As a check on our modeling we fit a classification tree using the rpart package (Therneau *et al.*, 2013) in R and use the built-in 10-fold cross-validation method to select our pruning parameter. Figure 3 depicts the tree as represented by R’s rpart.plot package (Milborrow, 2015).

Nodes are labeled according to the small numbers in boxes above the nodes. Each node’s split is shown in text (e.g. “AR_Prod < 20”), with observations for which the split predicate is true going to the left and those for which it is false going to the right. Counts underneath nodes show the distribution of the response variable at that node. So, for example, at the root (topmost) node the 399 training set observations include 118 for which the fill rate was >100 per cent and 281 for which it was greater. Because the second group is the larger, the root node is labeled with a 1 (shaded box) – that is, the tree uses a cutoff of 0.5. With that cutoff, the tree’s misclassification rate on the test set is slightly better than the logistic regression’s, at 23.6 per cent. It produces a slightly lower area under the ROC curve of 0.753, which still

Table IV.
Logistic regression
model coefficients

Independent variable	Coefficient	<i>p</i> -value
(Intercept)	-0.8035	0.0114
RQD	-0.0325	<0.001
Attrition	0.1506	<0.001
AR Prod	0.1388	<0.001
Region SE	1.5560	<0.001

Table V.
Actual versus
predicted values for
the logistic
regression model,
cutoff 0.5

	Actual value	Predicted value	
		0	1
0		39	19
1		33	108

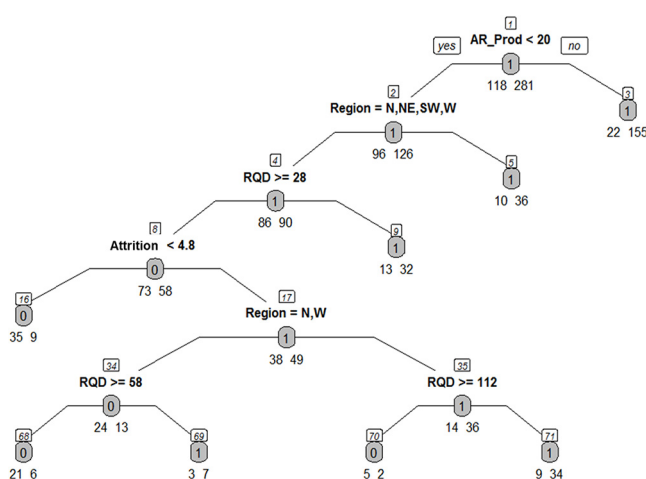


Figure 3.
Classification tree
model after pruning

meets the standard for acceptable discrimination (Hosmer *et al.*, 2013). Interestingly, the tree performance was better on the true 1's (for which the misclassification rate was 11.3 per cent) and worse on the true 0's (53.4 per cent) than the logistic regression.

The tree, shown in Figure 3, can yield additional insights. At Node 1, the model indicates that locations with higher *AR Prod* are more likely to have fill rates above 100 per cent. At Node 2, the model indicates that locations in *Region Southeast* are more likely to have fill rates above 100 per cent. Analogous relationships are observable at the nodes using *Attrition* and *RQD* as the splitting criteria. Of interest is that the classification tree model did not retain any of the more direct population demographic variables such as unemployment, obesity and QMA counts. It is possible that the inclusion of the Army Reserve production (*AR Prod*) variable sufficiently captures the impact of these factors, or that these factors have less impact on USAR staffing than one might assume.

Conclusion

To provide an improved screening methodology for USAREC and USAR, we developed statistical models to determine a candidate location's potential to meet a unit's SL1 staffing requirements. We aggregated demographic data from eight separate data sources resulting in a final data set that contained 17 demographic factors for each ZIP code within a 90-min drive of any reserve center. We expanded on a method to allocate the population in this radius to reserve centers to avoid over-counting (Mehay, 1989 and Parker, 2015). We conducted exploratory data analysis to gain insight into the relationship between the predictors and the response prior to developing logistic regression and classification tree models for use as a screening tool in the stationing process.

Both models provide insight into how the population demographics are likely to influence the ability of the reserve location to meet its staffing requirements. We find it intriguing that both models retain the following independent variables as significant: number of SL1 positions, TPU Attrition rate, Army Reserve Production and regional location of the unit. Future analysis could use these factors as a starting point for further research into a particular population's or market's ability to support changing USAR staffing requirements. We also find it noteworthy that none of the models retained *Unemployment* as a significant predictor. This area stands out for further analysis, as it is

contrary to the findings of several earlier research efforts into military recruiting (Brockett *et al.*, 1997; Warner, 1990; and Knowles *et al.*, 2002). It is possible that some aspect of the difference between active and reserve service reduces the impact that unemployment has on the ability of a population to meet recruiting demand, or that the retention of the *Region* variable is a sufficient proxy for the effects of unemployment.

Although both the classification tree and logistic regression models produce levels of accuracy that will provide valuable recommendations to decision makers involved in the USAR stationing process, we recommend the use of the logistic regression model. The logistic regression model displayed better performance on the task of identifying those locations most at risk. The use of this model will allow analysts involved in the stationing process to screen out locations with a low probability of meeting the TPU's staffing requirements.

