The role of benzathine penicillin G in predicting and preventing all-cause acute respiratory disease in military recruits: 1991–2017

Author for correspondence:
Jacob D. Ball, PhD, Army Public Health Center, CPHE, 5158 Blackhawk Road, Bldg E-1570, Aberdeen Proving Ground, MD 21010
E-mail: jball2@ufl.edu

Jacob D. Ball[1,2,3], Mattia A. Prosperi[1], Alfonza Brown[3], Xinguang Chen[1], Eben Kenah[4], Yang Yang[2,5], Derek A. T. Cummings[2,6,*] and Caitlin M. Rivers[3,7,*]

[1]Department of Epidemiology, University of Florida, Gainesville, USA; [2]Emerging Pathogens Institute, University of Florida, Gainesville, USA; [3]U.S. Department of Defense, Army Public Health Center, Army Medical Command, Edgewood, USA; [4]Department of Biostatistics, The Ohio State University, Columbus, USA; [5]Department of Biostatistics, University of Florida, Gainesville, USA; [6]Department of Biology, University of Florida, Gainesville, USA and [7]Center for Health Security and Department of Environmental Health & Engineering, Johns Hopkins University, Baltimore, USA

Abstract

The adenovirus vaccine and benzathine penicillin G (BPG) have been used by the US military to prevent acute respiratory diseases (ARD) in trainees, though these interventions have had documented manufacturing problems. We fit Poisson regression and random forest models (RF) to 26 years of weekly ARD incidence data to explore the impact of the adenovirus vaccine and BPG prophylaxis on respiratory disease burden. Adenovirus vaccine availability was among the most important predictors of ARD in the RF, while BPG was the ninth most important. BPG was a significant protective factor against ARD (incidence rate ratio (IRR) = 0.68; 95% confidence interval (CI) 0.67–0.70), but less so than either the old or new adenovirus vaccine (IRR = 0.39, 95% CI 0.38–0.39 and IRR = 0.11, 95% CI 0.11–0.11), respectively. These results suggest that BPG is moderately predictive of, and significantly protective against ARD, though to a lesser extent than either the old or new adenovirus vaccine.

Introduction

Acute respiratory diseases (ARD) represent a significant concern to the US military and thus national security, as they are responsible for between 12 000 and 27 000 missed training days among recruits, annually [1]. The US Army defines ARD syndromically; soldiers must present to a medical treatment facility with: oral temperature >100.5 °F, recent signs or symptoms of acute respiratory tract inflammation, and either having a limited duty profile or being removed from duty altogether to have ARD. As a result, both viral and bacterial pathogens, as well as non-pathogenic conditions can result in ARD. Adenoviruses, influenza viruses, respiratory syncytial viruses and human coronaviruses make up the majority of viral contributors to ARD, while *Streptococcus pneumoniae*, *Streptococcus pyogenes* ('strep') and *Mycoplasma pneumoniae* represent some of the major bacterial etiologic agents of ARD [2, 3].

Basic Combat Training (BCT) consists of a 10-week long training programme designed to prepare recruits for service in the Army. There are both physical training and practical training exercises, and BCT culminates with a physical fitness test that all recruits must pass in order to graduate from BCT. Recruits at BCT have been shown to be at higher risk of ARD compared with their civilian counterparts [2, 3]. This excess risk has been attributed to crowded barracks [4] and other environmental challenges [5], as well as to stress from physically and psychologically demanding conditions [3, 4, 6, 7]. There are four Forts that currently serve as BCT sites: Fort Jackson (in South Carolina), Fort Benning (in Georgia), Fort Leonard Wood (in Missouri) and Fort Sill (in Oklahoma). Fort Jackson is the largest BCT installation, which trains approximately 35 000 recruits annually.

In the 1960s, before the rollout of the adenovirus vaccine, approximately 80% of trainees would contract a respiratory infection during BCT. The majority of those infections were attributed to adenovirus types 4 and 7 [8]. Of those who contracted a respiratory infection, nearly 20% were hospitalised [4]. This excess risk has been attributed to crowded barracks [4] and other environmental challenges [5], as well as to stress from physically and psychologically demanding conditions [3, 4, 6, 7].

