Drug and Alcohol Dependence Reports 10 (2024) 100215

Contents lists available at ScienceDirect

Drug and Alcohol Dependence Reports

journal homepage: www.elsevier.com/locate/dadr

Causal factors in childhood and adolescence leading to anabolic-androgenic steroid use: A machine learning approach

James I. Hudson [a,b,*], Yaakov Hudson [c], Gen Kanyama [a,b], Jiana Schnabel [a], Kristin N. Javaras [b,d], Marc J. Kaufman [b,e], Harrison G. Pope Jr. [a,b]

[a] *Biological Psychiatry Laboratory and Psychiatric Epidemiology Research Program, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA*
[b] *Department of Psychiatry, Harvard Medical School, Boston, MA USA*
[c] *Israel Center for Entrepreneurship, Jerusalem, Israel*
[d] *Division of Women's Mental Health, McLean Hospital, Belmont, MA, USA*
[e] *McLean Imaging Center, McLean Hospital, Belmont, MA, USA*

HIGHLIGHTS

- Anabolic-androgenic steroid (AAS) use is a large public health problem.
- Causal factors for AAS use were sought using an existing dataset.
- Machine learning and causal inference theory were used to identify causal factors.
- Six potential causal factors emerged, with body image concerns the most prominent.
- Machine learning combined with causal inference has a wide range of applications.

ARTICLE INFO

Keywords:
Anabolic-androgenic steroids
Body-image disorders
Causal inference
Eating disorders
Machine learning
Risk factors

ABSTRACT

Background: Prior research has demonstrated associations between anabolic-androgenic steroid (AAS) use and features from several childhood and adolescent psychosocial domains including body image concerns, antisocial traits, and low levels of parental care. However, prior approaches have been limited by their focus on individual features and lack of consideration of the relevant causal structure.

Methods: We re-analyzed data from a previous cross-sectional cohort study of 232 male weightlifters aged 18–40, of whom 101 had used AAS. These men completed retrospective measures of features from their childhood and early adolescence, including body image concerns, eating disorder psychopathology, antisocial traits, substance use, and family relationships. Using an approach informed by principles of causal inference, we applied four machine-learning methods – lasso regression, elastic net regression, random forests, and gradient boosting – to predict AAS use.

Results: The four methods yielded similar receiver operating curves, mean area under the curve (range 0.66 to 0.72), and sets of highly important features. Features related to adolescent body image concerns (especially muscle dysmorphia symptoms) were the strongest predictors. Other important features were adolescent rebellious behaviors; adolescent feelings of ineffectiveness and lack of interoceptive awareness; and low levels of paternal care.

Conclusions: Applying machine learning within a causally informed approach to re-analyze data from a prior study of weightlifters, we identified six factors (most prominently those related to adolescent body image concerns) as proposed causal factors for the development of AAS use. Compared with the prior analyses, this approach achieved greater methodologic rigor and yielded stronger and broader findings.

* Corresponding author at: Biological Psychiatry Laboratory and Psychiatric Epidemiology Research Program, McLean Hospital, 115 Mill Street, Belmont, MA 02478, USA.

E-mail address: jhudson@mclean.harvard.edu (J.I. Hudson).

https://doi.org/10.1016/j.dadr.2023.100215
Received 21 August 2023; Received in revised form 30 November 2023; Accepted 26 December 2023
Available online 29 December 2023

1. Introduction

Anabolic-androgenic steroid (AAS) use, once restricted primarily to competitive athletes, has now become a major worldwide substance use disorder affecting tens of millions of individuals (Pope et al., 2014). AAS use may cause adverse cardiac (Baggish et al., 2017), neuroendocrine (Rasmussen et al., 2016), and other medical (Pope et al., 2014) and neuropsychiatric effects (Bjornebekk et al., 2017; Pope et al., 2021), and thus represents a growing public-health problem. Consequently, studies from our laboratory (Kanayama et al., 2003, 2018; Pope et al., 2012) and others (Bahrke et al., 2000; Brower et al., 1994; Handelsman and Gupta, 1997; Kindlundh et al., 1999) have sought to assess risk factors for use of AAS. These studies have typically identified body image concerns and antisocial traits as major risk factors.

To sharpen our understanding of these factors, we reanalyzed the data from our 2012 study of risk factors for AAS use (Pope et al., 2012) utilizing two advances in data analysis: *machine learning* and modern *causal inference*. Machine learning offers two main advantages over traditional approaches in settings such as this one. First, because it uses hold-out sets for testing against models optimized on training sets, it is less vulnerable to over-fitting and thus has greater out-of-sample generalizability. Second, one can apply several different machine-learning models, each with different assumptions, to the same data. If the results of the models are convergent, confidence in the validity of the findings is increased, whereas differences across models would prompt further investigation.

Modern causal inference, based on "counterfactual outcomes," often uses directed acyclic graphs (DAGs) to efficiently articulate the underlying causal structure of interest and then uses graphical analysis to guide the testing and interpretation of the data. Thus, we move from simple prediction to causation. This concept is expressed by Hernán and associates (Hernán et al., 2019) as follows:

> "*Prediction* is using data to map some features of the world (the inputs) to other features of the world (outputs)… *Counterfactual prediction* is using data to predict certain features of the world as if the world had been different, which is required in *causal inference* applications. An example of causal inference is the estimation of the mortality rate that would have been observed if all individuals in a study population had received screening for colorectal cancer vs. if they had not received screening."

We have previously published an outline of our theoretical approach using machine learning coupled with the principles of causal inference to potentially help identify causal factors for mental disorders (Brennan and Hudson, 2022) (although this approach falls well short of the ideal of a fully specified multivariate causal model). We hypothesized that this same approach, applied to the data from our 2012 study, would provide deeper and more rigorous insights into the causes of AAS use.

2. Methods and materials

2.1. Data acquisition

Our 2012 study of risk factors for AAS use (Pope et al., 2012) used a cross-sectional cohort design (Hudson et al., 2005) to assess 233 male weightlifters, aged 18–40, recruited by advertisements in gymnasiums. Details of the study design, recruitment methods, and the resulting sample of participants are provided in prior publications (Kanayama et al., 2018; Pope et al., 2012). For the present analysis, we excluded one participant with incomplete data, leaving 232 men, of whom 101 had used AAS at some point in their lives. Both our original study and the present analysis were approved by the Mass General Brigham Institutional Review Board; all participants provided written informed consent before any study procedures were performed.

In the 2012 study, we administered a battery of psychometrically established instruments asking participants to retrospectively report various child and adolescent attributes that we hypothesized to be plausible causal factors for AAS use. We chose these instruments on the basis of experience from our 2003 pilot study (Kanayama et al., 2003), together with the available literature on childhood and adolescent attributes shown to be associated with the development of adult substance use (Elkins et al., 2007; Fergusson et al., 2008; Hayatbakhsh et al., 2008; Tarter et al., 2004). These instruments included the Body Dysmorphic Disorder Modification of the Yale-Brown Obsessive Compulsive Scale (BDD-YBOCS) (Phillips et al., 1997), modified to assess participants' level of concern about their muscularity in early adolescence (age 13–16). Participants also completed a modified version of the Eating Disorders Inventory-2 (Garner, 1991) ("MEDI"), which also tapped adolescent body image concerns, together with other related features strongly associated with eating disorders. The MEDI was modified from the original Eating Disorders Inventory so that the questions were asked retrospectively about adolescent features rather than current (adult) features, and the items involving body dissatisfaction were rephrased to focus on muscularity rather than obesity, as described in previous studies from our center and elsewhere (Goldfield and Woodside, 2009; Kanayama et al., 2003; Blouin and Goldfield, 1995). For example, the item "I think my thighs are too big" was rephrased to read "I felt that my legs were too small." We also administered two other instruments focused on features from age 13 through 16: the Impulsive Sensation-Seeking Scale (McDaniel and Mahan, 2008) and the Adolescent Risk-Taking Questionnaire (ARQ) (Gullone et al., 2000), and two additional instruments focused on childhood experiences: the Wender Utah Rating Scale for the retrospective diagnosis of childhood attention deficit hyperactivity disorder (Ward et al., 1993) and the Parental Bonding Instrument (PBI) (Parker et al., 1979).

2.2. Conceptual framework and causal models

The scores on the individual scales and subscales above, which in the present paper are termed "*features*," represent psychometrically validated measures (e.g., derived from factor analysis) of 19 underlying latent factors that we hypothesized were plausible causal factors for AAS use. Although these hypothesized causal factors are latent and not directly measurable, the corresponding features represent measurable estimates of the levels of these factors. We also measured potential confounding variables.

