UTS UNIVERSITY OF TECHNOLOGY SYDNEY

The Metabolomic Investigation of Urinary Biomarkers for Progestogenic and Dopaminergic Substance Misuse in Racehorses

by Madysen Elbourne

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

Under the supervision of
Prof Shanlin Fu
Dr Adam Cawley
Dr Anjali Gupta

University of Technology Sydney
Centre of Forensic Science
Faculty of Science

February 2025

Certificate of Original Authorship

I, Madysen Elbourne, declare that this thesis is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Science at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: February 27th 2025

This page was intentionally left blank

Acknowledgements

The past four years of research have been a combination of problem solving, wonderful discoveries, devastating roadblocks and a lot of hard work. I would not have made it this far without the support of so many incredible people along the way, who I deeply appreciate for their kindness and generosity.

A big thank you to Glenys and Jaymie for their work conducting the administration studies analysed in this thesis, to the horses involved in the studies and to AgriFutures for their study funding. To my Shimadzu contacts, Chris, Paul, and Cedric. Thank you for your assistance with everything LC-HRMS, from my silly little questions to the big QTOF challenges we faced. To Chris Fitzgerald and Mal McLeod from ANU, for their assistance and guidance in all things sulfate synthesis and coding, I truly could not have done some of this work without your wisdom and patience. To Kireesan, Lauren D, Stacey, Paul, Ming, Nina, John, Lance, Josh, and all ARFL other staff members. Thank you for supporting my learning in the lab, asking about the progress of my research, and answering any questions I have had along the way. To all the students I've taught over the past four years. Thank you for always making my day brighter, and for allowing me the opportunity to assist you in your learning.

To my incredible supervisors, Prof Shanlin Fu and Dr Adam Cawley, who have genuinely been the best supervisors I could've ended up with. Thank you for your constant guidance, your prompt and constructive feedback, and your support in every research decision I've made over the years. I hope you're also proud of the research we've achieved together.

To Rin, Archie, Eathan, Lana, Beth, Kathy, and everyone else who have come and gone in the UTS Forensic Toxicology Research group, for being close friends, mentors and a sounding board to each other when we needed it, thank you. To everyone in the CFS Research students' group, but especially Ana, Ciara, Harry, Matt, and Teneil, thank you for being the best research support group a scientist could ever ask for. I came into this space uncertain in my abilities and am now closing this chapter reassured in my skills and knowledge, and surrounded by amazing friendships.

Finally, thank you to my friends and family who have supported me throughout my PhD. Particularly my mum, dad, and brother, who have endured my most anxious moments as well as my most triumphant ones, thank you for everything you've done to support me and my career.

“Progress, not perfection.”

Table of Contents

List of Figures

List of Tables

Acronyms

Abbreviation	Definition
AADC	Aromatic L-Amino Acid Decarboxylase
AAS	Anabolic Androgenic Steroids
ABP	Athlete Biological Passport
ACN	Acetonitrile
AI	Artificial Intelligence
ALT	Altrenogest
ANN	Artificial Neural Networks
AORC	Association of Official Racing Chemists
ARFL	Australian Racing Forensic Laboratory
ATD	Androsta-1,4,6-triene-3,17-dione
AUC	Area Under the Curve
BBB	Blood Brain Barrier
BP	Boiling Point
BW	Body Weight
CART	Classification and Regression Trees
CE	Collision Energy
CG	Castrated Male Horse (Gelding) from the Control Group
Class.Error	Classification Error
CM	Female Horse (Mare) from the Control Group
CNS	Central Nervous System
COMT	Catechol-O-Methyltransferase
CS	Cortisol Sulfate
CSV	Comma Separated Values

Abbreviation	Definition
D	Day
DCM	Dichloromethane
DDA	Data Dependent Acquisition
DEA	Diethylamine
DHEA	Dehydroepiandrosterone
DIA	Data Independent Acquisition
DL	Deep Learning
DMF	Dimethylformamide
DOPAC	3,4-Dihydroxyphenylacetic Acid
DT	Decision Trees
E	Epitestosterone
E1S	Estrone Sulfate
EAAS	Endogenous Anabolic Androgenic Steroids
EBP	Equine Biological Passport
EPO	Erythropoietin
ERC	Endogenous Reference Compound
ESI	Electrospray Ionisation
EtOAc	Ethyl Acetate
Etr	Epitrenbolone
eV	Electron Volts
FA	Formic Acid
FC	Fold Change
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FSH	Follicle-Stimulating Hormones

Abbreviation	Definition
FWHM	Full Width at Half Maximum
GC-MS	Gas Chromatography - Mass Spectrometry
h	Hour
H	Horse
HCA	Hierarchical Cluster Analysis
HIF	Hypoxia-Inducible Factor
HMDB	Human Metabolome Database
HQC	High Quality Control Level
HRMS	High Resolution Mass Spectrometry
HVA	3-Methoxy-4-Hydroxyphenylacetic Acid, Homovanillic Acid
Hz	Hertz
IABRW	International Agreement on Breeding, Racing, and Wagering
IFHA	International Federation of Horseracing Authorities
IM	Intramuscularly Injected
IQR	Interquartile Range
IR	Intensity Ratio
ISTD	Internal Standard
K-NN	K-Nearest Neighbours
kg	kilogram
kV	kilovolts
L-DOPA	Levodopa
LC-MS	Liquid Chromatography - Mass Spectrometry
LH	Luteinizing Hormones
LLE	Liquid-Liquid Extraction
LLOQ	Lower Limit of Quantification
LOD	Limit of Detection

Abbreviation	Definition
LOWESS	Locally Weighted Scatterplot Smoothing
LQC	Low Quality Control Level
LR	Logistic Regression
M	Molar
m/z	Mass-to-Charge Ratio
MA	Maximum Abundance
MAE	Mean Absolute Error
MAO-B	Monoamine Oxidase-B
MCCV	Monte-Carlo Cross Validation
MeOH	Methanol
mg	milligram
mg/kg	milligram per kilogram
mg/mL	milligrams per millilitre
min	Minutes
ML	Machine Learning
mL	millilitre
mM	millimolar
mm	millimetre
MoNA	MassBank of North America
MQC	Medium Quality Control Level
MRM	Multiple Reaction Monitoring
MRPL	Minimum Reporting Performance Limit
MS	Mass Spectrometry
MSE	Mean Square Error
MW	Molecular Weight
NaOH	Sodium Hydroxide

Abbreviation	Definition
NB	Naïve Bayes
ng/mL	nanograms per millilitre
NH_3	Concentrated Ammonia
NIST	National Institute of Standards and Technology
NMI	National Measurement Institute
NSW	New South Wales
NT	Northern Territory
OPLS-DA	Orthogonal Partial Least-Squares Discriminant Analysis
Opt	Optimised
PC	Principal Component
PCA	Principal Component Analysis
PD	Parkinson's Disease
PDF	Portable Document Format
PLS-DA	Partial Least-Squares Discriminant Analysis
PNMT	Phenylethanolamine N-Methyltransferase
PP	Protein Precipitation
PR	Precision Recall
Pr	Progesterone
Pred. Acc.	Predictive Accuracy
PregS	Pregnenolone Sulfate
QC	Quality Control
QQQ	Triple Quadrupole
QTOF	Quadrupole Time of Flight
RASL	Racing Analytical Services Limited
RE	Relative Error
RF	Random Forest

Abbreviation	Definition
rhEPO	Recombinant Human Erythropoietin
RMSE	Root Mean Squared Error
ROC	Receiver Operating Characteristic
RSD	Relative Standard Deviation
RT	Retention Time
S/N	Signal-to-Noise Ratio
SA	South Australia
SO_3•py	Sulfur Trioxide Pyridine Complex
SPE	Solid Phase Extraction
SVM	Support Vector Machines
T	Testosterone
Td	Trendione
TG	Castrated Male Horse (Gelding) from the Treated Group
TIC	Total Ion Chromatogram
TM	Female Horse (Mare) from the Treated Group
TN	True Negative
TP	True Positive
TPR	True Positive Rate
Tr	Trenbolone
TS	Testosterone Sulfate
Tyr	Tyramine
UHPLC	Ultra-High-Performance Liquid Chromatography
ULOQ	Upper Limit of Quantification
v/v	volume per volume
VIP	Variable Importance Projection
WA	Western Australia

Abbreviation	Definition
WADA	World Anti-Doping Agency
WAX	Weak Anion Exchange
x g	Gravity
XGBoost	Extreme Gradient Boosting
µg/mL	micrograms per millilitre
µL	microlitre
µm	micrometre
2-ME2S	2-Methoxy-Estradiol Sulfate
3-MT	3-Methoxytyramine
6-OXO	Androst-4-ene-3,6,17-trione

Research Communications

JOURNAL PUBLICATIONS

Elbourne, M., Cawley, A., Stanley, S., Bowen, C., & Fu, S. (2022). Intelligence benefit of the 3-methoxytyramine to tyramine ratio in equine urine. *Drug Testing and Analysis*, *14*(5), 936-942. https://doi.org/10.1002/dta.3264

Cawley, A., Keen, B., Tou, K., **Elbourne, M.,** & Keledjian, J. (2022). Biomarker ratios. *Drug Testing and Analysis*, *14*(5), 983-990. https://doi.org/10.1002/dta.3250

Fitzgerald, C. C. J., Bowen, C., **Elbourne, M.,** Cawley, A., & McLeod, M. D. (2022). Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers. *Journal of the American Society for Mass Spectrometry*, *33*(7), 1276-1281. https://doi.org/10.1021/jasms.2c00092

Elbourne, M., Fu, S., Cawley, A., Bowen, C., & Gupta, A. (2022). Metabolomic analysis of equine urine to determine route of altrenogest administration. *Toxicologie Analytique et Clinique*, 34(3, Supplement), S55-S56. https://doi.org/10.1016/j.toxac.2022.06.068

Elbourne, M., Keledjian, J., Cawley, A., Bowen, C., Fitzgerald, C., Gupta, A., Noble, G., Loy, J., & Fu, S. (2023). Metabolomic analysis of equine urine to determine administration route of altrenogest. 23rd Proc. Int. Conf. Racing Anal. Vets., Hong Kong.

Portions of this article and additional analysis are found in ***Chapter 3: Quantitative Analysis of Altrenogest Misuse***

Elbourne, M., Keledjian, J., Cawley, A., & Fu, S. (2024). Administration Route Differentiation of Altrenogest via the Metabolomic LC-HRMS Analysis of Equine Urine. *Molecules*, *29*(21). https://doi.org/10.3390/molecules29214988

Portions of this article and additional analysis are found in ***Chapter 5: Administration Route Differentiation of Altrenogest via the Metabolomic LC-HRMS Analysis of Equine Urine***

CONFERENCE PROCEEDINGS AND COMMUNICATIONS

(presenting author underlined)

Elbourne, M., Cawley, A., Stanley, S., Bowen, C., & Fu, S. GC-MS Profiling of Equine Urine for Longitudinal Assessments, presented at the Forensic & Clinical Toxicology Association Conference (FACTA), March 2022 – Brisbane, Australia

Elbourne, M., Cawley, A., Bowen, C., Noble, G., Gupta, A., & Fu, S. Metabolomic analysis of equine urine to determine route of altrenogest administration, presented at the 59th Annual Meeting of The International Association of Forensic Toxicologists Conference (TIAFT), September, 2022 - Versailles, France

Elbourne, M., Keledjian, J., Cawley, A., Bowen, C., Fitzgerald, C., Gupta, A., Noble, G., Loy, J., & Fu, S. Enhanced data analysis workflows for LC-HRMS used in forensic toxicology applications, presented at the 60th Annual Meeting of The International Association of Forensic Toxicologists Conference (TIAFT), August 2023 - Rome, Italy

Elbourne, M., Keledjian, J., Cawley, A., Bowen, C., Fitzgerald, C., Gupta, A., Noble, G., Loy, J., & Fu, S. Metabolomic analysis of equine urine to determine route of altrenogest administration, presented at the 23rd International Conference of Racing Analysts and Veterinarians (ICRAV), September 2023 - Hong Kong SAR

Elbourne, M., Cawley, A., Stanley, S., Bowen, C., & Fu, S. GC-MS profiling of equine urine for longitudinal assessments, poster presented at the 23rd International Conference of Racing Analysts and Veterinarians (ICRAV), September 2023 - Hong Kong SAR

Elbourne, M., Keledjian, J., Cawley, A., Bowen, C., Gupta, A., & Fu, S. Administration route determination of altrenogest through the metabolomic analysis of equine urine, presented at the 23rd Triennial International Association of Forensic Sciences (IAFS) meeting held in conjunction with the 26th Symposium of the Australian & New Zealand Forensic Science Society (ANZFSS), November 2023 - Sydney, Australia

Elbourne, M., Keledjian, J., Cawley, A., Bowen, C., Gupta, A., & Fu, S. An untargeted approach for the hypothesis generation of metabolomic biomarkers from an exogenous steroid administration, presented at the Forensic & Clinical Toxicology Association Conference (FACTA), March 2024 - Hobart, Australia

Elbourne, M., Cawley, A., Bowen, C., Keledjian, J., Gupta, A., & Fu, S. The development and validation of a routine LC-HRMS method for dopamine manipulation detection, presented at the Forensic & Clinical Toxicology Association Conference (FACTA), March 2024 - Hobart, Australia

Elbourne, M., Keledjian, J., Bowen, C., Cawley, A., & Fu, S. The indirect detection of dopaminergic manipulation in equine urine via an optimised routine and metabolomic-based LC-HRMS method, presented at the 25th International Mass Spectrometry Conference (IMSC), August 2024 - Melbourne, Australia

Abstract

This thesis involves extensive research concerning the direct quantification and metabolomic investigation of progestogenic and dopaminergic substance misuse in racehorses, and the implementation of data processing workflows and machine learning tools to improve biomarker discovery in an equine anti-doping context.

Anabolic androgenic steroids are some of the illegal compounds most commonly used in the sporting industry. The use of these compounds is prohibited in the equine thoroughbred racing industry due to their performance enhancing benefits. Altrenogest, also known as allyltrenbolone, is a synthetic form of progesterone used therapeutically to suppress unwanted symptoms of oestrus in female horses, and is commercially available in oral and injectable forms. When altrenogest is administered however, prohibited steroid impurities, trendione (as major impurity), trenbolone (minor), and epitrenbolone (from metabolism), can be detected. Current Racing NSW ruling dictates the allowed use of altrenogest must occur outside of one clear day of any racing, and administration permitted, orally, to female horses only. Therefore, a complementary qualitative and metabolomic analysis approach was developed for the potential to improve the detection of intramuscularly injected altrenogest, using features differing from oral altrenogest.

Samples obtained from a ten-mare administration, conducted by Charles Sturt University with ethics approval (A19050), were analysed involving both oral and intramuscular administration of altrenogest. The collection period spanned from day zero to 21 days post-administration, totalling 32 samples collected per horse. Quantification of trendione and epitrenbolone by liquid chromatography-tandem mass spectrometry was performed following enzyme hydrolysis of urine and liquid-liquid extraction. A bivariate analysis of trendione and epitrenbolone levels provided evidence for discerning oral and intramuscular administrations. Quantified levels of these impurities did not return below the 1 ng/mL reporting limit until 2 days post final intramuscular administration, whereas, in the orally administered samples, impurities levels had returned below this reporting limit within 24 hours after final administration. A complementary untargeted steroid sulfate method was adapted to analyse the intact sulfated steroids present in urine samples using liquid chromatography-high resolution mass spectrometry with data dependent acquisition. Liquid chromatography high-resolution mass spectrometry analysis of equine urine samples found five sulfated compounds, estrone sulfate, testosterone sulfate, 2-methoxyestradiol sulfate, pregnenolone sulfate, and cortisol sulfate, with the potential to differentiate between oral and intramuscularly

administered altrenogest using a random forest classification model. The best model results, comparing two horses' administration normalised peak area datasets, gave an AUC score of 0.965 with a confidence level of 95% (between 0.931 and 0.995). The differentiation between ALT administration types from an excreted urine sample will greatly benefit the thoroughbred racing industry, due to the nature of race day sampling, ALT as an allowed therapeutic substance, and the complicated situation regarding the steroid impurities present in commercial products.

Stalevo®, a levodopa containing therapeutic drug designed to treat Parkinson's disease in human patients, has been suspected to be used in the equine racing industry intended for dopaminergic manipulation. In response to the first known observation of dopaminergic manipulation in the racing industry in 1996, an international threshold for 3-methoxytyramine was established by Wynne *et al.* in 2004, to detect dopaminergic manipulation in racehorses. This threshold at 4 µg/mL (4,000 ng/mL) has the capability to correctly identify the administration of additional levodopa, however it has limitations in not being able to accurately detect more complex Parkinson's disease medication abuse, or levodopa micro-dosing within the threshold.

To address this concern, an LC-QTOF-MS method was optimised and validated for the urinary quantification of 3-methoxytyramine and tyramine to identify levodopa misuse. A complementary targeted and metabolomic approach was also implemented here to monitor catecholamines related to the levodopa metabolism in equine urine. This approach enabled the search for more robust urinary biomarkers in their intact sulfate and glucuronide conjugated forms, in addition to the free metabolite. Samples obtained from a 12-horse administration, conducted by Charles Sturt University with ethics approval (A20277) were analysed. Metabolomic screening obtained a group of related metabolites with the potential to identify exogenous levodopa administration. The potential for a machine learning model to extend the detection window for dopaminergic manipulation was also successfully implemented.

This research has successfully demonstrated the implementation of metabolomics for a complementary targeted/untargeted analytical approach to improve doping detection in complex cases. Additionally, multiple biomarkers and biomarker profile classification models have been proposed for the improved detection of progestogenic and dopaminergic substance misuse in racehorses.

Chapter 1: An Introduction of Metabolomics for Doping Detection in Equine Racing

1 AN INTRODUCTION OF METABOLOMICS FOR DOPING DETECTION IN EQUINE RACING

1.1 DOPING DETECTION IN HUMAN AND ANIMAL SPORTING INDUSTRIES

Doping is a term referring to the use of illicit substances or misuse of therapeutic substances by athletes. This is typically for the desired effect of performance enhancement or performance suppression in a sporting event. The act of doping creates an unfair advantage to those involved, and as a whole, threatens the integrity of the sporting industry [1]. Thus, the main goal of anti-doping research is to protect the welfare of athletes and the integrity of the sport.

The list of substances that are monitored for doping is controlled by the World Anti-Doping Agency (WADA, https://wada-ama.org/en) in the human athlete industry and by the International Federation of Horseracing Authorities (IFHA, https://ifhaonline.org) in the equine industry [2, 3]. However, unlike the human sporting field, the equine racing industry does not have a standardised set of rules and regulations that all participating countries and jurisdictions must follow. Within the local racing scene, rules and regulations are also dictated by state bodies. In Australia, for example, local ruling is conducted by Racing NSW (New South Wales, https://racingnsw.com.au), Racing Victoria (https://racingvictoria.com.au), Racing SA (South Australia, https://racingsa.com.au), Racing WA (Western Australia, https://racingwa.com.au), Thoroughbred Racing NT (Northern Territory, https://trnt.org.au), and Racing Queensland (https://racingqueensland.com.au). Foremost, these states abide by the ruling predicated by Racing Australia (https://racingaustralia.horse/home.aspx), who align with the IFHA international agreement. Each organisation has their published guidelines for racing and associated areas (allowed gear, rules around breeding, permitted betting practices, etc) that should be followed. The local ruling must always enforce the regulations set out by the IFHA in their “International Agreement on Breeding, Racing, and Wagering” (IABRW) [2]. Article 6A of this ruling denotes the types of doping substances that must be monitored for in a racing laboratory and a reporting limit or threshold that should be applied. The detection of these compounds is to be done though the analysis of the pre- or post-race collection of plasma or urine samples. A screening method is generally applied to the sample, dependent on the biological matrix collected.

1.1.1 Current Detection Methods

Generally speaking, there are currently two main detection methods undertaken for the detection of illegal drug use. These are the direct detection of the administered parent drug,

or alternatively, the targeted detection through a direct metabolite of such. Current practices in anti-doping detection align with targeting known drugs or direct metabolites of such through time-consuming targeted qualitative and/or semi-quantitative screening analysis. The current detection methods are advantageous in a number of scenarios/detection types; namely, when identifying exogenous substance doping, when the parent drug or direct metabolite has a relatively long half-life in the system prior to metabolism and excretion, and when the drug is present at high levels in the body. These could be considered best case scenario detections.

There are, however, some limitations to this current process; namely, the detection of endogenous compound doping is increasingly difficult, as well as drugs that are rapidly metabolised and quickly excreted from the body. Additionally, lower concentration compound detection, i.e., micro-dosing or sub-threshold dosing is increasingly difficult with these techniques. Therefore, there is a need for a more advanced analysis process to address these limitations. The implementation of metabolomic analysis workflows (explained in **Section 1.2: Metabolomics in Doping Analysis** below) could provide a solution to many of the current limitations mentioned.

1.1.2 Equine Anti-Doping: Biomarkers

The animal sporting industry differs somewhat in the doping context to human sports. Whilst there is the potential to improve performance in the equine industry, doping can also be utilised to inhibit performance. This results in a broader scope of prohibited substances requiring detection [4]. The discovery of novel biomarkers of doping may provide a significant improvement in the way substance misuse is currently detected in racehorses and potentially reduce or streamline the screening process of samples.

A biomarker is defined as “a molecule that indicates an alteration of the physiological state of an individual, in relation to health or disease state, drug treatment, toxins and other challenges of the environment” [5]. On a broader scale, it is considered to be any measurable parameter altered as a result of a challenge to the individual’s system [4]. When synthetic copies of endogenous (i.e., naturally occurring) compounds are misused, it can be difficult to detect as their presence in a biological sample is not unusual. To meet this challenge, biomarkers (endogenous metabolites that are peripheral to the target compound) can be monitored to detect abnormal levels that can provide intelligence of the intended performance enhancement or suppression [6]. This work aims to build on the current body of biomarkers used for doping detection in the equine racing industry.

1.1.3 Intelligence: Equine Biological Passport (EBP)

The idea for an individual baseline profile was adapted from a similar system put in place for human athletes by WADA, referred to as the Athlete Biological Passport (ABP) [7, 8]. The ABP program, first published in 2004, was designed to monitor selected biological variables over time that indirectly reveal the effects of doping, instead of attempting to detect the doping substances themselves [7-9]. WADA has also implemented the use of endogenous reference compounds (ERCs), which are substances that are metabolically related and analysed simultaneously with target compounds to reduce analytical and/or physiological variance, and therefore improve *intra*-individual longitudinal assessments that aim to identify change resulting from pharmacological manipulation. Currently, the main compounds being utilised by WADA as ERCs originate from progestins, corticosteroids, and other adrenal precursors [10].

The same principles have been applied by Racing NSW to the equine racing industry to develop the Equine Biological Passport (EBP) [11, 12]. Racing NSW established the world's first thoroughbred EBP to perform longitudinal profiling of samples collected from specific horses over time. Analysis of biomarkers by targeted analysis and untargeted feature extraction have been applied to discern pharmaceutical manipulation from physiological variation [12]. The EBP utilises individual reference limits to account for individual levels for specific biomarkers. The current EBP includes 25 biomarkers for monitoring, and aims to provide additional information to the stewards and Racing NSW officials in potential doping scenarios.

Biomarker ratios have also been used historically to monitor the use of testosterone and its precursors in geldings and mares. Prior to the current concentration threshold of 20 ng/mL in geldings, a testosterone to DHEA (T/DHEA) ratio greater than five (5) was proposed in 2018 for testosterone misuse in both geldings and mares, and a testosterone to epitestosterone (T/E) ratio of 12 was adopted as a threshold proposed in 2017 for the detection of testosterone misuse in mares [13, 14]. This T/E threshold was based on data collected from the analysis of 294 post-race urine samples and following an intramuscularly injected dose of 1000 mg of testosterone hexahydrobenzoate [14, 15]. Biomarker ratios have also been proposed in 2022 for implementation in the EBP by Elbourne *et al.* with the 3-methoxytyramine-to-tyramine (3-MT/Tyr) ratio for the detection of levodopa (L-DOPA) doping [16].

1.2 Untargeted Metabolomics in Doping Analysis

Metabolomics is the study of small molecules for the qualitative and quantitative investigation of metabolite changes in the body that are caused by a stimulus, which in this case, is a drug

intake. Introducing metabolomic processes, complementary to the current detection methods, will lead to greater biological interpretation of the results, in-turn, translating to more meaningful research outcomes. In this thesis, metabolomics was utilised as a technique to identify compounds of interest that could be investigated further as a biomarker or biomarker ratio with the potential for implementation into the EBP or current routine screening methods.

Metabolomics requires the use of high-resolution mass spectrometry (HRMS) instrumentation which, in turn, involves the collection of large amounts of data from its analysis. It can be conducted generally in two ways, as targeted or untargeted metabolomics, which dictates whether the compounds of interest are known or unknown prior to sample analysis. Targeted metabolomics aims to obtain information from, and quantify the presence of a pre-defined set of compounds, however, this is not classified as a true “omics” approach as it is limited in analyte scope [17, 18]. Whereas, an untargeted approach involves screening for all metabolites in a sample which may then be used to identify novel biomarkers that are associated with a particular physiological state. This is considered to be a true “omics” approach [18-20].

Commonly used sample extraction techniques applied to urine samples for metabolomic investigation include protein precipitation (PP), solid phase extraction (SPE), or liquid-liquid extraction (LLE) [21]. Often this is done in combination with enzyme of chemical hydrolysis to target the free compound and reduce the amount of phase I and II conjugated compounds. Common instrumentation and analytical techniques include gas and liquid chromatography coupled with a quadrupole time of flight (QTOF), mass spectrometry (or tandem MS/MS, HRMS). The extraction method and analytical technique is chosen dependent on the desired research outcome and the desired target compound class. Data processing tools and analysis programs/software are often required to comprehensively interpret the data obtained from high throughput HRMS analysis. A brief breakdown of these techniques provided below summarises a more comprehensive overview published by Keen *et al.* (2022) in their review of metabolomics in clinical and forensic toxicology, sports anti-doping and veterinary residues [19].

1.2.1 Protein Precipitation (PP)

PP methods generally involve the use of a small volume of biological fluid (~100 µL) before quenching (i.e., the addition of cold solvents, acids or rapid heating). to preserve biological activity [17, 18, 22]. This technique is beneficial for metabolomic style analyses as there is minimal opportunity for metabolite loss, it is also quick and requires very little work pre-analysis

[23]. However, this rapid technique may fail to remove all matrix contamination and result in increased chromatography noise [24].

1.2.2 Solid Phase Extraction (SPE)

SPE is a widely used sample preparation technique due to its high extraction yields and repeatability [25, 26]. One of the main benefits of SPE is compound class selectivity via washing (to remove matrix interferences) and elution (of desired analytes) [25, 27]. This can, however, introduce bias by the exclusion of compound classes [28].

1.2.3 Liquid-Liquid Extraction (LLE)

LLE uses immiscible solvents to transfer target compounds between aqueous (i.e., hydrophilic) and organic (i.e., hydrophobic) phases. This extraction method provides selection and isolation of target compounds with minimal matrix contamination [29].

1.2.4 Gas Chromatography - Mass Spectrometry (GC-MS)

GC-MS has excellent separation efficiency while maintaining reproducible retention times [30]. This chromatographic technique is considered the gold standard for analytical determinations with increased sensitivity and specificity, together with the use of spectral libraries [27, 31-33]. The major challenge for GC-MS is analysis of non-volatile, highly polar and thermally unstable compounds. Compounds are often subjected to chemical modification, but these derivatised compounds can display limited stability [30, 32].

1.2.5 Liquid Chromatography - Mass Spectrometry (LC-MS)

The use of LC-MS improves the ability to detect and characterise a broader range of analytes, particularly small polar compounds that are insufficiently volatile and/or too labile for GC-MS [31, 32].

1.2.6 High Resolution Mass Spectrometry (QTOF)

High resolution analytical instruments (i.e., QTOF and Orbitrap) are increasingly popular due to their advantage of acquisition in full-scan mode, high scanning speeds, accurate mass and high resolution [18, 32, 34-39]. LC-HRMS allows for the screening and confirmation of a large scope of organic compounds [35].

1.2.7 Data Processing and Analysis Platforms

Data processing tools and analysis programs/software is often required to comprehensively interpret the data obtained from high throughput HRMS analysis. The tool (or tools) chosen

are dependent on the analytical technique used and the method of data acquisition selected [30, 40, 41]. Using a combination of programs to streamline the interpretation of results is often referred to as a multi-tool workflow, or pipeline. The main benefits of these workflows are to automate time-consuming tasks, i.e., data pre-processing (chromatographic alignment, filtering and normalisation), feature extraction, statistical analysis and MS annotation [18, 40].

1.3 Examples of Complex Doping Detection

1.3.1 Steroid Markers

Like humans, the hormones of female horses (mares) fluctuate relative to their oestrus cycle (ovulation). With this hormonal fluctuation, mares often exhibit unwanted behavioural changes that could threaten the safety of themselves, as well as those working alongside the animals [42-44]. Namely the jockeys, trainers and handlers, as well as other horses. During oestrus, progesterone levels naturally decrease and production of estrogen spikes. To counteract this, progestogen containing products, commercially available as oral and injectable forms, are commonly used in mares to regulate and control ovulation for breeding purposes [42, 45].

Altrenogest (ALT, allyltrenbolone, (8S,13S,14S,17R)-17-hydroxy-13-methyl-17-prop-2-enyl 1,2,6,7,8,14,15,16-octahydrocyclopenta[a]phenanthren-3-one) is a synthetic steroidal progestin used to suppress oestrus in female horses and characteristically displays both anabolic and androgenic affects. ALT acts by decreasing levels of the endogenous gonadotrophin hormones, luteinizing (LH) and follicle-stimulating hormones (FSH). This results in more manageable mares for training and competition alongside male horses, while also improving the workplace safety of riders and handlers [42, 45, 46].

When ALT is synthesised, the steroid, trenbolone (Tr, (8S,13S,14S,17S)-17-hydroxy-13-methyl-2,6,7,8,14,15,16,17-octahydro-1H-cyclopenta[a]phenanthren-3-one), is produced as a side product and impurity. A second unwanted steroid, trendione (Td, (8S,13S,14S)-13-methyl-1,2,6,7,8,14,15,16-octahydrocyclopenta[a]phenanthrene-3,17-dione), can occur through oxidation of trenbolone in the bottle/synthesis process, and epitrenbolone (Etr, (8S,13S,14S,17R)-17-hydroxy-13-methyl-2,6,7,8,14,15,16,17-octahydro-1H-cyclopenta[a] phenanthren-3-one) can be formed from metabolism of trenbolone (Figure 1-1). These steroids (trenbolone, trendione, and epitrenbolone) are banned for use entirely in the thoroughbred racing industry as they're also anabolic androgenic steroids (AAS) and can have a performance enhancing effect on the horse. This has posed a problem as they can be detected in the urine when ALT is administered.

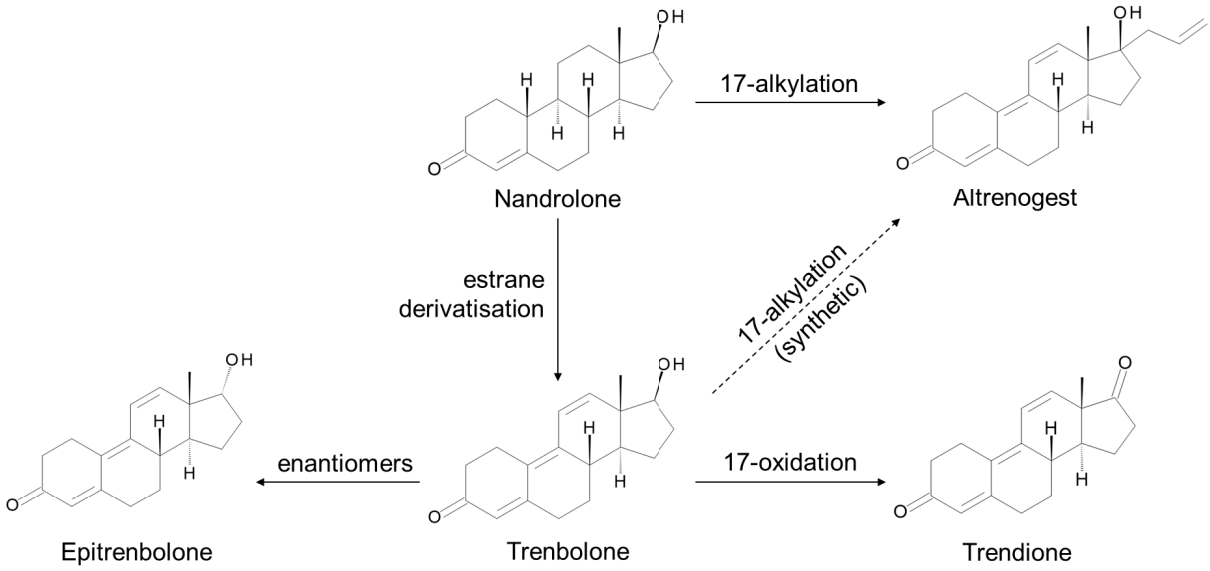


Figure 1-1: Synthesis of nandrolone to altrenogest (ALT) with steroid impurities (trenbolone, trendione and epitrenbolone) formed through incomplete transformation.

It is also understood that larger concentrations of these impurities are present in injected forms of the compound, in comparison to the oral form in the market (discussed further in **Chapter 3, Section 3.3.2**) [47]. Therefore, there is concern that the use of this therapeutic substance may cause an unfair advantage if higher levels of these steroids are in fact present. In response to the detection of low-levels of steroid impurities contained in commercially available ALT products and its use observed in the thoroughbred racing industry, Rule LR44A was introduced by Racing NSW in June 2018 [48]. This local ruling advised against the use of ALT in fillies and mares being administered orally within one clear day of racing (at least 24 hours), with a detection limit of 1 ng/mL for any steroid impurities present (trendione, trenbolone and epitrenbolone). The use of injectable forms of ALT was also prohibited (with a complete ban in male and gelded horses already in place). This ruling has since been rescinded (July 2023) as the detection limit at 1 ng/mL was deemed too low, and inappropriate for enforcement in all cases of ALT presence [49]. Removing this limit allowed more discretion to the stewards for further investigation and prosecution to exist on a case-by-case basis. However, the warning and recommendation to avoid injectable forms of ALT still remains and updates to the horseracing community were communicated to reflect this [47, 50].