To address this burden, live, enteric-coated oral adenovirus vaccines against types 4 and 7 (AdV-4 and -7, referred to collectively here as the 'old vaccine') were routinely used year-round beginning in 1971 [2]. The sole manufacturer of AdV-4 and -7 ceased production in 1994 due to low public demand for the product. The final doses were shipped in 1996, and distributed to Army recruits only during the winter months until the stockpile was depleted

in 1999 [9]. As a result of the phase out, ARD rates, particularly adenovirus-associated ARD rates, increased at BCT sites [4]. The increased disease burden cost the Army an estimated $10–$26 million in medical costs and lost training time annually. In 2011, live, enteric-coated oral vaccines against types-4 and -7 were approved for use in military personnel [10]. These 'new vaccines' were derived from strains and stocks used by the original manufacturer, and all protocols and logistical procedures were as close to those used in the old vaccine manufacturing line [11]. The subsequent year-round administration of the new vaccines resulted in drastic, sustained declines in ARD at all BCT installations [9, 10].

Benzathine penicillin G (BPG, trade name 'Bicillin') is a high-dose antibiotic injection administered to military recruits upon accession to BCT. Prophylactic use of BPG is intended to reduce the impact of imported bacterial infections on troop health upon entry to BCT in hopes of preventing further spread [12]. Multiple studies have suggested that BPG and other antibiotic prophylaxis have positive effects on recruit health and ARD rates [13, 14]. Injection with 1.2 M units of BPG has been shown to provide protection against infection with strep and other bacterial pathogens for up to 6 weeks [12], with other studies suggesting the duration may be shorter [15]. One such study found that the duration of protection may be as low as <2 weeks in Navy recruits [16]. Generally, BPG prophylaxis is given to all Army recruits upon entry to the 10-week long BCT programme. Additional doses are not administered to trainees after protection wanes.

Although it has been Army policy to administer BPG to new recruits since the 1950s, there has been some variability in administration. For reasons that are not clear, Fort Jackson has never given BPG prophylaxis. Also, multiple BPG manufacturing problems resulted in relatively minor and localised supply shortages, and some of these shortages were associated with increases in ARD burden. One of the most extensive BPG shortages occurred in April–July 2016 and affected all BCT installations. Fort Benning chose not to resume routine prophylaxis after the shortage ended, and thus has not administered BPG since March 2016. The variability in the use of BPG in time relative to variability in adenovirus vaccine administration at different sites provides an opportunity to conduct an observational study of the impact of BPG use on ARD.

Here we fit, k-fold cross-validated and externally validated random forest (RF) and Poisson regression (PR) models to weekly counts of ARD in Army recruits at BCT installations from January 1991 to April 2017 in order to evaluate the role of BPG prophylaxis at reducing ARD incidence. We evaluated the relative performance of the RF to the PR in predicting all-cause ARD, and quantified the relative importance of BPG and other covariates in this prediction. We hypothesised that BPG availability would be significantly associated with ARD. Specifically, we hypothesised that BPG would have a moderately protective effect on ARD incidence, and that its effect size would be lower than that of the new adenovirus vaccine.

Methods

Data

As mentioned above, a trainee is said to have ARD if he or she presents to a medical treatment facility with: oral temperature >100.5 °F, recent signs or symptoms of acute respiratory tract inflammation, and either having a limited duty profile or being removed from duty altogether. Weekly counts of all-cause ARD from January 1991 to April 2017 among recruits at the four BCT installations, recruit population size and training type (BCT or a longer programme called One Station Unit Training – OSUT – a combined BCT-advanced individual training (AIT) programme whereby both training programmes are completed in succession at the same installation rather than completing BCT at one installation and travelling to a second installation for AIT) were obtained from the Army ARD Surveillance Program [17]. BCT installations are required to report weekly counts of ARD to Army Public Health Center as part of the ARD Surveillance Program. Data on BPG prophylaxis and adenovirus vaccine availability (both the old and new vaccines) were extracted from the literature and cross-referenced against ARD Surveillance reports and Army memoranda.

