Prediction of upper respiratory illness using salivary IgA in youth athletes

Running head: Immune function and URI in youth athletes

Submission type: Original investigation

Thomas Sawczuk[1,2], Ben Jones[1,3,4,5,6], Mitchell Welch[6], Clive Beggs[1], Sean Scantlebury[1], Kevin Till[1,3,5]

[1] Institute for Sport, Physical Activity and Leisure, Leeds Beckett University, Leeds, United Kingdom

[2] Queen Ethelburga's Collegiate, Thorpe Underwood, York, United Kingdom

[3] Yorkshire Carnegie Rugby Club, Headingley Carnegie Stadium, Leeds, United Kingdom

[4] The Rugby Football League, Red Hall, Leeds, United Kingdom

[5] Leeds Rhinos Rugby Club, Headingley Carnegie Stadium, Leeds, United Kingdom

[6] School of Science and Technology, University of New England, Armidale, New South Wales, Australia

Corresponding Author:

Thomas Sawczuk: Room G12, Cavendish Hall; Institute for Sport, Physical Activity and Leisure; Leeds Beckett University, Headingley Campus; West Yorkshire; LS6 3QS

Email: t.sawczuk@leedsbeckett.ac.uk

Phone: (0044) 7530 945555

Word count: 3857

Abstract word count: 249

Table count: 0

Figure count: 4

Abstract

Purpose: To evaluate the relative importance and predictive ability of salivary IgA (s-IgA) measures with regards to upper respiratory illness (URI) in youth athletes. **Methods:** Over a 38-week period, twenty-two youth athletes (age 16.8; ±0.5 years) provided daily symptoms of URI and fifteen fortnightly passive drool saliva samples, from which s-IgA concentration and secretion rate were measured. Kernel smoothed bootstrapping generated a balanced dataset with simulated data points. The random forest algorithm was used to evaluate the relative importance (RI) and predictive ability of s-IgA concentration and secretion rate with regards to URI symptoms present on the day of saliva sampling (URI_{day}), within two weeks of sampling (URI_{2wk}) and within four weeks of sampling (URI_{4wk}). **Results:** The percentage deviation from average healthy s-IgA concentration was the most important feature for URI_{day} (median RI 1.74, interquartile range 1.41-2.07). The average healthy s-IgA secretion rate was the most important feature within URI_{4wk} (median RI 0.94, interquartile range 0.79-1.13). No feature was clearly more important than any other in URI_{2wk}. The median area under the curve values were 0.68, 0.63 and 0.65 for URI_{day}, URI_{2wk} and URI_{4wk} respectively. **Conclusions:** The RI values suggest the percentage deviation from average healthy s-IgA concentration may be used to evaluate the short-term risk of URI, whilst the average healthy s-IgA secretion rate may be used

to evaluate the long-term risk. However, the results show that neither s-IgA concentration nor secretion rate can be used to accurately predict URI onset within a four-week window in youth athletes.

Introduction

The negative impact of illness on performance, and the mechanisms by which athletic development could be hindered, are now well established[1–3]. Illness usually occurs within a sporting situation when the immune response to a pathogen is compromised[4–7]. It has recently been documented that psychological stress affects this immune response by a similar magnitude when compared to physical stress[8]. Given that youth athletes are subjected to a unique set of academic, maturational and social stressors alongside their sporting demands (i.e. their training/match load and associated travel requirements)[9], and that their stress coping skills do not fully develop until adulthood[10], there is significant scope for the combined effect of these physical and psychological stressors to place them at significant risk of illness incidence in high performance sporting environments[11,12]. Consequently, in an effort to maximise youth athlete development and performance, regular monitoring of the risk of upper respiratory illness (URI) has been encouraged[13]. Despite this, only one study has evaluated the risk of URI in youth athletes[14], as the majority of studies in the area have focussed on collegiate[5,7] or professional[4,6] athletes.

One method of assessing a youth athlete's risk of URI is via assessment of their salivary Immunoglobulin A (s-IgA)[7,15]. In recent years, the use of s-IgA as an URI monitoring strategy has

increased as it provides a relatively non-invasive and objective measure of athlete immune function[16,17]. IgA is the predominant antibody secreted by the mucosa as part of the 'first line of defence' against the vast majority of infections attempting to invade the human body[17,18], so it is unsurprising that some studies have associated reductions in s-IgA with an increased risk of URI incidence[4–6]. However, this association has not always been replicated[7,14] and the true predictive ability of s-IgA with regards to URI incidence remains unclear[5–7].