Future analysis in this area should focus on refining the population allocation method and the identification of additional unit and demographic factors that could be affecting staffing levels. One assumption implicit in this analysis is that the current reserve locations are representative of future reserve locations. Future work could seek to address this limitation of the current modeling approach. Additional research may find value in locating datasets that better represent the population likely to enlist in a USAR unit. As an example, if future researchers can obtain unemployment and underemployment data focused on those between 17 and 35 years of age they may be able to develop improved models.

References

- Bradford, R. and Hughes, T. (2007), *Army Reserve Stationing Study*, CAA-2006249, Center for Army Analysis, Fort Belvoir, VA.
- Brockett, P.L., Rousseau, J.J., Zhou, L., Golany, B. and Thomas, D.A. (1997), "Forecasting and allocation of US army recruiting resources", *Military Operations Research*, Vol. 3 No. 3, pp. 13-29.
- Center for Disease Control and Prevention (2010), "Community health status indicators (CHSI) to combat obesity, heart disease and cancer database", Object name RISKFACTORSANDACCESSSTOCARE, available at: www.healthdata.gov/data/dataset/community-health-status-indicators-chsi-combat-obesity-heart-disease-and-cancer (accessed 9 March 2015).
- Department of the Army (2005), "Assignments, attachments, details, and transfers (Army regulation 140-10)", Government Printing Office, Washington, DC.
- Department of the Army (2010), "Army unit status reporting and force registrations (Army regulation 220-1)", Government Printing Office, Washington, DC.
- Department of Defense (2011), "Comprehensive review of the future role of the reserve component: volume I executive summary and main report", Office of Vice Chairman of the Joint Chiefs of Staff and Office of Assistant Secretary of Defense for Reserve Affairs, Washington, DC.
- Deputy Secretary of Defense (1996), "Reserve component facilities programs and unit stationing", DOD Directive 1225.7, Deputy Secretary of Defense, Washington, DC.
- Dertouzos, J.N. and Garber, S. (2006), *Human Resource Management and Army Recruiting: Analyses of Policy Options*, RAND Corporation, Santa Monica, CA.
- Fair, M.L. (2004), "Geo-demographic Analysis In Support of the United States Army Reserve (USAR) Unit Positioning and Quality Assessment Model (UPQUAM)", Master's thesis, Naval Postgraduate School, Monterey, CA.
- Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013), *Applied Logistic Regression*, 3rd ed., Wiley & Sons, Hoboken, NJ.
- Klerman, J.A. (2009), *Rethinking the Reserves*, RAND Corporation, Santa Monica, CA.

- Knowles, J.A., Parlier, G.H., Hoscheit, G.C., Ayer, R., Lyman, K. and Fancher, R. (2002), "Reinventing army recruiting", *Interfaces*, Vol. 32 No. 1, pp. 78-92.
- Laurent, J.S. (2005), *An Integrated Plan Is Needed to Address Army Reserve Personnel and Equipment Shortages*, GAO-05-660, U.S. Government Accountability Office, Washington, DC.
- McDonald, J., White, E.C., Hill, R.R. and Pardo, C. (2017), "Forecasting U. S. Army enlistment contract production in complex geographical marketing areas", *Journal of Defense Analytics and Logistics*, Vol. 1 No. 1, pp. 66-84.
- Mehay, S.L. (1989), "An enlistment supply and forecasting model for the US Army reserve", USAREC Study Report 89-2. Naval Postgraduate School, Monterey, CA.
- Milborrow, S. (2015), "rpart.plot: Plot rpart models. an enhanced version of plot.rpart", R package version 1.5.2, available at: <http://CRAN.R-project.org/package=rpart.plot> (accessed 1 April 2015).
- Office of the Chief of Army Reserves (OCAR) (2015), "U.S. Army reserve at a glance: twice the citizen", Army Reserve Communications, Washington, DC, available at: www.usar.army.mil/resources/Media/ARAG%20V3%20Final%20PDF.pdf (accessed 10 February 2015).
- Parker, N. (2015), "Predicting US Army Reserve unit manning using market demographics", Master's thesis, Naval Postgraduate School, Monterey, CA.
- Pickup, S.L. (2009), *Army Needs to Reevaluate Its Approach to Training and Mobilizing Reserve Component Forces*, GAO-09-720, U.S. Government Accountability Office, Washington, DC.
- R Core Team (2013), "R: a language and environment for statistical computation. R foundation for statistical computing", Vienna, Austria, available at: www.R-project.org (accessed April 15, 2015).
- Therneau, T.M., Atkinson, E. and Ripley, B. (2013), "rpart: Recursive partitioning. R package version 4.1-3", available at: <http://CRAN.R-project.org/package=rpart> (accessed 15 April, 2015).
- USA Census Bureau (2014a), "2009-2013 5-Year american community survey, object name B14004", available at: http://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_5YR_B14004&prodType=table (accessed 9 March, 2015).
- USA Census Bureau (2014b), "2010 ZCTA to county relationship file", available at: www2.census.gov/geo/docs/maps-data/data/rel/zcta_county_rel_10.txt (accessed 21 February, 2015).
- Warner, J.T. (1990), "Military recruiting programs during the 1980s: their success and policy issues", *Contemporary Economic Policy*, Vol. 8 No. 4, pp. 47-67.

Corresponding author

Nathan Parker can be contacted at: nparker@nps.edu

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgrouppublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com