Analysis

Data cleaning and analysis were implemented using R version 3.2.1 (R Core Team 2014). Both an RF model and a PR model were used in a k-fold cross-validation framework (k = 5), with 3 years of data – one from each 'era' of adenovirus vaccine availability – held out for external validation. In k-fold cross-validation, the data are partitioned into k subsets ('folds') of roughly equal size. The model is then trained on k-1 folds and used to predict values in the kth fold. The process is iterative so that each of the folds acts as the test set and a predicted value is generated for every observation in the original dataset. Performance was evaluated by comparing mean absolute error (MAE) across folds within the same model and within each fold across models [18]. K = 5 was selected in order to conserve computational power, while also providing an 80/20 split between the training and testing set in each of the iterations. The superiority of the RF relative to the PR was calculated by

$$100 \times \left[1 - \left(\frac{\mathrm{MAE_{RF}}}{\mathrm{MAE_{PR}}} \right) \right], \qquad (1)$$

and statistical significance of these MAEs was determined through a Student's t test adjusted for sample overlap [19].

RF procedures and prediction metrics were run using the *randomForest* [20] and *forecast* [21] R packages, respectively. RF models are a non-linear ensemble of decision trees in which bootstrapped samples from the original dataset are used for model construction, while the remaining observations constitute the out-of-box sample on which to test the model's performance [22–24]. A single decision tree in the RF represents a hierarchical partition of the bootstrapped data grown by considering highly discriminative variables among randomly drawn subsets of the entire covariate set [23]. Each level of categorical variables was fed into the model as a binary variable in order to minimise bias in selecting variables to split on [25]. Mean squared error (MSE) was used as a metric for variable importance to assess the relative predictive ability of each covariate in the model, including BPG.

PR was used to estimate incidence rate ratios (IRRs) and corresponding confidence intervals (CIs) for BPG and each of the covariates on weekly ARD. Trainee population size was included in this model as an offset rather than as a fixed effect in order to adjust for how the size of the population at risk influences the likelihood of observing more cases.
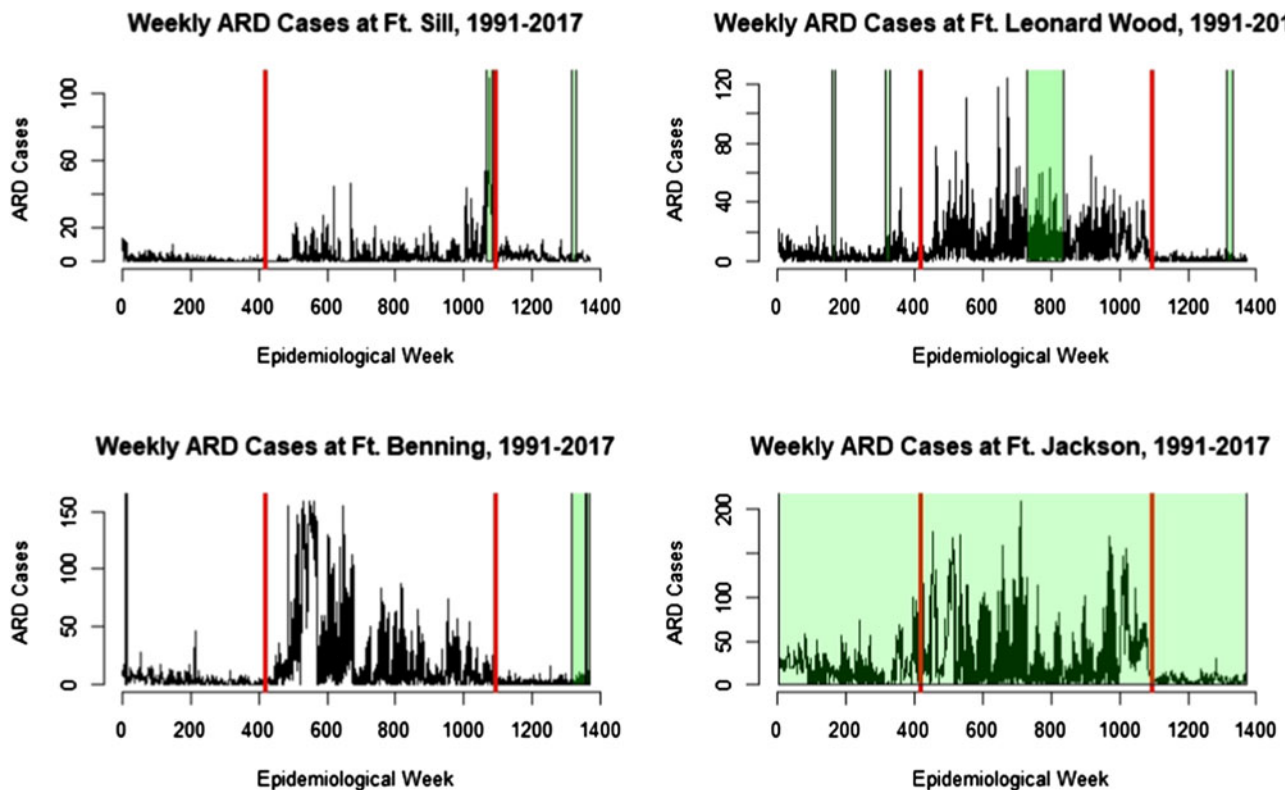
Fig. 1. Time series of ARD cases from 1991 to 2017 at Fort Sill (top left), Fort Leonard Wood (top right), Fort Benning (bottom left) and Fort Jackson (bottom right). The period between the red bars is the adenovirus vaccine shortage, with the time preceding the left bar representing the 'old' vaccine era, and the 'new' vaccine era to the right of the right bar. Green shading indicates BPG prophylaxis shortage. Fort Jackson never has used BPG for prophylaxis.