Previous research has investigated the pharmacokinetic relationship of ALT, and quantification of ALT in equine. However, research is currently lacking in the steroid impurities present in the ALT product, and their differences between administration route formulations. There has also been conflicting research and published opinions suggesting that ALT may provide anabolic effects to an administered horse, in part due to its steroid impurities, which led many governing racing bodies around the world to ban the therapeutic use of ALT outright

[42, 51]. Racing NSW were one of the few local jurisdictions to keep ALT off the prohibited list and instead adapted its ruling to restrict the use accordingly [49].

Hodgson *et al.* (2005) investigated the behavioural effects of prolonged use of altrenogest in mares, examining social hierarchy, activity budget, body-mass and body condition of 11 horses (5 treated, 6 control). The article reported that a prolonged oral administration for a period of eight (8) weeks at the recommended therapeutic dosage did not affect hierarchies, body-mass or body condition [42]. The authors also concluded that these results suggest no obvious anabolic or behavioural effects from ALT that might affect the race performance of a mare. Contrastingly, a paper published by Gillon *et al.* (2021) investigated potential androgenic activity of ALT in equine urine and plasma by a recently developed androgen receptor bioassay (HEK293-AR) [51]. ALT's bioassay activity was directly compared to testosterone and trenbolone due to their structural similarities to ALT and their own high activity potency. The bioassay results proved ALT's ability to activate the androgen receptor, and the authors proposed its potential for inducing anabolic effects in the horse. However, there are some limitations to this study and the authors acknowledged that a targeted study on the potential performance enhancing effects of long-term ALT use in female horses has not yet been conducted and is necessary to complete this work. The pharmacokinetics of ALT in horses was also investigated by Machnik *et al.* (2007) via a ten (10) horse elimination study [45]. Equine urine and plasma samples, collected up to 15 days past initial administration, were analysed by LC-MS/MS to determine ALT concentrations. Results showed after five (5) daily oral therapeutic doses of ALT, the substance was detectable up to 72 hours in plasma after the last administration. The clearance of ALT in urine was not stated. The quantification of the steroid impurities in ALT has not yet been conducted.

Therefore, the aim of this research was to quantify the steroid impurities in equine urine from orally and intramuscularly injected (IM) administered ALT samples, and to identify if it was possible to distinguish between these two different administration routes using metabolomics, in complement with the current detection techniques.

1.3.2 Dopamine Markers

L-DOPA (Levodopa, l-3,4-dihydroxyphenylalanine) is used to treat the effects of Parkinson's disease (PD); a progressive, degenerative neurological condition that affects the control of body movements [52, 53]. In suitable doses, L-DOPA reduces the symptoms of PD as it can cross the blood-brain barrier to produce more dopamine where it is needed [52]. Since discovery of the benefits of L-DOPA on human PD patients, it is suspected to have been misused in equine sports to increase locomotor activity [53, 54]. The first known observation

of dopaminergic manipulation in the racing industry is assumed to have occurred in 1996, where five post-race urine samples were found to contain suspiciously high levels of 3-MT following GC-MS analysis at Racing Analytical Services Limited (RASL) in Victoria [55]. At this time, there was also growing concern over the advancements in L-DOPA containing, therapeutic medications, which had inadvertently enabled an increased accessibility of these products. Research by Vine and Wynne et al. into the potential causes of these elevated 3-MT levels led to investigations of precursor loading via tyrosine administrations (tyrosine > L-DOPA > dopamine) and possible dietary sources of L-DOPA, dopamine and Tyr (tyramine, 4-(2-aminoethyl) phenol) [56, 57]. Their work demonstrated the viability of tyrosine precursor loading to access the L-DOPA pathway and observed a correlation between the up-regulation of L-DOPA biosynthesis and the excretion of related metabolites, 3,4-dihydroxyphenylacetic acid (DOPAC) and 3-methoxy-4-hydroxyphenylacetic acid (HVA) [56]. However, these metabolites were proven unsuitable as indicative biomarkers of dopaminergic manipulation due to inhibited formation with dopamine elevation (regarding HVA), and unreliable elevation with the administration of dopaminergic agents (regarding DOPAC) [57].

To control this, the L-DOPA and a direct dopamine metabolite 3-MT (3-methoxytyramine, 4-(2- aminoethyl)-2-methoxyphenol) are monitored in equine urine samples for compliance below an international threshold of 4 µg/mL (4,000 ng/mL) [2, 54]. This threshold, however, is deliberately conservative, and so, may be insensitive to some doping practices, i.e., sub-threshold doping [12, 58].

Because of this issue, further research was conducted in 2020 by Elbourne *et al.* to propose additional measures for the detection of sub-threshold doping that had been suspected to be occurring [16]. A lower 3-MT limit of 0.776 µg/mL (776 ng/mL), in conjunction with a 3-MT-to-Tyr ratio (3-MT/Tyr) limit of 5.3 was proposed and published in 2022 [16]. These limits were calculated based on a reference population study conducted with almost 2,500 horse samples to acquire baseline levels of 3-MT and Tyr. This work was modelled off the study design previously presented and published by Wynne *et al.* [54-57, 59]. To mitigate the risk of manipulating endogenous levels, the development of an EBP biomarker/ratio (3-MT/Tyr) was proposed to enable the longitudinal monitoring of biomarkers to improve the detection of doping [12, 60, 61].

This new area of research investigated the administration of a human PD drug that includes L-DOPA, carbidopa ((2S)-3-(3,4-dihydroxyphenyl)-2-hydrazinyl-2-methylpropanoic acid) and entacapone ((E)-2-cyano-3- (3,4-dihydroxy-5-nitrophenyl)-N, N-diethylprop-2-enamide) to assess biomarkers capable of identifying sub-threshold manipulations (Figure 1-2). Targeted

analysis with 3-MT and Tyr was conducted, and untargeted metabolomic analysis was done with L-DOPA and dopamine related metabolites.

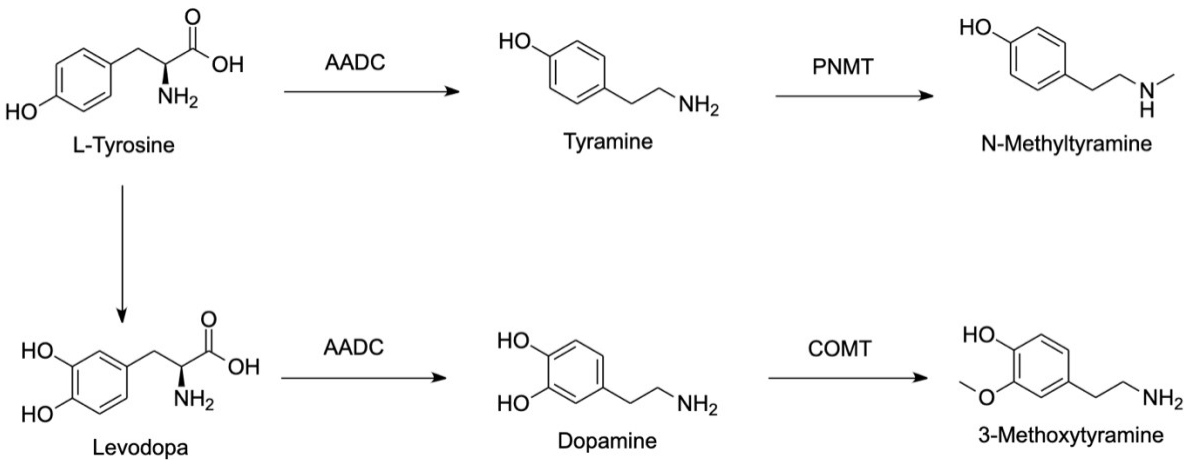


Figure 1-2: Compounds related to levodopa. AADC stands for aromatic L-amino acid decarboxylase, PNMT refers to phenylethanolamine N-methyltransferase and COMT is the abbreviated form of catechol-O-methyltransferase.

The addition of a catechol-O-methyltransferase (COMT) inhibitor, such as the synthetic compound entacapone, along with commonly administered carbidopa, stops L-DOPA from degrading peripherally, thereby slowing the decrease of L-DOPA in the bloodstream and reducing the formation of 3-MT to be excreted from the body. This dopamine to 3-MT pathway inhibition is of great concern to the current urinary 3-MT threshold performance for the detection of dopaminergic manipulation in racehorses. Previously published articles by Richards *et al.* (2014), McKinney *et al.* (2014), and Stanley *et al.* (2018) have voiced their concerns on this matter and the need for new biomarkers to circumvent this issue [58, 62, 63].

Previous research has looked into the abuse of COMT inhibitor compounds (nitecapone and tolcapone) in racehorses, and the excretion metabolites proceeding various COMT inhibitor compound (tolcapone, entacapone, nitecapone, and nebicapone) administration in rats, dogs, and humans [63-68]. However, research regarding the urinary effect of COMT inhibitors in combination with an aromatic L-amino acid decarboxylase (AADC) inhibitor (i.e., carbidopa) and L-DOPA in a PD therapeutic medication administered to racehorses is still lacking.

1.4 THESIS AIMS AND OUTCOMES

This thesis involves eight chapters concerning the direct quantification and metabolomic investigation of progestogenic and dopaminergic substance misuse in racehorses, and the implementation of data processing workflows and machine learning tools to improve biomarker discovery in an equine anti-doping context.

Chapter 2: The Current Applications and Future Potential Implementations of Machine Learning Strategies in Metabolomic-Based Anti-Doping Research

Chapter 2 reviews the advances in, and current use of, machine learning algorithms and models for metabolomic studies in the equine anti-doping and forensic toxicological space, and provides examples of where these models have been implemented. This will allow understanding for appropriate models used in proceeding chapters.

Part I: Altrenogest Administration Study

Chapter 3: Quantitative Analysis of Altrenogest Misuse

Chapter 3 aims to investigate the development and optimisation of a LC-MS/MS method for the urinary quantification of known steroid impurities (trendione, epitrenbolone and trenbolone) in the oral and IM administration study of the therapeutic synthetic progestin, ALT. This will provide a fit-for-purpose LC-MS/MS validated method for the quantification of steroid impurities in equine urine to be used in Chapter 4. Additionally, the results from this chapter provide an understanding of the difference in steroid impurity levels between administration routes and inform the methodology and analysis of Chapter 4 and Chapter 5.

Chapter 4: The Use of Progesterone as an Endogenous Reference Compound (ERC) for Altrenogest Administration Route Determination

Chapter 4 continues the investigation of the oral and IM administration study of altrenogest, quantifying urinary levels of endogenous progesterone. This chapter develops and optimises a profile classification machine learning model capable of differentiating between the administration routes involving the use of a biomarker ratio with progesterone acting as an ERC. This will provide confidence in the potential use of endogenous markers (metabolomics) in a multivariate approach to differentiate ALT administration routes in Chapter 5.

Chapter 5: Administration Route Differentiation of Altrenogest via the Metabolomic LC-HRMS Analysis of Equine Urine

Chapter 5 expands the investigation of the altrenogest administration study to a metabolomic workflow optimised for the analysis of intact conjugated steroids in their sulfate and glucuronide forms. This chapter also investigates the benefits of a classification model to differentiate between oral and IM administrations of altrenogest.

Part II: Levodopa Administration Study

Chapter 6: Quantitative Analysis of Dopaminergic Manipulation

Chapter 6 focuses on identifying levodopa misuse via the optimisation of a LC-HRMS method for the urinary quantification of 3-MT and Tyr in a *Stalevo*® (levodopa, carbidopa, entacapone) administration study. This provides a fit-for-purpose LC-HRMS validated method for the targeted detection of dopamine manipulation in equine urine, and reiterates the use of proposed limits from prior research to better detect sub-threshold doping. Additionally, the quantitative results of 3-MT and Tyr provides a better understanding of exogenous L-DOPA metabolism and informs potential target compounds in Chapter 7.

Chapter 7: Multivariate Approach for Dopaminergic Manipulation Detection

Chapter 7 continues the investigation of levodopa misuse using a metabolomics approach to search for more robust urinary biomarkers in their intact sulfate and glucuronide conjugated forms, in addition to the free metabolite. This chapter also details the potential for a machine learning model implemented to extend the detection window for dopaminergic manipulation.

This thesis investigates the relevance of urinary biomarkers to identify the presence and effect of progestogenic and dopaminergic substances that represent threats to the integrity of racing. These areas are considered two priority issues for Racing NSW and the complexity in their detection share common issues. Whilst the L-DOPA research was a natural progression of this PhD candidate, ALT was considered a higher priority to Racing NSW due to a lack of harmonisation for the detection of steroid impurities currently faced by the racing industry. Both areas focus on developing holistic metabolic profiles to identify biomarkers of prohibited administrations.

Chapter 2: The Current Applications and Future Potential Implementations of Machine Learning Strategies in Metabolomic-Based Anti-Doping Research

2 THE CURRENT APPLICATIONS AND FUTURE POTENTIAL IMPLEMENTATIONS OF MACHINE LEARNING STRATEGIES IN METABOLOMIC-BASED ANTI-DOPING RESEARCH

2.1 MACHINE LEARNING AND METABOLOMICS IN DOPING ANALYSIS

Artificial intelligence (AI) is a computer technology designed to mimic how the human brain thinks, with the desire to analyse large-scale data and process ever-growing complex problems. Machine learning (ML) is considered a subcategory within the AI umbrella, alongside computer vision, natural language processing, robotics and speech recognition [69]. ML is used to develop models that can deal with large-scale datasets and, through learning, can solve complex problems. AI and ML offer countless opportunities, with its various applications in understanding the structures or trends in vast amounts of data collected or generated from modern high-throughput experiments [70]. With the obvious grand potential available in ML, it has been adopted as a tool to tackle the processing and detailed analysis of large high-throughput datasets. Already increasing in popularity in the peripheral clinical and environmental fields, ML is also beginning to pick up traction in the forensic toxicological space, particularly regarding metabolomics and anti-doping research.

Metabolomics is the study of small molecules for the qualitative and quantitative investigation of metabolite changes in the body that are caused by a stimulus, for example, the administration of an illegal substance [71]. Recent advancements in the analytical techniques for the metabolomic analysis of biological samples has enabled the collection of increasingly large amounts of data and information. The tools required for the investigation of this data have therefore, needed to keep up in order to obtain meaningful interpretations of the results. ML has the capability of automating the data processing, analysis and interpretation of this data into meaningful research findings. Once tested and validated for robustness and reliability they may be translated into routine processes to improve current detection methods.

Whilst there are many ML tools available, the missing link is the education and understanding of what each algorithm is suited for, and which model is appropriate for specific datasets toward the desired outcome. This gap between a foreign concept and understanding AI and ML has so far limited its widespread implementation and potential advancements in this field. Bridging this gap is the first step to improving anti-doping research and enabling the possibility for new discoveries to be made as a result. This review details a guide on commonly used ML

models and how they can be implemented to advance anti-doping research and peripheral fields.

2.2 Types of Machine Learning Tools and How they could be implemented

ML algorithms commonly used in metabolomic research can be classified by three main categories, which are divided based on how the model interprets and learns from the provided dataset. These categories are supervised learning, unsupervised learning, and semi-supervised learning.

2.2.1 Supervised Learning

Supervised learning algorithms involve training a ML model on a labelled dataset, where each data point is paired with a corresponding label or output value [70]. The model learns to map the input data to the desired output, allowing it to make predictions for new, previously unseen data. The goal of this ML style is to predict outcomes from new data based on learned relationships from the labelled data fed to the model. This type of model is commonly used for classification and regression type problems, where data is often categorised or labelled and an output value is known.

Common supervised learning model types used in metabolomic investigations include linear regression, logistic regression (LR), decision trees (DT), random forests (RF), and support vector machines (SVM). These models are selected based on the type of dataset being trained and the desired outcome from the model.

2.2.1.1 Linear Regression

Linear regression models the relationship between dependent and independent variables by fitting a straight line (linear equation) to the observed data [72]. Predictions based on linear regression are relatively straight forward, where predictions are made on the basis of an observed data trend [73, 74]. Some assumptions apply when using a simple linear regression model, including, the linearity of the dataset, observational independence, constant error variance, and normal distribution [75]. Table 2-1 details the main advantages and disadvantages of a linear regression model.

Table 2-1: Advantages and disadvantages for the use of a linear regression machine learning model.

Advantages:	Disadvantages:
• Easily implemented and interpretable • Computationally efficient • Well-established algorithm	• Assumes linear relationship • Sensitive to multicollinearity • Assumes feature suitability to the model • Susceptible to overfitting and underfitting • Limited explanatory power for complex variable relationships

2.2.1.2 Logistic Regression (LR)

LR is a statistical model used to predict a binary outcome (one scenario out of two possible alternatives) based on a set of independent variables (those that influence the outcome) using a logarithmic odds scale [76]. Logistic regression categorises data into discrete classes by studying the relationship from a given set of labelled data. Logistic regression can be classified into three main types; binomial, multinomial, or ordinal, dependent on the dataset variables. Table 2-2 details the main advantages and disadvantages of a LR model.

Table 2-2: Advantages and disadvantages for the use of a logistic regression (LR) machine learning model.

Advantages:	Disadvantages:
• Easily implemented and interpretable • No assumptions made about class distribution • Easily extend to multiple classes (multinomial regression) • Fast to classify unknown data • No feature scaling required • No hyper-parameter tuning needed	• Susceptible to overfitting • Assumption of linearity between dependent and independent variables • Can only be used to predict discrete functions • Requires average or no multicollinearity between independent variables • Difficult to obtain complex relationships

	• Needs linearly related independent variables

2.2.1.3 Decision Trees (DT)

DT is a statistical decision support tool that uses a tree-like model of decisions and possible consequences in the model. Each tree is similar in structure to that of a flowchart. Each node represents a test, each subsequent branch represents the outcome of the test, and each leaf node represents a class label [77, 78]. Decision trees consist of three types of nodes: decision, chance, and end nodes [78]. Table 2-3 details the main advantages and disadvantages of a DT model.

Table 2-3: Advantages and disadvantages for the use of a decision tree (DT) machine learning model.

Advantages:	**Disadvantages:**
• Easily implemented and interpretable • Handles both numerical and categorical data without extensive pre-processing • Provides insights into feature importance for decision-making • Handles missing values and outliers • No feature scaling or normalisation required • Easy visualisation • Automatic feature selection • Applicable to classification and regression tasks	• Requires large datasets for quality results • Susceptible to overfitting • Sensitive to small changes in data • Potential bias with imbalanced data • Requires extensive training time

2.2.1.4 Random Forest (RF)

RF is a statistical classification method composed of an assembly of many decision trees constructed during the training phase. This model is sometimes also categorised as an ensemble learning model, which is denoted based on its used of a bagging technique, differing from a typical DT model. This model generally outperforms DT as it corrects the observed

overfitting. New objects are classified based on the attributes of the data. Each tree is classified and gives a vote for the chosen attribute. When used for classification, the classification with the most votes is chosen, and when used for regression purposes, the average votes are used [79, 80]. Table 2-4 details the main advantages and disadvantages of a RF model.

Table 2-4: Advantages and disadvantages for the use of a random forest (RF) machine learning model.

Advantages:	Disadvantages:
• High predictive accuracy • Generally resistant to overfitting (however, not impossible) • Easy handling of large datasets • Feature importance extraction • Applicable to classification and regression tasks • Can handle imbalanced or missing data	• Requires computational complexity • Extensive memory usage • Longer prediction times necessary • Lack of interpretability (dependent/independent variables) • Uncorrelated predictions • Sensitive to changes in data

2.2.1.5 Support Vector Machines (SVM)

SVM is a popular classification tool in many fields, where training observations are split into two classes by constructing a hyperplane, which is a decision boundary that separates the data points [81]. The distance between the hyperplane and the nearest data points of each class is called the margin, and the points onto which this margin hits are called the “support vectors”. Table 2-5 details the main advantages and disadvantages of an SVM model.

Table 2-5: Advantages and disadvantages for the use of a support vector machine (SVM) machine learning model.

Advantages:	Disadvantages:
• High-dimensional performance • Non-linear compatibility • Easily adaptable • Not easily biased by outliers • Not overly sensitive to overfitting	• Slow training time • Parameter tuning difficulty • Sensitive to noisy datasets • Limited interpretability • Feature scaling sensitivity

• Applicable to classification and regression tasks

2.2.1.6 Naïve Bayes (NB)

NB modelling use Bayes' Theorem to classify data based on the probabilities of different classes given the features of the data. There are three main types of an NB model; Gaussian – assumes normal distribution, Multinomial – features represent frequency, and Bernoulli – suited for binary features and classification tasks. Table 2-6 details the main advantages and disadvantages of an NB model.

Table 2-6: Advantages and disadvantages for the use of a Naïve Bayes (NB) machine learning model.

Advantages:	**Disadvantages:**
• Easy to implement and computationally efficient • Effective in cases with a large number of features • Performs well even with limited training data • It performs well in the presence of categorical features • For numerical features data is assumed to come from normal distributions	• Assumes that features are independent, which may not always hold in real-world data • Can be influenced by irrelevant attributes • May assign zero probability to unseen events, leading to poor generalisation

2.2.1.7 K-nearest Neighbours (K-NN)

K-NN modelling tackles classification and regression problems by finding the K nearest neighbours to a given data point based on a distance (or similarity) metric. The three commonly used distance metrics are; Euclidean distance, Manhattan distance and Minkowski distance [82]. The class or value of the data point is then determined by the majority vote or average of the K neighbours. This approach allows the algorithm to adapt to different patterns and make predictions based on the local structure of the data. Table 2-7 details the main advantages and disadvantages of a K-NN model.

Table 2-7: Advantages and disadvantages for the use of a K-nearest neighbours (K-NN) machine learning model.

Advantages:	Disadvantages:
• Easily implemented • Adapts smoothly to new datapoints • Minimal hyperparameter tuning required	• Lacks scalability • Unable to handle high dimensionality data • Susceptible to overfitting

2.2.1.8 Partial Least-Squares Discriminant Analysis (PLS-DA)

PLS-DA algorithms model the relationship between independent and dependent variables, by linking two data matrices X (i.e., raw data) and Y (i.e., groups, class). This model works to maximize the covariance between the independent variables X and the corresponding dependent variable Y of highly multidimensional data by finding a linear subspace of the explanatory variables [83]. Table 2-8 details the main advantages and disadvantages of a PLS-DA model [83].

Table 2-8: Advantages and disadvantages for the use of a partial least-squares discriminant analysis (PLS-DA) machine learning model.

Advantages:	Disadvantages:
• Easily interpretable • Handles highly collinear and noisy data • Variable Importance Projection identification (VIP scores) • Handles high dimensionality data • Robust classifier	• Requires a large dataset to achieve predictive accuracy • Requires model validation • Susceptible to overfitting • Predictive accuracy statistics (R^2 and Q^2) not reliable in classification model • Poor handling of missing values

Evaluation metrics are used to test a model's performance and can be utilised in the optimisation of the chosen algorithm. Some of the most common and useful evaluation metrics are listed and briefly explained in Table 2-9.

Table 2-9: A tabulated list of common evaluation metrics, with a brief explanation and related equations.

Evaluation Metric	Equations	Definition
Mean Square Error (MSE)	$MSE = \frac{\sum_{i=1}^{n}(y_{i,true} - y_{i,predicted})^2}{n}$	calculates the average of the squared differences between the actual and predicted values for all the data points
Mean Absolute Error (MAE)	$MAE = \frac{\sum_{i=1}^{n}\lvert y_{i,true} - y_{i,predicted}\rvert}{n}$	calculates the accuracy of a regression model. measures the average absolute difference between the predicted values and actual values
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{i,true} - y_{i,predicted})^2}{n}}$	how well the observed data points match the expected values, or the model's absolute fit to the data
Coefficient of Determination (R^2)	$R^2 = 1 - (\frac{RSS}{TSS})$ $RSS = \sum_{i=1}^{n}(y_{i,predicted} - \bar{y}_{true})^2$ The sum of squared differences between predicted data points (y_i) and the mean of the response variable ($\bar{y}$) $TSS = \sum_{i=1}^{n}(y_{i,true} - \bar{y}_{true})^2$ The sum of squared differences between individual data points (y_i) and the mean of the response variable ($\bar{y}$)	measure of the proportion of variance in the dependent variable that is explained the independent variables in the model
Accuracy	$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$	the proportion of correctly classified instances
Precision	$Precision = \frac{TP}{TP + FP}$	the accuracy of positive predictions

Recall **(Sensitivity or True Positive Rate-TPR)**	$Recall = \frac{TP}{TP + FN}$	the proportion of correctly predicted positive instances among all actual positive instances
F1 Score	$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$	the harmonic mean of precision and recall
Area Under the Curve – Receiver Operator Characteristics **(AUC-ROC)**	TPR v FPR	measures the area under the ROC curve, providing an aggregate measure of a model's performance across different classification thresholds
Area Under the Curve – Precision Recall **(PR-AUC)**	Precision v Recall	measures the area under the precision-recall curve, providing a summary of a model's performance across different precision-recall trade-offs

Table Definitions: TP is true positive, TN is true negative, FP is false positive, and FN is false negative. TPR is true positive rate, FPR is false positive rate.

Some model packages, run through Python or R scripts, have limited metrics built in to assist in model optimisation, inclusive of those listed, in addition to feature importance ranking. Feature importance ranking is another tool to measure and extract meaningful features from the dataset. This function acts to aid model interpretation for new uses.

2.2.2 Unsupervised Learning

In unsupervised ML, the algorithm learns patterns from unlabelled data. The algorithm takes a dataset with only inputs and attempts to find a structure in the data by grouping or clustering the data points [80]. In comparison to supervised learning, unsupervised models deal with unlabelled data, where the data points do not have associated labels or output values. The goal is to uncover hidden patterns, relationships or structures inherent in the data itself without the guidance of labelling. Common algorithm uses include dimensionality reduction, clustering and association models. Two models popular in metabolomic studies include principal component analysis (PCA), and *k*-means clustering, which are explained in further detail below.

2.2.2.1 Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique used to examine the interrelations among a set of variables. It does this whilst preserving the most important patterns or relationships between variables without any prior knowledge of target variables. PCA models are often considered the unsupervised version of the PLS-DA model, as they both tackle dimensionality reduction. Table 2-10 details the main advantages and disadvantages of a PCA model.

Table 2-10: Advantages and disadvantages for the use of a principal component analysis (PCA) machine learning model.

Advantages:	Disadvantages:
• Feature selection • Data visualisation • Multicollinearity • Noise reduction • Data compression • Outlier detection	• Difficult interpretation and explanation of principal components • Appropriate data scaling required • Potential for information loss • Assumes linear relationships • Requires computational complexity • Susceptible to data overfitting

2.2.2.2 K-means Clustering

k-means clustering uses unlabelled, unclassified data and assigns each datapoint to a cluster, that is dependent on their distance from the centre of the clusters. This sorting is based on patterns and variations found in the datapoints. The number of clusters in the model should be provided in advance [84]. Table 2-11 details the main advantages and disadvantages of a *k*-means clustering model.

Table 2-11: Advantages and disadvantages for the use of a k-means clustering machine learning model.

Advantages:	Disadvantages:
• Easily implemented • Computationally efficient • Able to handle large datasets • Quickly identifies trends, relationships and outliers • Application versatility	• Difficulty processing high-dimensional data • Need to specify clusters beforehand • Sensitive to cluster centroid positions

	• Assumptions made on cluster shape and size • Not suitable for missing or categorical data • Limited interpretability

2.2.2.3 Hierarchical Cluster Analysis (HCA)

Similar to *k*-means clustering, HCA is a connectivity-based clustering model that groups the data points together that are close to each other based on the measure of similarity or distance. However, unlike *k*-means, this model does not require a pre-specified number of clusters and instead builds an unfixed hierarchy of clusters. HCA methods exist in two forms; agglomerative or divisive, either the bottom-up approach starting with a large number of clusters and work to combine similar, or a top-down approach starting with one cluster and splitting clusters based on datapoint variables, respectively. Table 2-12 details the main advantages and disadvantages of an HCA model.

Table 2-12: Advantages and disadvantages for the use of a hierarchical cluster analysis (HCA) machine learning model.

Advantages:	**Disadvantages:**
• Easy implementation • Scalability • Handling non-linear datasets • Flexible to dataset typing • Multi-level output	• Sensitive to outliers • Requires computational complexity • Potentially difficult to interpret • Reliant on user's experience

2.2.2.4 Deep Learning (DL)

Deep Learning is a subset of ML that use artificial neural networks (ANN). These models construct complex relationships by connecting input variables to the outcome, going beyond standard ML classification techniques. DL algorithms, like ANN, use a layered architecture, composed of several layers of 'neurons', consisting of an 'input' layer, 'output' layer and intermediate 'hidden' layers. These layers link inputs to the desired output using manipulated input variables. The outputs of hidden layers are linearly combined and often passed through a classifier function to produce a singular output value. This model style is particularly useful

at tackling large-scale datasets with high dimensionality. Table 2-13 details the main advantages and disadvantages of a DL model.

Table 2-13: Advantages and disadvantages for the use of a deep learning (DL) machine learning model.

Advantages:	Disadvantages:
• Can handle large scale datasets • High-dimensionality reduction • Adaptive structure • Automatic feature learning • Handling non-linear relationships • Capable of handling structured and unstructured data • Handles missing data • Scalability	• Requires computational complexity • Lack of interpretability • Difficult to evaluate model • Minimal model control • Susceptible to overfitting • Dependent on data quality

2.2.3 Semi-Supervised Learning

Semi-supervised learning falls between unsupervised and supervised learning, combining a small amount of labelled data with a large amount of unlabelled data during the training process and using context to identify data patterns [80]. The goal of semi-supervised learning is to learn a function that can accurately predict the output variable based on the input variables, similar to supervised learning. However, unlike supervised learning, the algorithm is trained on a dataset that contains a combination of both labelled and unlabelled data.

Semi-supervised learning models may be implemented when regression, classification, clustering or dimensionality reduction is desired, and are particularly useful when there is a large amount of unlabelled data available, but it is difficult to label all of it. This type of learning is not seen as frequently in metabolomic studies compared to its supervised and unsupervised counterparts.

Table 2-14: Common examples of supervised, unsupervised, and semi-supervised machine learning models.

Supervised Learning	Unsupervised Learning
• Linear Regression • Logistic Regression (LR) • Decision Trees (DT) • Naïve Bayes (NB) • K-nearest Neighbours (KNN) • Support Vector Machine (SVM) • Random Forests (RF) • Partial Least Squares (PLS)	• Principal Component Analysis (PCA) • *k*-means Clustering • Hierarchical Cluster Analysis (HCA) • Deep Learning (DL) ○ Artificial Neural Network (ANN)

Semi-supervised Learning
• Support Vector Machine (SVM) • Artificial Neural Network (ANN) • Classification, Clustering, Regression, and Dimensionality Reduction algorithms

Choosing between supervised and unsupervised learning depends on the nature of the data and the specific problem to solve (Table 2-14, Figure 2-1). Supervised learning is suitable for tasks where the desired output is known, while unsupervised learning is suitable for tasks where the goal is to explore and understand the underlying structure of the data.

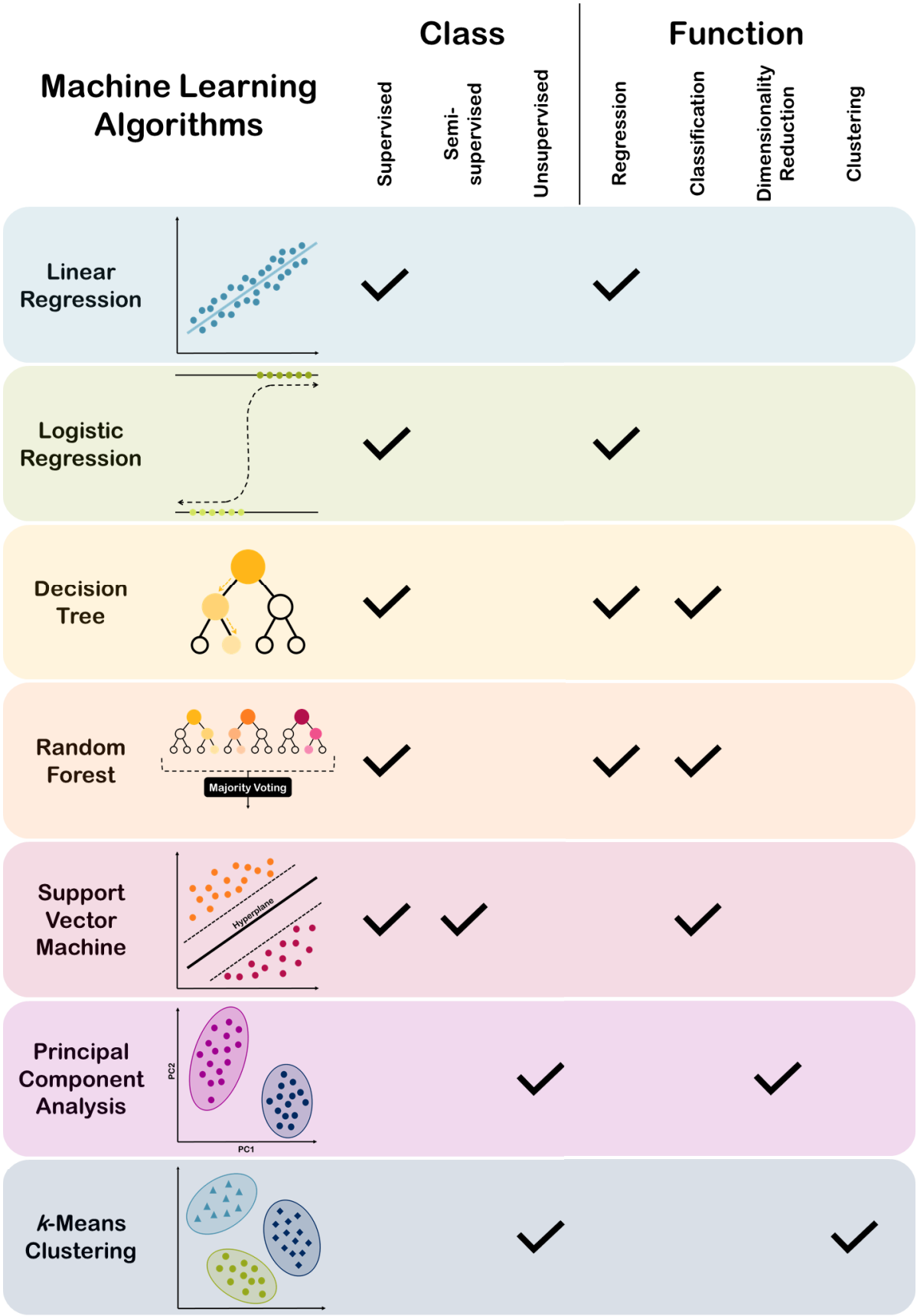


Figure 2-1: Commonly implemented machine learning (ML) algorithms in metabolomics tabulated with ML classes (supervised, semi-supervised, and unsupervised learning) and ML function (regression, classification, dimensionality reduction, and clustering) categorised.