Some of the discrepancy in findings between s-IgA studies can be accounted for by the data collection methods used. S-IgA can be measured as either a concentration[5,6] or a secretion rate[5,7], and it is possible that both measures are important at different time points prior to URI incidence. For example, a significant transient weekly reduction in s-IgA concentration in the three weeks prior to URI incidence has been reported in elite adult yachtsmen[6], but no significant difference in s-IgA secretion rate was observed on a similar weekly basis in female collegiate soccer players[7]. Similarly, although s-IgA secretion rate was associated with an increased risk of URI in collegiate American football players, s-IgA concentration was not in the four weeks or greater time period between sampling points[5]. Taken together, the results of these studies suggest that s-IgA concentration may be a more important short-term predictor of URI, with a

declining importance up to three weeks prior to URI onset. Conversely, s-IgA secretion rate may be a more important long-term predictor of URI (four weeks and greater prior to onset). However, no study has proven this hypothesis so a study considering the relative importance of s-IgA concentration and secretion rate as predictors of URI at different time points in the four weeks prior to onset is warranted.

Alongside a greater understanding of the relative importance of s-IgA concentration and secretion rate at different time points prior to URI onset, it would be beneficial to elucidate the true predictive ability of s-IgA with regards to URI incidence. Although a three week transient drop in s-IgA concentration below the individual's average healthy value (averaged from all those values where URI was not present) was associated with URI incidence in elite adult yachtsmen, the same study found that only 28% of the samples identifying a drop in s-IgA concentration below 70% of this value resulted in subsequent URI incidence[6]. Similarly, only 48% of the samples falling below 40% of the average healthy s-IgA concentration resulted in future URI incidence. Furthermore, although s-IgA secretion rate was the only significant measure in the regression model for collegiate American football players, the model over-predicted URI incidence at each sampling point and s-IgA secretion rate accounted for only 12-42% of the variance in URI incidence[5]. It

has recently been shown within injury prevention research that training load variables demonstrating strong associations with injury incidence are not necessarily strong predictors of injury timing[19,20], so it is pivotal that the true predictive ability of immune measures with regards to URI is fully evaluated.

To understand the true relative importance and predictive ability of s-IgA with regards to URI in youth athletes, it is necessary to shift from a data modelling or correlation-based approach (e.g. mixed effects models or regressions) to a machine learning or prediction-based approach (e.g. decision trees or random forests)[21]. The key difference between these two approaches is the method by which they are optimised. Correlation-based approaches are optimised to maximise how well the model fits the data, thus allowing risk factors of URI incidence to be identified. However, machine learning approaches are optimised by evaluating the predictive accuracy of the model on unseen data, allowing the importance of different variables as predictors to be evaluated[21]. Recently, the use of machine learning techniques to predict adverse outcomes to the training process has been advocated[22]. However, when test samples are unbalanced (e.g. with low URI incidence), these methods can produce inaccurate results as the classification model produced may achieve high levels of overall accuracy by predicting the majority class (e.g. 'not ill') with high accuracy at the cost of

poor accuracy for the minority class (e.g. 'ill'), rather than predicting both classes equally[23]. To resolve this issue, several bootstrapping methods have been identified. These methods create a balanced sample dataset by providing simulated data based on the statistical properties of the original dataset, thereby allowing machine learning techniques to provide accurate results[24]. This approach has not yet been considered within the s-IgA and URI literature in sport, so there is scope for its introduction as a method of identifying the true relative importance and predictive ability of s-IgA with regards to URI incidence in youth athletes.

Although the current literature suggests that s-IgA concentration may be a suitable short-term risk factor for URI incidence[6] and that s-IgA secretion rate may be a more appropriate long-term risk factor of URI[5], no study has evaluated the relative importance of these variables as predictors of URI onset at different time points prior to its occurrence using data previously unseen by the model. Furthermore, the true predictive ability of s-IgA with regards to URI can be questioned[5–7]. Consequently, the aims of this study were to 1) identify the relative importance of s-IgA concentration and secretion rate at different timepoints in the four weeks prior to URI onset; and 2) identify the true predictive ability of s-IgA with regards to URI

incidence in youth athletes using data previously unseen by the model.

Methods

Participants

Twenty-two youth athletes aged 16-18 years (age 16.8 ± 0.5 years, height 173.3 ± 6.5 cm, body mass 70.1 ± 10.8 kg) participated in this study. Participants were recruited from a local independent school in the United Kingdom (UK), where they were members of the school's sport scholarship programme. Basketball (*n*=2), cricket (*n*=3), football (*n*=5), hockey (*n*=1), netball (*n*=5) and rugby union (*n*=6) were represented by sixteen males and six females competing at club/school (*n*=9), professional academy (*n*=4), county/regional (*n*=6) and international (*n*=3) standard in their respective sports. In total, participants took part in 10.5 ± 3.4 hours of sporting activity per week alongside their academic commitments of 8.30am-4pm. All participants were made aware of the benefits and risks of the study, and written informed consent was provided by all participants and their parents prior to the study. Ethics approval was granted by Leeds Beckett University Ethics Committee.