The RF model was then re-run on the external validation set, and variable importance measures were extracted and compared with the average importance measures from the cross-validation. Similarly, a PR model was performed on the external validation set. The MAEs for the RF and PR models were compared using equation 1.

We then ran PR separately on the entire training dataset (used in the cross-validation) to obtain IRR point estimates and corresponding CIs for BPG and other covariates. These IRR estimates allowed us to test the hypothesis that BPG would have a significant protective effect on the rate of all-cause ARD, and to compare the magnitude of the effect with that of adenovirus vaccine.

Results

Prediction

Cross-validation (22.5 years)
Figure 1 presents the raw time series (1991–2017) for the four BCT installations, Fort Sill (top left), Fort Benning (top right), Fort Leonard Wood (bottom left) and Fort Jackson (bottom right).

The predicted values from the RF and PR fivefold cross-validation analyses are overlaid on the original time series in Figure 2. Prediction accuracy for the RF and PR *k*-fold cross-validation analyses are presented in Table 1. The RF model was able to explain on average 65.2% of the variance in the data across the folds, and the estimates across each testing fold were consistent ranging from 63.8% to 66.5%. The RF model, on average and for each individual fold, had a lower MAE than the PR, average

MAE of 6.5 (range 6.3–6.7) and 7.3 (range 7.1–7.7) for the RF and PF models, respectively. From equation 1, the average superiority of the RF model over the PR model was estimated at 11.6%, with a range of 10.9–12.1% across folds. This difference was statistically significant ($P < 0.001$).

Variable importance, as measured by per cent increase in MSE for each independent variable when left out of models, in each fold of the RF model is presented in Table 2. Of note, the new adenovirus vaccine was the most important variable in all five folds, with an average of 133% increase in MSE. Trainee class size in hundreds and the old adenovirus vaccine were the second and third most important variables with mean per cent increase MSE values of 83% and 75%, respectively. The month of January and training type were the next most important predictors, with per cent increase MSE values of 39% and 35%, respectively. Exclusion of an indicator of BPG use increased MSE by an average of 21.2%.

External validation
Both the PR and RF models were used to predict ARD case counts in an external validation set consisting of 3 years – one from each adenovirus vaccine 'era'. Figure 3 overlays the predicted values from each of the two models on the observed data. The RF model still out-performed the PR by 20.9%, with MAE values of 4.9 and 6.2 for the RF and PR, respectively (data not shown).

Table 3 presents the average variable importance measures for the covariates from the cross-validated RF model, and those from when the same RF model was applied to the external validation set. The five most important variables in the cross-validation analysis were also the five most important variables in the validation
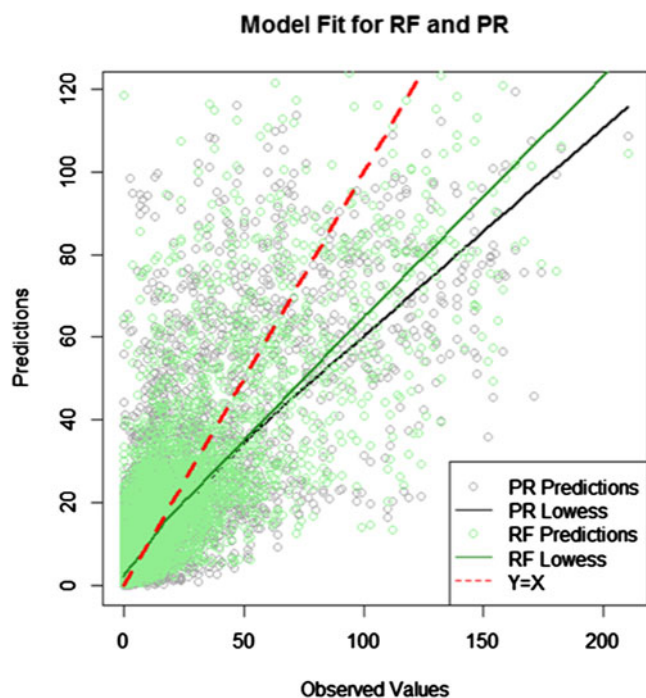
Model Fit for RF and PR