2.3 Doping in Human and Animal Sporting Industries

Doping is a term referring to the use of illicit substances of misuse of therapeutic substances by athletes. This is typically for the desired effect of performance enhancement or performance suppression in a sporting event. The act of doping causes an unfair advantage to those involved, and as a whole, threatens the health of the athlete and integrity of the sporting industry [1]. This issue does not exist in a vacuum and its flow-on effects should also be acknowledged when company endorsements or gambling is involved with the athlete and sporting event. When large amounts of money are involved with these events, companies who have sponsored may feel their brand is at risk by association, and individuals who have placed a bet may feel cheated of their 'rightful' winnings. These implications are additional considerations as to the unlawful and immoral nature of doping [85]. Thus, the main goal of anti-doping research is to protect the welfare of the athlete and the integrity of the sport.

There are currently two main detection methods undertaken for detecting illegal drug use, which are the direct detection of the administered parent drug, or alternatively, the targeted detection through a direct metabolite of such. Current practices in anti-doping detection align with targeting known drugs or direct metabolites of such through time-consuming targeted qualitative and/or semi-quantitative screening analysis. The current detection methods are advantageous in a number of scenarios/detection types; namely, when identifying exogenous substance doping, when the parent drug or direct metabolite has a relatively long half-life in the system prior to metabolism and excretion, and when the drug is present at high levels in the body. These could be considered best case scenario detections. There are, however, some limitations to this current process; namely, the detection of endogenous compound doping is increasingly difficult, as well as drugs that are rapidly metabolised and quickly excreted from the body. Additionally, lower concentration compound detection, i.e., micro-dosing or sub-threshold dosing is increasingly difficult with these techniques. Therefore, there is a need for a more advanced analysis process to address these limitations. The implementation of metabolomic analysis workflows that incorporate ML technologies could provide a solution to many of the current limitations listed above. This is mainly due to high-powered processing capabilities and potential for greater biological interpretation of findings.

2.4 Current Applications and Future Potential Implementation of Machine Learning in Anti-Doping Research (or related fields)

With the benefits and the increased usage of ML gaining a reputation in analytical sciences, some notable publications and resources have followed, detailing their recent applications,

future possibilities, and current limitations [17, 86-95]. In an effort to acknowledge what has already been written and to avoid repeating existing reviews, this section will summarise the findings of selected published reviews and articles.

2.4.1 Doping Detection in Athletes

In 2024, Yang *et al.* conducted a thorough review on the advantages of ML in doping detection with some examples of potential applications [96]. These four areas of improvement were regarding ML implementation in sample screening, automated doping recognition, the identification of doping metabolites, and indirect doping detection [96]. The potential integration of classification models into primary sample screening were proposed to reduce the time-consuming process of flagging suspicious samples for confirmatory testing. The examples referenced tested an individual-, or combination of PCA, LR, NB, SVM, K-NN, DT, RF, and ANN models, highlighting the potential of a supplementary screening tool in addition to conventional methods, with a reduction in time, costs and resources [97-99]. The development of an automated method for doping-type recognition based on spectral information, as opposed to the present heavy reliance on available reference standard substances was proposed to reduce detection process time and resources. Examples included the investigation of K-NN, SVM, RF, ANN, HCA and PCA models, among others, with impressive results [100-102]. Implementing ML to process and interpret high-throughput untargeted metabolomic information for the identification of doping-related metabolites, was proposed to assist in biomarker discovery, thereby reducing the likelihood of false positive identifications and the extensive time-consuming component of untargeted data analysis. Referenced examples involved the use of PCA, PLS, HCA, LR, RF, K-NN, ANN and SVM, among others, using evaluation metrics to determine model performance when comparing outcomes [103-105].

The last potential application focused on the advantages of implementing ML for the indirect detection of doping, particularly when distinguishing endogenous and exogenous sources of a substance in the body. Particular examples of interest, due to their endogenous natures, of substances banned and controlled by WADA were AAS and endogenous AAS (EAAS), as well as erythropoietin (EPO), (also including recombinant human EPO (rhEPO) and hypoxia-inducible factor (HIF) activators) commonly referred to as blood doping [106-109]. There is potential for ML models to improve the current ABP profiling system to increase its sensitivity and improve on the current biomarker selection, for better longitudinal profiling outcomes [9]. One example tested a combination of regression and classification models for the detection of blood doping, with a gaussian process regression model exhibiting the best performance [110].

In 2015, Chan *et al.* reported using a Orthogonal PLS-DA (OPLS-DA) model in a targeted metabolomic approach to detect the misuse of steroidal aromatase inhibitors in equine sports by biomarker profiling [111]. Urinary concentrations of 31 endogenous steroid compounds were obtained from two (2) resting and two (2) in-training gelded horses, treated with a steroidal aromatase inhibitor (either, androst-4-ene-3,6,17-trione (6-OXO) or androsta-1,4,6-triene-3,17-dione (ATD)), in addition to control samples taken from in-training horses (n = 28). OPLS-DA analysis highlighted seven (7) potential biomarkers that could be used to differentiate urine samples obtained from the control and the treated groups. A weighted equation was established with these biomarkers for screening purposes (screening criterion) and a criterion threshold calculated. On the basis of the OPLS-DA model and the resulting weighted equation, the administration of 6-OXO and ATD could be detected for up to four (4) and nine (9) days (equating to 2.1 times and 2.5 times longer detection periods compared to direct detection of metabolites), respectively.

Ohnuma *et al.* (2021) established a post-race biomarkers database and applied pathway analysis to assist in the identification of potential biomarkers in post-race equine plasma from comparison to resting equine samples [112]. Using MetaboAnalyst 4.0 for statistical analysis, pathway analysis, and Receiver Operating Characteristics (ROC) analysis, this article proposed two (2) lipid-based biomarker ratios (pentadecenoyl carnitine/ galactosylglycerol, P/G, and heptadecenoyl carnitine/ galactosylglycerol, H/G) that could differentiate between rest and post-race sample groups. The monitoring of these biomarkers was proposed to differentiate between true inter-/ intra-individual variances in the physiological status, and doping agents use in the horse, via up- or down-regulation. This work was also validated with a single-blind test to demonstrate the post-race biomarkers database applicability, and was proven as a strategy that could be extended further to implement identification of genuine doping samples for the identification of doping biomarkers [112].

2.4.2 Clinical Toxicology and Disease

Tan *et al*. also published an investigation of AI used for urine analysis in a clinical medicine application, with a focus on disease diagnosis and treatment. The article listed the most commonly used algorithms as DT, SVM, RF, ANN and DL, and identified SVM and RF as most prominent in clinical medicine research given their benefits in AI-assisted diagnosis technologies [69]. This paper focuses on the potential clinical applications for ML to improve current diagnosis systems. Example areas given included AI-assisted diagnosis tools to help reduce the risk of misdiagnosis, which is of particular concern in locations with poor medical systems in services. Another was the commonly acknowledged benefit of ML improving the efficiency and accuracy of processing large amounts of data through automation. This is

especially relevant to urine sample analysis due to the characteristic large volume, simple non-invasive collection, and information rich nature of urine. The authors also state the lack of evaluation standard for AI regarding effectiveness, clinical applicability and safety/patient privacy is a major concern that is still in need of being addressed [69].

Galal *et al.* reviewed the current usage of ML tools for disease modelling and classification in clinical toxicology and found that ML techniques have been used recently in the metabolomic analysis of numerous diseases which were categorised into “cancer” and “non-cancer conditions” [70]. Their review of ML applications for classification in metabolomic studies also agreed that RF and SVM were among the most popular algorithms used in metabolomic studies, with DL slowly increasing in use as datasets grow more complex and higher computing power becomes more readily available. The importance of selecting the most appropriate model for specific studies was emphasised, while reiterating the need for researchers to be educated in the benefits, limitations and suitability of common ML approaches [70]. Another article, by Shabana *et al.* (2020), proposed the use of ML models in the classification of appropriate drug administration route (i.e., oral, parenteral (injection), and topical) in pre-clinical trialling [113]. The tested models included LG, RF and DT, and they were trained with 21 unique pharmacological characteristic features. Results found that the LR model performed well on 21 features, however, RF outperformed this, only requiring five (5) features to achieve a predictive accuracy score of 97%. This work proved the implementation of AI techniques viable in the pre-clinical space, suggesting that hybrid techniques and further hyper tuning would further increase the classification accuracy of the proposed model [113].

2.4.3 Forensic Toxicology and Metabolomics

Steuer *et al.* posed the question of “where are we heading?” regarding untargeted metabolomics approaches in the clinical and forensic toxicological fields [71]. The review detailed examples of metabolomic analysis benefiting routine toxicological screening methods, often incorporating a ML model for biomarker discovery in samples with classification of treated/non-treated. Since 2015, Misra *et al.* has published a comprehensive overview of available metabolomic-based data processing and statistical visualisation tools and resources [114-118]. These programs can be implemented into an analysis workflow as a stand-alone system or in a multi-tool pipeline for high-throughput data analyses.

2.5 CHALLENGES AND LIMITATIONS OF ML IMPLEMENTATION

2.5.1 Interpretability and Transparency of ML Models

The interpretability of ML processes and results can often be complicated and become a shortcoming of its use. In response to this, “interpretable AI” has gained popularity as a potential solution. Its use has been showcased in research published by Zhang *et al.*, in the related clinical metabolomics field for its use in biomarker discovery for PD [119]. This study used an interpretable neural network (NN) framework to accurately predict disease and identify significant biomarkers using whole metabolomics data sets without a *priori* feature selection. The performance of the NN approach for predicting PD from blood plasma metabolomics data was significantly higher than other ML methods with a mean area under the curve of >0.995. PD-specific markers that predate clinical PD diagnosis and contribute significantly to early disease prediction were identified including an exogenous polyfluoroalkyl substance. The authors also inferred from this research that this approach could improve diagnostic performance for many diseases using metabolomics and other untargeted ‘omics methods [119].

The benefits and limitations of commercially available programs against open-source applications have also been compared in terms of their transparency in prior studies [71]. The general consensus is that while commercially available tools have a lower barrier to entry in their usability, they often lack transparency in the functionality and can be restricted in the customisation of models. Alternatively, open-source and freely available applications allow greater transparency in their workflows and offer a wide range of freedom in feature/parameter customisation. The limitation here arises with the knowledge required to operate such programs [71].

2.5.2 Validation of ML Algorithms for Anti-Doping Purposes

Whilst the use of AI and ML in analytical analysis has grown in popularity, there remains a lack of harmonisation regarding appropriate usage and the reporting of results. However, some guidelines have been proposed by corporations and individuals as a useful starting point to follow [92, 120, 121]. This can ensure appropriate methodology is used with consistency when implementing ML models and publishing related results.

2.6 CONCLUSION

Selecting which ML model is appropriate for specific research applications will depend on the problem to be solved and nature of data generated. Understanding what algorithms are

available and how they work is the first challenge for implementing ML into anti-doping research. The use of ML tools has the potential to improve the quality and depth of anti-doping research, in a complementary format to the current detection systems. The harmonisation of ML methods and guidelines for best practice are crucial for future work. The collaboration of researchers, sports organisations, and regulatory bodies should help in advancing anti-doping efforts with the assistance of the wide range of ML tools.

Part I: Altrenogest Administration Study

Chapter 3: Quantitative Analysis of Altrenogest Misuse

This chapter investigates the development, optimisation and validation of a (LC-MS/MS) method for the quantification of known urinary steroid impurities (trendione, epitrenbolone and trenbolone) in the oral and intramuscularly injected (IM) administration study of the therapeutic synthetic progestin, altrenogest (ALT).

The quantitative analysis is taken from a first author publication entitled “Metabolomic analysis of equine urine to determine administration route of altrenogest’’ published in the 23rd International Conference of Racing Analysts and Veterinarians (ICRAV) Conference Proceedings. The preparation of the initial draft was performed by M. Elbourne with manuscript edits provided by A. Cawley and S. Fu. The metabolomic analysis component was not included, as a dedicated metabolomic analysis chapter is presented in **Chapter 5**. Altrenogest product analysis is also included in **Chapter 3 Section 3.3.2**, which was not included in the initial publication of this work.

Chapter 4: The Use of Progesterone as an Endogenous Reference Compound (ERC) for Altrenogest Administration Route Determination

This chapter continues the previous chapter’s quantitative analysis of the ALT administration study and focuses on an investigation in the use of progesterone as an endogenous reference compound (ERC), along with trendione and epitrenbolone as ratios to differentiate oral administration from IM administration.

Chapter 5: Administration Route Differentiation of Altrenogest via the Metabolomic LC-HRMS Analysis of Equine Urine

This chapter expands the investigation of the altrenogest administration study to a metabolomic workflow optimised for the analysis of intact conjugated steroids in their sulfate

and glucuronide forms. This chapter also investigates the benefits of a classification model to differentiate between oral and IM administrations of altrenogest.

Components of this chapter were published in a first author publication entitled “Administration Route Differentiation of Altrenogest via the Metabolomic LC-HRMS Analysis of Equine Urine’’ in the journal *Moleculesˆ*. The preparation of the initial draft was performed by M. Elbourne with manuscript edits provided by A. Cawley and S. Fu.

At time of publishing, compounds 2-methoxyestradiol sulfate (2-ME2S) and cortisol sulfate (CS) still required synthesis to complete the verification of compound identification. This was subsequently completed and included in the chapter.

Chapter 3: Quantitative Analysis of Altrenogest Misuse

3 QUANTITATIVE ANALYSIS OF ALTRENOGEST MISUSE

3.1 INTRODUCTION

When ALT is administered, prohibited steroid impurities accumulated during commercial preparation of ALT, trendione (as major impurity), trenbolone (minor), and epitrenbolone (from metabolism) can be detected [42, 45]. These steroids, along with many other AAS, have performance limits, for example epitrenbolone at 1 ng/mL, suggested by the IFHA and enforced in the Australian equine racing industry [122]. Current Racing NSW ruling (at time of submission) dictates the allowed use of ALT must occur outside of one clear day of any racing, and administration permitted, orally, to female horses only [48, 49].

The development of a complementary approach allows for the potential to improve the detection of IM ALT, using features differing from oral ALT, such as an untargeted analysis to identify possible metabolomic features that may be influenced by an injected administration. Abnormal levels of endogenous substances can, therefore, provide an extended window of detection and give rise to the possibility of a biomarker or biomarker ratio to distinguish between oral and IM administration [60].

This research focused on the quantification of steroid impurities in mares administered with ALT via LLE of hydrolysed urine samples and LC-MS/MS analysis. A targeted approach was utilised to monitor urinary levels of AAS impurities; trendione, epitrenbolone, and trenbolone. Due to the Racing NSW ruling regarding the allowed use of oral ALT and prohibited use of IM ALT, the primary aim of this chapter was to distinguish oral administration from the IM route by the quantification of trendione, trenbolone, and epitrenbolone.

3.2 MATERIALS AND METHODS

3.2.1 Chemicals and Reagents

Acetonitrile (ACN), ammonium formate (99.995 %), anhydrous sodium sulfate salt (Na_2SO_4), dichloromethane (DCM), diethylamine (DEA), ethyl acetate (EtOAc), formic acid (FA), methanol (MeOH) and sodium hydroxide (NaOH) of LC-MS grade were purchased from Merck (Kilsyth, Australia). Aqueous ammonia solution was purchased from Chem-Supply (Gillman, Australia). Potassium di-hydrogen phosphate (KH_2PO_4) was purchased from LabServ provided by ThermoScientific (North Ryde, NSW, Australia). Di-sodium hydrogen phosphate (Na_2HPO_4) and sodium acetate (CH_3COONa) were obtained from Sigma Aldrich (Castle Hill,

NSW, Australia). The water used was ultrapure grade (18.2 MW.Ω.cm) obtained from a ThermoFisher Barnstead Smart2Pure system (ThermoScientific; Langenselbold, Hungary). The β-glucuronidase enzyme from abalone (BG100) was manufactured by Kura Biotech (Puerto Varas, Chile) and supplied by PM Separations (Capalaba, Queensland, Australia).

Phosphate buffer (0.1 M, pH 5.0) was prepared by dissolving 54.4 g of potassium di-hydrogen phosphate in approximately 3.0 L of purified water and pH adjusted to 5.0 with 1.0 M potassium hydroxide or orthophosphoric acid 85%. This was then diluted again to 4.0 L with purified water and mixed. Sodium acetate buffer (0.20 M, pH 5.0) was prepared by dissolving 16.4 g sodium acetate in approximately 800 mL of purified water, and adjusted to pH 5.0 ± 0.1 using 10.0 M sodium hydroxide and/or concentrated acetic acid. The solution was then diluted to 1.0 L with purified water and mixed. Sodium phosphate buffer (0.25 M, pH 9.5) was prepared by dissolving 30.0 g of di-sodium hydrogen phosphate in approximately 900 mL of purified water and pH adjusted to 9.5 ± 0.2 with 10.0 M sodium hydroxide or orthophosphoric acid 85%. This was then diluted to 1.0 L with purified water and mixed.

3.2.1.1 Reference Standards

Reference standards trendione (Td), trenbolone (Tr), epitrenbolone (Etr), and progesterone (Pr) were purchased from Sigma Aldrich (Castle Hill, NSW, Australia). Internal reference standards (ISTD) trenbolone-d_5, and progesterone-d_9 were purchased from Toronto Research Chemicals (Ontario, Canada).

3.2.1.2 Altrenogest Products

Small portions of *Readyserve*® Altrenogest oral solution (2.2 mg/mL) and *Readyserve*® Altrenogest injection solution (50 mg/mL) products used in the administration study were provided by Glenys Noble and Jaymie Loy from Charles Sturt University, Wagga Wagga, Australia for analysis (assumed supplier is Ceva Animal Health Pty Ltd, Glenorie, Australia).

3.2.2 Administration Study

The oral and IM ALT products used in the administration studies were analysed to determine the approximate percentage (%v/v) of steroid impurities present in the commercial preparations in comparison to reported altrenogest concentrations. The concentration of ALT present in the samples were assumed based off the dosage given on each packaging. The recommended therapeutic dosage for each product aligned with the bodyweight (BW) of the horse (Table 3-1).

Table 3-1: Therapeutic dosage values from oral and intramuscularly injected (IM) altrenogest (ALT) products. BW denotes body weight of the horse being administered.

	Oral Product	**IM Product**
Bottle Concentration of ALT	2.20 mg/mL	50.0 mg/mL
Therapeutic Dosage	1 mL/50 kg BW	1 mL/166 kg BW
Therapeutic Concentration of ALT Administered	0.044 mg/kg BW	0.3 mg/kg BW
i.e., 500 kg horse	22.0 mg administered	150 mg administered

The administration study was conducted using ten mares (5-16 years; 498 ± 39.3 kg) with oral (*n* = 5) and IM (*n* = 5) routes of ALT. The amount of ALT administered was dependent on the total BW of the individual horse; each oral administration dose was 0.044 mg/kg BW, whereas each IM administration dose was 0.3 mg/kg BW. This study administered daily oral doses for 14 days, and two long-acting IM injections of ALT into the muscle at the base of the neck, on day zero and 7 days. The calculated concentrations of total ALT given to each horse over the administration period is included in Table 3-2, alongside the individual horse's BW, age, horse breed, and administration route type.

Sample collection was conducted at 8 am daily, inclusive of the two weeks of administration, with an additional one week of sampling conducted after, totalling to 21 days of sample collection. Two frequent sampling days were also conducted at the first and final administration points for each route, where samples were collected every 2 hours until the 8-hour mark plus a 12-hour sample was also collected. A detailed representation of the samples collected can be seen in the appendices (Appendix 1 and Appendix 2). The individual frequent sample days were day zero and day 14 for oral, and day zero and day 7 for IM. Ethics approval was provided for this study by Charles Sturt University (A19050), with funding for the study provided by AgriFutures Australia as part of the Thoroughbred Horses Program.

Table 3-2: Administration horses categorised by route type (oral or intramuscular – IM), horse breed type (thoroughbred or standardbred), age, body weight (BW in kg, as at commencement of study), and the calculated total dosage of altrenogest (ALT) given over the two-week active administration period.

	Administration Route	Horse Breed	Age (years)	Body Weight (BW in kg)	Total ALT Administration Dosage (mg)
Horse 1	Oral	Thoroughbred	14	538	355
Horse 2	IM	Thoroughbred	12	440	265
Horse 3	Oral	Thoroughbred	10	500	330
Horse 4	IM	Thoroughbred	5	472	284
Horse 5	Oral	Thoroughbred	5	522	344
Horse 6	IM	Thoroughbred	12	458	275
Horse 7	Oral	Standardbred	11	502	331
Horse 8	IM	Standardbred	7	574	345
Horse 9	Oral	Standardbred	16	518	341
Horse 10	IM	Thoroughbred	9	460	277

3.2.3 Altrenogest Product Analysis

3.2.3.1 Oral Altrenogest Product

Oral ALT product (0.2 mL, *Readyserve*® oral solution (2 mg/mL)) was combined with sodium phosphate buffer (pH 9.5, 2 mL) and MilliQ water (3 mL) in a large screw-capped, glass tube and vortexed. The sample mixture was adjusted to pH 9.5 before LLE was conducted with DCM (7 mL) using a rotary mixer at slow speed for 15 minutes. After which, the sample was centrifuged for 15 minutes at 2093 x g. The top aqueous layer was discarded, and anhydrous sodium sulfate salt was added as a drying agent. The organic layer was transferred into a new screw cap glass tube, and sodium hydroxide (0.2 M, 4 mL) was added. The tube was returned to the rotary mixer for 15 minutes and centrifuged (2093 x g) for another 15 minutes. The aqueous layer was discarded, and the organic layer was transferred to a new glass tube, where the ISTD (trenbolone – d_5; 223.7 ng/mL, 37.5 µL, 40 ng/mL equivalent) was added and vortexed, before being evaporated to dryness under nitrogen at 80 °C.

3.2.3.2 Intramuscular Altrenogest Product

IM ALT product (0.45 mL, *Readyserve*® injection solution (50 mg/mL)) was combined with sodium acetate buffer (pH 5.0, 2 mL) and MilliQ water (3 mL) in a large screw-capped, glass tube and vortexed. The sample mixture was adjusted to pH 9.5 before LLE was conducted with DCM (5 mL) using a rotary mixer at slow speed for 15 minutes. After which, the sample was centrifuged for 15 minutes at 2093 x g. The top aqueous layer was discarded, and anhydrous sodium sulfate salt was added as a drying agent. The organic layer was transferred into a new screw cap glass tube, and sodium hydroxide (0.2 M, 2 mL) was added. The tube was returned to the rotary mixer for 15 minutes and centrifuged (2093 x g) for another 15 minutes. The aqueous layer was discarded, and the organic layer was transferred to a new glass tube, where the ISTD (trenbolone – d_5; 2.237 µg/mL, 8.0 µL, 40 ng/mL equivalent) was added and vortexed, before being evaporated to dryness under nitrogen at 80 °C.

3.2.3.3 Calibration and Quality Control Samples

A five non-zero datapoint calibration range, neat standard quality control (QC) samples and blank matrix (buffer) samples were prepared with trendione and trenbolone, and analysed alongside the product samples. The calibration samples were made up with 0.2 mL of phosphate buffer as the desired ‘blank’ matrix. Appropriate volumes of trendione and trenbolone standards (Table 3-3) were aliquoted before being taken through the sample extraction procedure for oral ALT product (**Section 3.2.3.1**).

Table 3-3: Calibration curve volumes added for trendione and trenbolone standards at each concentration level.

Calibration Level	Trendione Volume Added (µL, from 196.0 ng/mL working standard)	Trenbolone Volume Added (µL, from 196.6 ng/mL working standard)	Concentration (ng/mL)
1	5.10	5.00	5.00
2	10.2	10.1	10.0
3	20.4	20.3	20.0
4	51.0	50.8	50.0
5	81.6	81.3	80.0
Neat QC Standard	51.0	50.8	50.0

3.2.3.4 Reconstitution and Analysis

The oily residue was reconstituted with MeOH (0.8 mL) and vortexed well, before refrigeration at −20 °C. The sample was then centrifuged (2093 x g) and aliquoted (0.2 mL) into an LC vial for analysis using the Shimadzu 8060NX liquid chromatograph-triple quadrupole-tandem mass spectrometer (LC-QQQ-MS/MS) with a Shim-pack XR-ODS III column (1.6 µm, 2.0 mm x 75 mm, Shimadzu, Kyoto, Japan). Mobile solutions were A: 5 mM ammonium formate in ultrapure water (pH 3.0) and B: 0.1% FA in ACN. This method utilised a gradient of 0-1.00 minute (15% B), 1.01-11.00 minutes (15-20% B), 11.01-20.00 minutes (20% B), 20.01-23.00 minutes (100% B), and 23.01-25.50 minutes (15% B). The flow rate was 0.4 mL/min, the injection volume used was 5 µL, and the column temperature was 40 °C.

Data was extracted with LabSolutions (version 4.50 SP1) and analysed using Insight Explore (pre-release version). Statistical analysis was performed with Microsoft Excel (version 16.42) using essential statistical functions for evaluation purposes.

3.2.4 Urine Sample Preparation and Instrumental Analyses

Equine urine (3 mL) was combined with ISTD mix (trenbolone–d_5 and progesterone–d_9; 60 µL, 1 ng/mL equivalent) and pH adjusted to 5.0-5.5 in a large screw-capped, glass tube. Acetate buffer (2 mL) and β-glucuronidase enzyme (50 µL) were added and vortexed, after which the solution was incubated at 37°C overnight. The solution was readjusted to pH 9.0-9.5 before LLE with DCM (7 mL) using a rotary mixer at slow speed for 15 minutes. After this, the sample was centrifuged for 15 minutes at 2093 x g. The aqueous (urine) layer was discarded, and anhydrous sodium sulfate was added as a drying agent. The organic layer was transferred into a new screw cap glass tube, and sodium hydroxide (0.2 M, 2 mL) was added. The tube was returned to the rotary mixer for 15 minutes and centrifuged (2093 x g) for another 15 minutes. The aqueous layer was discarded, and the organic layer was transferred to a new glass tube, before being evaporated to dryness under nitrogen at 60 °C.

The dried residue was reconstituted with 0.1% FA in ultrapure water and 0.1% FA in MeOH (50 µL each, respectively) and aliquoted into LC vials for analysis using the Shimadzu 8060NX LC-QQQ-MS/MS with a Shim-pack XR-ODS III column (1.6 µm, 2.0 mm x 75 mm, Shimadzu, Kyoto, Japan). Mobile phase solutions were A: 5 mM ammonium formate in ultrapure water (pH 3.0) and B: 0.1% FA in ACN. This method utilised a gradient of 0-1.00 minute (15% B), 1.01-11.00 minutes (15-20% B), 11.01-20.00 minutes (20% B), 20.01-23.00 minutes (100% B), and 23.01-25.50 minutes (15% B). The flow rate was 0.4 mL/min, the injection volume used was 10 µL, and the column temperature was 40 °C.

Data was extracted with LabSolutions (version 4.50 SP1) and analysed using Insight Explore (pre-release version). Statistical analysis was completed with Microsoft Excel (version 16.42) using essential statistical functions and used to produce plots for evaluation purposes.

3.2.5 Method Validation Criteria

The LC-QQQ-MS/MS method was validated for the quantification of trendione, trenbolone and epitrenbolone. The parameters assessed were sensitivity, selectivity/specificity, linearity, accuracy, precision, recovery, and matrix effects. Validation criteria by Peters et al has been followed. [123-125]

Trendione, trenbolone and epitrenbolone were quantified using trenbolone-d_5 internal standard. Duplicate QC samples of each standard were also spiked at 1 ng/mL in a blank female horse urine matrix, previously tested for the absence of ALT, and followed the above extraction procedure. The optimised multiple reaction monitoring (MRM) transitions used for

the quantification of all compounds are presented as the mass-to-charge ratios (*m/z*) in the table below (Table 3-4).

*Table 3-4: List of multiple reaction monitoring (MRM) transitions (m/z) used to quantify the compounds of interest for targeted analysis. An * is indicative of the quantifier ion used for each compound.*

Compounds	Optimised MRMs (*m/z*)
Trenbolone–d_5 (ISTD)	275.50 → 204.15*, 258.30
Trendione	269.20 → 192.25*, 225.25, 169.05
Epitrenbolone	271.00 → 253.30*, 199.30, 238.30
Trenbolone	270.50 → 199.30*, 253.30, 227.20

Sensitivity

Sensitivity was tested by estimating the Limit of Detection (LOD); the lowest concentration of analyte in a sample that can be detected but not necessarily quantified as an exact value, and the Lower Limit of Quantification (LLOQ); the lowest amount of an analyte in a sample that can be quantitatively determined with suitable precision and accuracy. These are established using the two equations below [123].

$$LOD = \frac{3.3 \times standard\ deviation\ of\ the\ y\ intercept}{slope}$$

$$LLOQ = \frac{10 \times standard\ deviation\ of\ the\ y\ intercept}{slope}$$

Selectivity/Specificity

Selectivity and specificity are often used interchangeably as they have very similar meanings. However, strictly speaking, specificity refers to methods that produce a response for a single analyte, whereas, selectivity refers to methods that produce a response for a number of

chemical entities [126]. Therefore, this project will be testing selectivity as multiple chemical components were analysed.

Selectivity has been defined as “the ability of the bioanalytical method to measure unequivocally and to differentiate the analyte(s) in the presence of components which may be expected to be present” [126]. Therefore, selectivity was determined by mass differentiation of the compounds, their MRMs and their retention time (RT) noted in minutes on the column (MRMs - Table 3-4, RTs - Table 3-5).

Linearity

Linearity was assessed between the LLOQ and the Upper Limit of Quantification (ULOQ); the maximum analyte concentration of a sample that can be quantified with acceptable precision and accuracy (bias). As a general rule, the ULOQ is identical to the concentration of the highest calibration standard. A calibration range of 0.05 ng/mL to 5 ng/mL was implemented for trendione, whilst 0.05 ng/mL to 2 ng/mL was used for trenbolone and epitrenbolone.

Accuracy

Accuracy was determined by comparing mean values from QC levels against expected values as a percentage, known as the relative error percentage. The QC concentration used in this method was 1 ng/mL of trendione, epitrenbolone, and trenbolone.

$$Relative\ Error\ (\%RE) = \frac{calculated\ concentration - mean\ actual\ concentration}{mean\ actual\ concentration} \times 100$$

Acceptable criteria for accuracy were ± 15% relative error (RE) of the accepted reference value and ± 20% near LLOQ.

Precision

This was determined by calculating the standard deviation and mean values of seven (7) equine urine spiked replicates at the QC level. The precision is described by the percent relative standard deviation (%RSD) equation below.

$$Relative\ Standard\ Deviation\ (\%RSD) = \frac{standard\ deviation}{mean\ concentration} \times 100$$

The acceptable limit for precision was 15% RSD for all values except for LLOQ which was 20% RSD.

Recovery

Recovery is considered the percentage of the analyte response after sample preparation compared with that of a solution containing the analyte at a concentration corresponding to 100% recovery. The acceptable recovery range criteria were 80% to 120%. This was assessed from seven (7) equine urine (non-hydrolysed) spiked replicated pre- and post- LLE, at a QC level of 1 ng/mL, using the equation below.

$$Recovery\ (\%) = \left(\frac{pre\ extraction}{post\ extraction}\right) \times\ 100$$

Matrix Effects

Matrix effects were determined by comparing the analyte response of seven (7) neat standards to seven (7) replicate post- LLE spiked matrix samples. This was then calculated using the equation below.

$$Matrix\ Effects\ (\%) = \left(\frac{post\ extraction}{neat\ standard}\right) \times\ 100$$

The matrix effects were assessed using the theory that if the result was above 100% the compound is experiencing ion enhancement and if the result was below 100% the compound ionisation is being suppressed.

3.3 RESULTS AND DISCUSSION

3.3.1 Validation Results

Sensitivity (Table 3-5) was evaluated using a calculated LOD of 3.3*standard deviation of the intercept/slope, and a calculated LLOQ of 10*standard deviation of the intercept/slope. Selectivity was assessed by averaged retention time of seven (7) injections of each compound to give the values in the table below. Accuracy, precision, recovery, and matrix effects results were calculated using a 1 ng/mL concentration spike. Recovery and matrix effects values were calculated using the peak area of each compound from pre and post LLE sampling, and neat spikes.

Table 3-5: Validation results for trendione (Td), epitrenbolone (Etr) and trenbolone (Tr) in a urine matrix.

	Trendione	**Epitrenbolone**	**Trenbolone**
LOD (ng/mL)	0.02	0.01	0.01
LLOQ (ng/mL)	0.07	0.03	0.04
RT (minutes, averaged)	20.36	19.66	19.06
Correlation coefficient (R^2)	0.9999	0.9999	0.9999
Accuracy (%RE)	-1.42	0.08	1.11
Precision (%RSD)	4.5	5.8	9.2
Recovery (%)	101.5	81.2	86.8
Matrix effects (%)	14.2	27.8	16.8

3.3.2 Altrenogest Product Results

Validation results for the ALT product were obtained for the sensitivity, selectivity/specificity, and linearity, due to the change up in matrix being used, i.e., phosphate buffer solution, rather than equine urine. Sensitivity was calculated using the equations provided above (**Chapter 3, Section 3.2.4.1.1**) with data from the calibration curves used for the trendione and trenbolone quantification. Some slight retention time shifting can be observed between Table 3-5 and

Table 3-6, which can be explained due to the time difference between when the samples were analysed (i.e., 2 years apart) and the age of the column used. Sensitivity results (LOD and LLOQ) obtained for trendione and trenbolone are comparable between validation tests, with little discernible difference.

Table 3-6: Validation results for the analysis of trendione and trenbolone concentrations in oral and intramuscularly injected altrenogest products.

	Trendione	**Trenbolone**
LOD (ng/mL)	0.03	0.05
LLOQ (ng/mL)	0.09	0.15
RT (minutes, averaged)	20.1	18.8
Correlation coefficient (R^2)	0.9997	0.9991

The amount of trendione and trenbolone present were 0.00061% and 0.00001% in the intramuscular product, and 0.00007% and 0.00004% in the oral product, respectively (to five decimal places) for each compound (Table 3-7).

Table 3-7: Concentration results for trendione and trenbolone levels in oral and intramuscularly injected (IM) altrenogest (ALT) products used in the administration study.