Our causal modeling is best presented using directed acyclic graphs (DAGs), which provide a rigorous, yet intuitive, framework for examining causality (Diemer et al., 2021; Greenland et al., 1999; Hernán and Robins, 2022; Pearl, 1995). We briefly present this causal modeling approach here. Additionally, in the supplementary materials associated with this paper (Section S1), we provide an expanded and detailed technical explanation of this methodology.

The causal model for each feature, prior to the application of machine learning, is depicted in DAG 1 (Fig. 1). Using machine learning, the association between this feature and AAS use is then assessed while conditioning on all other measured variables (indicated by enclosing these variables in boxes), as depicted in DAG 2 (Fig. 1). If the assumptions of DAG 2 hold, then any association observed between the feature and AAS use represents a direct causal effect of that feature. Of course, the interpretation of the association observed between a given feature and AAS use as a direct causal effect is valid only if the DAG is correctly specified. Therefore, when evaluating the findings, one must carefully consider the potential for bias due to other sources of association that are not accounted for. In Section 4.2 below (and in an expanded fashion in the supplementary materials (Section S1)), we use DAGs to assist in the evaluation of potential unmeasured, or so-called "residual" confounding.

2.3. Machine learning analysis

For machine learning analysis, the dependent (outcome) variable

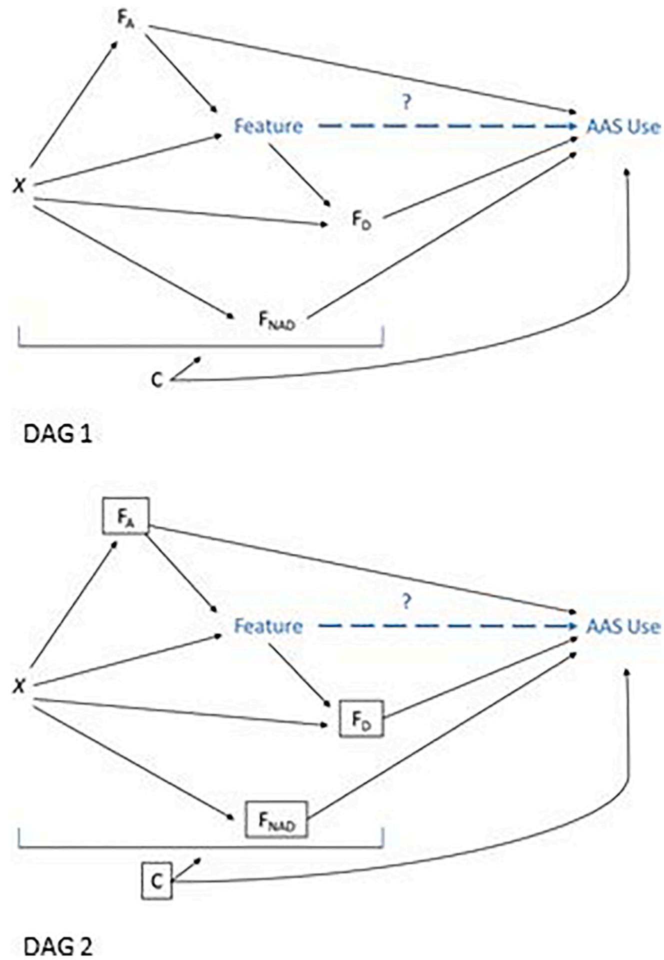

Fig. 1. DAG (directed acyclic graph) 1 is the causal model for the relationship of a given feature (Feature) to the outcome of developing AAS use (AAS Use) and to other measured and unmeasured variables. F_A is a set of measured features that are *ancestors* of Feature, and F_D is a set of measured features that are *descendants* of Feature; that is, they are, respectively, causes and effects (possibly mediated in part by other intermediate variables) of Feature (see technical definition in the glossary in the supplementary materials, Table S1). F_{NAD} is a set of measured features that are non-ancestors and non-descendants of Feature; *X* is a set of unmeasured common causes for all pairs within the set of all features in the model (i.e., {Feature, F_A, F_D, F_{NAD}}); C is a set of measured confounders that cause AAS Use and also cause one or more of the set of measured features and *X* (i.e., {Feature, F_A, F_D, F_{NAD}, *X*}). The dotted line from Feature to AAS Use represents the primary hypothesis of the study; namely, that the given feature has a direct causal effect on development of AAS use.
DAG 2 represents the model used in this study to evaluate the association between a given feature and AAS use, which conditions on (controls for) F_A, F_D, F_{NAD}, and C, as designated by a box around these variables. If the assumptions of DAG 2 hold, then any association observed between a given feature and AAS use is attributable to an independent (that is, not mediated through other features in the model) causal effect of the feature on AAS use. Further details of the assumptions of the DAG and their justification, particularly that the observed associations between any features and AAS use are causal and not due to forms of confounding, are presented in the supplementary materials, Section S1.

was AAS use. The predictor variables were of two types: (1) *demographic/design* variables (age, birth cohort, race, and the study evaluation site (California, Florida, or Massachusetts), which were either design variables or ones that likely influenced selection into the study, and thus had to be conditioned upon (and considered conceptually as potential confounders); and (2) the 19 *feature* variables, introduced above, that represented candidate causal factors for AAS use. Summary statistics for both types of variables in the AAS-user and non-user groups are presented in Table 1, and further detailed in the original paper (Pope et al., 2012). Our rationale for variable selection is also further detailed elsewhere (Brennan and Hudson, 2022; Diemer et al., 2021).

We entered the predictor variables into models representing four methods of supervised machine learning (Hastie et al., 2017; James et al., 2021) *lasso regression* (Tibshirani, 1986), *elastic net regression* (Zou and Hastie, 2005), *random forests* (Breiman, 2001), and *gradient boosting* (Friedman, 2001).

We chose lasso and elastic net as representative of so-called regression "shrinkage methods," which are extensions of parametric generalized multivariate linear regression models that add shrinkage or penalty parameters to reduce the magnitude of the estimated coefficients in a manner that chooses the model with the lowest out-of-sample prediction error. Elastic net differs from lasso only in that it adds a second shrinkage parameter. These shrinkage methods can be easily compared with traditional logistic regression, and the estimated coefficients associated with each feature have a straightforward familiar interpretation as regression coefficients that are also a measure of standardized effect size.

We chose random forests and gradient boosting as representative of non-parametric ensemble methods that are not only well-established, but also easily implemented (e.g., have relatively few hyperparameters to specify) so that they are transparent and readily replicated by other investigators. Unlike regression shrinkage methods, these methods are based on decision trees, and do not assume a generalized multivariate distribution. Thus, they perform well even if these distributional assumptions are violated. Random forests averages predictions over multiple decision trees of differing depths and sets of input variables. It has become the most popular and well-studied alternative to the parametric shrinkage models such as lasso and elastic net (Couronné et al., 2018). Gradient boosting, which is representative of so-called "boosting" methods, starts with an initial weak decision tree and makes many small iterative changes that each slightly reduce prediction error. All models were fit in Python using *Scikit-learn* (Pedregosa et al.,

Table 1
Characteristics of anabolic-androgenic steroid users and non-users.

Variables	AAS Users $N = 101$	Non-users $N = 131$
Demographic/Design Variables		
Age, median [IQR]	30 [25, 35]	27 [23, 32]
White race, *N* (%)	88 (86.3)	89 (67.9)
State		
Florida	42	53
Massachusetts	38	32
California	21	46
Features		
Muscle dysmorphia symptoms, median [IQR]	4 [1, 8]	2 [0, 3]
Wendler Utah Rating Scale, median [IQR]	33 [18, 47]	28 [14, 44]
Parental Bonding Instrument		
Paternal care, median [IQR]	19 [13, 24]	23 [18, 29]
Paternal overprotection, median [IQR]	13 [8, 18]	12 [5, 16]
Maternal care, median [IQR]	26 [20, 32]	29 [23, 34]
Maternal overprotection, median [IQR]	15 [10, 20]	14 [9, 21]
Impulsive Sensation Seeking Scale, median [IQR]	13 [8, 16]	12 [8, 15]
Adolescent Risk-Taking Questionnaire		
Antisocial behaviors, median [IQR]	6 [4, 9]	6 [4, 7]
Rebellious behaviors, median [IQR]	10, 5, 15]	7 [3, 12]
Reckless behaviors, median [IQR]	7 [3, 10]	4 [2, 7]
Thrill-seeking behaviors, median [IQR]	9 [6, 11]	9 [7, 12]
Modified Eating Disorders Inventory		
Body dissatisfaction, median [IQR]	7 [3, 11]	4 [1, 8]
Bulimia, median [IQR]	0 [0, 2]	0 [0, 1]
Drive for muscularity, median [IQR]	1 [0, 4]	0 [0, 2]
Ineffectiveness, median [IQR]	4 [1, 9]	1 [0, 5]
Interoceptive awareness, median [IQR]	2 [0, 6]	1 [0, 3]
Interopersonal distrust, median [IQR]	5 [2, 8]	3 [1, 6]
Maturity Fears, median [IQR]	3 [1, 5]	3 [1, 5]
Perfectionism, median [IQR]	5 [2, 8]	7 [4, 10]

AAS, anabolic-androgenic steroid.