Fig. 2. Predicted and observed values from the fivefold cross-validation.

Table 1. Comparing prediction accuracy for fivefold cross-validation by mean absolute error for random forest and Poisson regression

	% Variance explained random forest	MAE random forest	MAE Poisson regression	1 – Ratio of MAE
Fold 1	65.37	6.53	7.33	10.91
Fold 2	66.49	6.72	7.66	12.27
Fold 3	63.75	6.44	7.28	11.54
Fold 4	65.83	6.46	7.31	11.63
Fold 5	64.39	6.27	7.13	12.06
Average	65.17	6.48	7.34	11.68

Table 2. Variable importance (% increase mean squared error when left out) for random forest model (*k*-fold cross-validation) using trainee class size

Variable	Mean	S.D.	Minimum	Maximum
Population in hundreds	83.4	5.4	74.9	89.7
Fort Sill	22.9	2.9	20.0	26.8
Fort Leonard Wood	18.2	0.8	17.4	19.5
Fort Benning	17.9	1.3	16.4	19.3
Fort Jackson	21.8	1.2	20.0	23.4
January	39.0	3.4	33.8	42.6
February	23.0	2.1	20.9	25.7
March	14.4	4.9	9.4	21.1
April	9.3	1.6	7.4	11.4
May	15.3	3.1	11.7	19.6
June	8.2	3.4	5.2	13.6
July	10.6	1.5	9.3	12.9
August	11.4	3.3	6.6	15.2
September	4.3	3.8	0.5	10.3
October	3.0	1.2	2.0	4.8
November	8.6	4.0	3.0	13.6
December	−1.4	7.3	−8.4	9.5
BPG	20.8	1.4	19.2	22.3
Old adenovirus vaccine	75.3	3.2	71.7	80.1
New adenovirus vaccine	133.1	5.3	127.6	141.8
BCT/OSUT	35.1	2.4	32.7	38.7

analysis: the new adenovirus vaccine availability, trainee population size, availability of the old adenovirus vaccine, the month of January and training type. In both analyses, BPG availability was ranked as the ninth most important variable.

Inference on BPG effectiveness

PR model results from the training dataset used in the cross-validation procedure are presented in Table 4 to obtain IRRs. BPG prophylaxis availability was found to be significantly protective against all-cause ARD (IRR = 0.68, 95% CI 0.67–0.70). Similarly, both the old and new adenovirus vaccines were significantly protective against ARD when compared with the adenovirus vaccine shortage, with IRRs of 0.39 (95% CI 038–0.39) and 0.11 (95% CI 0.10–0.11), respectively. Compared with Fort Jackson, Forts Benning and Leonard Wood were significant risk factors for ARD

(IRR = 1.17, 95% CI 1.13–1.20 and IRR = 1.29, 95% CI 1.26–1.33, respectively). Fort Sill was significantly protective compared with Fort Jackson (IRR = 0.67, 95% CI 0.65–0.69). BCT training was a significant risk factor compared with OSUT. Furthermore, a 100 person increase in trainee population size was associated with a 3% increase in ARD incidence rate.

Discussion

In this study, we sought to build a predictive model for all-cause ARD and assess how important BPG and other covariates are in its performance. Furthermore, we hypothesised that BPG was significantly associated with all-cause ARD – specifically, that BPG would have a significant, moderately predictive of and protective against all-cause ARD. We also hypothesised that the magnitudes of its association and predictive ability would be lower than those of the new adenovirus vaccine.