	Oral Product	**IM Product**
Trendione	7.208×10^{-5} %v/v (1.6 ng/mL)	6.088×10^{-4} %v/v (304.4 ng/mL)
Trenbolone	4.502×10^{-5} %v/v (1.0 ng/mL)	1.121×10^{-5} %v/v (5.6 ng/mL)

The trend of trendione as the main impurity and trenbolone (with some conversion to epitrenbolone via metabolism) as a secondary by-product in ALT products align with what was observed in the administration study samples, presented in the following sections (**Chapter 3, Section 3.3.3** and **Section 3.3.4**). Additionally, higher concentrations of both impurities were observed in the IM ALT product, comparatively to the oral treatment. The concentration of

trendione in the IM product is considered an estimate as the value (304.4 ng/mL) fell outside the calibration range (5 - 80 ng/mL). Dilution and reanalysis were not possible at the time due to the high response of ALT by LC-MS/MS risking contamination to the instrument.

The extraction methodologies for each ALT product differ slightly due to a number of factors. The routine Australian Racing Forensic Laboratory (ARFL) method used 1 mL for sample analysis, however, less than this amount was available for the IM product. This was because the sample was taken from the remaining residue of a bottle used in this administration study for an accurate depiction of the injected steroid impurity concentrations. Additionally, the first extraction with 1 mL of oral product left a large amount of oily product post dry-down under nitrogen and did not allow for a clean reconstitution. A second extraction with only 0.2 mL was performed instead, in an attempt to improve the analysis. The extracted samples were also cooled in a −20 °C freezer to separate the oil residue from the MeOH extract, however, this was difficult to control as the two layers would immediately start to recombine once the sample was removed from the freezer for aliquoting into LC vials.

3.3.3 Oral Administration

The data collected from the ten mares analysed showed that the most significant change of these impurities occurred for trendione, and the minor impurity elevation was epitrenbolone. Trenbolone was detected at times in low concentrations, often below the LLOQ. However, the method may not account for all trenbolone present due to the potential for trenbolone to be excreted in urine as the sulfate conjugate, which is poorly hydrolysed by the β-glucuronidase enzyme used in this method [127]. For appropriate quantification of the trenbolone impurity present in ALT administration samples, chemical hydrolysis methods may provide a more effective way of deconjugating both glucuronide and sulfate conjugates of trenbolone, an important topic for future investigation.

From the trendione and epitrenbolone excretions, these compounds displayed obvious elevation during the frequent sampling days and consistently low or non-detectable levels out of these periods (Figure 3-1).

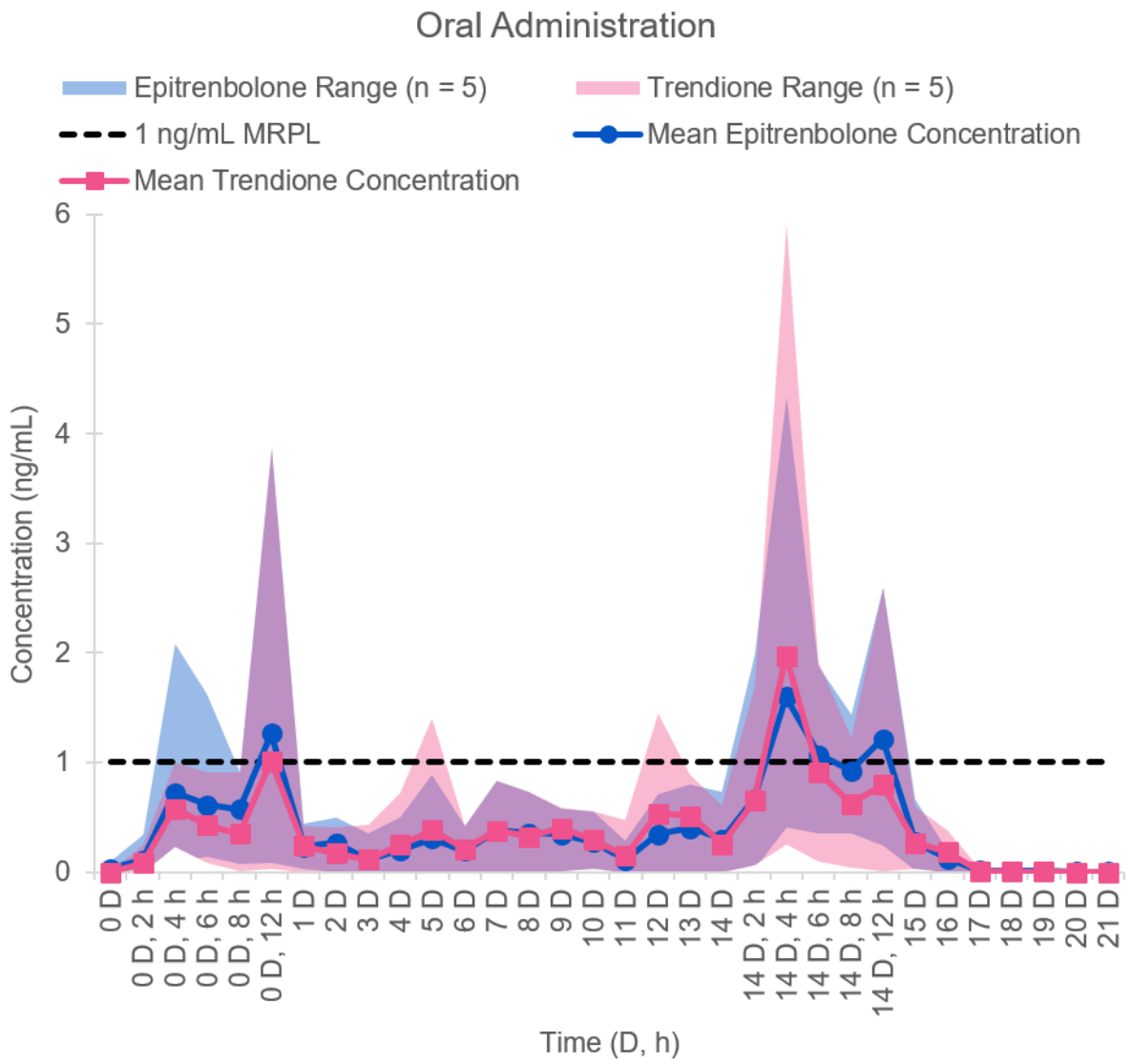


Figure 3-1: Mean urinary excretion profile concentrations (ng/mL) of trendione and epitrenbolone over 21 days in equine urine with oral altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL) for racing laboratories to detect the presence of anabolic androgenic steroids (AAS) in equine urine.

The black dotted line in this figure indicates a 1 ng/mL minimum reporting performance level (MRPL) for racing laboratories to detect the presence of AAS in equine urine and has been implemented here as a baseline for quantification and assumed effect [122, 128, 129]. This plot shows the mean concentrations of both trendione and epitrenbolone for five horses over the 21-day collection period as a scatterplot, and the range of concentrations given in the shaded regions as an area plot. For further details of these results, separated mean excretion plots of trendione and epitrenbolone and individual horse plotted results per steroid impurity can be found in the appendices (see Appendix 3 to Appendix 7)

The maximum individual concentrations for trendione and epitrenbolone were 5.9 ng/mL (horse 9) and 4.3 ng/mL (horse 9), respectively, at time point 14 days and 4 hours (14 D, 4 h) for both compounds. As the maximum concentration for trendione was found to be greater than 5 ng/mL (the highest validated calibration point), this point can only be considered an estimate. Future research may consider validation of an extended calibration range or suitable dilution experiments in order for this value to be accurately quantified.

The oral administration of ALT demonstrated low levels of impurities, only breaching the MRPL of 1 ng/mL at peak administration times of 8 h and 14 D, 4 h for trendione. After one clear day of administration for the initial and two-week administration time points of therapeutic oral treatment, both impurities were below the MRPL.

3.3.4 Intramuscular Administration

Reviewing the below graphs regarding the IM administration route, elevated levels of impurities were again seen in the initial and second frequent sampling days at day zero and 7 D (Figure 3-2). The concentrations at these periods were greater than those observed following the oral administration. The maximum individual concentrations for trendione and epitrenbolone after IM administration were 5.9 ng/mL (horse 4) and 3.3 ng/mL (horse 2), respectively, at time points 9 days (9 D) and 12 hours (12 h), respectively. Similarly, to oral administration results, the maximum concentration for trendione was calculated as greater than 5 ng/mL, so this point can also only be considered an estimate.

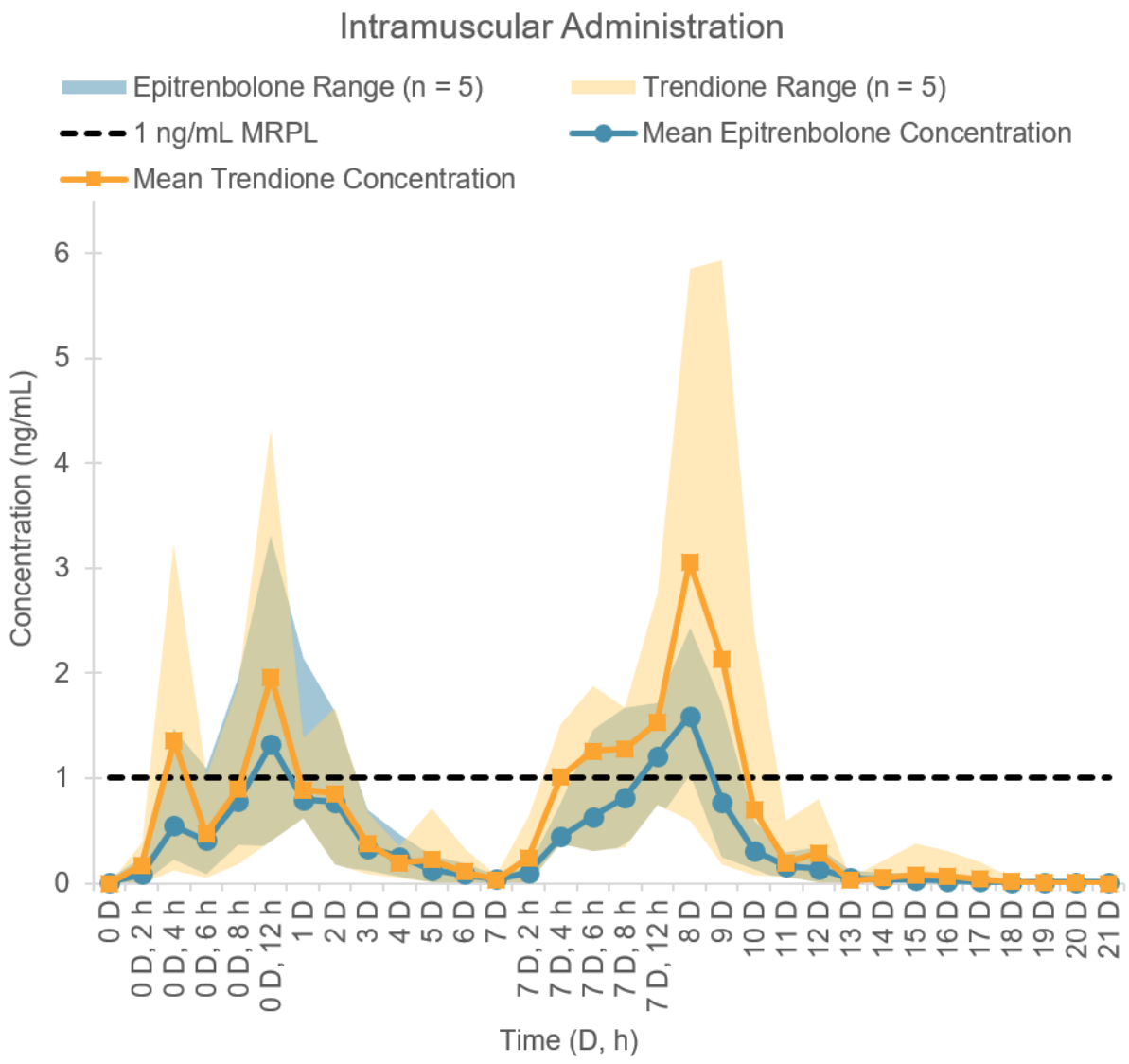


Figure 3-2: Mean urinary excretion profile concentrations (ng/mL) of trendione and epitrenbolone over 21 days in equine urine with the administration of intramuscularly injected altrenogest (two injections administered at 0 D and 7 D), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates the 1 ng/mL minimum reporting performance level (MRPL).

The IM administration of ALT showed more concentrated levels of the prohibited steroid impurities and remained elevated for a greater period of time. Trendione remained elevated above the MRPL up to two days post-administration of ALT, whilst at 9 D, some epitrenbolone samples had fallen back below the MRPL. These more concentrated and longer elevated level findings indicate higher concentrations of the impurities in the IM administrated product and potential absorption in the system due to the administration route used.

The appearance of trendione, epitrenbolone, or trenbolone in a raceday sample may, therefore, be indicative of injected use of ALT or the direct use of one or more of these prohibited steroids. Impurities surpassing the MRPL would be used for both intelligence and possible misuse/allowed use identification purposes. This distinction is dependent on when

the sample was taken, i.e., raceday or out of competition sampling. Further details of these results can be found in the appendices (see Appendix 8 to Appendix 12), separated mean excretion plots of trendione and epitrenbolone and individual horse plotted results per steroid impurity.

3.3.5 Bivariate Scatterplot

A bivariate scatterplot (Figure 3-3) of the targeted quantification data provided the potential to discriminate between the oral and IM administration routes. This plot contains a total of 220 individual day (24 h) samples from the 10 horses studied. Concordance of urinary trendione and epitrenbolone concentrations being measured greater than 1 ng/mL may be indicative of an injected administration. Samples with trendione and epitrenbolone levels that have fallen outside this 1 ng/mL range have been labelled with the horse (i.e., H2, H4, H6, H8, or H10) and the timepoint in hours (h) or days (D) post-administration. The epitrenbolone and trendione concentrations have also been given as the (x,y) coordinates, respectively, in the plot.

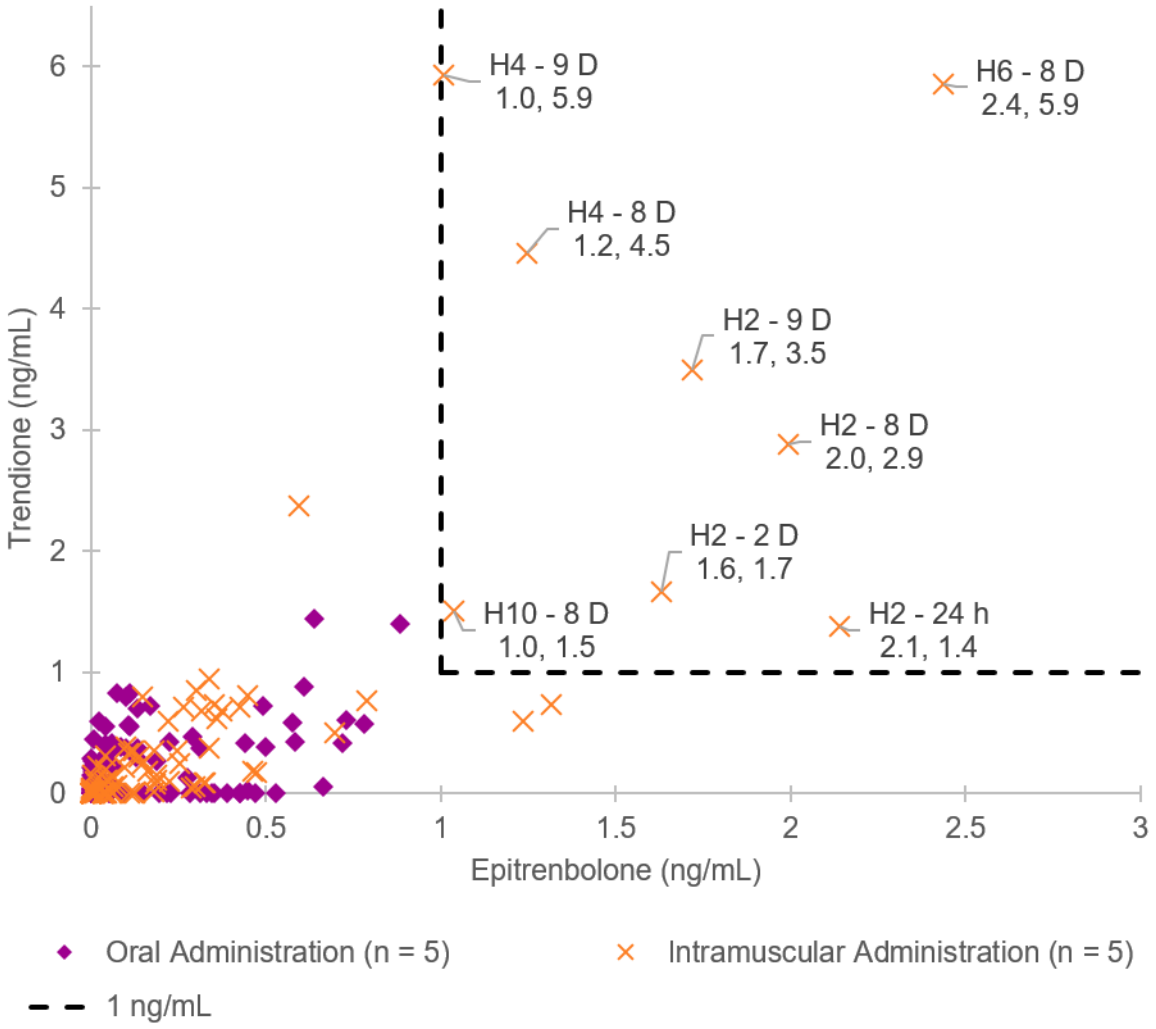


Figure 3-3: Bivariate scatter plot of trendione and epitrenbolone concentrations (ng/mL) following oral, purple diamonds, and intramuscular (IM), orange crosses, altrenogest administration. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL). H identifies the administration horse, and D or h denotes the day of hour timepoint the sample relates to. The x, y data labels indicate the epitrenbolone concentration, and trendione concentration of that sample, respectively.

3.4 Conclusions and Future Work

Both the oral and IM ALT products contained the two expected impurities and confirms the cause of their presence in the administered horses. Concentrations of impurities in the ALT products align with the excreted concentrations found in the treated horses. The trend of trendione as the main impurity and trenbolone (with some conversion to epitrenbolone via metabolism) as a secondary by-product in ALT products aligns with what was observed in the administration study samples. Additionally, higher concentrations of both impurities were observed in the IM ALT product, comparatively to the oral treatment.

This research aimed to analyse phase I and II metabolites, quantitatively and qualitatively, in mares impacted by differing ALT administration routes. A validated method was developed for the targeted quantification of steroid impurities by LC-MS/MS in equine urine. No evidence of elevation in these steroid impurities was observed outside 24 hours of oral ALT administration. The IM quantified steroid impurities, however, exceeded 1 ng/mL up to two days post-final administration. It is important to note that the concentration of trenbolone in the ALT administration samples may be significantly underestimated due to the use of only β-glucuronidase enzyme failing to effectively hydrolyse its sulfate conjugates. Future studies should be conducted using alternative deconjugation methods such as a sulfatase enzyme or chemical hydrolysis in order to verify the results and conclusions made from this preliminary investigation. Further research into the possibility of differing ALT administration routes from those tested here would be advantageous.

Chapter 4: The Use of Progesterone as an Endogenous Reference Compound (ERC) for Altrenogest Administration Route Determination

4 THE USE OF PROGESTERONE AS AN ENDOGENOUS REFERENCE COMPOUND (ERC) FOR ALTRENOGEST ADMINISTRATION ROUTE DETERMINATION

4.1 INTRODUCTION

As previously discussed (**Chapter 1, Section 1.3.1**), progesterone levels naturally decrease and production of estrogen spikes during the oestrus cycle of a female horse [130]. To counteract this, progestogen containing products including ALT, are commercially available and commonly used to regulate and control ovulation for breeding purposes [131, 132]. The administration of ALT will mimic the higher progesterone levels seen in the out of oestrus (or often referred to as "heat") phase, and leads to a natural suppression of estrogen production and endogenous progesterone excretion (Figure 4-1) [43, 133].

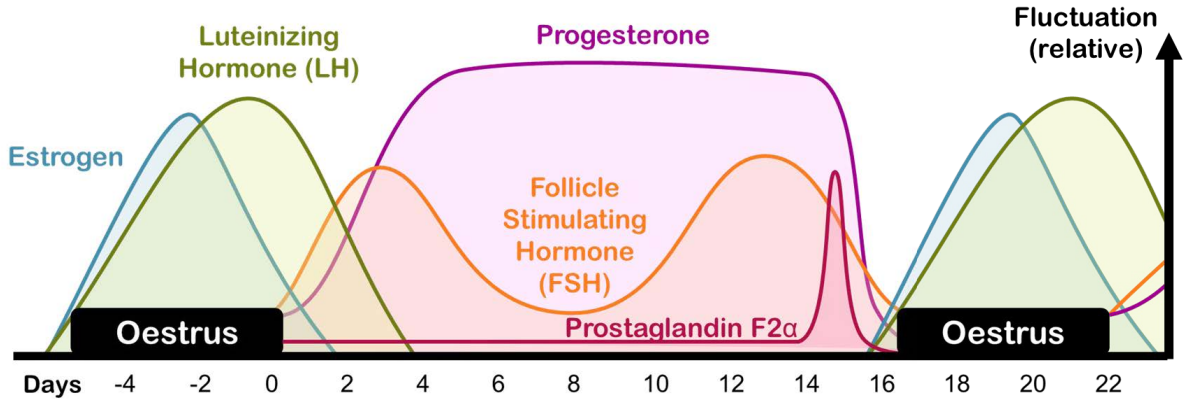


Figure 4-1: Oestrus cycle in female horses, with days denoted on the x-axis and relative fluctuation shown on the y-axis. Visually adapted from equine-reproduction.com.

The administration of ALT causes a down regulated influence on levels of progesterone excreted. This influence appears to differentiate between the oral and IM administrations of ALT in the studied horses, with more sustained down-regulation following IM administrations compared to oral administrations. Due to this, the use of progesterone as an ERC was investigated, along with trendione and epitrenbolone as ratios to differentiate oral administration from IM administration.

4.2 MATERIALS AND METHODS

4.2.1 Chemicals and Reagents

All chemicals and reagents were used as outlined in **Chapter 3, Section 3.2.1**.

4.2.2 Administration Study

The administration study parameters and details were as outlined in **Chapter 3, Section 3.2.2**.

4.2.3 Sample Preparations and Instrumental Analyses

The sample preparations and instrument analyses were as outlined in **Chapter 3, Section 3.2.4**.

4.2.4 Method Validation Criteria

Full validation parameters are available in **Chapter 3, Section 3.2.5**.

A calibration range of 0.05 ng/mL to 5 ng/mL was implemented for progesterone. Progesterone was investigated to monitor biological effects on the equine system, using progesterone-d_9 as the internal standard and matching calibration range to trendione. Duplicate QC samples of each standard were also spiked at 1 ng/mL in a blank female horse urine matrix, previously tested for the absence of ALT, and followed the above extraction procedure. The optimised MRM transitions used for the quantification of all compounds are presented in the table below (Table 4-1).

*Table 4-1: List of MRM transitions (m/z) used to quantify the compounds of interest for targeted analysis of progesterone. An * is indicative of the quantifier ion used for each compound.*

Compounds	Optimised MRMs (*m/z*)
Progesterone – d_9 (ISTD)	323.90 → 113.30*, 100.20
Progesterone	314.90 → 96.90*, 109.00, 297.40

4.2.5 Data Processing and Machine Learning

Machine learning was conducted on the peak height data extracted from exported results on Shimadzu's LabSolutions Insight. The dataset incorporated the $\log_{10}$ transformed peak height values from trendione, trenbolone, epitrenbolone, and progesterone, along with the ratio of

epitrenbolone-to-progesterone (Etr/Pr) and ratio of trendione-to-progesterone (Td/Pr), which were also $\log_{10}$ transformed. Peak height data was chosen over peak area due to the inconsistency regarding chromatographic noise and peak shape with lower levels of progesterone. Data points from 9 horses (oral *n* = 4, IM *n* = 5) at 24-hour timepoints only, provided a total of 188 averaged timepoints. The data was labelled based on breed type (0 = thoroughbred or 1 = standardbred) and administration type (0 = oral or 1 = IM) (Table 4-2).

Table 4-2: Administration route type (oral or intramuscularly injected (IM)) and horse breed type (thoroughbred or standardbred) of all ten horses involved in the altrenogest (ALT) administration study. Horse 3 samples were not included in this chapter regarding progesterone levels.

Horse	Administration Type	Horse Breed Type
Horse 1 (H1)	Oral (0)	Thoroughbred (0)
Horse 2 (H2)	IM (1)	Thoroughbred (0)
Horse 3 (H3)	Oral (0)	Thoroughbred (0)
Horse 4 (H4)	IM (1)	Thoroughbred (0)
Horse 5 (H5)	Oral (0)	Thoroughbred (0)
Horse 6 (H6)	IM (1)	Thoroughbred (0)
Horse 7 (H7)	Oral (0)	Standardbred (1)
Horse 8 (H8)	IM (1)	Standardbred (1)
Horse 9 (H9)	Oral (0)	Standardbred (1)
Horse 10 (H10)	IM (1)	Thoroughbred (0)

The machine learning model used in this chapter involved a python script written by T. Durant, accessed through GitHub [134]. A PDF version of the script used in this chapter has been provided in the appendices (Appendix 17). The code utilised the XGBoost, *NumPy, pandas, os, and Matplotlib* packages and libraries, all of which are openly accessible and free to use.

The model package used in this chapter was called XGBoost which stands for extreme gradient boosting. XGBoost is an optimised distributed gradient boosting library designed by Chen and Guestrin [135], to be highly efficient, flexible and portable. It implements machine learning algorithms under the Gradient Boosting framework using parallel tree boosting. Gradient boosting is a machine learning technique based on boosting that targets *pseudo*-residuals (also known as prediction errors), instead of residuals in traditional boosting. It gives a prediction model in the form of an ensemble of weak prediction models, which are models that make very few assumptions about the data, and are typically simple decision trees (Figure 4-2) [136].

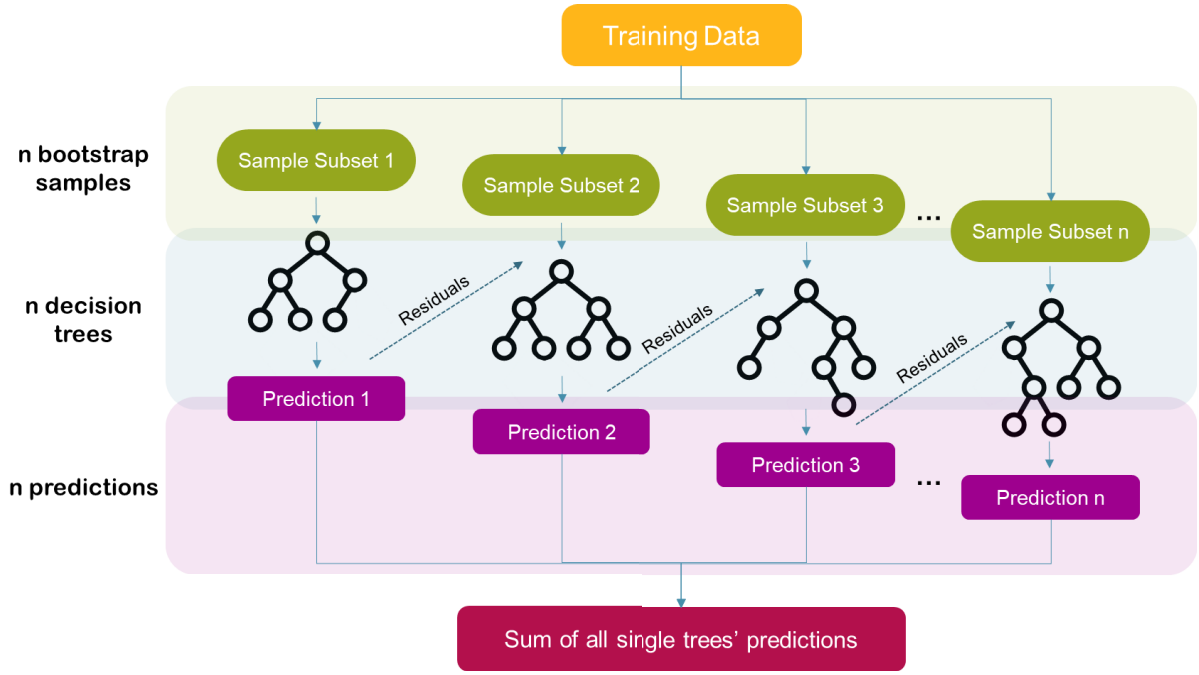


Figure 4-2: Extreme gradient boosting (XGBoost) workflow. Adapted from Mohammadi et al. [137].

Numerous libraries were also used for their own data manipulation, analysis and visualisation functions. These include: *NumPy*; a powerful Python library used for working with arrays and performing numerical computations [138]. *Pandas*; a powerful and open-source Python library designed for data manipulation and analysis. It provides fast, flexible, and expressive data structures that make working with "relational" or "labelled" data both easy and intuitive [139]. *Matplotlib*; is a comprehensive library for creating static, animated, and interactive visualizations in Python [140]. *Matplotlib's* primary purpose is to provide tools and functionality to represent data graphically, allowing for easier understanding and analysis. *Scikit-learn*; a popular machine learning library in Python that provides simple and efficient tools for data mining and data analysis [141]. A module called *os* was also utilised for Python to access and read administration study data in comma separated values (csv) filetype.

The XGBoost Classifier model parameters were optimised to return improved precision recall area under the curve (PR-AUC) scores and predictive accuracy scores using the functions listed in Table 4-3 and Table 4-4 below. All other parameters remained as default settings.

Table 4-3: Initial parameters specified for the XGBoost model.

Specifying Model Parameters	Explanation	Setting
XGBClassifier	A function allowing XGBoost models to be trained for classification tasks	precision_recall_curve
Evaluation metric	Metric interacts with early stopping to determine best iteration of the model	'logloss' (binary log loss – provides probabilistic measure)
Train	The portion of the dataset used for training the model	60%
Test	The portion of the dataset used to test the model	30%
Validate	The portion of the dataset used to validate the model	10%

Table 4-4: XGBoost model parameters for optimisation in Python. Default and optimised settings along with the allowed range and a brief explanation are given. Opt denotes Optimisation. ∞ in the range column indicates no upper numerical limit.

Parameters	Description	Default	Range	Opt
n_estimators	Size of the forest to be trained	100	20 - 1000	N/A
learning_rate	Learning rate of the model	0.3	0 - 1	0.3
max_depth	Maximum depth of a tree	6	3 - ∞	3
reg_alpha	L1 regularisation term on weights. Increasing this value will make model more conservative.	0	0 - ∞	N/A
reg_lambda	L2 regularisation term on weights. Increasing this value will make model more conservative.	1	1 - ∞	N/A
subsample	Subsample ratio of the training dataset	1	0 - 1	N/A
min_child_weight	Minimum sum of instance weight (hessian) needed in a child	1	0 - ∞	1
colsample_bytree	Subsample ratio of columns when constructing each tree	1	0 - 1	N/A
early_stopping_rounds	Activates early stopping. Finds the optimal number of boosting rounds	0	0 - ∞	10
random_state	Random number seed.	0	0 - ∞	42

4.3 RESULTS AND DISCUSSION

4.3.1 Validation Results

Sensitivity (Table 4-5) was evaluated using a calculated LOD of 3.3*standard deviation of the y-intercept/slope, and a calculated LLOQ of 10*standard deviation of the y-intercept/slope. Selectivity was assessed by averaged retention time of seven (7) injections of each compound to give the values in the table below. Accuracy, precision, recovery, and matrix effects results were calculated using a 1 ng/mL concentration spike. Recovery and matrix effects values were calculated using the peak area of each compound from pre and post LLE sampling, and neat spikes.

Table 4-5: Validation results for progesterone in a urine matrix.

	Progesterone
LOD (ng/mL)	0.06
LLOQ (ng/mL)	0.18
RT (minutes, averaged)	22.39
Correlation coefficient (R^2)	0.9996
Accuracy (%RE)	119.2
Precision (%RSD)	2.8
Recovery (%)	102.0
Matrix effects (%)	38.6

Due to the endogenous nature of progesterone in equine urine, the accuracy, precision and linearity results may fall outside the accepted range of variation (± 15 or 20%), which is expected when spiked into blank urine that already contains an unknown amount of progesterone (linearity intercept at 3.33 ng/mL indicates an approximate concentration of progesterone in the blank urine).

4.3.2 Oral Administration

Regarding the assessment of progesterone concentrations, whilst LOD and LLOQ validation returned acceptable results with a blank urine matrix, authentic equine urine samples provided chromatography with higher baselines that made integration and subsequent concentration determination difficult. In addition to this, the following results will demonstrate suppression of progesterone levels that was observed in the administration study, assumed to be caused by the introduction of the ALT products, also contributing to lower concentrations of progesterone in the samples.

To overcome these difficulties, the peak area ratio with the internal standard (progesterone-d_9) was used instead and showed a noticeable change from day zero. An area ratio was used here as the endogenous nature of progesterone, combined with the ALT's purpose of suppressing progestin production, caused progesterone levels to decrease over the three-week period, from day zero samples. Whilst this does not improve the validity of the data where levels may have dipped below the LLOQ, it was suitable as an aid in visually demonstrating the differences between the oral and IM excretion. It should also be noted that oral administration data is averaged from four (4) horses, rather than five (5), as one of the five (5) horses was analysed prior to the addition of progesterone quantification in the study (Figure 4-3). Timewise plots of the individual orally administered horses can be found in the appendices (Appendix 13).