Salivary IgA

Over 38-weeks, between August and April, participants provided 15 saliva samples. Each sample was separated by two

weeks unless external factors (e.g. school holidays) dictated otherwise. Consequently, samples were collected three weeks apart twice and four weeks apart once. On each occasion, saliva samples were collected on the same weekday (Wednesday) within a 30-minute period, between 10:15 and 10:45. Participants provided samples prior to their first exercise of the day and following a 2-hour period without eating or drinking. Upon arrival, 50 ml of water was provided for participants to rinse out their mouths. Saliva sampling (via the passive drool method) commenced 10 minutes later. Participants were instructed to sit upright with their head tilted downwards and provide an unstimulated saliva sample of approximately 1.5 ml into a 3.5ml cryogenic vial. The volume of the sample and the time taken to produce it were recorded prior to freezing at -80˚C for storage within 30 minutes of completion. All standardisation and measurement methods mirrored those of previous studies[5,6]. Saliva samples were assessed in duplicate for s-IgA concentration by a service laboratory (ARU Biomarker Laboratory, Norwich, UK) using enzyme-linked immunosorbent assay in line with the manufacturer's protocols (Secretory IgA EIA Kit Research, Catalogue Number 1-1602-SAL, Salimetrics LLC, State College, PA). S-IgA results were calculated as s-IgA concentration (µg/mL) and s-IgA secretion rate (s-IgA concentration divided by the time taken to provide the sample in minutes; µg/min).

In order to ensure that URI episodes were not systematically missed (e.g. through non-responding when ill if not at school), s-IgA samples were only included in analysis if a minimum number of daily URI questions were completed before the next sampling time point. The thresholds for inclusion were 7/14 for two weeks, 10/21 for three weeks and 14/28 for four weeks. The between-day CV was calculated for s-IgA concentration and secretion rate from the thirteen participants who did not suffer from URI at either week 0 or 2.

URI questionnaire

Participants provided URI details seven days per week as part of their usual wellness monitoring programme. Details were provided using an online platform (Sports-train, Nice) similar to a previous study[25]. The question to report URI was: *"Have you continuously experienced any of the following over the last 24 hours? Sore throat, runny nose, cough, scratchy throat, nasal congestion, headache, fever, hoarseness, sneezing and/or body aches and pains (not related to delayed onset muscle soreness)?"* based on previously identified symptoms of URI[26] and similar to other self-report questions used within sport as part of wellness monitoring programmes[27]. Participants also identified the number of days the symptoms were present. URI episodes were only included in the analysis if a minimum of one symptom

persisted continuously for 72 hours or longer and were not considered to be recurrent (i.e. those URI episodes which began within seven days of the previous episode finishing[6]). Those URI episodes identified as occurring due to an allergy, established via discussion with the athlete in question, were also removed from analysis.

Statistical analyses

In order to assess the predictive ability of s-IgA with regards to URI, the collected data was split into three datasets. One evaluated URI incidence on the day of saliva sampling (URI_{day}), one evaluated URI incidence within the 2 weeks following sampling (URI_{2wk}) and one evaluated URI incidence within the 4 weeks following sampling (URI_{4wk}). These timings were chosen to best replicate a previous study considering deviations from the average healthy s-IgA concentration[6], within the fortnightly sampling limitation of this study.

Based on previous research, three features were included in each dataset for s-IgA concentration and secretion rate: the individual's average healthy measure, the percentage deviation from the individual's average healthy measure, and the raw value of the measure[5,6]. The individual's average healthy measure was calculated by averaging all of his/her s-IgA concentration or secretion rate values over a minimum of four weeks before and

two weeks after URI incidence, matching as closely as possible to the three and one week periods previously used[6]. To account for other potential confounding variables, the three classification features of sport group, sex and environment were also added to the model. All three classification features had previously shown relationships with URI incidence[28,29]. Due to the small sample size for each sport, the sport group variable clustered athletes into court-based (basketball and netball), field-based (football and hockey), contact (rugby) and summer (cricket) sports. Sex split the athletes based on their sex (female for netball and hockey, male for all others), and environment split the sports into indoor (basketball and netball) or outdoor (all others) sports. The outcome variable was a binary URI or no URI.