The RF model explained on average over 65% of the variance in the data, and outperformed the PR model *k*-fold cross-validation by 11.6% (Table 1). In the external validation, the RF model's predictive accuracy was over 20% superior to that of the PR model.

BPG, in the PR on the full dataset, was associated with a 32% reduction in ARD incidence (IRR = 0.68; 95% CI 0.67–0.70). In
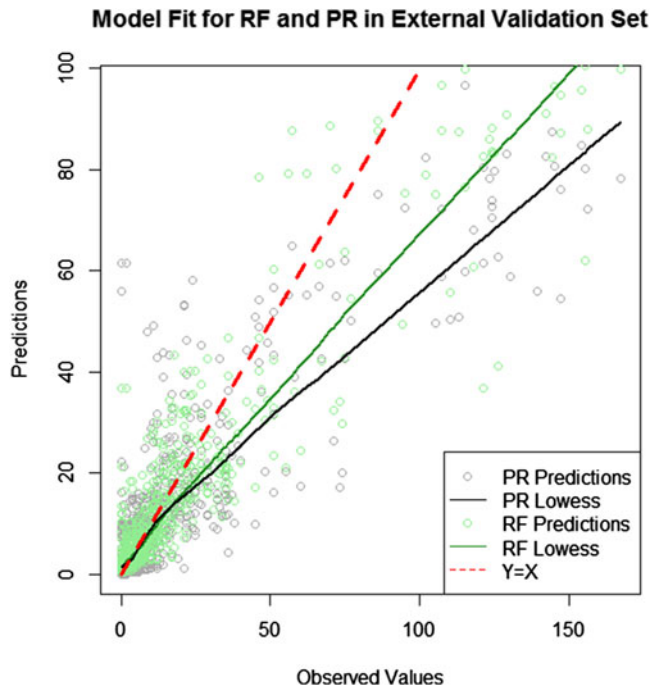
Model Fit for RF and PR in External Validation Set

Fig. 3. Predicted and observed ARD case counts in external validation set.

Table 3. Comparison of variable importance between the cross-validation set and the external validation set from RF models

Variable	Average % increase MSE CV	% Increase MSE EV
New adenovirus vaccine	133.09	131.46
Trainee population size in hundreds	83.398	82.34
Old adenovirus vaccine	75.262	71.67
January	38.998	40.71
Training type (BCT/OSUT)	35.146	35.69
February	23	25.17
Fort Sill	22.926	21.75
Fort Jackson	21.794	20.85
BPG availability	20.806	19.96
Fort Leonard Wood	18.226	19.51
Fort Benning	17.928	16.7
May	15.274	13.75
March	14.39	9.76
August	11.404	9.68
July	10.56	9.54
April	9.258	9.26
November	8.566	9
June	8.16	7.34
September	4.262	4.84
October	3.002	3.03
December	−1.424	1.95

the RF *k*-fold cross-validation, BPG was found to be the seventh most important variable, with a 21% increase in MSE. This indicates that while BPG availability is not highly predictive of all-cause ARD, it does have a significantly protective effect. This result is consistent with a previous study which found that BPG prophylaxis was associated with a broad and persistent protection against not only group A streptococcal infections, but ARD in general [26].

It is well established that the development of the new adenovirus vaccine was associated with substantial declines in ARD [9, 10]. Our finding that the new adenovirus vaccine was associated with an 89% lower incidence rate compared with when there was no adenovirus vaccine further supports its effectiveness. The old adenovirus vaccine, which was thought to be effective [27] but not to the same degree as the new vaccine [9], was similarly statistically significantly protective against ARD, with a 61% reduced IRR compared with weeks when there was no adenovirus vaccine available.

Trainee population size in hundreds was the second most important variable in the RF model, representing an 84% mean increase in MSE across the five folds. This result was expected, since larger trainee class sizes means that more people are at risk of contracting ARD. In the Poisson model, log trainee population size in hundreds was included as an offset in order to constrain its effect on the remaining variables' coefficient estimates.

Temporal covariates such as the months of January and February were also important variables, associated with an average of 42% and 25% increase in RMSE, respectively. However, January and February both are shown to be statistically significant protective factors (IRR = 0.53 and 0.81, respectively) on ARD. January comes after what is usually an extended winter holiday when trainees are permitted to leave the installation. Leaving the psychologically and physically stressful BCT environment may result in improved temporary immune function upon their return [8]. An additional explanation could be that summer months are when the most recruits receive BCT, and therefore more transmission is expected compared with other months.