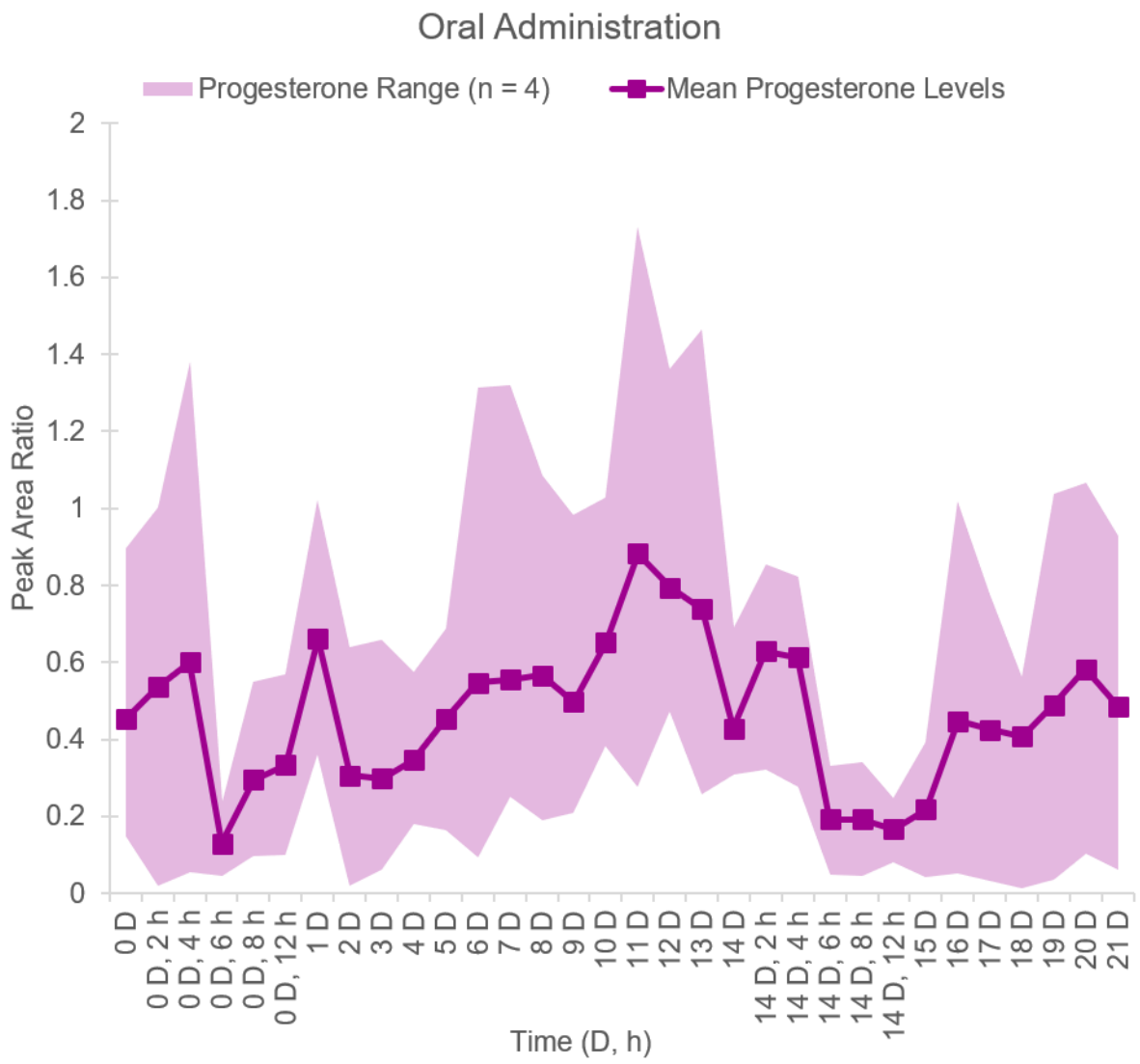


Figure 4-3: Mean urinary excretion profile of progesterone as a peak area ratio (progesterone/progesterone-d_9), over 21 days in equine urine with an oral altrenogest administration (daily for the initial 14 days), averaged from four (4) different horses. Shaded area indicates the range of individual horse samples collected from each timepoint.

Whilst each timepoint fluctuated in response to the daily administration of ALT, suppression was clearly observed in the two frequent sampling periods at the 4-hour mark, after a small elevation occurred post administration. These findings showed promise in implementing a complementary biomarker approach for progestin administration.

4.3.3 Intramuscular Administration

In addition to the oral administration findings, progesterone displayed prolonged suppression over the 21-day period on average in IM administration, with regard to the individual's day zero ratio value (Figure 4-4). This suppression could be cause for concern, indicating use of the injectable product. Timewise plots of the individual orally administered horses can be found in the appendices (Appendix 15).

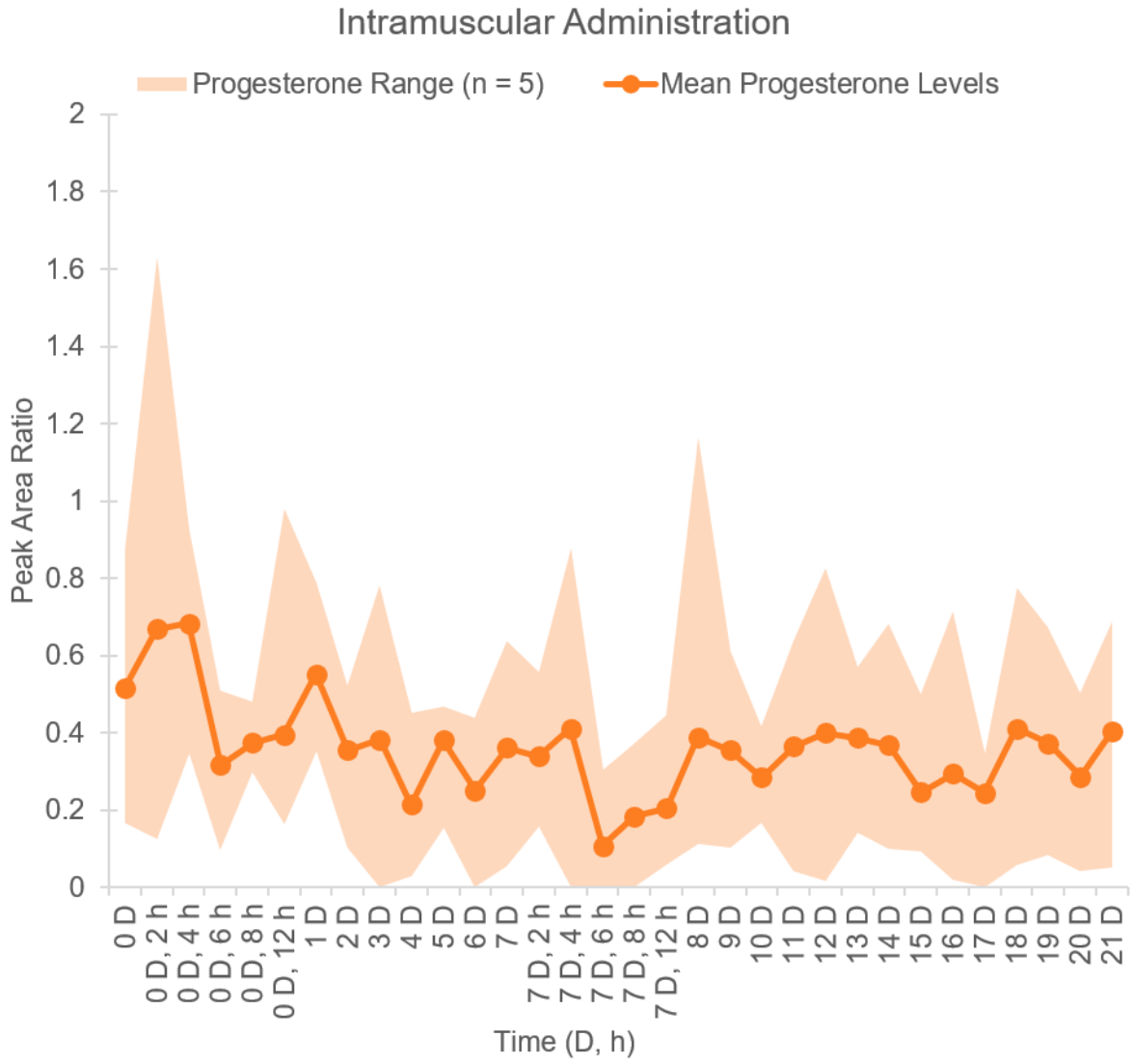


Figure 4-4: Mean urinary excretion profile of progesterone as a peak area ratio (progesterone/progesterone-d_9), over 21 days in equine urine with the administration of intramuscularly injected altrenogest (two injections administered at 0 D and 7 D), averaged from five (5) different horses. Shaded area indicates the range of individual horse samples collected from each timepoint.

Figure 4-5 presents the mean fold change, calculated from day zero, of all horses sampled from the oral and IM administered ALT over 21 days. With the administration of oral ALT, levels of progesterone increase and decrease throughout the three weeks of sample

collection. In contrast, with the administration of IM ALT, the levels of progesterone were continuously suppressed throughout the three weeks of sample collection.

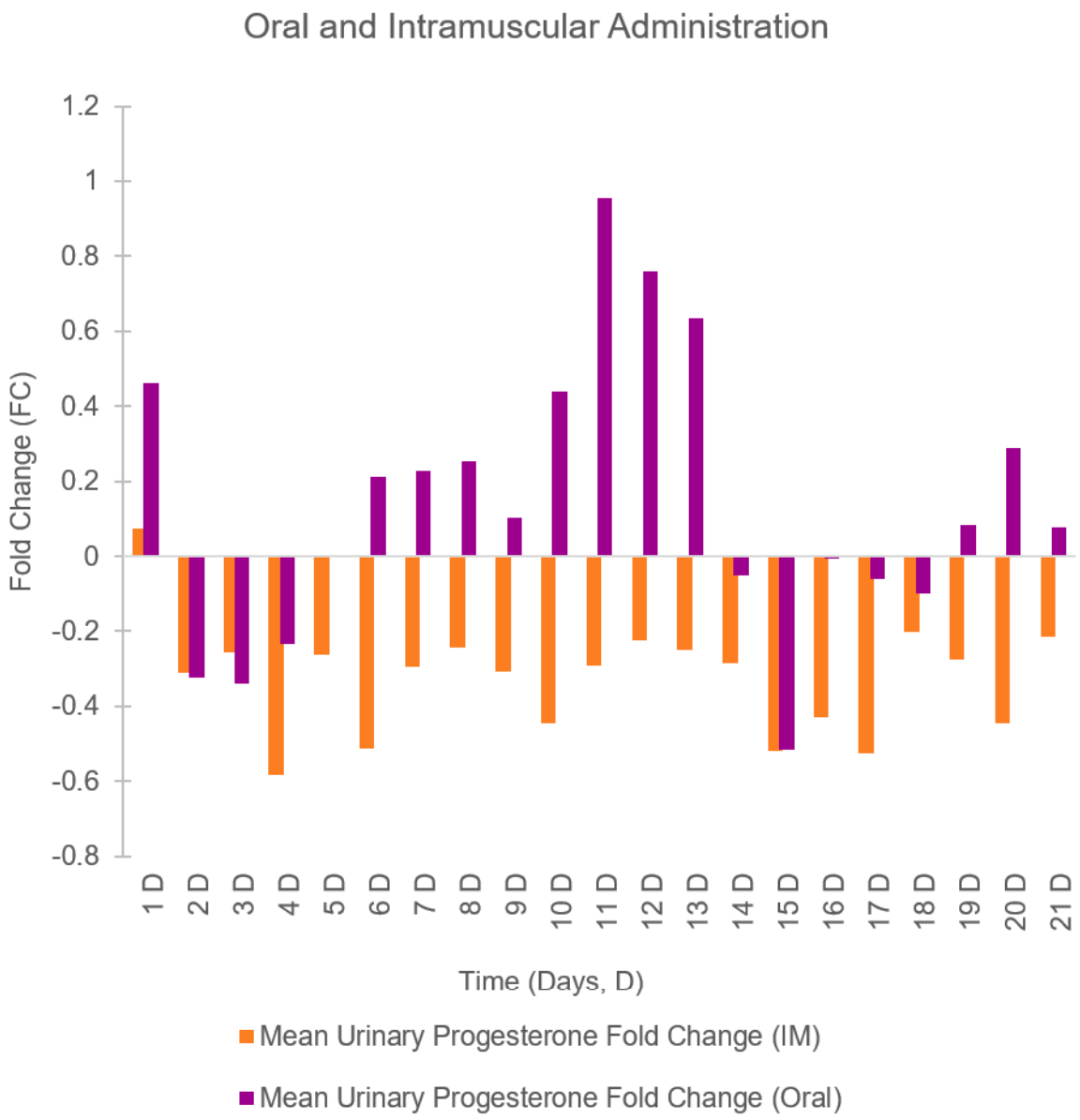


Figure 4-5: Mean urinary progesterone fold change (from day zero) plot, comparing oral (purple) and intramuscular (orange) administrations of altrenogest in equine urine over 21 days.

This administration route differentiation can be observed more clearly using a fold change graph comparing oral and IM progesterone fluctuation from day zero. Figure 4-5 shows a visual difference in suppression of progesterone from the IM administration when compared to its oral counterpart. While oral levels fluctuate, mimicking a “normal” unmedicated progesterone production in the system, the IM levels did not return to baseline after day zero and, at some sample points continue to decrease further.

It should be noted here that, naturally, progesterone production does show some diurnal variation in the horse [132, 142]. However, a major limitation of this administration study was the lack of placebo/control horse samples that may have indicated whether the observed excretion pattern differs from an un-administered horse. With this in mind, an assumption was made that the observed results (with comparison between responses to the administration routes) were as a result of ALT administration.

4.3.4 Machine Learning

Optimisation of the original GitHub code (explained in detail in **Section 4.2.5**, originally from reference [134], dataset provided in Appendix 18) was done on validation size (from 0.20 to 0.10) and XGBClassifier parameters (max_depth, min_child_weight, learning_rate, and n_estimators), using the GridSearchCV function from Scikit-Learn, a popular machine learning library in Python. The optimised model parameters produced better training curve results with a training and validation loss plotted close together, showing less likely overfitting than the un-optimised results. Overfitting is denoted in a learning curve plot by the closeness of the training and validation lines as their loss is plotted per iteration. If the validation loss separates from the training loss as the iterations increase, this indicates that overfitting has occurred with the dataset. Overfitting reduces a model's accuracy and reliability. The training curve here indicates the model has not succumb to overfitting (Figure 4-6). A total of 35 iterations were used to train and validate the model using this dataset.

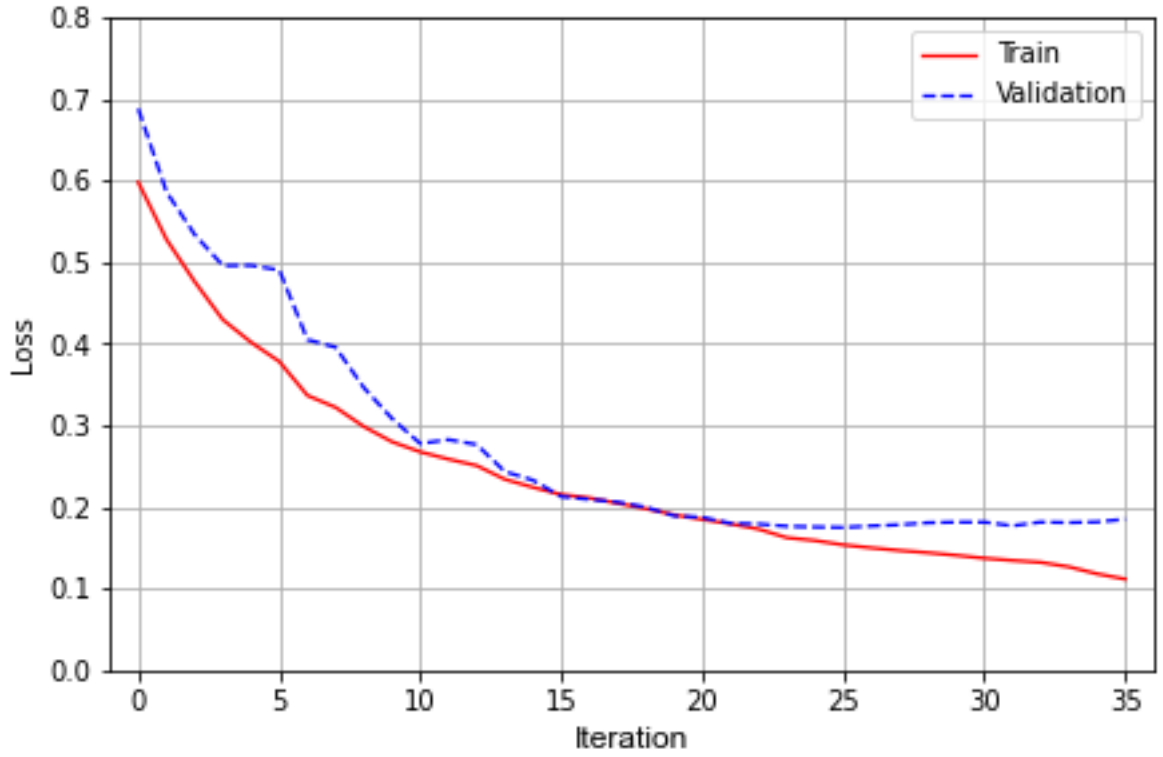


Figure 4-6: Learning curve for progesterone ratio model using XGBoost.

The classification model was trained, tested and validated based on known values, i.e., dataset with known classes and a PR-AUC score of 0.947. PR-AUC represents the models' overall performance in distinguishing oral samples from IM samples based on the information given, the closer the score is to 1, the better the model's performance. This can be displayed by a Precision Recall (PR) ROC plot in Figure 4-7.

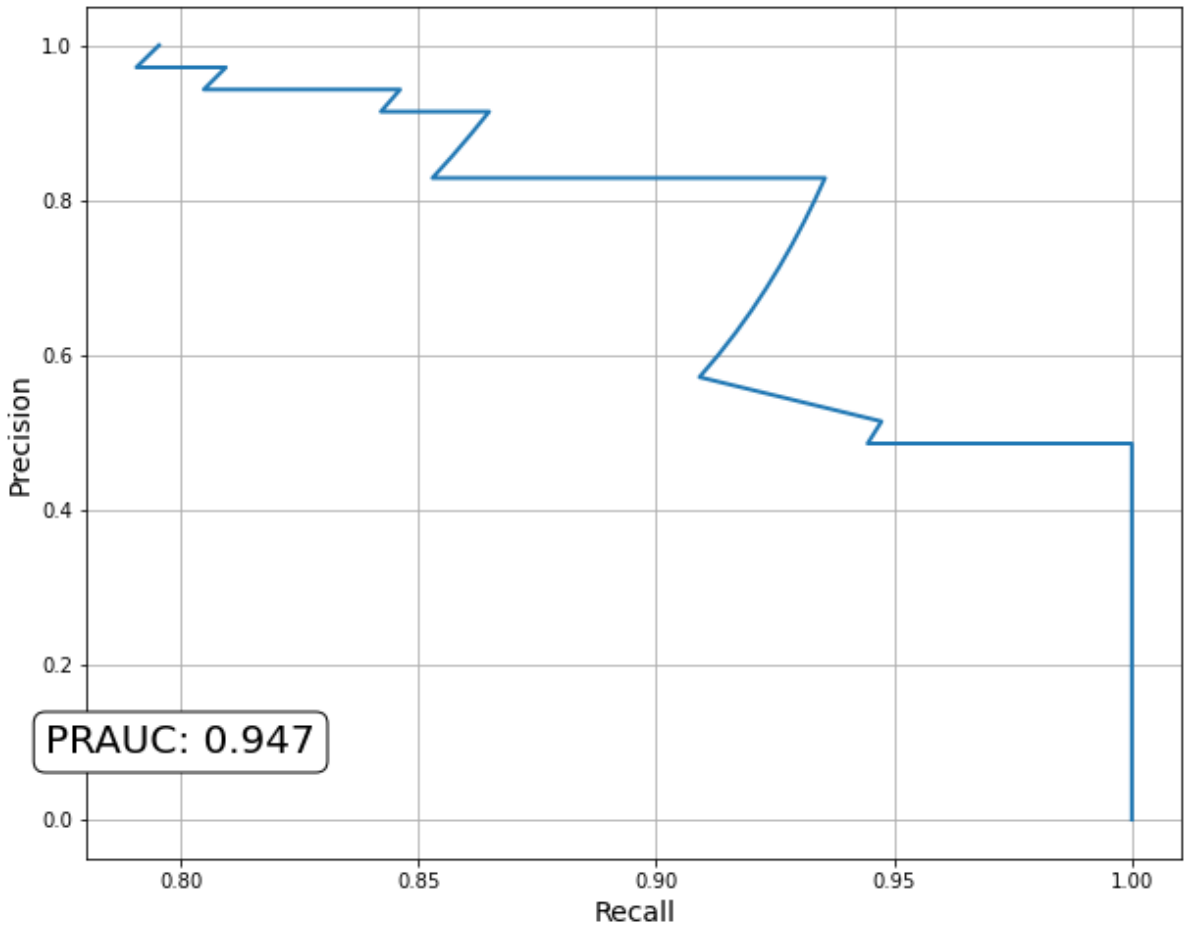


Figure 4-7: Precision Recall (PR) area under the receiver operating characteristic (AU-ROC) plot obtained from optimised XGBoost classification model.

Predictive accuracy is the measure of all correct classifications, relating to both true positive and true negative classifications. This model returned a predictive accuracy score of 84.21% in differentiating IM administration from oral administration, indicating appropriate model performance for the intended classification purpose.

Log transformation of the dataset was conducted to minimise the variability in results from outer influences, unrelated to the administration route being targeted, i.e., instrument variance and matrix effects. This transformation of the data made no negative impact on the PR-AUC result and positively increased the predictive accuracy score slightly (Table 4-6).

Table 4-6: A comparison of XGBoost model outcomes (PR-AUC and predictive accuracy scores) for untransformed and log transformed peak height progesterone ratio datasets, used to classify intramuscular (IM) altrenogest (ALT) administration route from the oral route sample.

	Untransformed Data	**Transformed Data**
PR-AUC Score	0.947	0.947
Predictive Accuracy	82.46%	84.21%

Statistical testing was also carried out in the script to verify the models' performance by calculating the classification error, mean absolute error (MAE) and an F1 score (Table 4-7). Helpful tutorials on how these metrics could be executed was found online [143]. Classification error is a commonly used metric when determining a model's performance and measures the proportion of instances where the predicted class label does not match the true class label. This is calculated from the number of misclassified samples divided by the total number of samples. Here a lower classification error indicates superior model performance, suggesting fewer mistakes made in the predictions. The MAE is the measure of the average absolute difference between the predicted and actual values. A lower MAE value indicates better model performance, suggesting that the prediction results are closer to the true values. The F1 score considers both the precision and recall performance of the model, where precision measures the proportion of true positive predictions in all positive predictions, and recall measured the proportion of true positive predictions within all actual positive predictions. The F1 score is commonly used as it represents the mean of precision and recall values, with a value closer to one and further from zero as the best possible score.

Table 4-7: XGBoost profile classification model scores and statistical analysis.

	Statistical Test/Result	Definition
Binomial Classification Accuracy	84.21%	the measure of all correct classifications, relating to both true positive and true negative classifications
PR-AUC	0.947	represents the models' overall performance in distinguishing oral samples from IM samples based on the information given
Feature Importance	Most important feature identified as **Td** (f1)	a form of visualising the scores assigned to input features based on their contribution to the model's predictions
Permutation Feature Importance	**Etr** (feature_2)	
Classification Error	0.1579	measures the proportion of instances where the predicted class label does not match the true class label
Mean Absolute Error (MAE)	0.1579	the measure of the average absolute difference between the predicted and actual values
F1 Score	0.8800	represents the mean of precision and recall values

Table Note: refer to Table 4-8 for explanation on important feature labels.

Feature importance plot is a form of visualising the scores assigned to input features based on their contribution to the model's predictions. This function provides valuable insights into which features contribute the most to the model and which may be considered redundant, which offers greater model interpretability and allows for further feature selection to occur. Features with higher importance scores are considered more influential in determining the output, while features with lower scores have less impact. Figure 4-8 was produced using a built-in feature importance function in XGBoost and plots the frequency (F score) at which each feature was used to split the data across all trees in the model. The feature labels ranked

in this importance plot and the following permutation importance plot correspond to the following data given to the model (Table 4-8).

Table 4-8: Corresponding feature labels to features in profile model. Horse breed types are grouped as 0 for thoroughbred and 1 for standardbred, compound labels Td, Etr, Tr, and Pr indicate trendione, epitrenbolone, trenbolone and progesterone, respectively. Height refers to the captured peak height data for each standard.

Importance Label	Model Features
feature_0 = f0	horse breed type (categorical – 0 or 1)
feature_1 = f1	Log_{10} Td height
feature_2 = f2	Log_{10} Etr height
feature_3 = f3	Log_{10} Tr height
feature_4 = f4	Log_{10} Pr height
feature_5 = f5	Log_{10} Td/Pr height ratio
feature_6 = f6	Log_{10} Etr/Pr height ratio

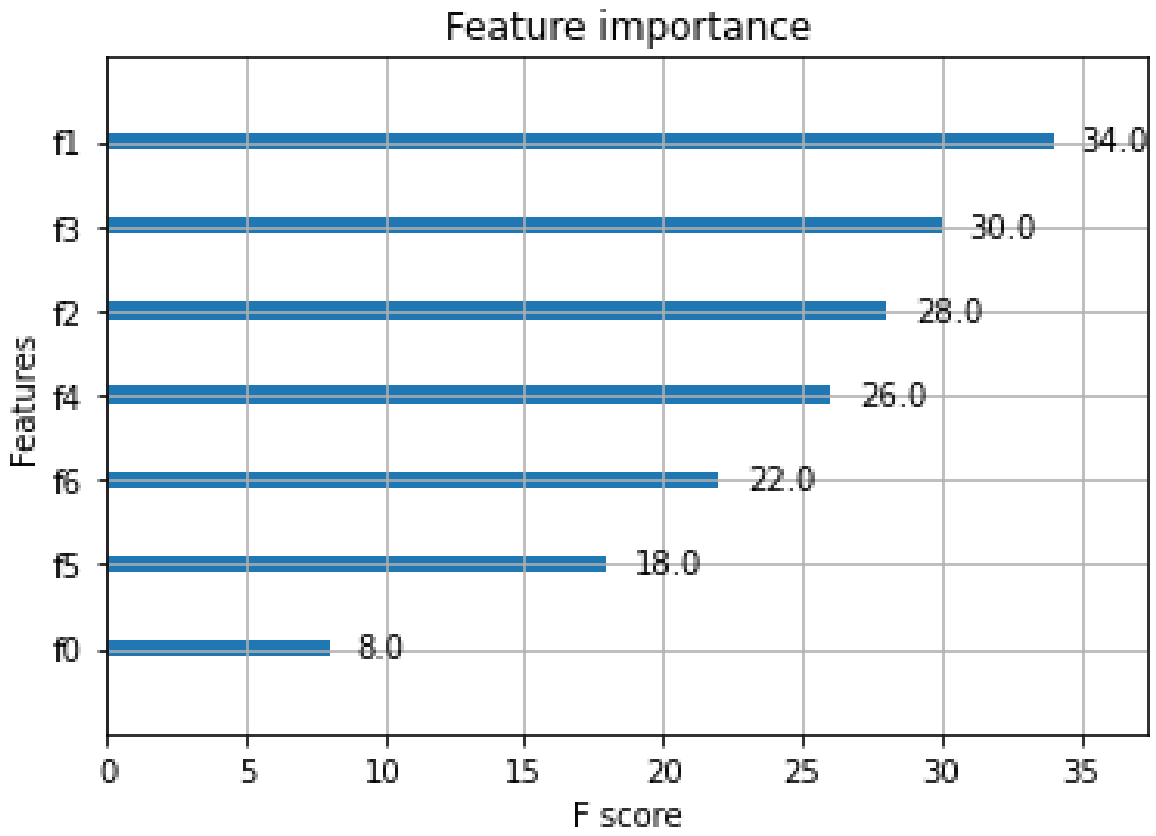


Figure 4-8: Feature importance plot using the internal extreme gradient boosting (XGBoost) model function of visualising which model features are most important in correctly classifying the samples.

Permutation testing was applied for feature importance to determine which markers were most effective in correctly classifying samples as oral or IM. This test was done using an external function through Scikit-Learn as it is understood that greater interpretation can be made from this test, as opposed to the internal XGBoost feature importance plot in Figure 4-9. The model identified epitrenbolone as the most influential in differentiating the administration groups, whilst the Etr/Pr values were considered as the least important, comparatively.

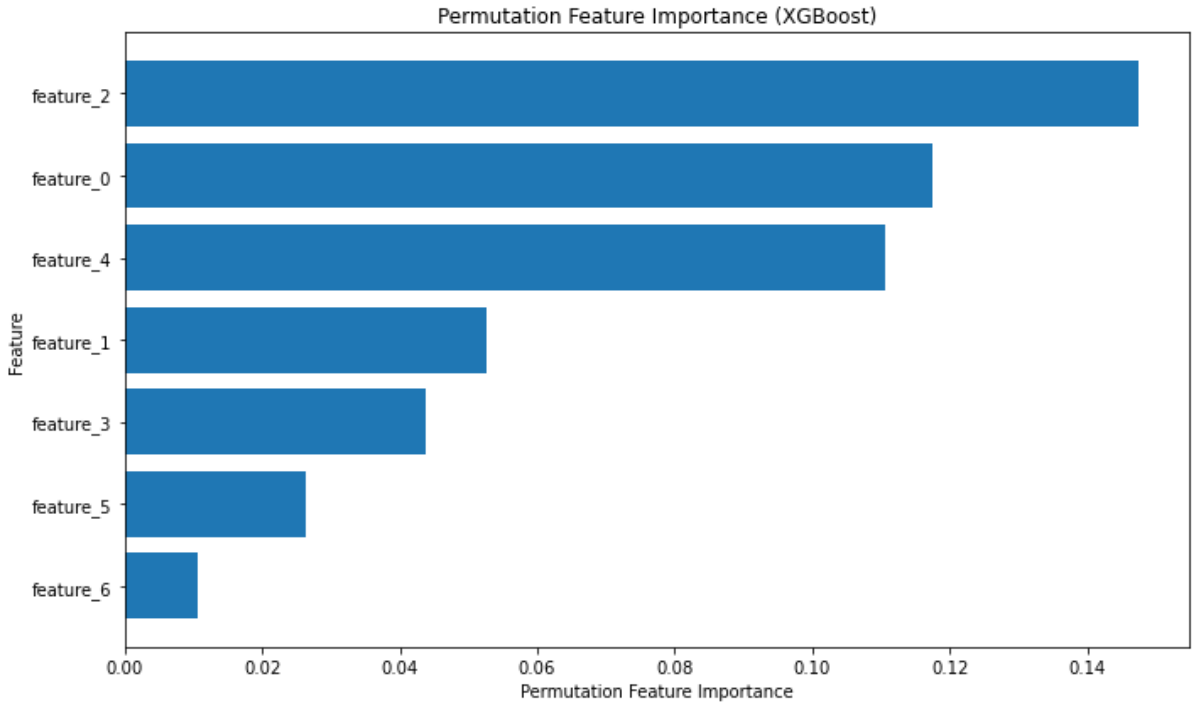


Figure 4-9: Permutation Feature importance, testing which features are most influential in classifying samples between oral and intramuscularly injected (IM) administration route.

Confusion matrix was then applied to determine the spread of true, and false, positive and negative classifications made by the model (Table 4-9). This information can help identify areas where the model is limited and guide further improvements or model selection decisions. The classification of a label (i.e., oral or IM) is determined by the calculated probability of the sample being IM, equating to a probability score of 1 for full confidence, and lower scores indicating a reduction in the degree of confidence the model has. The default threshold for a sample classification is greater than 0.5 (> 0.5) for an IM label and less than 0.5 (< 0.5) for an oral grouping.

Table 4-9: Confusion matrix data relating to extreme gradient boosting (XGBoost) model results classifying intramuscularly injected (IM) administered altrenogest (ALT) equine urine samples to orally administered ALT equine urine samples. Original output can be found in Appendix 19.

		Predicted Labels	
		Oral (0)	**IM (1)**
True Labels	**Oral (0)**	15	7
	IM (1)	2	33

These model confusion matrix results indicate that a low amount of false negative predictions occurred with a value of two (2), out of a possible 35 IM samples. However, this model illustrated a higher prevalence for false positive outcomes with seven (7) samples incorrectly identified as IM when they were in fact oral samples.

As this model incorporates samples from the full 21 days administration study and has shown good predictive accuracy for these samples, it is evident that detection and classification beyond the two days post final administration is achievable. The addition of a machine learning tool, complementary to the current routine screening process has enabled greater classification between the oral and IM administration route samples. This is important for identifying IM administered ALT in racehorses as the guidelines advise against its use. The results in **Chapter 3, Section 3.3.4** also support this finding with higher concentrations of steroid impurities present and remaining in the horses' system for a longer period of time, comparative to the oral administration type.

4.3.5 Model Verification

A blind sample from a past routine raceday doping control that returned a positive screening result for ALT was extracted and analysed in duplicate against the developed XGBoost Classifier model. Using the predict function, the new sample data was imported into the script and classified based on the same six features and both datasets returned a prediction of class 1, indicating against an allowed oral administration. This prediction also returned a probability score of 0.9799, demonstrating relatively high confidence in this prediction. Due to impartiality of the ARFL performing routine sample analysis, this verification was limited in having no additional metadata or intelligence from the case to conclusively provide evidence of using an injectable ALT product or use of an oral ALT product outside of the guidelines in relation to dose or time prior to racing.

The trial implementation and performance of this model has proven the monitoring of endogenous compound excretions has potential benefits in differentiating ALT administration routes in study samples and routinely acquired samples. Next steps for the implementation of this model would require working alongside ARFL to add the quantitative analysis of progesterone into the current screening method for ALT use, as well as validating this model in its current form on additional equine samples (that were unrelated to the original administration study) where the ALT administration route is truly unknown.

4.3.6 Study Limitations and Future Work

As the initial quantitative findings are based on changes in the peak area ratio, rather than true quantification of progesterone excretion, some discretion is advised when implying their biological significance. There are limitations in the accuracy of the observed progesterone suppression, particularly when levels fell below the calculated LLOQ and LOD, which may have falsely impacted the validity of results, and the proceeding modelling performance.

This model uses an extreme gradient boosting (XGBoost) algorithm that is an ensemble style of machine learning that can be prone to overfitting, when compared to other model styles, i.e., bagging or stacking. However, the boosting style was chosen over the other types due to its robustness when addressing non-evenly weighted datasets, meaning the un-even sized oral and IM datasets was not an issue.

The dataset used to train this model incorporated raw peak height and raw peak height ratio values that were log transformed to improve the models' accuracy. Testing with the same dataset prior to log transformation gave the same AUC score with a slightly lower predictive accuracy result of 82.46% proving that the transformation did not negatively affect the precision and recall scores, whilst positively impacting the models' predictability. This choice of data type could, however, be deemed a limitation for routine implementation. This is due to the variability that chromatography can display over long periods of time and susceptibility to change dependent on column (changing column type or replacing like-for-like), solvent gradients, tuning and maintenance cleaning of the instrument. Whilst log transformation was applied to reduce this variability, further normalisation and scaling could be implemented to mitigate false positive results due to one of the reasons mentioned above.