2011).

For each method, we first divided the sample randomly into quintiles using a given seed value. We then performed five-fold cross-validation based on this division, holding out one quintile (the "test set") and optimizing the model on the remaining four quintiles (the "training set"). Optimization was achieved by selecting the model with the best fit, as measured by the area under the curve (AUC), from among models that were fitted with different hyperparameter settings specific to each model (see supplementary materials, Section S2). We then fitted this optimized model to the test set and calculated its receiver operating characteristic (ROC) curve and AUC.

We repeated the five-fold validation procedure for two additional initial seed values and repeated the optimization and testing as described above for each of the two additional sets of quintiles. Thus, we obtained 15 ROC curves and values of AUC from the five tests on each of the three sets of quintiles. Our primary measure of fit was the ROC curve. Our primary measure of accuracy was the mean AUC of these 15 values. For comparison, we also calculated the mean AUC for the optimized model on the training set; that is, the model with the set hyperparameters that performed best on the training set.

Each of the 15 models quantified the relative importance of each feature. For lasso and elastic net, these measures represented the standardized effect sizes of the penalized regression coefficients; for random forests and gradient boosting, they are the mean decrease in impurity within each tree (https://scikit-learn.org/stable/modules/ensemble.html#l2014).

To compare the relative importance of different features, we ranked each feature (i.e., excluding demographic/design variables) within each method, based on the absolute value of the measure of feature importance (averaged over the 15 values), and then computed the mean rank across the four methods.

We performed two sensitivity analyses. First, to evaluate potential effects induced by the demographic/design variables, we fitted models *restricted to features of interest.* Second, to evaluate the potential effects of collinearity, we fitted models that *reduced collinearity*; specifically, for pairs of features that exhibited a correlation higher than a given threshold, we removed the feature that had the lower correlation with AAS use in favor of the feature with the higher correlation.

3. Results

3.1. Primary analyses

3.1.1. Machine learning implementation

For each of the four methods, the specific values of the 15 sets of hyperparameter values selected for testing against the test set, obtained through the process of hyperparameter tuning, are presented in the supplementary materials, Table S2. The ROC curves for each of the 15 testing iterations of each machine learning method, as well as the mean ROC curve, are presented in Fig. 2. The mean AUCs were similar for the

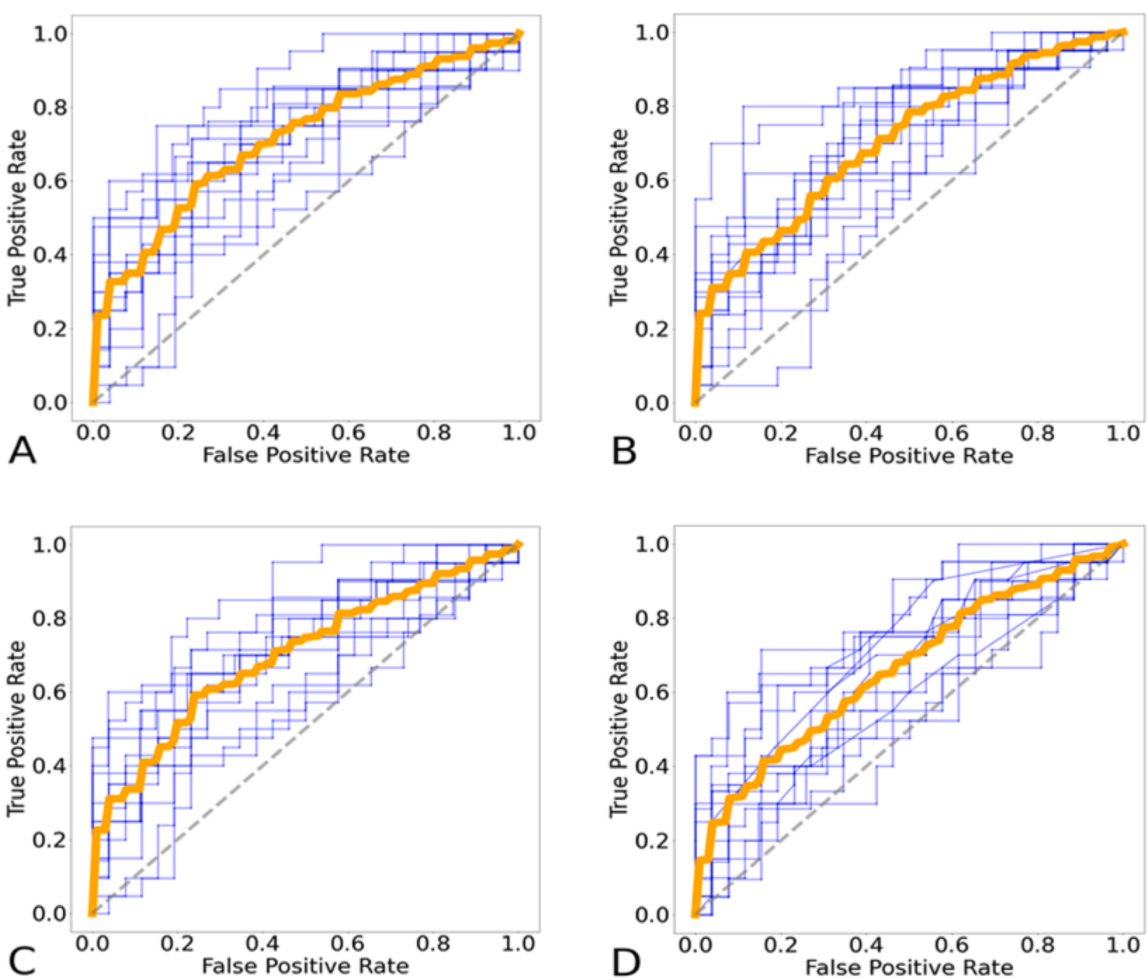

Fig. 2. Receiver operating characteristic (ROC) curves and area under the curve (AUC) for the four machine learning methods applied to the test set: (A) lasso; (B) elastic net; (C) random forests; (D) gradient boosting. Blue lines: ROC curves for each of the 15 model iterations; Orange lines: mean ROC curves; Dotted diagonal line: ROC value (AUC = 0.5), indicates chance expectation (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

four methods, ranging from 0.73 to 0.75 on the training set, and from 0.68 to 0.72 on the test set (Table 2).

3.1.2. Feature importance

The mean importance of features within each model is displayed in Fig. 3, and the mean rank of feature importance across all models is presented in Table 3. We assessed the association of a given feature with AAS use as: "strong" for six features: increased muscle dysmorphia symptoms, decreased MEDI-Perfectionism, decreased PBI-Paternal Care, increased ARQ-Rebellious Behaviors, increased MEDI-Ineffectiveness, and increased MEDI-Body Dissatisfaction (mean rank 1.0 to 5.8); "moderate" for increased MEDI-Interoceptive Awareness (mean rank 7.3); and "negligible" for the remaining features (mean rank $\geq$ 10.5).

3.2. Sensitivity analyses

In models *restricted to feature variables* (that is, excluding demographic/design variables (sensitivity analysis 1)), the AUC's and relative feature importance remained almost identical to those from the models including demographic/design variables (Table S2 and Fig. 4).

In models fitted to *reduce collinearity,* we found that the highest correlation coefficient for any pair of features across all models was 0.74, with only three additional pairs showing coefficients greater than 0.60 (Fig. 5) – suggesting that collinearity among features was modest and unlikely to be problematic.