Fort Leonard Wood and Fort Benning were both found to be significant risk factors for ARD compared with Fort Jackson (IRRs = 1.29 and 1.17 with 95% CIs of 1.26–1.33 and 1.13–1.20, respectively). Oppositely, Fort Sill was found to be significantly protective against ARD (IRR = 0.67, 95% CI 0.65–0.69). Each base may have a different baseline risk for the various pathogens that make up ARD due to varying environmental conditions across sites, and the protocol for assigning new recruits to BCT sites. Recruits from different parts of the country may have different immune profiles and thus may be more or less susceptible to infection from pathogens vulnerable to BPG.

Limitations

This study is subject to a number of limitations. First, BPG prophylaxis availability does not mean that BPG was not available or used reactively to control an ongoing outbreak. Situations where BPG was used to control an outbreak were not explicitly captured in the ARD Surveillance Program. Additionally, only those in the impacted trainee class or barrack would receive the antibiotic, and not necessarily all trainees in the installation. Similarly, adenovirus vaccine availability does not necessarily reflect use. There could be variability in administration in the few years between when the final adenovirus vaccine shipments and when the installations' respective stockpiles were depleted.

Table 4. Incidence rate ratios for independent variables in the Poisson regression model fit to the training data (22.5 years) with a population offset

Variable	IRR	95% CI	*P*-value
Installation			
Fort Sill	0.67	(0.65–0.69)	<0.001
Fort Benning	1.17	(1.13–1.20)	<0.001
Fort Leonard Wood	1.29	(1.26–1.33)	<0.001
Fort Jackson	REF	REF	REF
Month			
January	0.55	(0.53–0.57)	<0.001
February	0.81	(0.78–0.83)	<0.001
March	1.06	(1.03–1.08)	<0.001
April	1.07	(1.04–1.10)	<0.001
May	1.11	(1.08–1.14)	<0.001
June	0.93	(0.90–0.95)	<0.001
July	1.02	(0.99–1.04)	0.211
August	1.08	(1.06–1.11)	<0.001
September	REF	REF	REF
October	1.03	(1.00–1.05)	0.031
November	1.11	(1.08–1.14)	<0.001
December	1.04	(1.01–1.07)	0.018
BPG availability			
Available	0.68	(0.67–0.70)	<0.001
Shortage	REF	REF	REF
Adenovirus vaccine availability			
New vaccine	0.11	(0.11–0.11)	<0.001
Old vaccine	0.39	(0.38–0.39)	<0.001
Shortage	REF	REF	REF

Second, ARD has a syndromic case definition that encompasses viral and bacterial respiratory pathogens. Since BPG is an antibiotic, it is only effective against specific Gram-positive bacteria, including some *Streptococci* and *Treponema*. Meanwhile, it is thought that the majority of ARD in this population is caused by viral pathogens [2, 3]. BPG effectiveness in a given week is a function of the composition of respiratory pathogens circulating in the recruit population; this is not captured as part of the ARD Surveillance Program. Military treatment facilities are encouraged to test for strep in at least 60% of recruits with ARD, but true strep test compliance rates are around 30%. We opted to use ARD rather than a strep-specific outcome due to low, potentially differential testing compliance rates across installations over time.

In addition to strep surveillance, it is known that overall ARD surveillance practices differ by installation as well as over time. Differences in surveillance between installations are controlled for explicitly in our model, but the effect of surveillance bias cannot be completely accounted for. As a result, care should be given in making between-installation comparisons. Within-installation temporal variation in surveillance (such as in response to an outbreak) was not controlled for in the model. The only type of temporal variation taken into account was month, and not week,

season or year. Taking into account week of the year would have resulted in estimating an additional 52 parameters, which would have been computationally intensive and would have significantly decreased our statistical power. Year was not included in the analysis so that the resulting models could be then used to forecast ARD for prospective outbreak detection. Season was not included because within-season variation in weather patterns, as well as overall seasonal trends, should have been captured by the finer temporal scale.

The results of this work suggest that BPG prophylaxis reduces the rate of all-cause ARD by 32%. A randomised controlled trial to test the effectiveness of different dosing regimens of this intervention at reducing ARD would provide additional data on the modes of action and overall effectiveness.