Chapter 5: Administration Route Differentiation of Altrenogest Via the Metabolomic LC-HRMS Analysis of Equine Urine

5 ADMINISTRATION ROUTE DIFFERENTIATION OF ALTRENOGEST VIA THE METABOLOMIC LC-HRMS ANALYSIS OF EQUINE URINE

5.1 INTRODUCTION

The detection of steroids of either endogenous or exogenous origin has been frequently performed by measuring free steroids following enzymatic hydrolysis of equine urine [6]. It is known that enzyme-catalysed hydrolysis is less effective in hydrolysing the phase II steroid sulfate conjugates [6, 144], leading to possible inaccurate detection of some steroid metabolites. To overcome the complexity associated with hydrolysis techniques for urinary AAS [145], advances in LC-HRMS have enabled the possibility of identifying new and interesting features, which includes the direct analysis of intact phase II conjugated metabolites [146].

The development of a complementary approach has the potential to improve injected ALT detection, using features differing from oral ALT. Abnormal levels of endogenous substances can provide an extended window of detection and give rise to the possibility of a biomarker or biomarker ratio to distinguish between oral and IM administration [60]. LC-MS was utilised to analyse intact phase II conjugates without derivatisation. Sulfate conjugate analysis was the focus of this research, due to the nature of excretion of these steroids of interest, particularly relevant to equine metabolism [6, 144, 147].

Additionally, recent advancements in machine learning tools and multi-tool workflows for the processing and analysis of high throughput mass spectrometry data have enabled a more advanced investigation of metabolomics studies. This has led to improved/streamlined outcomes for biomarker discovery and anti-doping detection in the clinical and toxicological fields [71, 90]. This study adopted some of these novel tools to assess their benefit discerning between two administration scenarios (oral and IM). Whilst there has been substantial research using these methods to discern the presence or absence of drug administration [89], their potential use to differentiate between two administration routes of the same substance, at similar concentrations, does not appear to be as well explored.

This chapter focuses on an untargeted screening of the metabolomic profile in mares administered with ALT. The primary aim was to identify potential biomarkers that could assist in distinguishing oral administration from the IM route. An untargeted metabolomic approach

was utilised to monitor urinary levels of AAS and other metabolically related endogenous and exogenous steroids as both phase I (unconjugated) and phase II (sulfate conjugates), where possible. The analytical technique chosen for this work was LC-QTOF-MS due to its ability to analyse urinary compounds without the need for derivatisation, and capture large amounts of low abundance MS data.

5.2 MATERIALS AND METHODS

5.2.1 Reference Standards

Internal reference standards in the metabolomic ISTD mix includes, testosterone sulfate-d_3 (triethylammonium salt) obtained from National Measurement Institute (NMI, North Ryde, Australia), and epitestosterone glucuronide-d_4 and pregnenolone-d_4 purchased from Sigma Aldrich (Castle Hill, Australia). Details on trenbolone-d_5, and progesterone-d_9 were listed in **Chapter 3, Section 3.2.1.1**. Additional reference standard information can be found in Appendix 20.

5.2.2 Chemicals and Reagents

Ammonium formate (99.995%), diethylamine (DEA), ethyl acetate (EtOAc), methanol (MeOH), and sodium hydroxide (NaOH) of LC-MS grade were purchased from Merck (Kilsyth, Australia). Dimethylformamide (DMF), 99.8% was purchased from Chem-Supply (Gillman, Australia). Sulfur trioxide pyridine complex (SO_3•py) and concentrated ammonia (NH_3) were purchased from Sigma-Aldrich (Castle Hill, Australia). The water used was ultrapure grade (18.2 MW.Ω.cm) obtained from a ThermoFisher Barnstead Smart2Pure system (ThermoScientific; Langenselbold, Hungary).

Aqueous ammonia solution prepared by adding 2% v/v of concentrated ammonia to purified water. i.e., in 5 mL purified water, 0.1 mL concentrated ammonia was added.

5.2.3 Administration Study

This research utilised samples collected from six (6) horses (5–14 years; 488 ± 49 kg) from a study administering oral (n = 3) and IM (n = 3) routes of ALT. The amount of ALT administered was the advised therapeutic dosage on the product labels, which is dependent on the total BW of the individual horse. Each oral administration dose was 0.044 mg/kg BW from a 2.2 mg/mL oral solution, whereas each IM administration dose was 0.3 mg/kg BW from a 50 mg/mL injectable solution. This study administered daily oral doses for 14 days and two long-acting IM injections of ALT into the muscle at the base of the neck on day zero and at 7 days.

Sample collection was conducted at 8 am daily, inclusive of the two weeks of administration, with an additional one week of sampling conducted after the conclusion of administration, totalling 21 days of sample collection. Two frequent sampling days were also conducted at the first and final administration points for each route where samples were collected every two (2) hours until the 8-h mark, plus an additional 12-h sample. The individual frequent sample days were day zero and day 14 for oral and day zero and day 7 for IM administration groups. Samples were stored at −20 °C prior to sample extraction. Ethics approval was provided for this study by Charles Sturt University (A19050, Wagga Wagga, NSW, Australia).

5.2.4 Sulfated Compound Synthesis Methodology

SO_3•py (50 mg) was added to a solution of free steroid (5 mg) in DMF (0.5 mL) and stirred at room temperature for three (3) hours on the *Ratek Controls* rotary wheel. After stirring, the reaction was quenched with purified water (10 mL) and SPE was conducted. A Sep-Pak C18 Vac 3 cc SPE cartridge (from Waters, Rydalmere, Australia) was conditioned with MeOH (3 mL) and purified water (3 mL) before sample (10 mL) is loaded. The cartridge and sample were washed with aqueous ammonia solution (2 mL) and purified water (3 mL) before being eluted into a large glass vial with MeOH (9 mL; *3 mL lots repeated thrice*). The eluant was then split into multiple 2.5 mL Eppendorf tubes and dried down in the Genevac EZ-2Plus Evaporator (low BP setting for methanol, two (2) hours). The end product was accuractely weighed before being transferred into a glass LC vial for use and storage. This procedure was adapted from previously published methodologies, for further details see references provided [148, 149].

5.2.5 Sample Preparations and Instrumental Analyses

The untargeted steroid sulfate extraction and analysis was adapted from a method developed and published by Fitzgerald et al. [148, 150]. Briefly, urine aliquots (0.6 mL) were diluted with phosphate buffer (0.3 mL, 100 mM, pH 5.4) and centrifuged at 2093 x g for 10 min. The supernatant was then combined with an internal standard mix (testosterone sulfate-d_3, epitestosterone glucuronide-d_4, pregnenolone-d_4, trenbolone-d_5, progesterone-d_9; 20 µL) in a fresh Eppendorf tube, before being loaded into an Oasis weak anion exchange (WAX) 3 cc SPE cartridge (from Waters, Rydalmere, Australia) conditioned with MeOH (3 mL) and ultrapure water (3 mL). The cartridge was then consecutively washed with NaOH (2 mL, 0.1 M), phosphate buffer (3 mL, 100 mM, pH 5.4), and ultrapure water (3 mL) before steroid sulfates were eluted with a solution of EtOAc: MeOH: DEA (3 mL, 25:25:1% *v*/*v*). The extract was transferred into spin filters (from PhaseSep, Doncaster East, Australia) before being dried

under nitrogen at 60 °C and reconstituted with 10% MeOH in ultrapure water (0.25 mL). Extracted samples were stored at 4 °C prior to instrument analysis.

Samples were analysed using a Shimadzu LC-40 ultra-high-performance liquid chromatography coupled to a 9030-quadrupole time of flight-mass spectrometer (UHPLC-QTOF-MS). The column was a Shim-pack Velox C-18 (1.8 µm, 2.1 mm × 100 mm, Shimadzu, Kyoto, Japan). Mobile solutions included A: 20 mM ammonium formate in 100% ultrapure water and B: 20 mM ammonium formate in 100% MeOH. This method utilised an elution gradient of 0–0.50 minutes (10% B), 0.51–20.50 minutes (10–100% B), 20.51–23.00 minutes (100% B), and 23.01–27.50 minutes (10% B). The flow rate was 0.4 mL/min, the injection volume used was 5 µL, and the column temperature was 40 °C. Data-dependent acquisition (DDA) analysis was performed in positive and negative electrospray ionisation (ESI) modes. MS conditions included a full scan range of *m*/*z* 100–1000 and MS/MS range of *m*/*z* 50–1000 with an intensity threshold of 500. The collision energy (CE) spread utilised was 60 ± 15 eV, with a 2.50 kV spray voltage, resolution of 30,000 FWHM, and approximately 19 Hz scan speed.

5.2.6 Statistical Multiplatform Workflow

The processing method was adapted from C. Fitzgerald et al. [148, 150]. Further information regarding the workflow can be found in the appropriate references, cited above. The computer used to conduct the following data processing was a Windows Processor 12th Gen Intel® Core ™ i5-10500, 2667 MHz, running six (6) cores, and 16 gigabytes of memory. Sample data was exported from LabSolutions (version 4.50 SP1) in the raw file format (lcd) and imported to MS-DIAL (version 4.90) where it was processed, aligned, and normalised. This resulted in the generation of a Microsoft Excel (version 16.42) spreadsheet containing the normalised peak area intensities for all present features in a batch. This was manually refined by true MS/MS collection, signal-to-noise ratio (*S/N*), and minimum presence in batch (denoted by Fill%) thresholds before being imported to R for statistical analysis. Simultaneously, the raw LC files (converted to mascot generic format (mgf) filetype) were run through two Python scripts, using Jupyter Notebook (version 6.3.0), that extracted the MS/MS peak data from each file and amalgamated all the data into one spreadsheet. This data was then carried into RStudio (RStudio–version 2022.07.1, R–version 4.3.2), and statistical analysis was performed with a custom R script. A steroid sulfate searcher function was also created and utilised with the simple ‘if else’ request from the *dplyr* package against a prearranged list of 26 potential *m/z* of interest. This element of the script resulted in information collection of possible significant steroid sulfate peaks.

In addition to the above workflow, the results from the steroid sulfate searcher function were analysed in MetaboAnalyst 6.0 (https://www.metaboanalyst.ca) using the Biomarker Analysis tool to determine key features for differentiation between the oral and IM groups. A RF classification model was then tested on five (5) selected features.

5.3 RESULTS

5.3.1 MS Annotation and Statistical Analysis

Metabolomic data was acquired using UHPLC-HRMS/MS (Shimadzu, Kyoto, Japan) with DDA following SPE sample clean-up of equine urine samples collected from mare horses administered with ALT, either orally (n = 3) or intramuscularly (n = 3). Succeeding data acquisition, LC-MS files were exported and input into MS-DIAL, where the MS data underwent data alignment and locally weighted scatterplot smoothing (LOWESS) normalisation of total ion chromatograms (TIC) by applying tolerances to both retention time (±3 s) and mass accuracy (Δm/z ± 5 ppm), alongside the recommended settings for DDA-HRMS systems [151]. Jupyter Notebook was used to read two custom Python scripts that extracted and appended MS/MS data for each aligned feature containing a sulfate-derived fragment (negative ions m/z 79.9573 (SO_3^-), 80.9652 (HSO_3^-), 95.9523 (SO_4^-), 96.9601 (HSO_4^-), and neutral losses 79.9568 Da (SO_3) and 97.9674 Da (H_2SO_4)), before realigning it to the MS-DIAL normalised peak area output. A custom R script, run through RStudio, was then used to conduct statistical analysis. This started with a k-means clustering function to sort metabolic features [152]. Clustering was performed based on the normalised data of eight unique parameters, all derived from the MS/MS spectral information. These included the six (6) characteristic sulfate-derived fragments, as mentioned above, as well as two calculated parameters, i.e., intensity ratio (IR), the total sum of sulfate-derived fragments divided by the sum of all fragments, and maximum abundance (MA), the normalised relative abundance (%) of the largest sulfate-derived fragment. This allowed clustering to occur into non-sulfate and sulfate groupings due to the abundant nature of steroid sulfate fragmentation peaks (Figure 5-1).

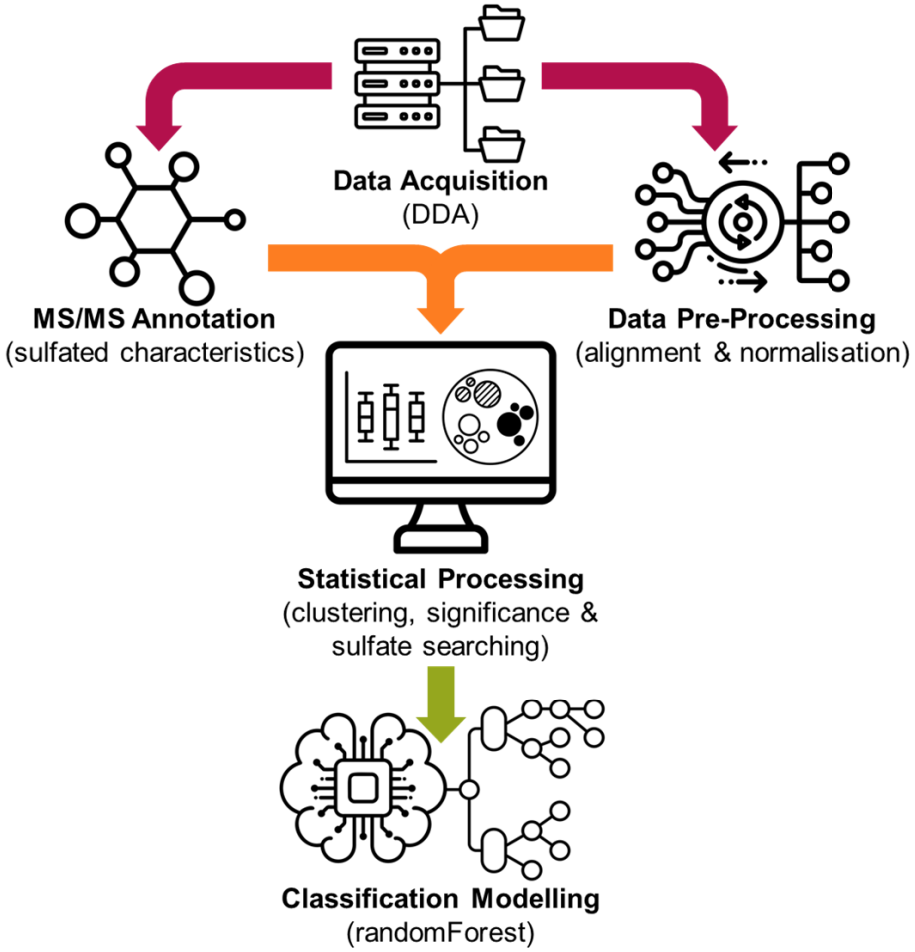


Figure 5-1: An overview of the presented metabolomic data processing workflow. Icons sourced from Noun Project©.

High throughput differential metabolite level analysis and data visualisation was then performed in R using the *limma* and *ggplot2* packages, respectively. The output gives a final curated list of metabolic features with associated statistical results, UHPLC-HRMS/MS data (RT and *m/z*), and product ion data for any sulfate-derived fragments detected by MS/MS. A novel steroid sulfate searcher function, coded into the R script, that extrapolated potential steroid sulfate features from the dataset was also used.

Two-dimensional ellipses PCA plotting was conducted on all analysed batches. Figure 5-2 shows a representative dataset from one of the IM administered horses with 32 sequential time points (0 h–504 h) of urine samples analysed in technical triplicates. The three smaller points of a single colour indicate the replicate data points, while the larger point of the same colour signifies the mean response. Pooled QC samples, prepared by combining a small amount of each sample from a single administered horse, can be seen in burgundy/purple, and their close clustering is indicated with an ellipse (labelled 'QC'). Two (2) other sample groups are coincidentally also included in the cluster but considered unrelated. Two (2)

external control samples (Control 1 and Control 2, coloured and circled in turquoise green and deep green, respectively) were also analysed alongside each batch using pre-administration (T0) samples from other horses in the study.

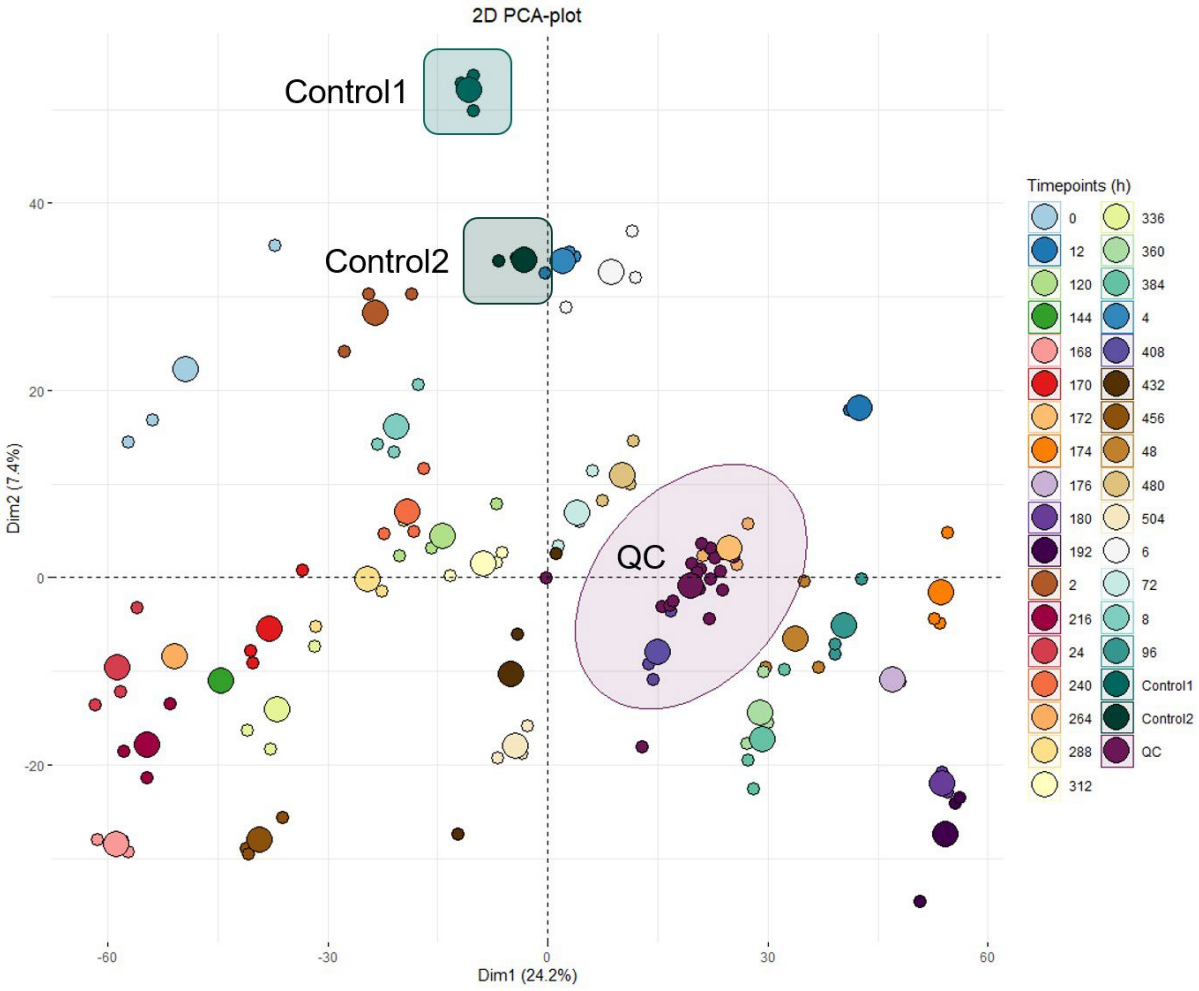


Figure 5-2: A two-dimensional ellipses principal component analysis (PCA) plot containing a representative dataset from horse 6 with 32 time points (h) of urine samples analysed in technical triplicates. The three smaller points of a single colour indicate the replicate data points, while the larger point of the same colour signifies the mean response. Pooled QC samples are noted in burgundy/purple, and their clustering is indicated with a circle. Control 1 and Control 2 samples are coloured and circled in turquoise green and deep green, respectively.

The *k*-means clustering of the data is represented by the scatterplot (Figure 5-3) separating sulfated (orange dots, $n = 859$) and non-sulfated (black dots, $n = 3796$) features labelled during data pre-processing. Separation of features occurs due to the predictability of sulfated fragmentation and is well represented by the calculated characteristics MA and IR. As mentioned above, high values of MA and IR for a feature imply a higher likelihood of a sulfated feature, whereas lower values imply the opposite. From here, each feature is assigned a cluster label 1 or 2, indicating sulfated or non-sulfated, respectively, to assist in the separate analysis of sulfated features in the proceeding steps.

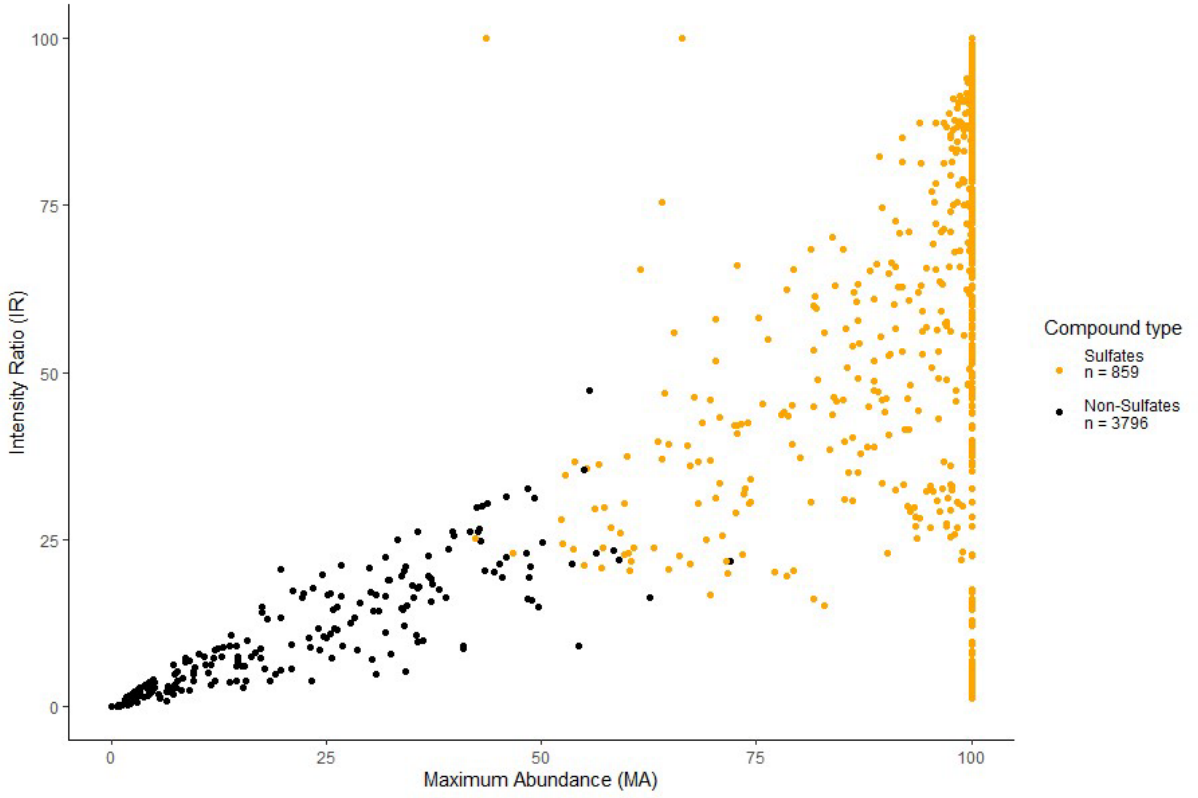


Figure 5-3: A k-means clustering scatterplot separating sulfated (orange dots, n = 859) and non-sulfated (black dots, n = 3796) features labelled during data processing. Separation of features is defined by the maximum abundance (MA) and intensity ratio (IR) calculated during the pre-processing of the dataset.

Volcano plots aid the visual identification of significantly changed features (from pre-administration (0 h) samples) in the metabolic dataset. In Figure 5-4, plot A includes all features identified by MS-DIAL in the pre-processing of the data, whilst plot B only displays sulfate-labelled data points. Non-significant values denoted in grey have a log-transformed adjusted *p*-value of less than 1. Blue-coloured features have an adjusted *p*-value greater than 1 and a fold change (FC) value less than ±8 (adjusted log FC less than ±3) and are classified as mid-significant. Red-coloured features are considered significant with an FC value greater than ±8 (adjusted log FC greater than ±3).

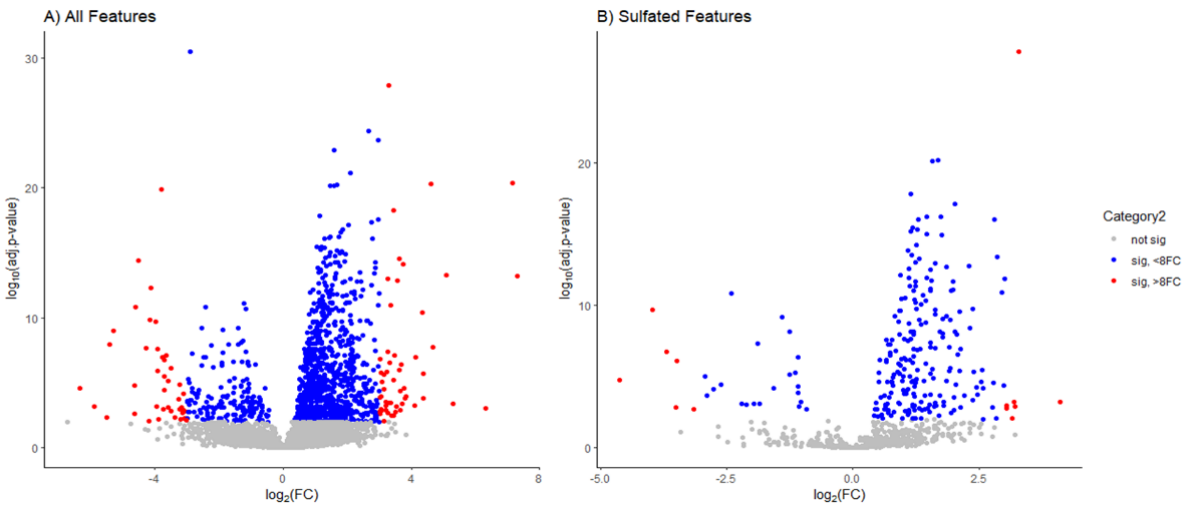


Figure 5-4: Volcano plots for visual identification of significantly changed features in the metabolic dataset. Plot (A) is inclusive of all features identified in the pre-processing of the data, whilst plot (B) only concerns sulfate-labelled data points.

In this example (Figure 5-4), plot A identifies 3567 non-significant, 993 mid-significant, and 95 significant features. The significant features are inclusive of 41 down-regulated (FC > −8) and 54 up-regulated (FC > +8) features of interest. Plot B shows 615 non-significant, 232 mid-significant, and 13 significant features, with the significant grouping containing six (6) down-regulated and seven (7) up-regulated features for further investigation.

The *m/z* of all sulfate-labelled features were then compared to a predetermined list of theoretical steroid sulfate compounds (26 in total), mostly endogenous and present in the steroid hormone metabolism pathway. From the 859 sulfated features extracted from the *k*-means scatterplot, the 'if else' search function identified 31 potential features that had a Δ*m/z* within ±5 ppm to 14 of the listed metabolites. It is worth noting that while multiple features can be preliminarily allocated the same theoretical metabolite, manual screening of the output was then conducted to investigate collected MS/MS and retention time for isolation of most-likely true identifications. This manually filtered list of theoretically identified features was then compared across all administration horses to account for consistent metabolite presence. This gave confidence that the selected compounds for classification modelling would be reliably identifiable in any unrelated urine sample.

Table 5-1 presents a combined list of selected potential sulfate-conjugated steroid metabolites extracted from the steroid sulfate searcher function, as explained above. The volcano plot results of these features were considered during the selection process; however, some more up-regulated features from the plots were omitted due to the lack of consistent appearance in all six horse administrations or their non-steroid-like structure, as indicated by MS data, including MS/MS fragmentation.

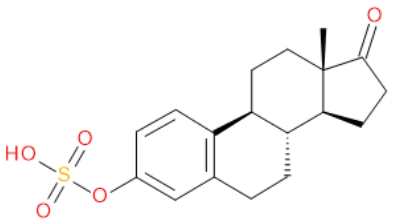
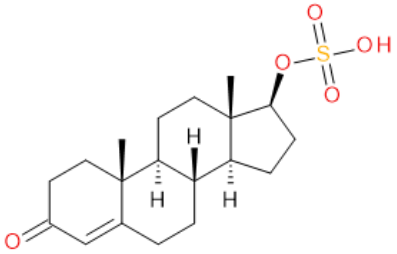
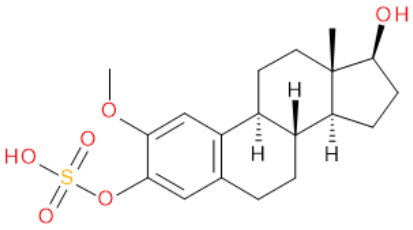
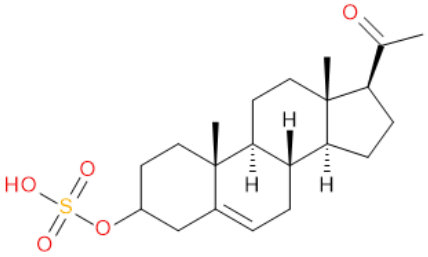
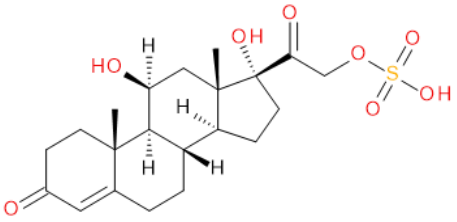
Table 5-1: Results obtained from the steroid sulfate searcher function included in the statistical processing R script. Deprotonated ions are given (in m/z) and their assigned retention times (in minutes).

	Deprotonated Molecule *(m/z)*	**Retention Time (min)**
(1)	349.11209	9.333
(2)	367.15909	10.108
(3)	381.13840	8.462
(4)	395.19061	12.454
(5)	441.15961	4.963

5.3.2 Compound Identification

Verification of feature metabolite identification is necessary to proceed with the biomarker proposal/investigation. Shimadzu's Insight Explore Assign feature was utilised to gain confidence in putative identifications before reference standards were purchased and/or synthesised (by the author or provided by C.C.J. Fitzgerald [148, 150]) for MS/MS fragmentation and retention time matching. Spectra acquired from an administration sample was compared to the theoretical fragmentation of 1000 potential compounds with matching *m/z*, extracted from ChemSpider, and returned a similarity score report for these potential identifiers based on the proposed MS/MS fragmentation. Mass accuracy tolerance could be adjusted and was set at $\Delta m/z \pm 5$ ppm for this investigation. From Table 5-1, features (1), (2), (3), (4), and (5) were putatively identified as estrone sulfate (E1S), testosterone sulfate (TS), 2-methoxyestradiol sulfate (2-ME2S), pregnenolone sulfate (PregS), and cortisol sulfate (CS), respectively (Table 5-2), using the steroid sulfate searcher function, and in Assign, returned convincing similarity reports to corresponding proposed MS/MS fragmentation.

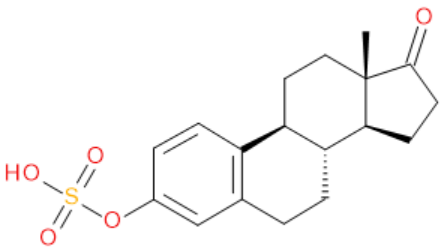
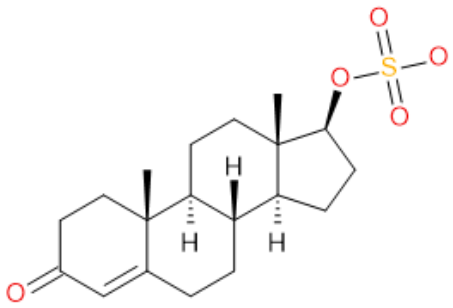
Table 5-2: Putatively identified compounds with chemical formulas and structural information.

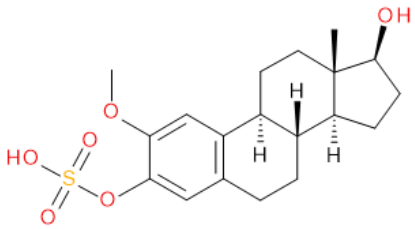
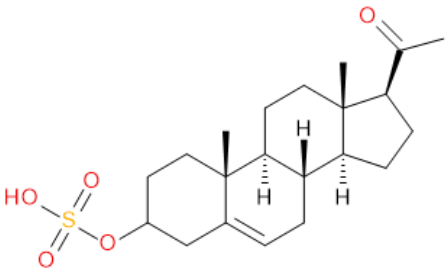
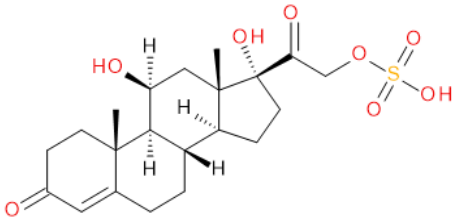
	Compound	Formula	Structure
(1)	Estrone Sulfate (E1S)	$C_{18}H_{22}O_5S$	
(2)	Testosterone Sulfate (TS)	$C_{19}H_{28}O_5S$	
(3)	2-Methoxy-Estradiol Sulfate (2-ME2S)	$C_{19}H_{26}O_6S$	
(4)	Pregnenolone Sulfate (PregS)	$C_{21}H_{32}O_5S$	
(5)	Cortisol Sulfate (CS)	$C_{21}H_{30}O_8S$	

Whilst Assign is beneficial in providing an indication of potential compound similarity from unknowns, these results are hypothesis-based and require further investigation with reference standards to confirm results. Synthesised reference standards of all features were obtained and analysed in matching conditions on the same instrument to determine true *m/z*, RT, and likely fragmentation patterns.

Table 5-3 provides the found *m/z*, RT and some characteristic MS/MS fragmentation observed from all synthesised steroid compounds (synthesised E1S, TS, and PregS were provided by Christopher C.J Fitzgerald and Malcom McLeod). These compounds were analysed neat (without urine matrix) to provide MS/MS, so a small amount of RT shift can, therefore, be expected. Detailed methodology of the synthesis of these compounds and similar compound types has been previously published and can be obtained from **Section 5.2.4**, and the referenced articles' supplementary material [148, 150].

Table 5-3: HRMS (-ESI) details for obtained synthesised reference compounds. Retention time (given in minutes) originates from synthesised samples spiked in blank urine. Additional information can be found in Appendix 21.