To establish the relative importance and predictive ability of s-IgA concentration and secretion rate with regards to URI in youth athletes, the random forest machine learning algorithm was used. To provide a robust assessment of the predictive ability of the features in URI, URI_{2wk} and URI_{4wk}, 1000 trials were conducted. In each trial, new randomized, stratified samples were generated from the original datasets for the training (50%), validation (25%) and testing (25%) subsets and the random forest algorithm was run. The training subset was z-score normalised and kernel smoothed bootstrapping[24] was used to balance it, providing equal numbers of positive and negative

URI samples. Neither the validation nor the testing subsets were subjected to this resampling process. Both the training and validation subsets were used for parameter selection and the predictive ability of the models was assessed using the testing subset as a measure of out-of-sample predictive ability. The results are presented via box and whisker plots, which provide the median, interquartile range (IQR) and outlying results from the 1000 trials evaluated for each model.

The relative feature importance was measured using the permutation method with mean decrease in accuracy reported as the metric for feature importance[30]. The mean decrease in accuracy is measured in arbitrary units (AU) and provides a relative measure of the importance for each feature to the performance of each model. A larger number is representative of a greater importance to the prediction of URI, but these values cannot be compared between models. The predictive ability of the models was evaluated using the area under the curve (AUC). The AUC is calculated based on the number of correct and incorrect predictions made by the model, providing a value between 0 and 1, where 0.5 indicates that the model has a 50% chance of predicting the right value. An AUC value greater than 0.75 was used as the threshold for a *good* predictive model[31]. All analyses were completed using bespoke scripts in MATLAB version 9.4 (MathWorks, Natick, Massachusetts).

Results

Of the 3,311 URI responses included in the study, 372 (11.2%) were classified as URI from 45 unique episodes. Participants reported a median of 2 URI episodes (range 0-7) and the average length of URI was 17 ± 18 days. The between-day CV was 31.1% for s-IgA concentration and 61.7% for s-IgA secretion rate.

Figures 1, 2 and 3 show the relative feature importance for the random forest models for URI_{day}, URI_{2wk} and URI_{4wk}. The positive values of the features across all three models indicate that all features improved the predictive ability of the model. However, from URI_{day} to URI_{4wk} there is a clear difference in the relative importance of two key variables: the percentage deviation from average healthy s-IgA concentration and the average healthy secretion rate. For the URI_{day} model, the percentage deviation from average healthy s-IgA concentration was the most important predictive feature (median relative importance 1.74, IQR 1.41-2.07), whereas for URI_{4wk}, the average healthy secretion rate was most important (median relative importance 0.94, IQR 0.79-1.13). No feature was clearly more important than any other in URI_{2wk}. Figure 4 provides the AUC for all models and shows that the median AUC is not

greater than the 0.75 practically meaningful threshold set in any of the models.

INSERT FIGURES 1, 2, 3 AND 4 AROUND HERE

Discussion

The aims of this study were to 1) identify the relative importance of s-IgA concentration and secretion rate as predictors of URI at different timepoints in the four weeks prior to URI onset; and 2) identify the true predictive ability of s-IgA with regards to URI incidence in youth athletes. Using a prospective longitudinal research design, our findings demonstrate that the relative importance of s-IgA measures differs depending upon the time before URI onset. At URI_{day}, the percentage deviation from the average healthy s-IgA concentration was the most important predictive feature of the model, whereas at URI_{4wk} the average healthy s-IgA secretion rate was the most important. All features held similar importance in the URI_{2wk} model. In terms of predictive ability, the AUC of all three models fell below the 0.75 AU threshold for a *good* predictor indicating that s-IgA measures cannot be used to accurately predict URI at three individual time points within four weeks of the URI episode (URI_{day}, URI_{2wk} and URI_{4wk}).

The key finding of this study is the differentiation between the importance of s-IgA concentration and s-IgA secretion rate as predictors of URI at different timepoints in the four weeks prior to URI onset. It has previously been shown that the percentage deviation from the average healthy s-IgA concentration transiently drops in the three weeks prior to URI[6] and is significantly related with URI incidence on the day of URI[4,6]. Conversely, s-IgA secretion rate has previously shown a greater association with URI than s-IgA concentration over periods of four weeks and greater[5]. Our predictive models add further support to these differences by identifying the percentage deviation from the average healthy s-IgA concentration as the most important feature for URI_{day} (median relative importance 1.74, IQR 1.41-2.07), and the average healthy s-IgA secretion rate as the most important feature for URI_{4wk} (median relative importance 0.94, IQR 0.79-1.13). However, at URI_{2wk} there appears to be a crossover in the importance of the two measures and there is no feature which is clearly more important than any other. It is possible that this can also be explained by the previously established three-week transient drop in s-IgA concentration prior to URI incidence[6]. In that study, a significant difference in the percentage deviation from average healthy s-IgA concentration was only observed when URI values were compared to four-weeks pre-incidence and measures of one and two weeks post-URI[6]. It is therefore possible that although the

percentage deviation from average healthy s-IgA concentration is important on the day of SRI, it may have little practical importance in the weeks leading up to it. Unfortunately, due to the fortnightly sampling employed in this study, we are unable to confirm the importance of the measure at one and three weeks prior to URI onset.