References

1. **Armed Forces Health Surveillance Center (AFHSC)** (2015) Surveillance snapshot: illness and injury burdens among U.S. military recruit trainees, 2014. *MSMR* **22**, 26.
2. **Gray GC et al.** (1999) Respiratory diseases among U.S. military personnel: countering emerging threats. *Emerging Infectious Diseases* **5**, 379–385.
3. **Sanchez JL et al.** (2015) Respiratory infections in the U.S. military: recent experience and control. *Clinical Microbiology Reviews* **28**, 743–800.
4. **Breese B, Stanbury J and Upham H** (1945) Influence of crowding on respiratory illness in a large naval training station. *War Medicine* **7**, 143–146.
5. **Russell KL** (2006) *Respiratory Infections in Military Recruits. Recruit Medicine*. Washington, DC: TMM Publications, pp. 227–253.
6. **Gleeson M** (2007) Immune function in sport and exercise. *Journal of Applied Physiology* **103**, 693–699.
7. **Cohen S, Tyrrell DAJ and Smith AP** (1991) Psychological stress and susceptibility to the common cold. *New England Journal of Medicine* **325**, 606–612.
8. **Dudding BA et al.** (1973) Acute respiratory disease in military trainees: the adenovirus surveillance program, 1966–1971. *American Journal of Epidemiology* **97**, 187–198.
9. **Clemmons NS et al.** (2017) Acute respiratory disease in US Army trainees 3 years after reintroduction of adenovirus vaccine1. *Emerging Infectious Diseases* **23**, 95–98.
10. **Radin JM et al.** (2014) Dramatic decline of respiratory illness among US military recruits after the renewed use of adenovirus vaccines. *Clinical Infectious Diseases* **59**, 962–968.
11. **Hoke CH and Snyder CE** (2013) History of the restoration of adenovirus type 4 and type 7 vaccine, live oral (adenovirus vaccine) in the context of the Department of Defense acquisition system. *Vaccine* **31**, 1623–1632.
12. **Morris AJ and Rammelkamp CH** (1957) Benzathine penicillin G in the prevention of streptococcic infections. *Journal of the American Medical Association* **165**, 664–667.
13. **Davis JC and Schmidt WC** (1957) Benzathine penicillin G. *New England Journal of Medicine* **256**, 339–342.
14. **Gray GC et al.** (1998) Weekly oral azithromycin as prophylaxis for agents causing acute respiratory disease. *Clinical Infectious Diseases* **26**, 103–110.
15. **Kassem AS et al.** (1996) Rheumatic fever prophylaxis using benzathine penicillin G (BPG): two-week versus four-week regimens: comparison of two brands of BPG. *Pediatrics* **97**, 992–995.
16. **Broderick MP et al.** (2011) Serum penicillin G levels are lower than expected in adults within two weeks of administration of 1.2 million units. *PLoS ONE* **6**, e25308.
17. **Lee S, Eick A and Ciminera P** (2008) Respiratory disease in army recruits: surveillance program overview, 1995–2006. *American Journal of Preventive Medicine* **34**, 389–395.
18. **Arlot S and Celisse A** (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys* **4**, 40–79.

19. **Bouckaert RR and Frank E** (2004) *Evaluating the Replicability of Significance Tests for Comparing Learning Algorithms. Advances in Knowledge Discovery and Data Mining.* Berlin, Heidelberg: Springer, pp. 3–12.
20. **Liaw A and Wierer M** (2002) Classification and regression by randomForest. *R News* **2**, 18–22.
21. **Hyndman RJ** (2017) *forecast: forecasting functions for time series and linear models* R package version 8.1. Available at http://github.com/robj-hyndman/forecast.
22. **Breiman L** (1996) Bagging predictors. *Machine Learning* **24**, 123–140.
23. **Breiman L** (2001) Random forests. *Machine Learning* **45**, 5–32.
24. **Cutler DR *et al.*** (2007) Random forests for classification in ecology. *Ecology* **88**, 2783–2792.
25. **Strobl C *et al.*** (2007) Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* **8**, 25.
26. **Gunzenhauser JD *et al.*** (1992) Broad and persistent effects of benzathine penicillin G in the prevention of febrile, acute respiratory disease. *The Journal of Infectious Diseases* **166**, 365–373.
27. **Russell KL *et al.*** (2006) Vaccine-preventable adenoviral respiratory illness in US military recruits, 1999–2004. *Vaccine* **24**, 2835–2842.