Our first model, using a correlation coefficient cut-off of 0.70 (sensitivity analysis 2), deleted ARQ-Reckless Behaviors. The second model, using a cutoff of 0.60 (sensitivity analysis 3), additionally deleted MEDI-Interoceptive Awareness, MEDI-Drive for Muscularity, and ARQ-Antisocial Behaviors. The relative feature importance changed little in either model (Table 3 and Fig. 4). The only change of consequence affecting any of the "strong" or "moderate" associations identified above was that in the second model, MEDI-Interoceptive Awareness was dropped in favor of MEDI-Ineffectiveness.

4. Discussion

Using a dataset from a prior study of 232 male weightlifters, we applied four machine learning methods to perform multivariate analyses of retrospectively rated features in childhood and early adolescence that represented candidate causal factors for subsequent AAS use. Our original univariate analyses of these data in 2012 had identified symptoms of muscle dysmorphia and conduct disorder as the two primary risk factors for AAS use. The current analysis, utilizing machine learning and causal inference, yielded generally similar, but more rigorous and fine-grained results. Specifically, we identified seven features, each showing a strong or moderate independent association with AAS use across all four machine-learning methods.

4.1. Identified features

The first and most influential feature was Muscle Dysmorphia Symptoms (ranked first in predictor importance in all four machine-learning methods). Also among the seven features identified was MEDI-Body Dissatisfaction – another body-image factor associated with AAS use even after conditioning on Muscle Dysmorphia Symptoms. These findings are consistent with our prior reports based on the same sample (Kanayama et al., 2018; Pope et al., 2012), but the present analyses bring adolescent body image into even sharper relief amidst the other predictors. These findings are also consistent with an earlier preliminary study from our group (Kanayama et al., 2006), as well as reports from many other centers (Brower et al., 1991; Buckley et al., 1988; Cafri et al., 2005; Choi et al., 2002; Cole et al., 2003; Dodge et al., 2008; Goldfield and Woodside, 2009; Hildebrandt et al., 2010; Murray et al., 2016; Olivardia et al., 2000; Rachon et al., 2006; Rohman, 2009; Blouin and Goldfield, 1995). Indeed, recent surveys suggest that personal appearance now greatly exceeds athletic aspirations as a motive for using AAS (Kanayama and Pope, 2012; Parkinson and Evans, 2006; Ip et al., 2011).

The third and fourth features were *ineffectiveness*, and *lack of interoceptive awareness*, also derived from subscales of the MEDI. The Ineffectiveness subscale includes items such as "I felt inadequate," "I felt insecure about myself" and "I had a low opinion of myself;" and the Interoceptive Awareness subscale refers to difficulty in identifying internal physiological or emotional states, as illustrated by items such as "I worried that my feelings would get out of control," or "when I was upset, I didn't know if I was sad, frightened, or angry." Note that, counterintuitively for interpretation, *high* scores on MEDI-Interoceptive Awareness indicate *lack* of interoceptive awareness (Garner, 1991; Polivy and Herman, 2002). Although these two subscales differ conceptually, they showed a moderately high correlation in our analyses ($r = 0.68$).

Notably, in an earlier preliminary study of 49 AAS users and 41 nonusers (Kanayama et al., 2006), our group found that EDI-Ineffectiveness and EDI-Interoceptive Awareness were not significantly associated with AAS use, but we had less power than the present study to detect differences. Also, we asked only about current, rather than retrospectively rated adolescent symptomatology. By contrast, Blouin and Goldfield (Blouin and Goldfield, 1995) found higher levels of EDI-Ineffectiveness and EDI-Interoceptive Awareness in 43 bodybuilders (most of whom had used AAS) as compared to 48 runners and 48 martial artists reporting little use of AAS.

The fifth feature, again from the MEDI, involved *low levels of perfectionism,* as reflected on the MEDI-Perfectionism subscale. Although attaining the second highest mean rank of the features, this finding should be regarded with caution for two reasons. First, this association was not hypothesized *a priori*; indeed, our general hypothesis was that all subscales of the MEDI, including Perfectionism, would be elevated among AAS users. Second, to our knowledge, there is no support in the literature for the hypothesis that low levels of perfectionism are associated with AAS use. Our earlier preliminary study (Kanayama et al.,

Table 2
Mean area under the curve for the main analysis and sensitivity analyses using four machine learning methods.

Mean Area Under the Curve (AUC)	Main Analysis		Sensitivity Analysis 1[a] (no design/ demographic variables)		Sensitivity Analysis 2[b] (correlation cut-off of $\geq$0.7)		Sensitivity Analysis 3[b] (correlation cut-off of $\geq$0.6)	
Method	Test Set[c]	Optimized Training Set[d]	Test Set[c]	Optimized Training Set[d]	Test Set[c]	Optimized Training Set[d]	Test Set[c]	Optimized Training Set[d]
Lasso	0.72	0.76	0.72	0.75	0.71	0.75	0.72	0.75
Elastic net	0.70	0.75	0.72	0.75	0.71	0.75	0.72	0.75
Random forests	0.71	0.73	0.70	0.72	0.72	0.74	0.70	0.73
Gradient boosting	0.66	0.74	0.66	0.74	0.66	0.75	0.68	0.75

[a] Model without design/demographic variables that represent potential confounders (see text).
[b] Model that deletes one variable from each pair of variables above a specified cut-off of correlation coefficient, $\geq$0.7 or $\geq$0.6 (see text).
[c] Primary measure, which is model optimized over hyperparameters using training set, and then applied to hold-out test set.
[d] Model optimized over hyperparameters using training set, and applied to the training set (prior to application to hold-out test set).

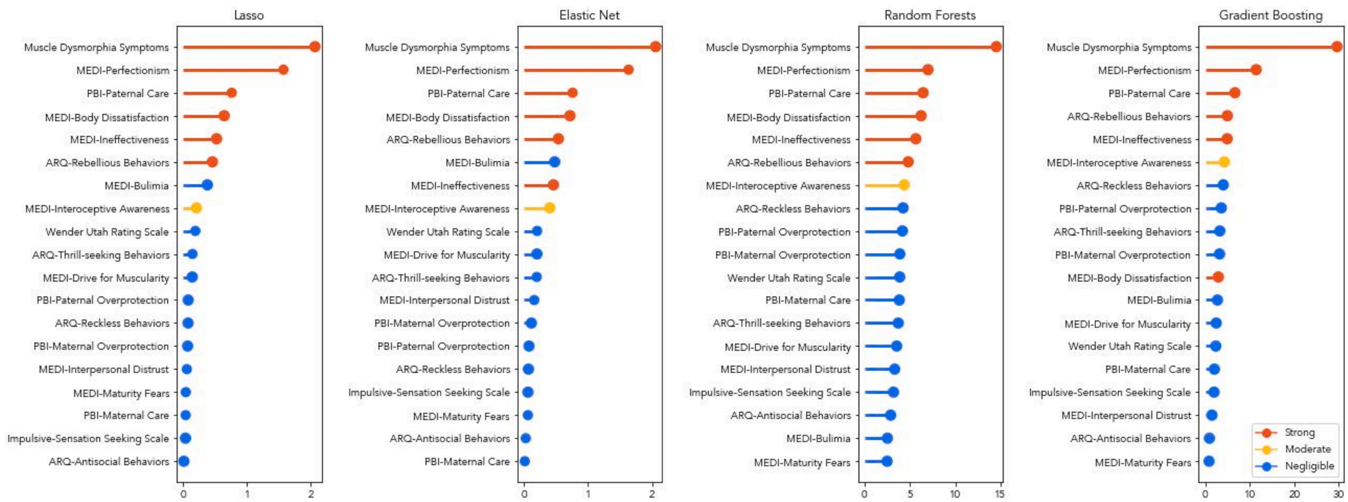

Fig. 3. Relative predictor importance for each of the four machine learning methods. Predictor color coding is based on the mean importance ranks across all methods (see text). Note that lasso and elastic net use the same scale, whereas the scale for random forests and gradient boosting is different. ARTQ – Adolescent Risk-Taking Questionnaire; MEDI – Modified Eating Disorders Inventory; PBI – Parental Bonding Instrument. (For interpretation of the color-coded strength of association (strong, moderate, and negligible) in this figure legend, the reader is referred to the web version of this article.)

2006) found that AAS users and non-users exhibited similar current self-ratings on this same measure. Also, Scarth et al. (2022) reported that AAS users with moderate to high symptoms of AAS dependence displayed *higher* levels of "rigid perfectionism" than AAS users with low symptoms of dependence. Subsequent studies will be needed to resolve these disparate findings.