	Reference Compound	Structure	Found *m/z* and RT (min)	(-)ESI MS/MS (*m/z*)
(1)	Estrone Sulfate (E1S) $C_{18}H_{22}O_5S$		*m/z* 349.1106, 9.203 min	81.21106 145.06341 183.73195 269.15411 349.11068
(2)	Testosterone Sulfate (TS) $C_{19}H_{28}O_5S$		*m/z* 367.1576, 10.090 min	79.95511 96.95901 177.02073 273.87398 352.13628 367.15753

(3)	2-Methoxy-Estradiol Sulfate (2-ME2S) $C_{19}H_{26}O_6S$		*m/z* 381.1420, 8.462 min	79.95812 95.95404 96.96006 109.02649 137.02532 337.4573
(4)	Pregnenolone Sulfate (PregS) $C_{21}H_{32}O_5S$		*m/z* 395.1888, 12.454 min	80.41009 96.95900 271.53139 290.36310 395.18897
(5)	Cortisol Sulfate (CS) $C_{21}H_{30}O_8S$		*m/z* 441.1639, 4.963 min	60.11081 96.96049 411.15526

5.3.3 Classification Modelling

The use of a RF classification model was tested with the five mentioned features, comparing one oral (*n* = 90) horse dataset and one IM (*n* = 93) horse dataset as the best-case scenarios, and resulted in strong differentiation between the oral and IM administration routes. This combined dataset totalled 183 normalised peak area data points incorporating 61 post-administration urine samples (i.e., time zero/pre-administration samples removed), analysed as technical triplicates. The non-uniform sample size per horse is due to the orally administered horses' missing time point of 480 h. All model types were trialled in MetaboAnalyst's Biomarker Analysis tool, including linear SVM and PLS-DA in addition to RF (Table 5-4). However, RF was found to be the most successful, with higher prediction accuracy results compared to SVM and PLS-DA. The dataset used for model testing has been provided in Appendix 22.

Table 5-4: A comparison of Area Under the Curve (AUC) and predictive accuracy scores obtained from Support Vector Machines (SVM), Partial Least-Squares Discriminant Analysis (PLS-DA), and Random Forest (RF) model testing using one oral administered horse sample set and one intramuscularly injected (IM) administered horses sample set.

	Tests	AUC Score	Predictive Accuracy
SVM	1	0.956	86.7%
	2	0.940	84.4%
	3	0.953	86.1%
PLS-DA	1	0.900	83.1%
	2	0.891	82.4%
	3	0.883	81.5%
RF	1	0.965	89.6%
	2	0.970	89.5%
	3	0.973	89.8%

These five potential markers, E1S, TS, 2-ME2S, PregS, and CS, were selected based on their consistent presence in the equine steroid hormone metabolic pathway and the direction of log_2-fold change values calculated in the previous statistical analysis. Within MetaboAnalyst, the dataset was filtered for any missing values (which, by default, were replaced with a value of 20% of the minimum peak intensity recorded for that feature). No data scaling was applied. Data transformation was also not selected due to the model type (RF) used in this study.

Classical ROC curve analysis of 2-ME2S (Figure 5-5), TS (Figure 5-6), E1S (Figure 5-7), CS (Figure 5-8), and PregS (Figure 5-9), with AUC scores and 95% confidence level in parentheses on plot are presented below. Boxplots are also included to the right of the ROC curve and indicates the abundance spread of each individual feature in the oral (0) and IM (1) groups, respectively.

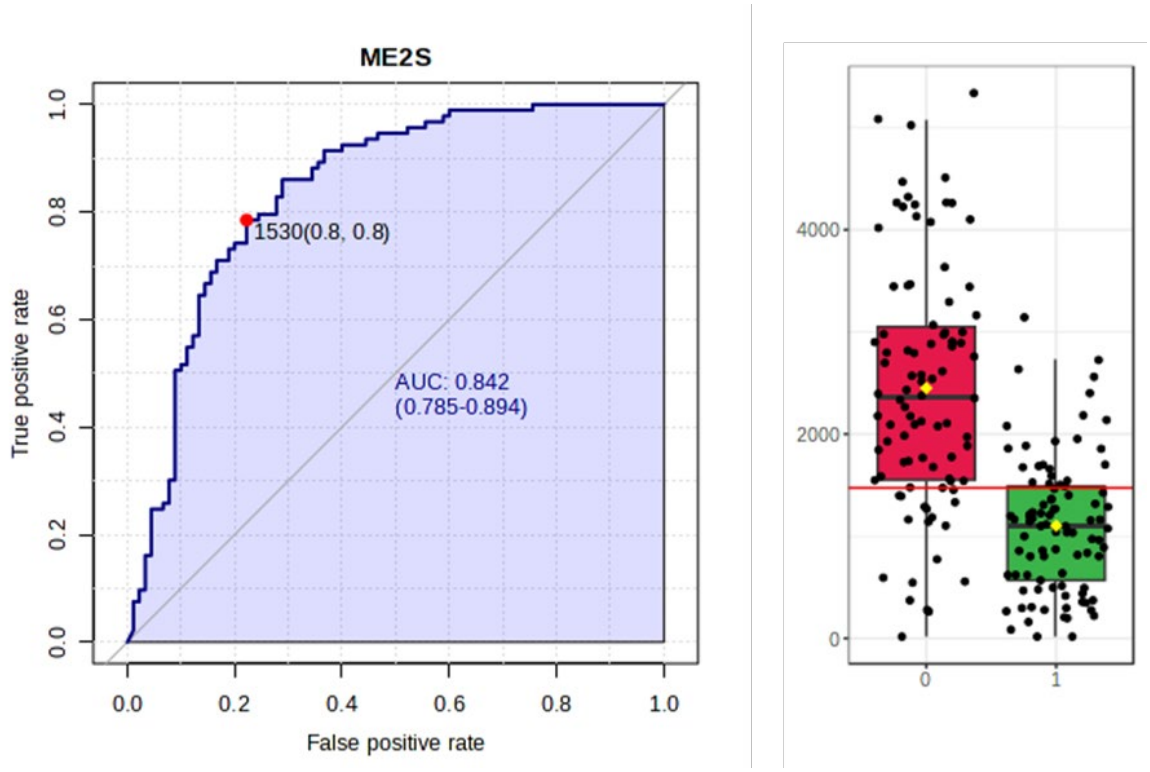


Figure 5-5: Classical ROC curve and box plot for 2-methoxy-estradiol sulfate (2-ME2S). Area under the curve (AUC) score is given in the ROC plot on the left, and box plot on the right presents oral (0) samples in red and intramuscularly injected (IM, 1) samples in green.

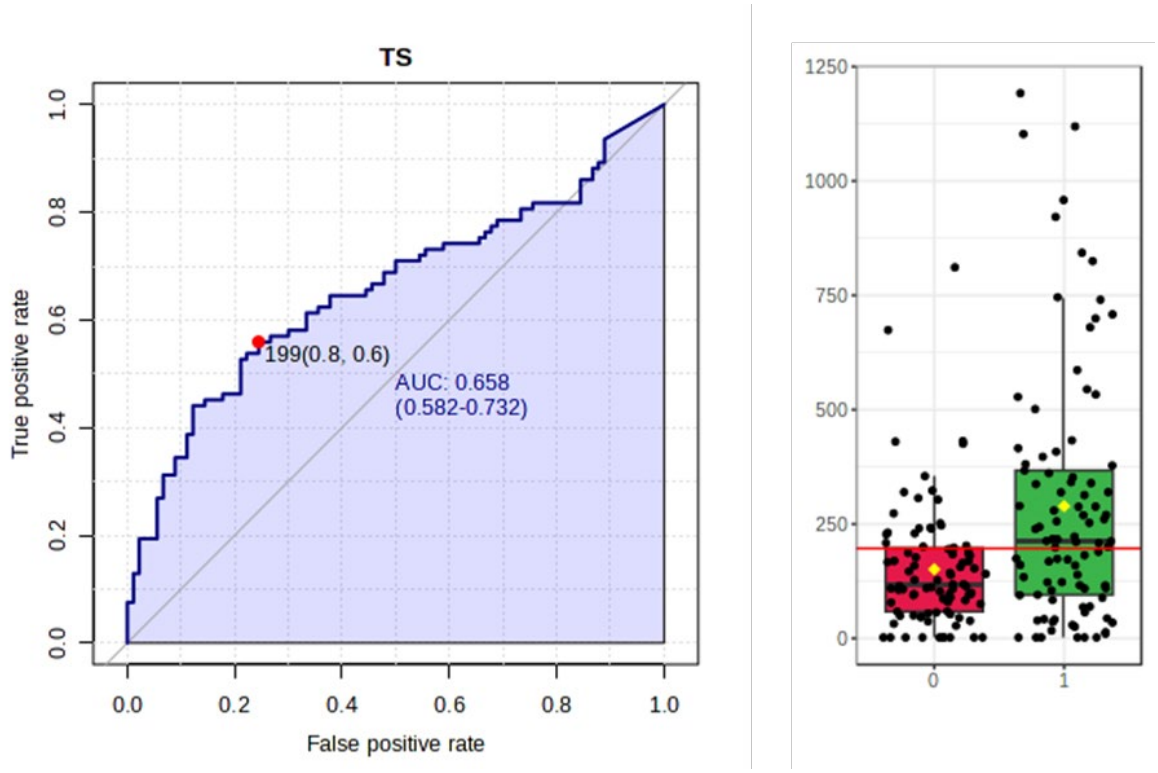


Figure 5-6: Classical ROC curve and box plot for testosterone sulfate (TS). Area under the curve (AUC) score is given in the ROC plot on the left, and box plot on the right presents oral (0) samples in red and intramuscularly injected (IM, 1) samples in green.

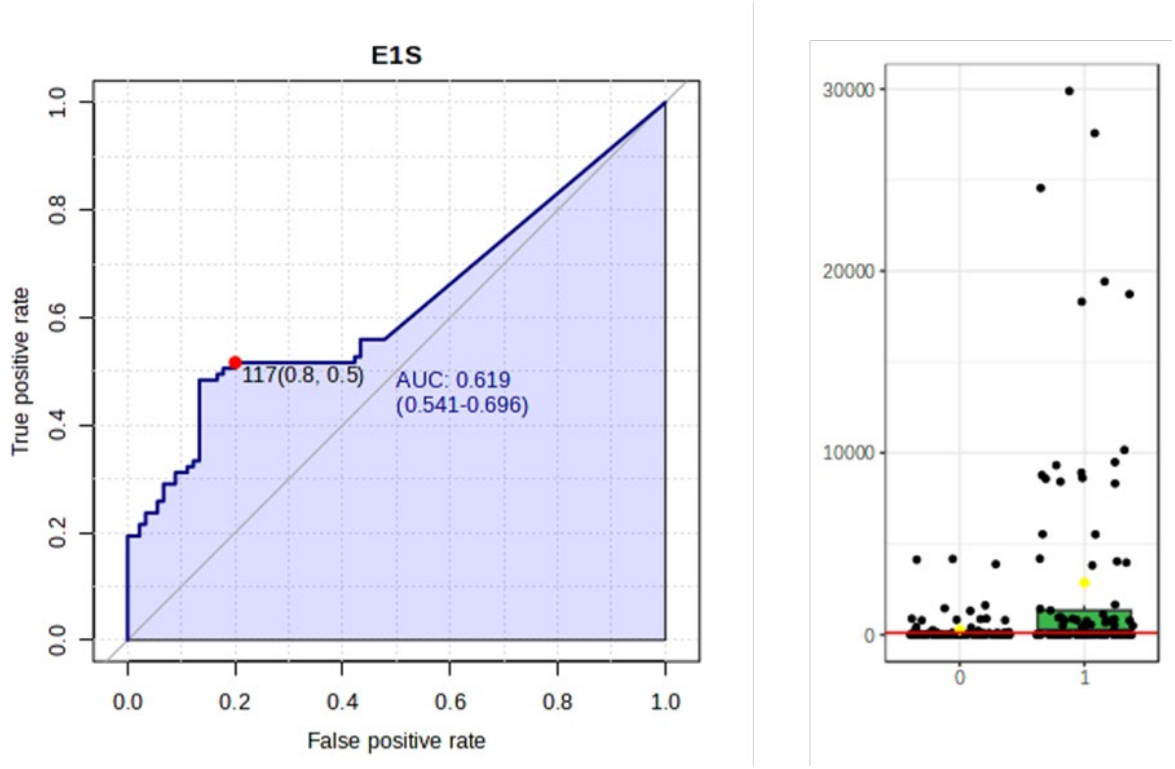


Figure 5-7: Classical ROC curve and box plot for estrone sulfate (E1S). Area under the curve (AUC) score is given in the ROC plot on the left, and box plot on the right presents oral (0) samples in red and intramuscularly injected (IM, 1) samples in green.

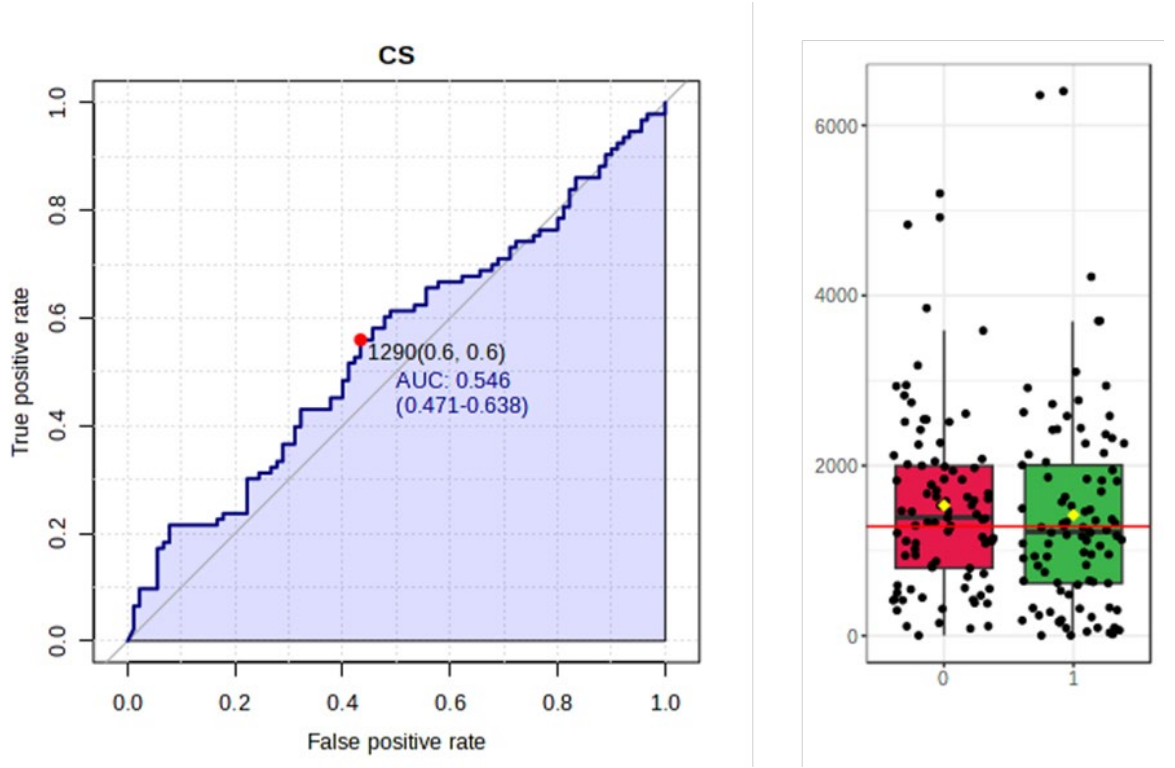


Figure 5-8: Classical ROC curve and box plot for cortisol sulfate (CS). Area under the curve (AUC) score is given in the ROC plot on the left, and box plot on the right presents oral (0) samples in red and intramuscularly injected (IM, 1) samples in green.

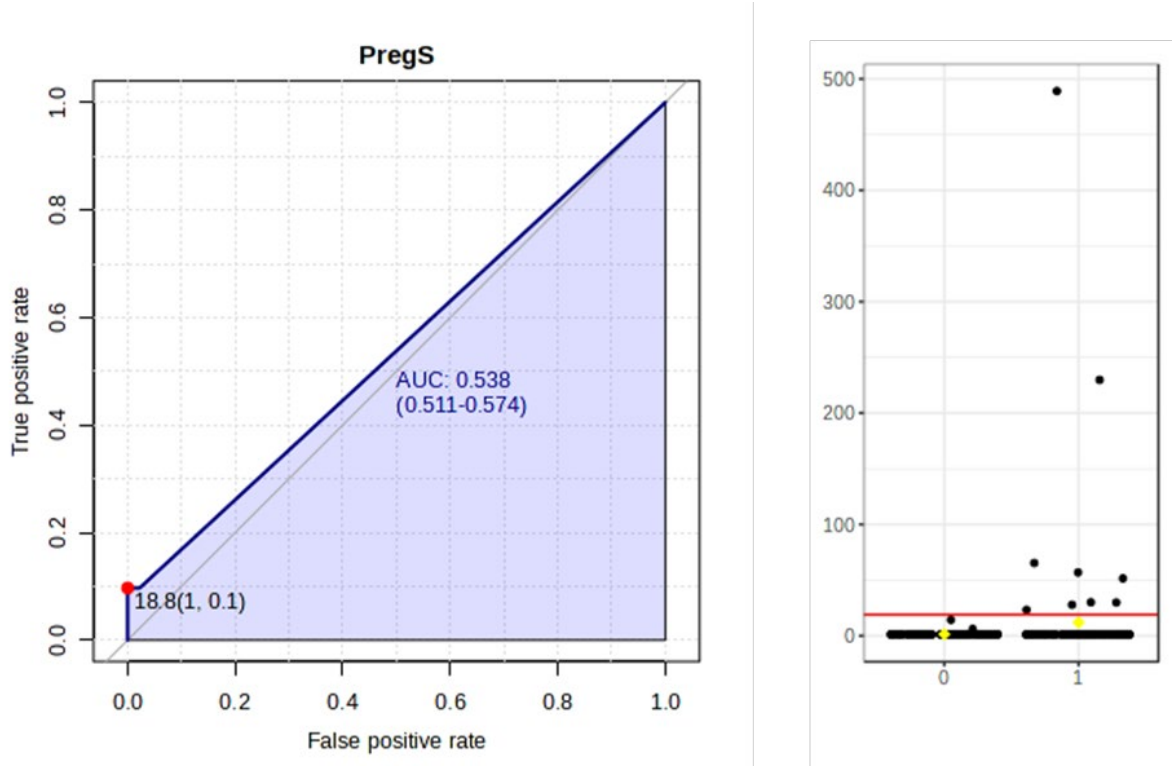


Figure 5-9: Classical ROC curve and box plot for pregnenolone sulfate (PregS). Area under the curve (AUC) score is given in the ROC plot on the left, and box plot on the right presents oral (0) samples in red and intramuscularly injected (IM, 1) samples in green.

Table 5-5 presents the classical univariate analysis of each compound and its individual biomarker potential. Of the five features, 2-ME2S produced the highest AUC result of 0.840. This AUC score implies that, on its own, 2-ME2S would provide reasonable classification between oral and IM samples.

Table 5-5: Univariate receiver operating characteristics (ROC) curve statistical results output from the MetaboAnalyst Biomarker Analysis function.

	AUC	***p*-Value**	**Log_2FC**	**Clusters**
2-ME2S	0.840	1.70×10^{-17}	1.15	1
TS	0.657	2.82×10^{-5}	−0.936	4
E1S	0.620	9.62×10^{-5}	−3.35	3
CS	0.550	5.05×10^{-1}	0.109	2
PregS	0.538	7.95×10^{-2}	−3.05	4

Table Note: Statistical information (rounded down to three significant figures) is listed as AUC (area under the curve) scores, p-values (also titled t-test results in MetaboAnalyst Report), Log_2FC (log fold change), and cluster identifier numbers denoting which features are clustering together/considered to behave in a similar manner in the dataset. Abbreviations used for listed biomarkers are 2-methoxyestradiol sulfate (2-ME2S), testosterone sulfate (TS), estrone sulfate (E1S), cortisol sulfate (CS), and pregnenolone sulfate (PregS).

The inclusion of all five features in a multivariate RF model, however, allows for greater robustness in the model. This is shown by the ROC curve in Figure 5-10. The dark blue line in the plot is inclusive of all five features and gave an AUC score of 0.965 with a confidence level of 95% (between 0.931 and 0.995). This translates to a predictive accuracy of 86.7%. This ROC curve is generated through Monte-Carlo cross validation (MCCV), where 66.6% (two-thirds, 2/3) of the dataset are used to evaluate feature importance (i.e., train) in the model, and the remaining 33.3% (one third, 1/3) of the data points are used to validate the model [87, 153]. The ranking method used by MetaboAnalyst in this model was RF, 'mean decrease in accuracy', with a subsampling (nRuns) value of 30. The number of trees created in this RF model was set at a default of 300. Reference [154] provides more information on the RF model implemented in MetaboAnalyst and links to the open-access GitHub© page with MetaboAnalystR 3.0 script.

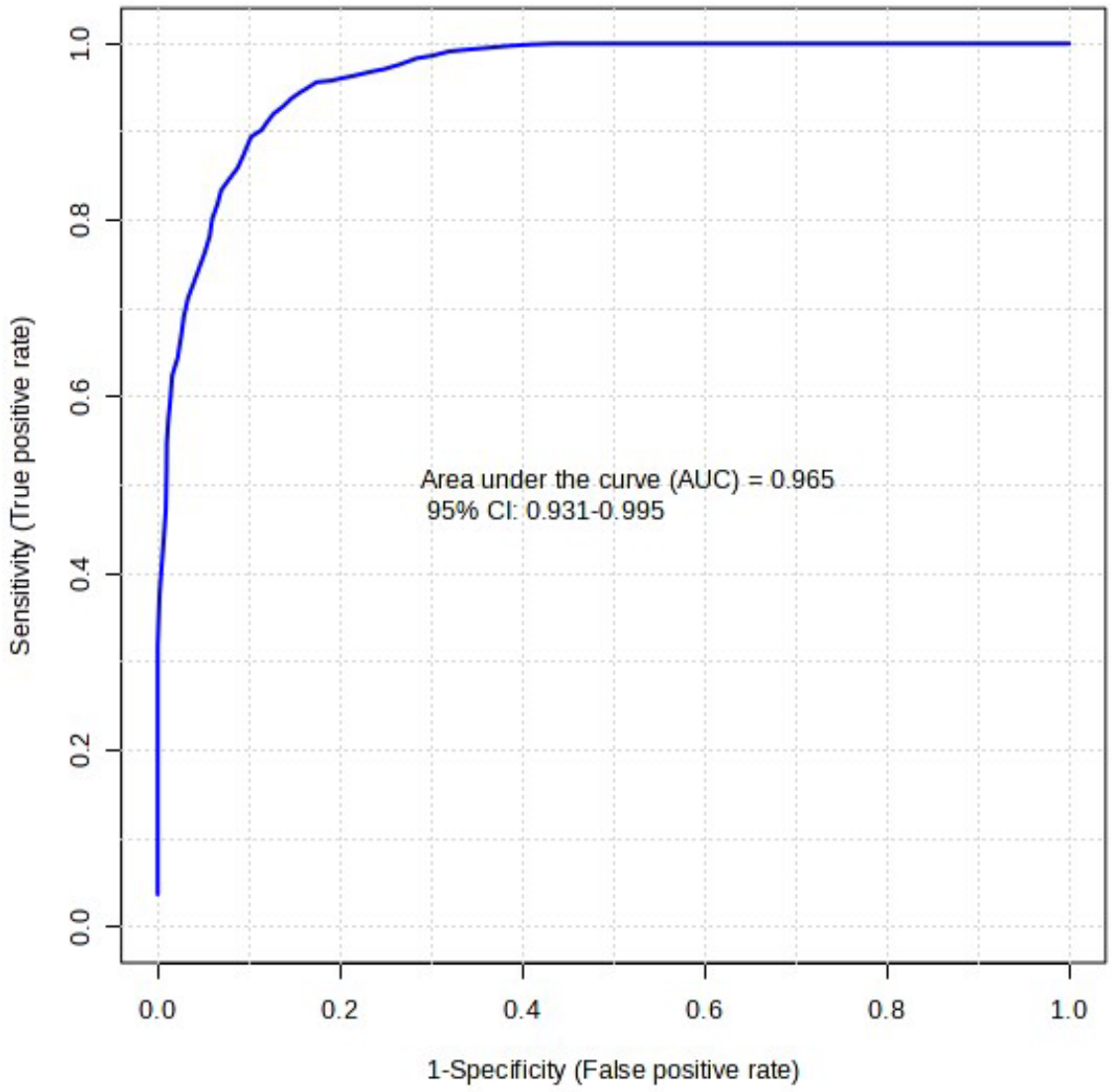


Figure 5-10: ROC curve plot for a random forest analysis of five biomarkers of interest. The dark blue line is inclusive of all five features, giving an AUC score of 0.965 with a confidence level of 95% (between 0.931 and 0.995).

From 93 IM samples and 90 oral samples (183 total samples), the model was able to correctly identify IM administration in 83 out of 93 samples (Table 5-6). This correlates to a true positive result of 89%. Likewise, oral administration was accurately identified 75 times out of 90 occurrences, which corresponds to a true negative score of 83%.

Table 5-6: Confusion matrix results from the multivariate ROC curve. The dataset consisted of 183 samples, with 93 intramuscular (IM) administration samples (positive outcome) and 90 oral administration samples (negative outcome).

		Predicted	
		Positive (IM)	**Negative (Oral)**
Actual	**Positive (IM)**	83	10
	Negative (Oral)	15	75

5.3.4 Model Verification

A blind control sample from a past case that returned a positive screening result for ALT was analysed against this RF sulfate biomarkers model. Using the new samples function in MetaboAnalyst, the new sample data was included in the original dataset, without administration information. The sample was then classified by the model, based on the same five features as the original dataset. The unknown sample was extracted and analysed in triplicate for reliability, and all samples returned a prediction of class 1, indicating against an allowed oral administration. This prediction also returned a probability score of 0.61, demonstrating prediction confidence above 50% for this sample.

5.4 DISCUSSION

The RF model presented above, involving five biomarkers, E1S, TS, 2-ME2S, PregS, and CS, showed promising classification of IM ALT administration distinct from oral administration of ALT. The selected group of biomarkers may be considered advantageous in targeting various areas of endogenous steroid metabolism and exogenous steroid excretion and influence (Figure 5-11). This work should be considered a preliminary study, due to the limited sample size of the test dataset, with further validation required from additional administration studies and/or equine athlete (race day) sampling scenarios.

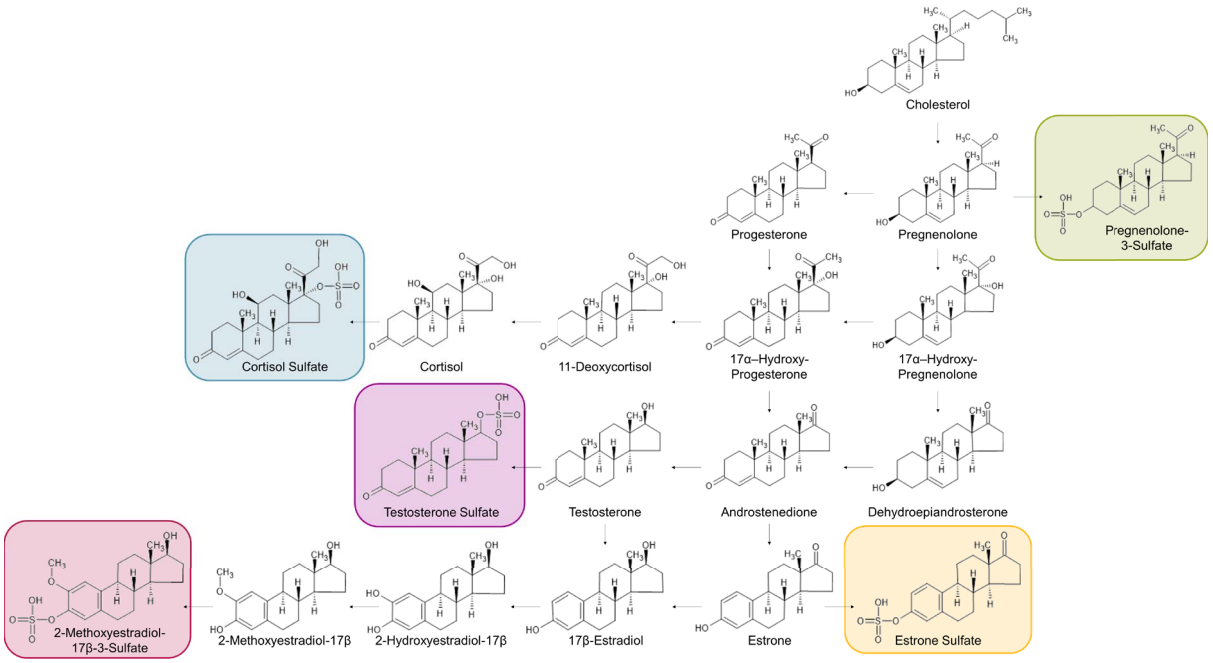


Figure 5-11: Steroid hormone excretion pathway. The five biomarkers incorporated in the random forest (RF) model, estrone sulfate, testosterone sulfate, 2-methoxyestradiol sulfate, pregnenolone sulfate, and cortisol sulfate, are highlighted.

RF modelling (Figure 5-12), developed by Leo Breiman in 2001, is a group of regression trees made from the random selection of samples of the training data [155, 156]. According to previous literature, RF has the advantages of high prediction classification accuracy, strong generalisation, and fast training speed, and can be used for classification, regression, and feature importance analysis. RF is often considered a more robust tool when compared to other similar tree-based classifier models (DTs, also known as classification and regression trees (CART)). Recently, RF has become popular as a biomarker detection tool in various metabolomics studies as it boasts the strength to deal with missing data and overfitting issues [157, 158]. Additionally, it can also tackle high-dimensional data sets without requiring feature elimination [158].

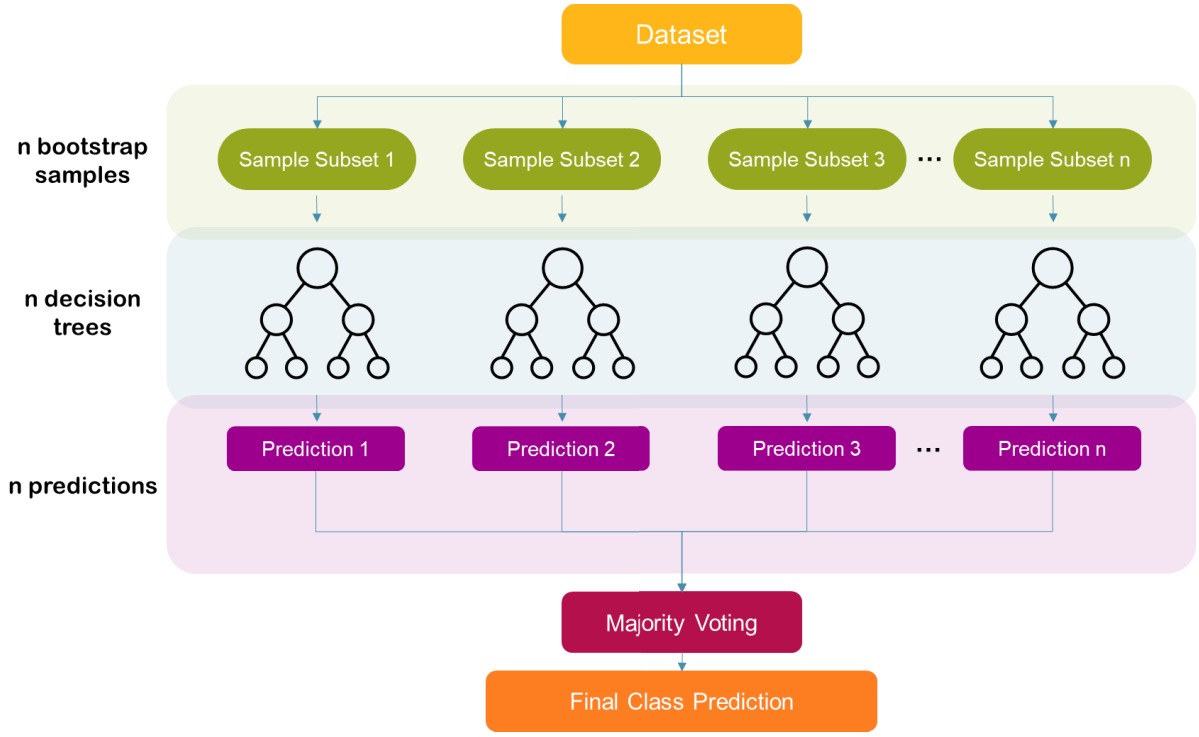


Figure 5-12: Random Forest (RF) model workflow.

There are many previously discussed benefits of implementing a multi-tool workflow into metabolomic study for biomarker discovery [91]. The primary goal in this scenario is to provide a preliminary screening tool of the collected metabolomic data for markers of interest. Additionally, these tools automate tedious and time-consuming tasks when processing and statistically analysing an extensive dataset [91]. Incorporating a workflow also allows the ability to target a specific area of the metabolomic pathway, if this is desired, with relative ease. With the increased popularity of this processing and analysis technique over the last decade, additional tools have become available to a wider research demographic with open-source accessibility, as well as increased customisation of already published script-based tools for optimisation to a specific research question [148]. In relation to the dataset presented above, MS-DIAL, custom Python and R scripts, and MetaboAnalyst formed the multi-platform workflow utilised to acquire the presented results.

This work has shown that the analysis of phase II conjugated steroids may provide superior detection of an exogenous AAS doping scenario, where the use of free or phase I metabolites to detect doping has been limited in success [159, 160]. Prior methods utilised enzyme or chemical hydrolysis techniques that aimed to cleave the phase II conjugated sites, leaving only the free steroid to detect [148]. However, it has been understood that hydrolysis is not ideal in many scenarios where the target compound structures may also metabolise further as a result of the added enzyme or chemical process, leading to potential inaccuracies in reported

compound concentrations [127, 161, 162]. In part, as a result of advancements in analytical analysis techniques, and the understanding that the majority of steroids are excreted from human and animal as their phase II conjugated form, there has been an increase in published methodologies favouring the qualitative and quantitative analysis of phase II conjugated steroids via LC-MS [145, 163, 164]. This work focussed on targeting sulfate conjugated steroids as it is understood to be the dominant excretion pathway in the equine species. Sulfated compounds are also believed to be slightly delayed in their metabolism and may be excreted after the glucuronide conjugate equivalent, leading to extended detection windows with a sulfated biomarker [165-167].