The second notable finding of this study is that none of our models' AUCs were greater than the 0.75 threshold set for a *good* predictive model of URI in youth athletes. This observation confirms that while s-IgA can be used as a risk factor for URI, it cannot predict URI incidence alone. Such a finding is perhaps unsurprising given the multifactorial risk factors of URI, including low energy availability, which can compromise immunity within two days[32], poor sleep, which influences immune function via the hypothalamic-pituitary-adrenal axis[33], environmental conditions, such as sharing water bottles and exposure to foreign pathogens from team-mates[28,34] and the individual's stress/recovery balance[28,34]. However, the lack of predictive ability of immune measures with regards to URI may also be related to their poor reliability. Our results indicated that the between day CV was 31.1% for s-IgA concentration and 61.7% for s-IgA secretion rate. With a between day CV of 31.1%, a s-IgA concentration sample with a true value of 100 µg/ml could hold values from 68.9-131.1 µg/ml based on the

test's error. This wide variation in the potential values of s-IgA concentration observed relative to its true value limits its predictive ability as the observed value could be much greater or smaller than the true value, resulting in numerous false positive or negative predictions of URI.

Although our manuscript uses novel and advanced machine learning based algorithms to identify the relative importance and true predictive ability of s-IgA with regards to URI incidence in youth athletes, the results are subject to limitations. The first of which is the lack of training load, sleep and psychological stress data, which could have explained some of the reductions in s-IgA levels and, if accounted for, may have helped provide a more accurate predictive model. Secondly, the use of self-reported data may have influenced the accuracy of the prediction model. The wide range of URI lengths (17 $\pm$ 18 days) is potentially indicative of different participants having different individual thresholds for when they are suffering from symptoms, which may have had a negative effect on the predictive ability of the model. Finally, the sample should be considered. Although this study was conducted in youth athletes, it was necessary to compare the research to adult athletes based on the current state of the literature[4–7]. It is important to note that the different demands of the two cohorts outlined in the introduction (different psychological stressors and coping mechanisms[8–10])

may ensure that comparison between the two cohorts proves unsuitable. Future studies should consider the true predictive ability of s-IgA with regards to URI using the machine learning based methods considered in this study to clarify whether there are any differences between the two cohorts. It should also be noted that this study used a small and unbalanced sample (16 males, 6 females), which may have influenced the results. Despite these limitations, our novel statistical analyses have provided results of significant merit within this field.

Practical applications

The results of this study are important for practitioners considering the use of s-IgA measures as predictors of URI incidence. Previous correlation-based analyses have identified significant relationships between s-IgA concentration, s-IgA secretion rate and URI incidence[5–7]. However, although the relative importance of these measures within our analyses confirms their usefulness as risk factors for URI incidence at different timepoints in the lead up to incidence, the predictive ability of the models indicate that s-IgA cannot be used to predict URI onset in youth athletes. It is therefore pivotal that s-IgA measures are only considered as risk factors for URI incidence and should be considered alongside other risk factors such as the sport evaluated, life stressors, nutritional deficiencies and energy

availability[28] when considering changes to an athlete's training schedule.

Conclusion

This study is the first to use advanced machine learning techniques to provide a robust assessment of the relative importance and predictive ability of s-IgA with regards to URI in youth athletes. In summary, our results show that there is a differentiation in the roles of s-IgA concentration and secretion rate in the four weeks prior to URI incidence. The percentage deviation from the average healthy s-IgA concentration was identified as the most important feature in URI_{day}, indicating it may be a useful short-term risk factor for URI incidence. However, its usefulness as a short-term risk factor in practice may be limited by its unclear importance relative to other features in URI_{2wk}. The average healthy s-IgA secretion rate was identified as the most important feature in URI_{4wk}, indicating that the athlete's healthy s-IgA secretion rate could be taken as a useful long-term risk factor for URI. However, none of the models produced were capable of accurately predicting URI onset, limiting the use of any of the features measured to risk factors rather than true predictors of URI incidence.

References

1. Raysmith BP, Drew MK. Performance success or failure is influenced by weeks lost to injury and illness in elite Australian track and field athletes: A 5-year prospective study. *J Sci Med Sport*. 2016;19(10):778-783. doi:10.1016/j.jsams.2015.12.515

2. Weidner TG, Sevier TL. Sport, exercise, and the common cold. *J Athl Train*. 1996;31(2):154-159.

3. Friman G, Wesslen L. Special feature for the Olympics: Effects of exercise on the immune system: Infections and exercise in high-performance athletes. *Immunol Cell Biol*. 2000;78(5):510-522.