The sixth feature, *adolescent rebellious behaviors*, was reflected by high scores on the ARQ-Rebellious Behaviors subscale (which comprised smoking, underage drinking, staying out late, getting drunk, and taking drugs). Note, however, that the ARQ-Rebellious Behaviors subscale displayed a moderately high correlation with two other ARQ subscales – ARQ-Reckless Behaviors ($r = 0.74$) and ARQ-Antisocial Behaviors ($r = 0.61$). Thus, it appears that rebellious behaviors cannot be teased apart from a broader set of behaviors that also includes reckless and antisocial behaviors. This interpretation is supported by an extensive literature reporting an association between AAS use and other substance use, criminality, and violence (Bahrke et al., 2000; Beaver et al., 2008; Buckley et al., 1988; Dodge and Hoagland, 2011; DuRant et al., 1993; Hallgren et al., 2015; Handelsman and Gupta, 1997; Hauger et al., 2021; Kindlundh et al., 1999; Klotz et al., 2006; Pope et al., 2021; Skarberg et al., 2009, 2010; Thiblin and Parlklo, 2002) – although most of these studies performed current assessments of these traits in individuals who had already initiated AAS, rather than retrospective assessments of these traits prior to AAS onset.

The final feature was having an *uncaring father*, as reflected by low levels of PBI-Paternal Care. This finding is congruent with studies such as that of Enns and colleagues (2002) showing an association between low paternal care on the PBI and an elevated lifetime risk of many psychiatric disorders in males, including particularly substance use disorders and antisocial personality disorder. As noted by these authors, psychodynamic and developmental theories have also emphasized the role of a distant or disengaged father in the development of externalizing disorders. Consequently, one might speculate that deficient paternal care could cause a boy to develop deficient self-esteem surrounding issues of masculinity, including masculine body image, thereby leading to AAS use to "treat" this deficit.

4.2. Identified features: independent causal features vs. non-causal features whose association with AAS use is attributable to bias?

As discussed above in Section 2.2, if DAG 2 is correct, then the seven features we have identified as independently associated with AAS use would represent features that each have a direct causal effect on AAS use; that is, a causal effect not mediated by other features in the model. However, it is possible that the DAG is mis-specified and hence has not accounted for all of the sources of the association between a given feature and AAS use. In such a case, the observed association between the feature and AAS use is a biased estimate of the direct causal effect. Bias could even reach a level whereby a feature would have no direct causal effect on AAS use at all (i.e., it would be a non-causal feature), but nevertheless would exhibit a non-negligible association with AAS use in machine learning testing.

As with any study (and with observational studies especially), it is important to evaluate the potential for all sources of bias, such as selection bias, information bias, and confounding. In our study, the threat posed by selection bias and information bias (including "recall bias") can be readily analyzed rigorously without the assistance of DAGs. Thus, we have not used DAGs to illustrate these forms of bias, but rather consider them below in the limitations presented in Section 4.3.

The analysis of potential confounding, however, is more complicated, and DAGs are particularly well-suited for this type of analysis. We present a brief discussion of confounding here, and additionally provide a more detailed technical analysis of this issue using DAGs in the supplementary materials, Section S1.3.

To assess the likelihood of bias due to confounding in this study, we must judge the plausibility of what we have categorized, for purposes of exposition, as two types of confounding: "external confounding" and "internal confounding." External confounding is the familiar type of confounding that arises from one or more unmeasured (or technically, uncontrolled for) variables outside or *external* to the set of features in the model. An external confounding variable is *both* (a) a cause for the given feature and (b) a cause for AAS use (see supplementary materials, Section S1.3 and DAG 3 in Fig. S3). Because this type of confounding is relatively straightforward to analyze, we consider it further in the limitations presented in Section 4.3 below.

Internal confounding arises when we (a) include in the model a feature that is not a cause of AAS use, *and also* (b) fail to include a feature that is a cause of that feature and a direct cause AAS use (see supplemental materials, Section S1.3 and DAG 4 in Fig. S4). For example, suppose that a model of causal factors for AAS use (a) included "weightlifting" as a feature, and additionally (b) failed to include "body-image concerns" (which, for purposes of this example, we are assuming to be a cause of weightlifting). In this situation, "weightlifting" would emerge as a feature independently associated with AAS use, and hence

Table 3
Mean importance rank across the four machine learning methods for 19 features.

Predictor[a]	Mean importance Rank	Strength of association with AAS Use	Mean effect size (Lasso and Elastic Net)[b]
↑ Muscle Dysmorphia Symptoms	1.0	Strong	2.8
↓ MEDI-Perfectionism	2.0	Strong	1.6
↓ PBI-Paternal Care	3.0	Strong	0.8
↑ ARQ-Rebellious Behaviors	5.3	Strong	0.5
↑ MEDI-Ineffectiveness	5.5	Strong	0.5
↑ MEDI-Body Dissatisfaction	5.8	Strong	0.7
↑ MEDI-Interoceptive Awareness[c]	7.3	Moderate	0.3
↑ ARQ-Reckless Behaviors	10.5	Negligible	0.1
↓ ARQ-Thrill-seeking Behaviors	10.8	Negligible	0.2
↑ MEDI-Bulimia	10.8	Negligible	0.4[d]
↓ Wendler Utah Rating Scale	11.0	Negligible	0.2
PBI-Maternal Overprotection	11.8	Negligible	<0.1
PBI-Paternal Overprotection	11.8	Negligible	<0.1
↑ MEDI-Drive for Muscularity	12.0	Negligible	0.2
MEDI-Interpersonal Distrust	14.8	Negligible	0.1
PBI-Maternal Care	15.8	Negligible	<0.1
Impulsive-Sensation Seeking Scale	16.5	Neglible	<0.1
MEDI-Maturity Fears	17.8	Neglible	<0.1
ARQ-Antisocial Behaviors	18.0	Neglible	<0.1

AAS, anabolic-androgenic steroid; ARQ, Adolescent Risk-Taking Questionnaire, MEDI, modified Eating Disorders Inventory; PBI, parental bonding instrument.
[a] Direction of arrows indicates positive (↑) or negative association (↓) with AAS use on lasso and elastic net; no arrows are given for variables with mean effect size of <0.1.
[b] Mean of standardized regression coefficients from lasso and elastic net models only; random forests and gradient boosting models are nonparametric and therefore do not yield this measure.
[c] Increased values of MEDI-Interoceptive Awareness subscale indicates low levels of interoceptive awareness.
[d] Higher mean effect size than other variables with negligible association due to lower rankings from random forests and gradient boosting models.

could be misinterpreted as a causal feature, when in fact it was merely a non-causal marker. However, our domain knowledge would likely prevent such a mistake. Specifically, we would be unlikely to (a) commit an error of commission in considering “weightlifting” as a candidate cause for AAS (because we would recognize that its association with AAS use was almost certainly due to a common ancestor with AAS use); while (b) simultaneously committing an error of omission by failing to include a measure of “body image concerns” (see expanded treatment in supplemental materials, Section S1.3 and DAGs 5 and 6 in Figs. S5 and S6).

Applying these principles to our study, we could not envisage any obvious conjunction of an identified feature with an unmeasured causal factor for AAS use that could induce one or more non-causal associations sufficient to account for a substantial part of the association between the identified feature and AAS use. However, because no observational study can fully exclude residual confounding as a source of bias, our interpretations regarding the causal effects of identified features must remain tentative.

4.3. Strengths and limitations

A strength of the present study is that it achieves greater methodologic rigor than prior analyses of potential causal factors for AAS use, including our previous univariate analyses of this same dataset, as discussed in Section 4 above. The consistent findings across four different machine learning methods, with different modeling assumptions and procedures, support the robustness of the findings, and suggest that they are not idiosyncratic to a single method. Also, choosing the model that is optimized over key hyperparameters on the training sets offers assurance that near-optimal models for each method were used. Finally, use of a hold-out test set yields a set of important features less vulnerable to overfitting, and thus more likely to generalize to other samples than with previous multivariate techniques. Moreover, by employing an explicit causal model based on domain knowledge and selecting features representing psychometrically established measures of plausible causal factors for AAS use, we reduced threats to the validity of our causal inferences relative to “agnostic” machine-learning classification methods (Brennan and Hudson, 2022; Hernán et al., 2019; Pearl, 2018).