4. Cunniffe B, Griffiths H, Proctor W, Davies B, Baker JS, Jones KP. Mucosal immunity and illness incidence in elite rugby union players across a season. *Med Sci Sports Exerc*. 2011;43(3):388-397. doi:10.1249/MSS.0b013e3181ef9d6b

5. Fahlman MM, Engels H-J. Mucosal IgA and URTI in American college football players: A year longitudinal study. *Med Sci Sport Exerc*. 2005;37(3):374-380. doi:10.1249/01.MSS.0000155432.67020.88

6. Neville V, Gleeson M, Folland JP. Salivary IgA as a risk factor for upper respiratory infections in elite professional athletes. *Med Sci Sports Exerc*. 2008;40(7):1228-1236.

doi:10.1249/MSS.0b013e31816be9c3

7. Putlur P, Foster C, Miskowski JA, et al. Alteration of immune function in women collegiate soccer players and college students. *J Sport Sci Med*. 2004;3(4):234-243.

8. Edwards JP, Walsh NP, Diment BC, Roberts R. Anxiety and perceived psychological stress play an important role in the immune response after exercise. *Exerc Immunol Rev*. 2018;24:26-34.

9. Mountjoy M, Armstrong N, Bizzini L, et al. IOC consensus statement: "Training the elite child athlete." *Br J Sports Med*. 2008;42(3):163-164. doi:10.1136/bjsm.2007.044016

10. Compas BE, Jaser SS, Dunbar JP, et al. Coping and emotion regulation from childhood to young adulthood: Points of convergence and divergence. *Aust J Psychol*. 2014;66(2):71-81. doi:10.1111/ajpy.12043.Coping

11. Pedersen A, Zachariae R, Bovbjerg DH. Influence of psychological stress on upper respiratory infection - A meta-analysis of prospective studies. *Psychosom Med*. 2010;72(8):823-832.

12. Moreira A, Arsati F, Lima-Arsati YB de O, Simões AC, Araújo VC de. Monitoring stress tolerance and occurrences of upper respiratory illness in basketball players by means of psychometric tools and salivary biomarkers. *Stress Heal*. 2011;27(3):e166-e172.

doi:DOI: 10.1002/smi.1354

13. Bergeron MF, Mountjoy M, Armstrong N, et al.
International Olympic Committee consensus statement
on youth athletic development. *Br J Sports Med*.
2015;49(13):843-851. doi:10.1136/bjsports-2015-
094962

14. Orysiak J, Witek K, Zembron-Lacny A, Morawin B,
Malczewska-Lenczowska J, Sitkowski D. Mucosal
immunity and upper respiratory tract infections during a
24-week competitive season in young ice hockey
players. *J Sports Sci*. 2017;35(13):1255-1263.
doi:10.1080/02640414.2016.1218039

15. He CS, Tsai ML, Ko MH, Chang CK, Fang SH.
Relationships among salivary immunoglobulin A,
lactoferrin and cortisol in basketball players during a
basketball season. *Eur J Appl Physiol*. 2010;110(5):989-
995. doi:10.1007/s00421-010-1574-8

16. Gleeson M. Immune function in sport and exercise. *J
Appl Physiol*. 2007;103(2):693-699.
doi:10.1152/japplphysiol.00008.2007.

17. Brandtzaeg P. Role of secretory antibodies in the
defence against infections. *Int J Med Microbiol*.
2003;293(1):3-15. doi:10.1078/1438-4221-00241

18. Brandtzaeg P. Secretory immunity with special reference
to the oral cavity. *J Oral Microbiol*. 2013;5(epub).

doi:10.3402/jom.v5i0.20401

19.   Fanchini M, Rampinini E, Riggio M, Coutts AJ, Pecci C, McCall A. Despite association, the acute:chronic work load ratio does not predict non-contact injury in elite footballers. *Sci Med Footb*. 2018;epub. doi:10.1080/24733938.2018.1429014

20.   McCall A, Dupont G, Ekstrand J. Internal workload and non-contact injury: A one-season study of five teams from the UEFA Elite Club Injury Study. *Br J Sports Med*. 2018;epub. doi:10.1136/bjsports-2017-098473

21.   Breiman L. Statistical modeling: The two cultures. *Stat Sci*. 2001;16(3):199-309. doi:10.1214/ss/1009213726

22.   Robertson SJ. Improving load/injury predictive modelling in sport: The role of data analytics. *J Sci Med Sport*. 2014;18(S1):e25-e26. doi:10.1016/j.jsams.2014.11.198

23.   Hulse J Van, Khoshgoftaar TM, Napolitano A. Experimental perspectives on learning from imbalanced data. In: *Proceedings of the 24th International Conference on Machine Learning*. ; 2007:935-942.