We acknowledge several limitations. First, looking at aspects of study design, there is the potential for selection bias. Although we attempted to minimize this bias by recruiting AAS-using and non-using weightlifters from the same gymnasiums, unmeasured differences between AAS users and non-users may have nevertheless influenced enrollment into the study, as discussed previously (Kanayama et al., 2003; Pope et al., 2012). Second, as with any observational study, there is the potential for confounding. We included demographic/design variables in our models to reduce this threat. For the regression-based models lasso and elastic net, there is sound theoretical support for including these variables to control for confounding, but for random forests and gradient boosting, the ability of these variables to control for confounding is unclear. Reassuringly, however, we found almost no difference in results between models with and without inclusion of demographic/design variables, suggesting that the threat of confounding from these variables was minimal. Although the possibility of unmeasured confounding variables remains, we believe that this threat is also low, given the careful consideration of this issue in the implementation of the cross-sectional cohort design, as discussed in our original paper (Pope et al., 2012). Third, there is the threat of information bias, especially in the form of “recall bias,” in that participants were asked to recall experiences from childhood and adolescence, which for many participants was two or more decades earlier. Further, because measurements of the features were derived from retrospective self-reports of participants, some of whom had developed AAS use and others who had not, the presence of the AAS use or non-use may have differentially influenced the reporting of prior attributes. For example, an AAS user might recall adolescent body dissatisfaction more vividly than a non-user, even if their true levels of body dissatisfaction had been similar. Fourth, as mentioned in the methods section, because our model is not a fully specified causal model, it cannot be used to evaluate the total effect of a given feature (direct effect plus indirect effect mediated by other variables), nor the structure of causal relationships among the features. Unless one makes very strong and difficult to verify assumptions, a fully specified causal model would require some combination of more domain knowledge and longitudinal measurements of features over time. Fifth, we chose four machine learning methods, but other popular methods, such as support vector machines and neural networks (Hastie et al., 2017), could have been used and might have yielded different results. However, the observed consistency across the four models suggests that other methods would yield similar results. Sixth, the sample size and number of features were small relative to many applications of machine learning. A larger dataset might yield more differences between methods in fit. Seventh, high correlations among features (collinearity) renders models unstable and attenuates estimates of the associations between the outcome and these correlated features. However, our features displayed little collinearity, and when we eliminated one feature from each pair of features showing correlation coefficients greater than 0.60, we found little effect on the results. Eighth, although we applied models developed on

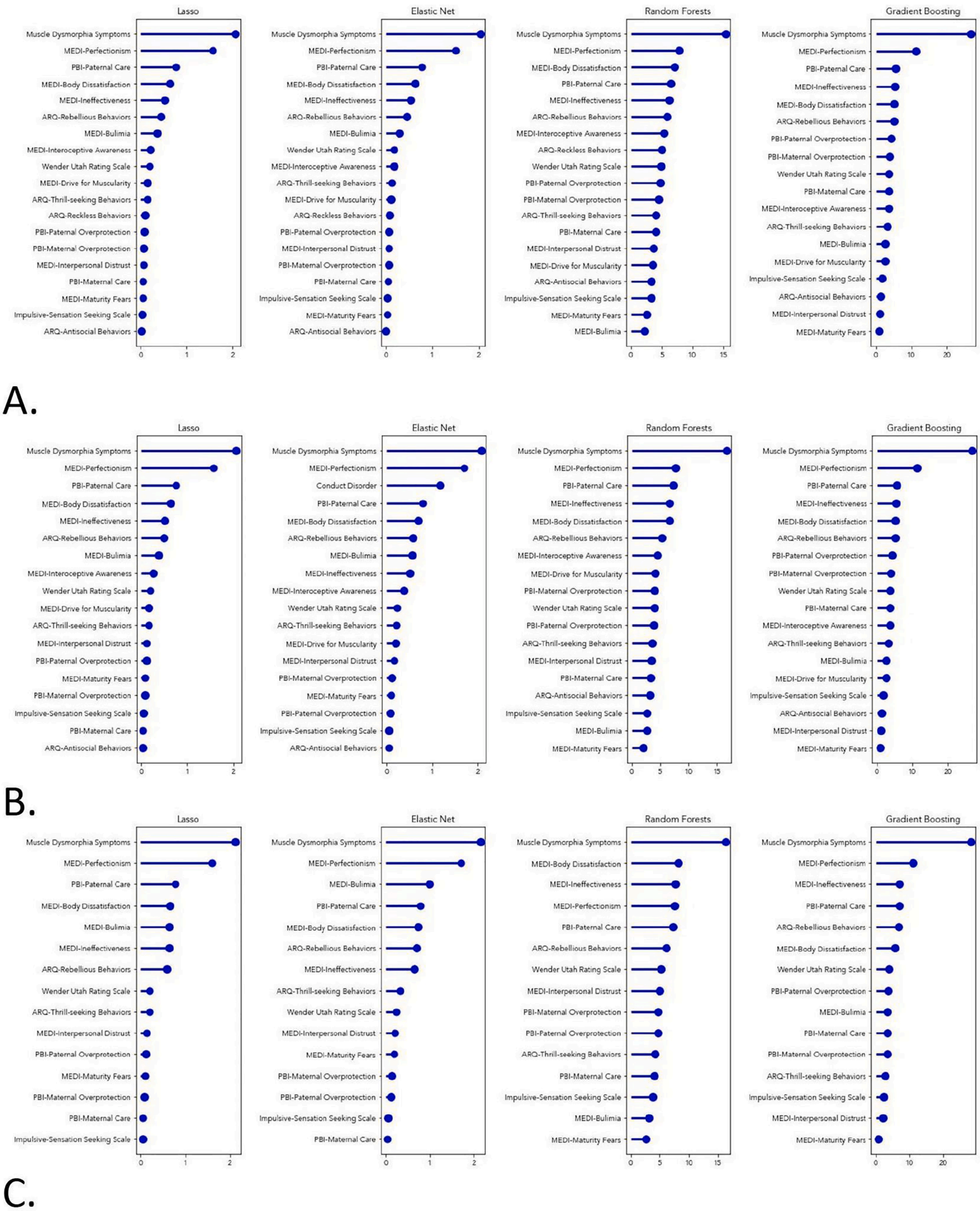

Fig. 4. Relative predictor importance for each of the four machine learning methods for sensitivity analyses. Note that lasso and elastic net use the same scale, whereas the scale for random forests and gradient boosting is different. Panel A. Sensitivity analysis 1: model restricted to feature variables; that is, excluded demographic/design variables. Panel B. Sensitivity analysis 2: correlation coefficient cut-off of 0.70, which deleted ARQ-Reckless Behaviors. Panel C. Sensitivity analysis 3: correlation coefficient cut-off of 0.60, which deleted ARQ-Reckless Behaviors, MEDI-Interoceptive Awareness, MEDI-Drive for Muscularity, and ARQ-Antisocial Behaviors. ARTQ – Adolescent Risk-Taking Questionnaire; MEDI – Modified Eating Disorders Inventory; PBI – Parental Bonding Instrument.

training sets to hold-out sets within the sample, they were not tested against separate outside samples from other investigations, and thus it is uncertain how well the results would generalize to other samples.

4.4. Conclusions

Applying four machine learning methods to data from a study of male weightlifters, we identified six factors in childhood or adolescence that represented potential causal factors for the development of AAS use: symptoms of muscle dysmorphia; body dissatisfaction; ineffectiveness; lack of interoceptive awareness; rebellious behavior; and poor paternal care. These finding potentially offer a more fine-grained profile of the type of individual at greatest risk for AAS use, and suggest interventions targeted at reducing that risk. Future studies will be necessary to explore