24.   Silverman B. *Density Estimation for Statistics and Data Analysis*. New York, NY: Routledge; 2018.

25.   Sawczuk T, Jones B, Scantlebury S, Till K. The influence of perceptions of sleep on wellbeing in youth athletes. *J Strength Cond*. 2018;epub.

26.   Spence L, Brown WJ, Pyne DB, et al. Incidence, etiology and symptomatology of upper respiratory illness in elite athletes. *Med Sci Sport Exerc*. 2007;39(4):577-586. doi:10.1249/mss.0b013e31802e851a

27.   Thornton HR, Delaney JA, Duthie GM, et al. Predicting self-reported illness for professional team-sport athletes. *Int J Sports Physiol Perform*. 2016;11(4):543-550. doi:10.1123/ijspp.2015-0330

28.   Walsh NP. Recommendations to maintain immune health in athletes. *Eur J Sport Sci*. 2018;epub. doi:10.1080/17461391.2018.1449895

29.   He CS, Bishop NC, Handzlik MK, Muhamad AS, Gleeson M. Sex differences in upper respiratory symptoms prevalence and oral-respiratory mucosal immunity in endurance athletes. *Exerc Immunol Rev*. 2014;20:8-22. doi:10.1007/s00438-014-0936-z

30.   Breiman L. Random Forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1017/CBO9781107415324.004

31.   Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *Can J Emerg Med*. 2006;8(1):19-20.

32.   Bermon S, Castell LM, Calder PC, et al. Consensus statement: Immunonutrition and exercise. *Exerc Immunol Rev*. 2017;23:8-50.

33.   Irwin MR. Why sleep is important for health: A

psychoneuroimmunology perspective. *Annu Rev Psychol*. 2015;66:143-172. doi:10.1146/annurev-psych-010213-115205.Why