	Muscle dysmorphia symptoms	MEDI-Body dissatisfaction	MEDI-Bulimia	MEDI-Drive for muscularity	MEDI-Ineffectiveness	MEDI-Interpersonal	MEDI-Interoceptive awareness	MEDI-Maturity fears	MEDI-Perfectionsism	ARQ-Antisocial behaviors	ARQ-Rebellious behaviors	ARQ-Reckless behaviors	ARQ-Thrill-seeking behaviors	Wendler-Utah Rating Scale	Impulsive Sensation Seeking Scale	PBI-Paternal care	PBI-Paternal overprotection	PBI-Maternal care
MEDI-Body disatifaction	0.43																	
MEDI-Bulimia	0.11	0.31																
MEDI-Drive for muscularity	0.20	0.39	0.62															
MEDI-Ineffectivness	0.38	0.53	0.42	0.42														
MEDI-Interpersonal	0.36	0.31	0.11	0.07	0.52													
MEDI-Interoceptive	0.36	0.36	0.56	0.42	0.68	0.46												
MEDI-Maturity fears	0.13	0.24	0.19	0.09	0.33	0.21	0.25											
MEDI-Perfectectionism	0.03	0.02	0.09	0.17	0.07	-0.01	0.18	0.00										
ARQ-Antisocial behaviors	0.22	0.26	0.32	0.15	0.37	0.25	0.50	0.09	0.07									
ARQ-Rebellious behaviors	0.12	0.21	0.20	0.09	0.34	0.24	0.40	0.09	-0.07	0.61								
ARQ-Reckless behaviors	0.15	0.19	0.19	0.15	0.29	0.24	0.43	0.02	0.00	0.62	0.74							
ARQ-Thrill-seeking behaviors	-0.02	0.02	0.09	-0.03	0.03	-0.06	0.09	0.06	0.17	0.42	0.35	0.35						
Wendler Utal Rating Scale	0.23	0.32	0.23	0.20	0.58	0.46	0.48	0.16	0.00	0.56	0.44	0.44	0.21					
Impulsive Sensation Seeking Scale	0.12	0.13	0.12	0.00	0.20	0.32	0.27	-0.04	0.04	0.53	0.56	0.42	0.33	0.44				
PBI-Paternal care	-0.21	-0.18	-0.14	-0.17	-0.31	-0.30	-0.22	0.05	0.04	-0.06	-0.06	-0.13	0.16	-0.30	-0.03			
PBI-Paternal overprotection	0.12	0.19	0.17	0.17	0.31	0.36	0.29	0.15	0.15	0.29	0.14	0.26	0.11	0.33	0.20	-0.11		
PBI-Maternal care	-0.08	-0.11	-0.04	-0.07	-0.22	-0.20	-0.16	0.04	0.13	0.01	0.01	-0.06	0.13	-0.19	0.05	0.43	-0.04	
PBI-Maternal overprotection	0.12	0.18	0.09	0.16	0.33	0.28	0.27	0.04	0.16	0.14	0.06	0.11	0.11	0.27	0.11	-0.09	0.47	-0.19

Fig. 5. Heatmap for correlation between features. Values represent correlation coefficients between pairs of features. Darker shades of blue indicate greater positive correlation; darker shades of red indicate greater negative correlation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the relationships among the identified factors, quantify their effect on the development of AAS use, and develop strategies to change AAS use trajectories.

CRediT authorship contribution statement

James I. Hudson: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Yaakov Hudson:** Conceptualization, Formal analysis, Writing – review & editing. **Gen Kanyama:** Writing – review & editing. **Jiana Schnabel:** Writing – review & editing. **Kristin N. Javaras:** Writing – review & editing, Methodology. **Marc J. Kaufman:** Funding acquisition, Resources, Writing – review & editing. **Harrison G. Pope:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

James Hudson has received consulting fees from Idorsia and Otsuka; and has received research grant support from Boehringer-Ingelheim and Idorsia. Harrison Pope, Jr. has provided expert testimony approximately once per year on cases involving anabolic-androgenic steroids. Yaakov Hudson is an employee of Banco BS2 (São Paulo, Brazil). Kristin Javaras has owned equity shares in Sanofi and Centene Corporation, served on the Clinical Advisory Board for Beanbag Health, and received research funding from the National Institute of Diabetes and Digestive and Kidney Diseases. Gen Kanayama, Jiana Schnabel, and Marc Kaufman have no disclosures.

Role of funding source

The content is solely the responsibility of the authors. The funding sources had no role in the content of the paper nor in the preparation of the manuscript for publication.

Acknowledgments

Funding for this study was provided in part by National Institute on Drug Abuse of the National Institutes of Health grant R01DA041866. Kristin Javaras' work on this project was also supported by the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health (principal investigator: KNJ; Award Number K23-DK120517). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dadr.2023.100215.

References

Baggish, A.L., Weiner, R.B., Kanayama, G., Hudson, J.I., Lu, M.T., Hoffmann, U., Pope Jr., H.G., 2017. Cardiovascular toxicity of illicit anabolic-androgenic steroid use. Circulation 135, 1991–2002.

Bahrke, M.S., Yesalis, C.E., Kopstein, A.N., Stephens, J.A., 2000. Risk factors associated with anabolic-androgenic steroid use among adolescents. Sports Med. 29, 397–405.

Beaver, K.M., Vaughn, M.G., Delisi, M., Wright, J.P., 2008. Anabolic-androgenic steroid use and involvement in violent behavior in a nationally representative sample of young adult males in the United States. Am. J. Public Health 98, 2185–2187.

Bjornebekk, A., Walhovd, K.B., Jorstad, M.L., Due-Tonnessen, P., Hullstein, I.R., Fjell, A. M., 2017. Structural brain imaging of long-term anabolic-androgenic steroid users and nonusing weightlifters. Biol. Psychiatry 82, 294–302.
Blouin, A.G., Goldfield, G.S., 1995. Body image and steroid use in male bodybuilders. Int. J. Eat. Disord. 18, 159–165.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brennan, B.P., Hudson, J.I., 2022. Applications of machine learning to improve diagnosis, advance treatment, and identify causal factors for mental disorders. Biol. Psychiatry Cogn. Neurosci. Neuroimaging 7, 635–637.
Brower, K.J., Blow, F.C., Hill, E.M., 1994. Risk factors for anabolic-androgenic steroid use in men. J. Psychiatr. Res. 28, 369–380.
Brower, K.J., Blow, F.C., Young, J.P., Hill, E.M., 1991. Symptoms and correlates of anabolic-androgenic steroid dependence. Br. J. Addict. 86, 759–768.
Buckley, W.E., Yesalis 3rd, C.E., Friedl, K.E., Anderson, W.A., Streit, A.L., Wright, J.E., 1988. Estimated prevalence of anabolic steroid use among male high school seniors. JAMA 260, 3441–3445.
Cafri, G., Thompson, J.K., Ricciardelli, L., McCabe, M., Smolak, L., Yesalis, C., 2005. Pursuit of the muscular ideal: physical and psychological consequences and putative risk factors. Clin. Psychol. Rev. 25, 215–239.
Choi, P.Y., Pope Jr., H.G., Olivardia, R., 2002. Muscle dysmorphia: a new syndrome in weightlifters. Br. J. Sports Med. 36, 375–376 discussion 77.
Cole, J.C., Smith, R., Halford, J.C., Wagstaff, G.F., 2003. A preliminary investigation into the relationship between anabolic-androgenic steroid use and the symptoms of reverse anorexia in both current and ex-users. Psychopharmacology 166, 424–429 (Berl).
Couronné, R., Probst, P., Boulesteix, A.L., 2018. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 19, 270.
Diemer, E.W., Hudson, J.I., Javaras, K.N., 2021. More (adjustment) Is not always better: how directed acyclic graphs can help researchers decide which covariates to Include in models for the causal relationship between an exposure and an outcome in observational research. Psychother. Psychosom. 90, 289–298.
Dodge, T., Hoagland, M.F., 2011. The use of anabolic androgenic steroids and polypharmacy: a review of the literature. Drug Alcohol Depend. 114, 100–109.
Dodge, T., Litt, D., Seitchik, A., Bennett, S., 2008. Drive for muscularity and beliefs about legal performance enhancing substances as predictors of current use and willingness to use. J. Health Psychol. 13, 1173–1179.
DuRant, R.H., Rickert, V.I., Ashworth, C.S., Newman, C., Slavens, G., 1993. Use of multiple drugs among adolescents who use anabolic steroids. N. Engl. J. Med. 328, 922–926.
Elkins, I.J., McGue, M., Iacono, W.G., 2007. Prospective effects of attention-deficit/hyperactivity disorder, conduct disorder, and sex on adolescent substance use and abuse. Arch. Gen. Psychiatry 64, 1145–1152.
Enns, M.W., Cox, B.J., Clara, I., 2002. Parental bonding and adult psychopathology: results from the US National Comorbidity Survey. Psychol. Med. 32, 997–1008.
Fergusson, D.M., Boden, J.M., Horwood, L.J., 2008. The developmental antecedents of illicit drug use: evidence from a 25-year longitudinal study. Drug Alcohol Depend. 96, 165–177.
Friedman, J., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232.
Garner, D.M., 1991. Eating Disorders Inventory-2. Psychological Assessment Resources. Odessa, Florida.
Goldfield, G.S., Woodside, D.B., 2009. Body image, disordered eating, and anabolic steroids in male bodybuilders: current versus former users. Phys. Sportsmed. 37, 111–114.
Greenland, S., Pearl, J., Robins, J.M., 1999. Causal diagrams for epidemiologic research. Epidemiology 10, 37–48.
Gullone, E., Moore, S.M., Moss, S., Boyd, C., 2000. The adolescent risk-taking questionnaire: development and psychometric evaluation. J. Adolescent. Res. 15, 231–250.
Hallgren, M., Pope Jr., H.G., Kanayama, G., Hudson, J.I., Lundin, A., Kallmen, H., 2015. Anti-social behaviors associated with anabolic-androgenic steroid use among male adolescents. Eur. Addict. Res. 21, 321–326.
Handelsman, D.J., Gupta, L., 1997. Prevalence and risk factors for anabolic-androgenic steroid abuse in Australian high school students. Int. J. Androl. 20, 159–164.
Hastie, T., Tibshirani, R., Friedman, J., 2017. The Elements of Statistical Learning: Data Mining, Inference, and Production, Second Edition, Corrected, 12th Printing. Springer-Verlag. New York.
Hauger, L.E., Havnes, I.A., Jorstad, M.L., Bjornebekk, A., 2021. Anabolic androgenic steroids, antisocial personality traits, aggression and violence. Drug Alcohol Depend. 221, 108604.
Hayatbakhsh, M.R., Mamun, A.A., Najman, J.M., OCallaghan, M.J., Bor, W., Alati, R., 2008. Early childhood predictors of early substance use and substance use disorders: prospective study. Aust. N. Z. J. Psychiatry 42, 720–731.
Hernán, M.A., Hsu, J., Healy, B., 2019. A second chance to get causal inference right: a classification of data science tasks. Chance 32, 42–49.
Hernán, M.A., Robins, J.M., 2022. Causal Inference: What If. Chapman & Hall/CRC. Boca Ratom, FL.
Hildebrandt, T., Alfano, L., Langenbucher, J.W., 2010. Body image disturbance in 1000 male appearance and performance enhancing drug users. J. Psychiatr. Res. 44, 841–846.
Hudson, J.I., Pope Jr., H.G., Glynn, R.J., 2005. The cross-sectional cohort study: an underutilized design. Epidemiology 16, 355–359.
James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning: With Applications in R, 2nd ed. Springer. New York.
Ip, E.J., Barnett, M.J., Tenerowicz, M.J., Perry, P.J., 2011. The Anabolic 500 survey: characteristics of male users versus nonusers of anabolic-androgenic steroids for strength training. Pharmacotherapy 31, 757–766.
Kanayama, G., Pope Jr., H.G., 2012. Misconceptions about anabolic-androgenic steroid abuse. Psychiatr. Ann. 42, 371–375.
Kanayama, G., Barry, S., Hudson, J.I., Pope Jr., H.G., 2006. Body image and attitudes toward male roles in anabolic-androgenic steroid users. Am. J. Psychiatry 163, 697–703.
Kanayama, G., Pope Jr., H.G., Cohane, G., Hudson, J.I., 2003. Risk factors for anabolic-androgenic steroid use among weightlifters: a case-control study. Drug Alcohol Depend. 71, 77–86.
Kanayama, G., Pope, H.G., Hudson, J.I., 2018. Associations of anabolic-androgenic steroid use with other behavioral disorders: an analysis using directed acyclic graphs. Psychol. Med. 48, 2601–2608.
Kindlundh, A.M., Isacson, D.G., Berglund, L., Nyberg, F., 1999. Factors associated with adolescent use of doping agents: anabolic-androgenic steroids. Addiction 94, 543–553.
Klotz, F., Garle, M., Granath, F., Thiblin, I., 2006. Criminality among individuals testing positive for the presence of anabolic androgenic steroids. Arch. Gen. Psychiatry 63, 1274–1279.
McDaniel, S.R., Mahan, J.E., 2008. An examination of the ImpSS scale as a valid and reliable alternative to the SSS-V in optimum stimulation level research. Pers. Individ. Diff. 44, 1528–1538.
Murray, S.B., Griffiths, S., Mond, J.M., Kean, J., Blashill, A.J., 2016. Anabolic steroid use and body image psychopathology in men: delineating between appearance- versus performance-driven motivations. Drug Alcohol Depend. 165, 198–202.
Olivardia, R., Pope Jr., H.G., Hudson, J.I., 2000. Muscle dysmorphia in male weightlifters: a case-control study. Am. J. Psychiatry 157, 1291–1296.
Parker, G., Tupling, H., Brown, L.B., 1979. A parental bonding instrument. Br. J. Med. Psychol. 52, 1–10.
Parkinson, A.B., Evans, N.A., 2006. Anabolic androgenic steroids: a survey of 500 users. Med. Sci. Sports Exerc. 38, 644–651.
Pearl, J., 1995. Causal diagrams for empirical research. Biometrika 82, 669–688.
Pearl, J. . 2018. Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution. Technical Report R-475. Available online at: 475 (http://ftp.cs.ucla.edu/pub/stat_ser/r475.pdf).
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Phillips, K.A., Hollander, E., Rasmussen, S.A., Aronowitz, B.R., DeCaria, C., Goodman, W. K., 1997. A severity rating scale for body dysmorphic disorder: development, reliability, and validity of a modified version of the Yale-Brown Obsessive Compulsive Scale. Psychopharmacol. Bull. 33, 17–22.
Polivy, J., Herman, C.P., 2002. Causes of eating disorders. Annu. Rev. Psychol. 53, 187–213.
Pope Jr., H.G., Kanayama, G., Hudson, J.I., Kaufman, M.J., 2021. Review article: anabolic-androgenic steroids, violence, and crime: two cases and literature review. Am. J. Addict. 30, 423–432.
Pope Jr., H.G., Wood, R.I., Rogol, A., Nyberg, F., Bowers, L., Bhasin, S., 2014. Adverse health consequences of performance-enhancing drugs: an Endocrine Society scientific statement. Endocr. Rev. 35, 341–375.
Pope, H.G., Kanayama, G., Hudson, J.I., 2012. Risk factors for illicit anabolic-androgenic steroid use in male weightlifters: a cross-sectional cohort study. Biol. Psychiatry 71, 254–261.
Rachon, D., Pokrywka, L., Suchecka-Rachon, K., 2006. Prevalence and risk factors of anabolic-androgenic steroids (AAS) abuse among adolescents and young adults in Poland. Soz Praventivmed. 51, 392–398.
Rasmussen, J.J., Selmer, C., Ostergren, P.B., Pedersen, K.B., Schou, M., Gustafsson, F., Faber, J., Juul, A., Kistorp, C., 2016. Former abusers of anabolic androgenic steroids exhibit decreased testosterone levels and hypogonadal symptoms years after cessation: a case-control study. PLOS One 11, e0161208.
Rohman, L., 2009. The relationship between anabolic androgenic steroids and muscle dysmorphia: a review. Eat. Disord. 17, 187–199.
Scarth, M., Havnes, I.A., Jorstad, M.L., McVeigh, J., Van Hout, M.C., Westlye, L.T., Torgersen, S., Bjornebekk, A., 2022. Severity of anabolic steroid dependence, executive function, and personality traits in substance use disorder patients in Norway. Drug Alcohol Depend. 231, 109275.
Skarberg, K., Nyberg, F., Engstrom, I., 2009. Multisubstance use as a feature of addiction to anabolic-androgenic steroids. Eur. Addict. Res. 15, 99–106.
Skarberg, K., Nyberg, F., Engstrom, I., 2010. Is there an association between the use of anabolic-androgenic steroids and criminality? Eur. Addict. Res. 16, 213–219.
Tarter, R.E., Kirisci, L., Habeych, M., Reynolds, M., Vanyukov, M., 2004. Neurobehavior disinhibition in childhood predisposes boys to substance use disorder by young adulthood: direct and mediated etiologic pathways. Drug Alcohol Depend. 73, 121–132.
Thiblin, I., Parlklo, T., 2002. Anabolic androgenic steroids and violence. Acta Psychiatr. Scand. 412, 125–128.
Tibshirani, R.J., 1986. Regression shrinkage and regression *via* the lasso. J. R. Stat. Soc. Ser. B 58, 267–288.
Ward, M.F., Wender, P.H., Reimherr, F.W., 1993. The Wender Utah rating scale: an aid in the retrospective diagnosis of childhood attention deficit hyperactivity disorder. Am. J. Psychiatry 150, 885–890.
Zou, H., Hastie, T., 2005. Regularization and variable selection *via* the elastic. J. R. Stat. Soc. Ser. B 67, 301–320.