34.   Drew MK, Vlahovich N, Hughes D, et al. A multifactorial evaluation of illness risk factors in athletes preparing for the Summer Olympic Games. *J Sci Med Sport*. 2017;20(8):745-750. doi:10.1016/j.jsams.2017.02.010
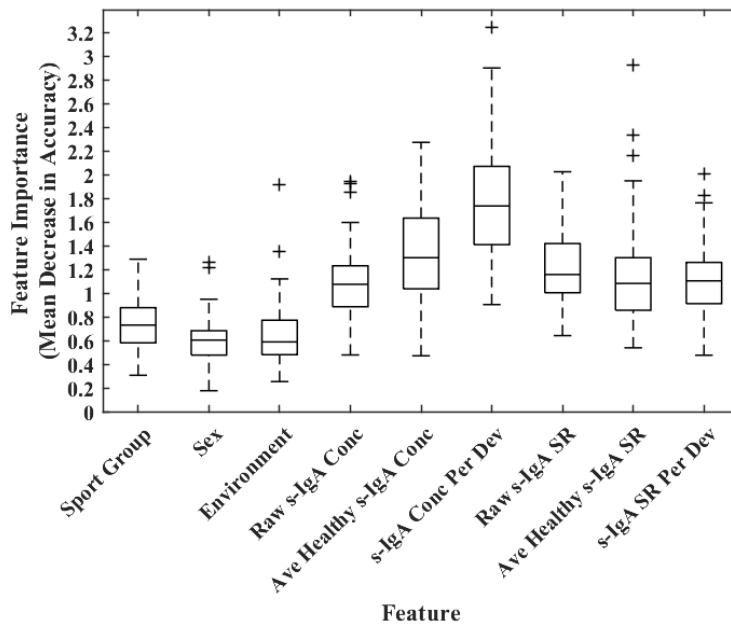
Figure 1: Box and whisker feature importance plots for URI$_{day}$. The box provides the median, lower and upper quartiles for the 1000 trials completed. The whisker extends to the minimum and maximum values, excluding outliers (denoted by crosses). NB: Conc = concentration; SR = secretion rate; s-IgA Conc Per Dev = percentage deviation from average healthy s-IgA concentration; s-IgA SR Per Dev = percentage deviation from average healthy s-IgA secretion rate.
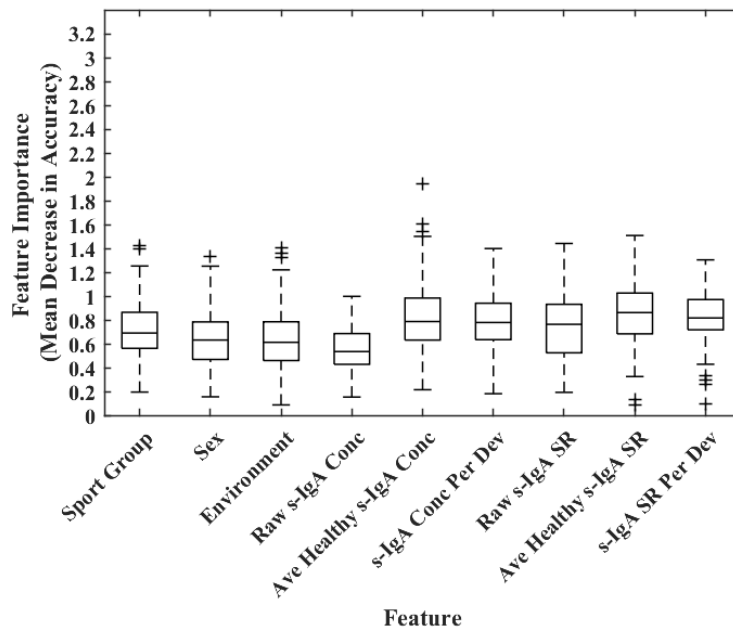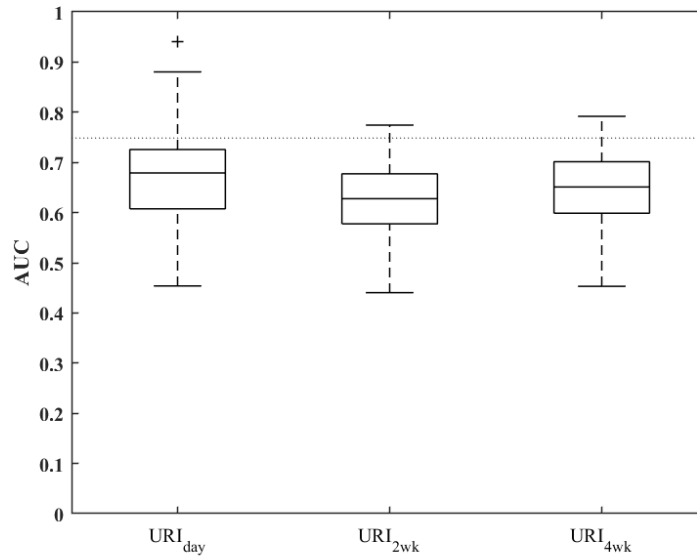
Figure 2: Box and whisker feature importance plots for URI$_{2wk}$. The box provides the median, lower and upper quartiles for the 1000 trials completed. The whisker extends to the minimum and maximum values, excluding outliers (denoted by crosses). NB: Conc = concentration; SR = secretion rate; s-IgA Conc Per Dev = percentage deviation from average healthy s-IgA concentration; s-IgA SR Per Dev = percentage deviation from average healthy s-IgA secretion rate.
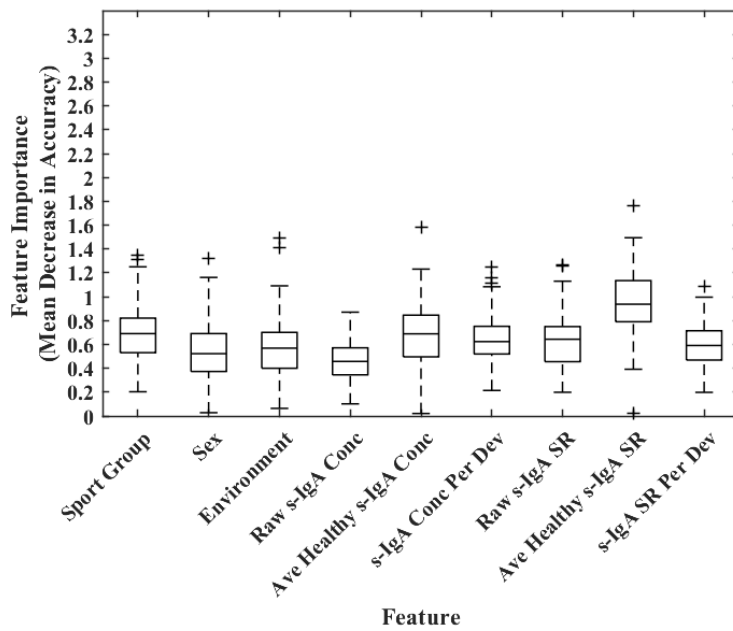
Figure 3: Box and whisker feature importance plots for URI4wk. The box provides the median, lower and upper quartiles for the 1000 trials completed. The whisker extends to the minimum and maximum values, excluding outliers (denoted by crosses). NB: Conc = concentration; SR = secretion rate; s-IgA Conc Per Dev = percentage deviation from average healthy s-IgA concentration; s-IgA SR Per Dev = percentage deviation from average healthy s-IgA secretion rate.

Figure 4: Box and whisker plots for the AUC of each model. The box provides the median, lower and upper quartiles for the 1000 trials completed. The whisker extends to the minimum and maximum values, excluding outliers (denoted by crosses). Dashed line represents 0.75 AU threshold for *good* predictor.