After the initial screening of metabolomic data, as demonstrated in this work, it is necessary to verify the identity of the potential biomarker, or group of markers, extracted. There is a range of tools available to assist at this stage of putative identification before final confirmation with purchased reference standards. MS spectral databases such as Human Metabolome Database (HMDB, https://hmdb.ca), MassBank of North America (MoNA, https://massbank.us), ChemSpider (https://chemspider.com), and National Institute of Standards and Technology (NIST) Chemistry WebBook (https://webbook.nist.gov/chemistry) are all easily accessible and relatively user-friendly resources. MS annotation tools are also an alternative that investigates the untargeted dataset directly, typically to extract fragmentation patterns relating to a subset of compounds predetermined by the user [19]. This study utilised the Assign feature offered by Shimadzu to add weight to the suspected identification of features before purchasing and synthesising the appropriate reference standards for confirmation.

Identified compounds should then be subjected to robustness testing to assess their suitability as a biomarker in the desired context. Consistency and reproducibility in their presence or absence in appropriate samples (influenced by oral or IM administrations) and detectable peak intensity are necessary for a successful and reliable biomarker [4, 19]. While this group of steroids has been proposed as having the potential to discriminate between an oral and IM administration of altrenogest in mares, this trial was conducted on a limited dataset of administration horses. Ultimately, additional testing is required before implementation would be possible. Such testing includes application to race day samples collected from active racehorses, as it has been previously established that it is possible for athletic horses to have differing steroid excretion profiles to stationary/grazing horses [111].

The use of a group of biomarkers is largely considered more favourable over a single marker detection or a biomarker ratio of two features. The adoption of a multivariate analysis over a uni- or bivariate presents some advantages to solving complex issues, due to its own more

complex nature [91]. This does, however, also act as a limitation to analysts with beginner or intermediate computational and coding skills, which, in developing, can cost the user large amounts of time, especially when grasping more “complex” statistical/machine learning and modelling tools. Additionally, the choice of biomarkers identified in this study provides a broad overview of the excreted urinary steroid hormone metabolism pathway previously acknowledged in the equine species [148].

Previously, research has focused on the anabolic and androgenic nature and pharmacodynamic effects of ALT on male, female, and gelded horses [43, 45]. However, that has not addressed the detection of ALT based on the type of administration route, which has recently become a focal point in the equine industry. The differentiation between ALT administration types from an excreted urine sample will greatly benefit the thoroughbred racing industry, due to the nature of race day sampling, ALT as an allowed therapeutic substance, and the complicated situation regarding the steroid impurities present in commercial products.

5.4.1 Model Verification Discussion

It should be recognised that the probability score (0.61) assigned to the unknown sample, from this model is considerably lower that the score given to the same samples in the XGBoost progesterone ratio classification model, where a score of 0.9799 was obtained. This discrepancy between the scores given is due to a number of differences between the model types and the datasets used for training the models. With this in mind, both models have indicated that the samples administration route was not consistent with an allowed oral administration that would be in line with the guidelines put in place by Racing NSW.

5.4.2 Study Limitations and Future Work

A fluctuation in the number of steroid sulfates captured with the coded steroid sulfate search function and in prior *k*-means clustering could indicate variability in the number of features captured between administration horses and should be considered a limitation of this study. This is most likely due to any small sample preparation variability or a change in wait time between extraction and instrument availability (due to instrument failures), for which extracted samples were stored at 4 °C until they were able to be analysed.

This work should be considered a preliminary study with further validation required from external administration studies, preferably with a larger quantity of collected samples from a variety of horses to validate inter- and intra-horse variability, and/or equine athlete (race day) sampling scenarios. Ideally, this model needs to be verified with samples unrelated to the batch used to identify the five chosen features.

5.5 CONCLUSIONS

A metabolomic approach to the differentiation of two equine administration routes of the steroidal progestin altrenogest was investigated. Liquid chromatography high-resolution mass spectrometry analysis of equine urine samples found five sulfated compounds with the potential to differentiate between oral and intramuscularly injected altrenogest administration using a random forest classification model. Using estrone sulfate, testosterone sulfate, 2-methoxyestradiol sulfate, pregnenolone sulfate, and cortisol sulfate, the best model results gave an AUC score of 0.965 with a confidence level of 95% (between 0.931 and 0.995). These compound identifications were confirmed with assistance from the Shimadzu Insight Explore Assign feature, as well as MS/MS spectrum and retention time matching with purchased and synthesised reference standards.

Part II: Levodopa Administration Study

Chapter 6: Quantitative Analysis of Dopaminergic Manipulation

This chapter focuses on identifying levodopa misuse via the optimisation of a (LC-HRMS) method for the urinary quantification of 3-methoxytyramine (3-MT) and tyramine (Tyr) in a *Stalevo*® (levodopa, carbidopa, entacapone) administration study.

Chapter 7: Multivariate Approach for Dopaminergic Manipulation Detection

This chapter continues the investigation of levodopa misuse using a metabolomics approach to search for more robust urinary biomarkers as their intact sulfate and glucuronide conjugated forms, in addition to the free molecule. This chapter also details the potential for a machine learning model to extend the detection window for dopaminergic manipulation.

Chapter 6: Quantitative Analysis of Dopaminergic Manipulation

6 QUANTITATIVE ANALYSIS OF DOPAMINERGIC MANIPULATION

6.1 INTRODUCTION

Stalevo®, a L-DOPA containing therapeutic drug, designed to treat PD in human patients, has been suspected to be used in the equine racing industry intended for dopaminergic manipulation. PD is defined as a movement and mood disorder, which occurs gradually as the brain slows and ceases its ability to naturally produce dopamine. There are many forms of PD treatments on the market, however, L-DOPA containing drugs are one of the most common and notably most effective treatment methods (Appendix 23). This is because L-DOPA is the direct precursor of dopamine and is able to pass through the blood brain barrier (BBB) to artificially induce dopamine production as seen in Figure 6-1.

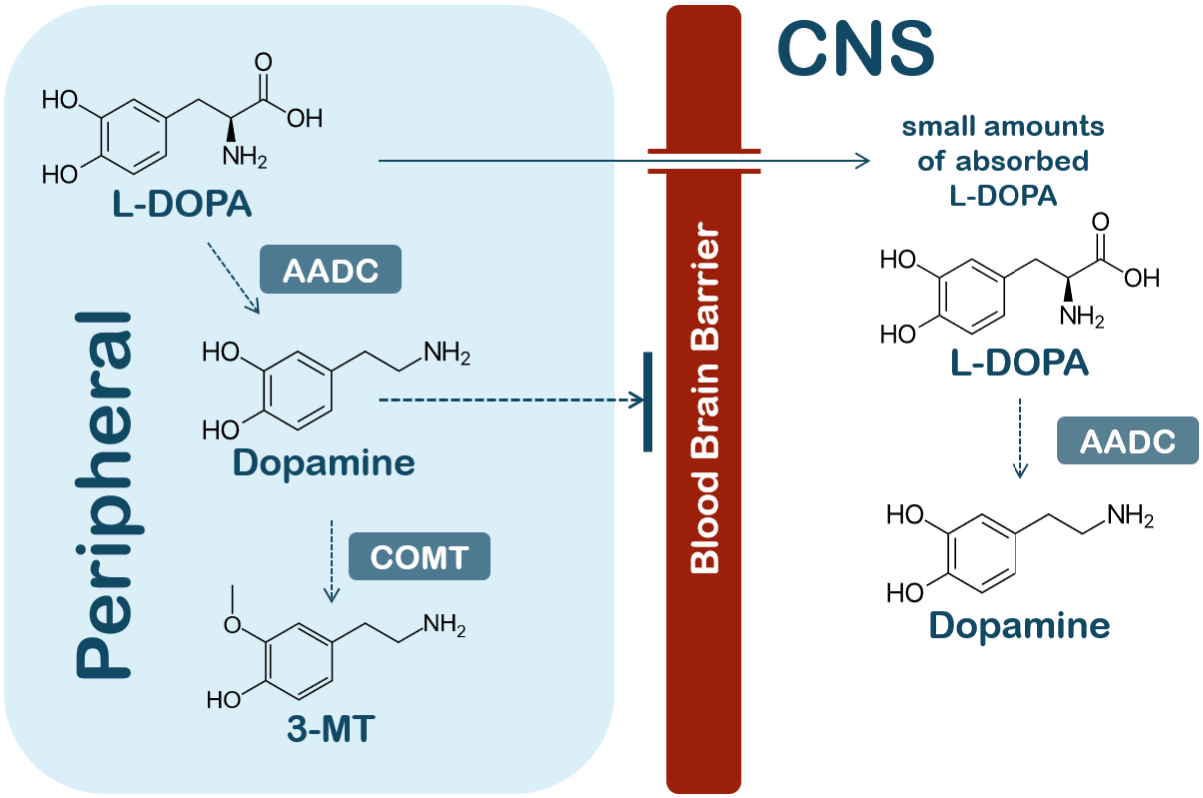


Figure 6-1: Transformation of L-DOPA to dopamine in the body. Acronym definition; L-DOPA for levodopa, 3-MT for 3-methoxytyramine, CNS for central nervous system, AADC for aromatic amino acid decarboxylase, COMT for catechol-O-methyltransferase.

The first known observation of dopaminergic manipulation in the racing industry is assumed to have occurred in 1996, where five post-race urine samples contained suspiciously high

levels of 3-MT following GC-MS analysis at RASL in Victoria. At this time, there were also growing concern over the advancements in L-DOPA containing therapeutic medications, and subsequent increased accessibility of these products. In response to this, by 2004 an international threshold for 3-MT had been established by Wynne *et al.*, to detect dopaminergic manipulation in racehorses [2, 54]. This threshold at 4 µg/mL (4,000 ng/mL) has the capability to correctly identify the administration of additional L-DOPA, however it has limitations in not being able to accurately detect more complex PD medication abuse, or L-DOPA micro-dosing within the threshold. PD medications now come in a variety of compound combinations to improve the concentration of L-DOPA that is able to pass through the BBB and avoid being metabolised peripherally and wasted [26, 53, 66].

The most common substances found in PD medications are; L-DOPA, which is a dopamine precursor, dopamine agonists (carbidopa) that mimic the effects of dopamine, COMT inhibitors, which increase the amount of L-DOPA that can work in the brain, monoamine oxidase-B (MAO-B) inhibitors to reduce the breakdown of dopamine in the brain, amantadine used by people who have developed abnormal movements (known as dyskinesia), and anticholinergic medicines to help with the tremor but which cause side effects that lead to their limited use [168].

Because of this, further research was conducted in 2020 by Elbourne *et al* to propose additional measures of identifying sub-threshold doping that had been suspected to be occurring [16]. A 3-MT limit of 0.776 µg/mL (776 ng/mL), in conjunction with a 3-MT/Tyr ratio limit of 5.3, were calculated based on a reference population study with almost 2,500 equine urine samples to acquire baseline levels of 3-MT and Tyr. This work was modelled from the study design previously presented and published by Wynne *et al.* [54-57, 59].

This chapter aims to identify L-DOPA misuse via the optimisation of an LC-MS method for the urinary quantification of 3-MT and Tyr using a 12-horse administration study with the human PD medication, *Stalevo*®. LC-QTOF-MS was used with an IMTAKT amino acid column for the purpose of monitoring catecholamines related to the L-DOPA metabolism pathway. LC-MS was chosen over the traditionally used GC-MS due to the advancements in sensitivity from newer LC instrumentation. Additionally, this was desired to negate the current need for hydrolysis and derivatisation of urine samples prior to routine GC analysis.

6.2 MATERIALS AND METHODS

6.2.1 Chemicals and Reagents

Acetonitrile (ACN) and methanol (MeOH) were purchased from Merck (Darmstadt, Germany). Potassium dihydrogen phosphate (KH_2PO_4) was purchased from LabServ provided by ThermoScientific (North Ryde, NSW, Australia). Disodium hydrogen phosphate (Na_2HPO_4) and methanolic hydrochloric acid (methanolic HCl) were obtained from Sigma Aldrich (Castle Hill, NSW, Australia). The β-glucuronidase enzyme from abalone (BG100) was manufactured by Kura Biotech (Puerto Varas, Chile) and supplied by PM Separations (Capalaba, Queensland, Australia).

The 3-methoxytyramine HCl standard used for calibration and QC was manufactured by HPC Standards (Cunnersdorf, Germany). The ISTD 3-methoxytyramine-d_4 was manufactured by CDN Isotopes (Quebec, Canada) with the deuterated sites at positions 1,1, 2, and 2. Tyramine HCl was purchased from Sigma-Aldrich (Castle Hill, Australia).

Methanolic HCl (0.1 M) was prepared by adding 10 mL 1.0 M methanolic HCl to 90 mL anhydrous MeOH and mixed. Phosphate buffer (0.1 M) at pH 5 was prepared by dissolving 54.4 g of potassium di-hydrogen orthophosphate in approximately 3 L of purified water and pH adjusted to 5.0 with 1.0 M potassium hydroxide or orthophosphoric acid 85%. This was then diluted again to 4 L with purified water and mixed.

6.2.2 Administration Study

A 12-horse administration study was completed where eight (8) *Stalevo® 100* tablets (Sandoz Pty Ltd, Australia, 800 mg L-DOPA, 200 mg carbidopa monohydrate, 1600 mg entacapone) were administered orally by nasogastric tube to each treated horse (a single treated mare (TM4) was required drenching, i.e., made to swallow tablets with a drug solution, as nasogastric tubing wasn't possible). Of the 12 horses, six (6) were mares and six (6) were geldings with two (2) control horses in each gender group (Table 6-1). Urine samples were collected at 24-hour intervals from 7 days prior to administration, immediately pre-administration (T0), then 2, 4, 6, 8, 12, 24 hours, and then in 24-hour intervals up until 7 days post-administration (Table 6-2). A total of 20 urine samples were taken per horse and stored at -20 °C until analysis.

Table 6-1: The amount of levodopa (L-DOPA), carbidopa, and entacapone administered to each horse in the administration study. Calculated by the concentration (in mg) of the compound in the eight tablets, divided by the horse body weight (in kg).

Sex	Administration Type	Abbreviation	L-DOPA (mg/kg BW)	Carbidopa (mg/kg BW)	Entacapone (mg/kg BW)
Mare (M)	Control (C)	CM1	-	-	-
		CM6	-	-	-
	Treated (T)	TM2	1.64	0.411	3.29
		TM3	1.50	0.375	3.00
		TM4	1.52	0.380	3.04
		TM9	1.53	0.382	3.05
Gelding (G)	Control (C)	CG1	-	-	-
		CG9	-	-	-
	Treated (T)	TG2	1.43	0.357	2.86
		TG3	1.47	0.366	2.93
		TG4	1.38	0.334	2.68
		TG10	1.34	0.334	2.68

Table 6-2: Time points for each urine sample taken along with the abbreviated label. Negative (-) values indicate collection occurred before the drug administration, while positive values indicate sampling after the administration was conducted.

Timepoint Abbreviation	Administration Days
-168	7 days prior to administration
-144	6 days prior to administration
-120	5 days prior to administration
-96	4 days prior to administration
-72	3 days prior to administration
-48	2 days prior to administration
-24	1 day prior to administration
0	Immediately prior to administration
1	1 hour after administration*
2	2 hours after administration
4	4 hours after administration
6	6 hours after administration
8	8 hours after administration
12	12 hours after administration
24	1 day after administration
48	2 days after administration

72	3 days after administration
96	4 days after administration
120	5 days after administration
144	6 days after administration
168	7 days after administration

**Table Note: a singular one-hour post administration sample was collected for treated gelding 10 (TG10).*

One freeze-thaw cycle was conducted to randomise the samples into 10 batches for sample extraction and analysis. A small amount (100 µL) from each sample (241 total acquired samples) was combined to create a 2 mL pooled QC sample to be refrozen with each randomised sample batch.

Analysis was completed in technical triplicates for each sample due to the small amount of urine collected at each timepoint. Ethics approval for the administration study was provided by the Charles Sturt University Animal Care and Ethics Committee (A20277), with funding for the study provided by AgriFutures Australia as part of the Thoroughbred Horses Program.

6.2.3 Sample Preparations and Instrumental Analyses

Sample extraction processes were adapted from Keen *et al.*, where an LC-MS method was optimised for dopamine related compounds analysed in equine plasma [169]. Urine (1 mL) was combined with an acetate buffer solution (1 mL) and enzyme (β-glucuronidase, 50 µL), pH adjusted to 5 – 5.5 and incubated at 37 °C for 12 hours (overnight). An aliquot (200 µL, equivalent to 100 µL of urine) of the combined urine and buffer solution was mixed with ACN (300 µL) and internal standard (40 µL, 4 µg/mL equivalent concentration) in a microcentrifuge tube (650 µL capacity, Cat. No. 3208; Costar; Salt Lake City, UT, USA). The solution was capped and vortexed before being micro-centrifuged using a Beckman Coulter Microfuge 20 (with inserts to hold tubes in place) at 18,000 x g for 5 minutes. The samples were then transferred to small glass test tubes before adding one drop (approximately 20 µL) of methanolic HCl and drying under a gentle flow of nitrogen at 50 °C. Reconstitution of the sample involved 20 mM ammonium formate in 80% ACN (200 µL), vialed, and stored at 4 °C prior to LC-QTOF-MS analysis.

LC-QTOF-MS analysis utilised an IMTAKT (Kyoto, Japan) Intrada Amino Acid column (2 mm x 100 mm, 3 µm) and a Phenomenex (Torrance, CA, USA) C18 SecurityGuard (2 mm x 4 mm) operated at 40 °C with a gradient elution, using the Shimadzu 9030 LC-QTOF-MS (Kyoto, Japan) instrument. Mobile phases include A: 100 mM ammonium formate in water, and B: 0.3% FA in 100% ACN. The gradient used was: 0 minutes A-B (25:75 v/v), 0-6 minutes A-B (30:70 v/v), 6-9 minutes A-B (100:0 v/v) with a 2-minute post run equilibration time. The flow rate was constant at 0.3 mL/min. All injection volumes were 5 µL.

Positive electrospray ionisation (+ESI) data was acquired using product ion scan mode to target compounds for quantification. The interface temperature was 350 °C with a gas flow of 11.0 L/min. The nebulising gas flow was set to 3.0 L/min. The CE spread was set at 20 ± 15 eV. Table 6-3 presents retention times, precursor ion and product ions used to identify each standard. The chemical formula of each product ion is also given.

Table 6-3: Retention times (in minutes), precursor ions $(M+H)^+$, product ions (m/z) and their associated chemical formulas for each compound of interest for the quantitative Stalevo administration detection method. ISTD denotes the internal reference standard used.

Analyte	Retention Time (min)	Flag	Precursor Ion (*m/z*) $[M+H]^+$	Product Ions (*m/z*)	Chemical Formula
3-Methoxytyramine–d_4	4.578	ISTD	172.1270	155.1001	$C_9H_7D_4O_2$
				140.0792	$C_8H_8D_2O_2$
				123.0742	$C_7H_7O_2$
3-Methoxytyramine	4.580	Target	168.1019	151.0748	$C_9H_{11}O_2$
				119.0483	C_8H_7O
				91.0540	C_7H_7
Tyramine	4.767	Target	138.0920	121.0545	C_8H_9O
				91.0544	C_7H_7
				77.0388	C_6H_5

The data generated was reviewed using LabSolutions (version 4.50 SP1), and then analysed using Insight program (pre-release version). LabSolutions Insight was utilised for alignment, integration and chromatographic processing of LC data. Microsoft Excel (version 16.42) was

incorporated to assist data amalgamation, processing, statistical analysis and timewise plotting of samples from each administration horse.

6.2.4 Method Validation Criteria

The LC-QTOF-MS method was validated for the quantification of 3-MT and Tyr. The parameters assessed were sensitivity, selectivity/specificity, linearity, accuracy, precision, recovery, and matrix effects. Validation criteria by Peters *et al*. has been followed [123-125].

Sensitivity

Sensitivity was tested by estimating the LOD; the lowest concentration of analyte in a sample that can be detected but not necessarily quantified as an exact value, and the LLOQ; the lowest amount of an analyte in a sample that can be quantitatively determined with suitable precision and accuracy.

The LOD and LLOQ were estimated visually by observing a peak signal to noise ratio of greater than 3 to 1, and 10 to 1, respectively [123].

Selectivity/Specificity

Selectivity and specificity are often used interchangeably as they can have similar meanings. However, strictly speaking, specificity refers to methods that produce a response for a single analyte, whereas, selectivity refers to methods that produce responses for a number of chemical entities [126]. Therefore, this project will be testing selectivity as multiple chemical components were analysed.

Selectivity has been defined as “the ability of the bioanalytical method to measure unequivocally and to differentiate the analyte(s) in the presence of components which may be expected to be present” [126]. Therefore, selectivity was determined by mass differentiation of the precursor and product ions and the retention time noted in minutes on the column (Table 6-3).

Linearity

Linearity was assessed between the LLOQ and the ULOQ; the maximum analyte concentration of a sample that can be quantified with acceptable precision and accuracy (bias). As a general rule, the ULOQ is identical to the concentration of the highest calibration standard. This method uses two calibration ranges to account for the expected up-regulation of 3-MT and down-regulation of Tyr following L-DOPA administration. The low calibration

range included concentration levels of 50, 250, 500, 800, and 1000 ng/mL (= 1 µg/mL). The high calibration points added for higher 3-MT levels included concentrations of 2, 4, 6, and 8 µg/mL. However, method validation was conducted on a calibration range with additional datapoints of 2, 3, 4, 5, 6, 7, and 8 µg/mL.

Accuracy

Accuracy was determined by comparing mean values from QC levels against expected values as a percentage, known as the relative error percentage. The low, medium and high QC concentration levels used in this method were 2, 4, 6 µg/mL of 3-MT.

$$Relative\ Error\ (\%RE) = \frac{calculated\ concentration - mean\ actual\ concentration}{mean\ actual\ concentration} \times 100$$

Acceptable criteria for accuracy were ± 15% RE of the accepted reference value and ± 20% near LLOQ.

Precision

This was determined by calculating the standard deviation and mean values of seven (7) equine urine spiked replicates at the QC levels. The precision is described by the %RSD equation below.

$$Relative\ Standard\ Deviation\ (\%RSD) = \frac{standard\ deviation}{mean\ concentration} \times 100$$

The acceptable limit for precision was determined to be 15% RSD for all values except for LLOQ which was 20% RSD.

Recovery

Recovery is the percentage of the analyte response after sample preparation compared with that of a solution containing the analyte at a concentration corresponding to 100% recovery. The acceptable recovery range criteria were 80% to 120%. This was assessed from seven (7)

equine urine (non-hydrolysed) spiked replicated pre- and post- protein precipitation, at LQC, MQC and HQC levels, using the equation below.

$$Recovery\ (\%) = \left(\frac{pre\ extraction}{post\ extraction}\right) \times 100$$

Matrix Effects

Matrix effects were determined by comparing the analyte response of seven (7) neat standards to seven (7) replicate post- protein precipitation spiked matrix samples. This was then calculated using the equation below.

$$Matrix\ Effects\ (\%) = \left(\frac{post\ extraction}{neat\ standard}\right) \times 100$$

The matrix effects were assessed using the theory that if the result was above 100% the compound is experiencing ion enhancement and if the result was below 100% the compound ionisation is being suppressed.

6.3 RESULTS

6.3.1 Method Validation Results

Sensitivity, linearity, accuracy, precision, recovery and matrix effects results noted in Table 6-4. LQC, MQC and HQC levels were spiked at 2, 4 and 6 µg/mL. All presented results sit within the acceptable validation criteria range, with exception to the 3-MT recovery results. However, due to the endogenous nature of 3-MT in urine, its presence in the blank urine (non-hydrolysed) used as a spiking matrix was not unexpected. These validation tests were also conducted in a water matrix to ensure spiked matrix suitability (these validation results can be found in the appendices, Appendix 24).

Table 6-4: Method validation results for the quantification of 3-methoxytyramine (3-MT) and tyramine (Tyr). Validations parameters include sensitivity, linearity (low and high calibration ranges), accuracy, precision, recovery and matrix effects.

		3-MT	Tyr
Sensitivity	LOD	< 50 ng/mL	< 50 ng/mL
	LLOQ	< 100 ng/mL	< 100 ng/mL
Linearity (R^2)	[a]Low Range	0.9960	0.9504
	[b]High Range	0.9945	0.9922
Accuracy (%RE)	LQC	-3.8	8.8
	MQC	5.4	-2.1
	HQC	11.4	-7.9
Precision (%RSD)	LQC	12.4	10.8
	MQC	3.1	3.7
	HQC	1.9	2.9
Recovery (%)	LQC	140.1	118.7
	MQC	129.5	108.6
	HQC	141.6	116.9
Matrix Effects (%)	LQC	12.2	17.9
	MQC	14.8	20.8
	HQC	14.4	19.3

Linearity Definitions: [a]Low Range refers to a calibration range of 0.05 – 1 µg/mL, whereas, [b]High Range refers to a calibration range of 2 – 8 µg/mL. LQC, MQC, and HQC spiked concentrations were 2, 4, and 6 µg/mL, respectively.

6.3.2 Quantitative Administration Study Results

Figures 6-2 to 6-6 present the 3-MT and Tyr concentrations, and 3-MT/Tyr ratio results for all 12 horses in the *Stalevo*® administration study. Each concentration plot includes a solid black line that indicates the current established 3-MT threshold at 4 µg/mL, and a dashed black line indicating the proposed 3-MT limit at 0.776 µg/mL (776 ng/mL). Each ratio plot contains a small dotted black line that represents the proposed 3-MT/Tyr ratio limit of 5.3. Individual horse timewise plotting for each targeted compound has been included in the appendices (Appendix 25 to Appendix 27).

Figure 6-2 shows the 3-MT and Tyr values calculated for the control horse group (n = 4) over the 20 timepoints. Both compounds remained stable for the administration period with no significant change between the week prior to administration and the week following administration. This assists in ruling out the potential of any variation in the 3-MT and Tyr concentrations occurring from external environmental factors. The highest concentrations identified for 3-MT and Tyr were 0.41 and 0.46 µg/mL, respectively.

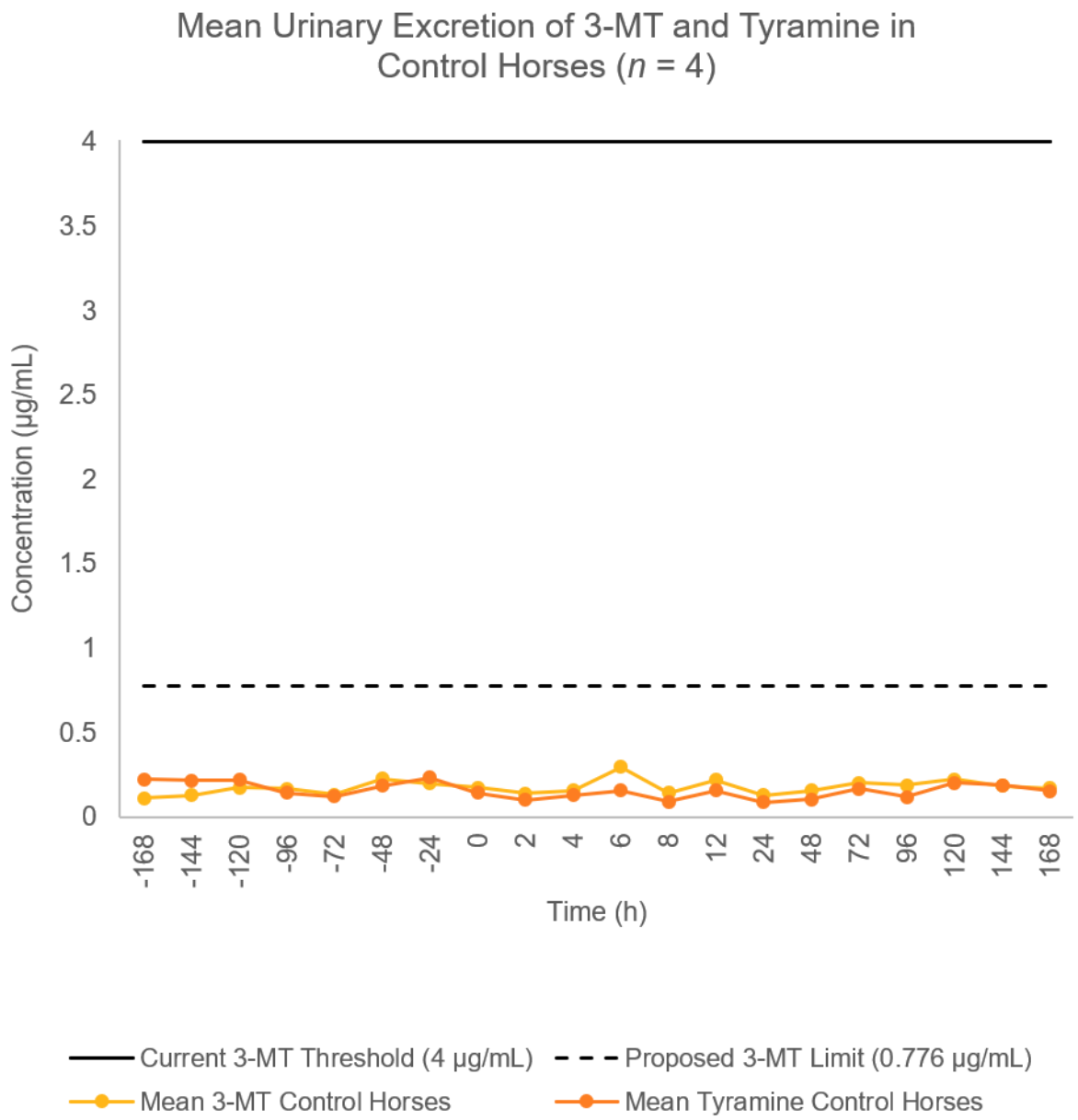


Figure 6-2: Timewise plotted concentration values for 3-methoxytyramine (3-MT) for individual control horses (n = 4). The black solid line represents the current 3-MT threshold (4 µg/mL) and the black dashed line indicates the proposed 3-MT intelligence limit (0.776 µg/mL).

Figure 6-3 includes the mean 3-MT and Tyr levels calculated for the treated horse group (*n* = 8) over the 20 timepoints. These results display an increase in 3-MT levels after the administration, and an unchanged excretion of Tyr. The deep blue line corresponds with the mean concentration of 3-MT, with a maximum mean concentration of 2.3 µg/mL, identified at 4 hours post-administration, with a range of 0.90 to 4.8 µg/mL.

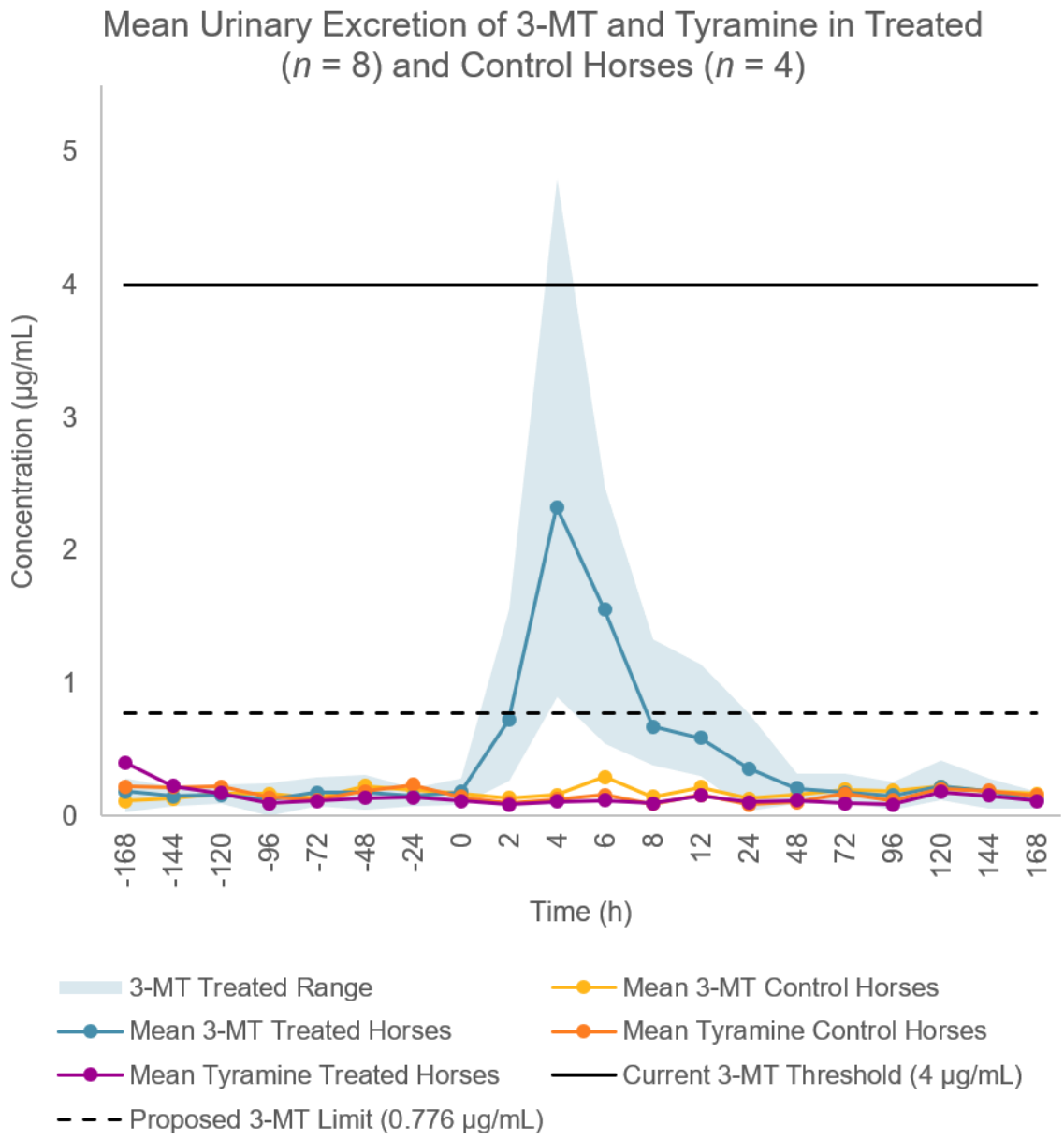


Figure 6-3: Timewise plotted concentration values for 3-methoxytyramine (3-MT) for individual control (n = 4) and treated (n = 8) horses. The black solid line represents the current 3-MT threshold (4 µg/mL) and the black dashed line indicates the proposed 3-MT intelligence limit (0.776 µg/mL).

A closer look at the proposed 3-MT limit in Figure 6-4, shows the mean concentration values pass below between the 6- and 8-hour timepoints. This indicates a detection period of at least 6 hours for dopaminergic manipulation based on the proposed 3-MT limit.

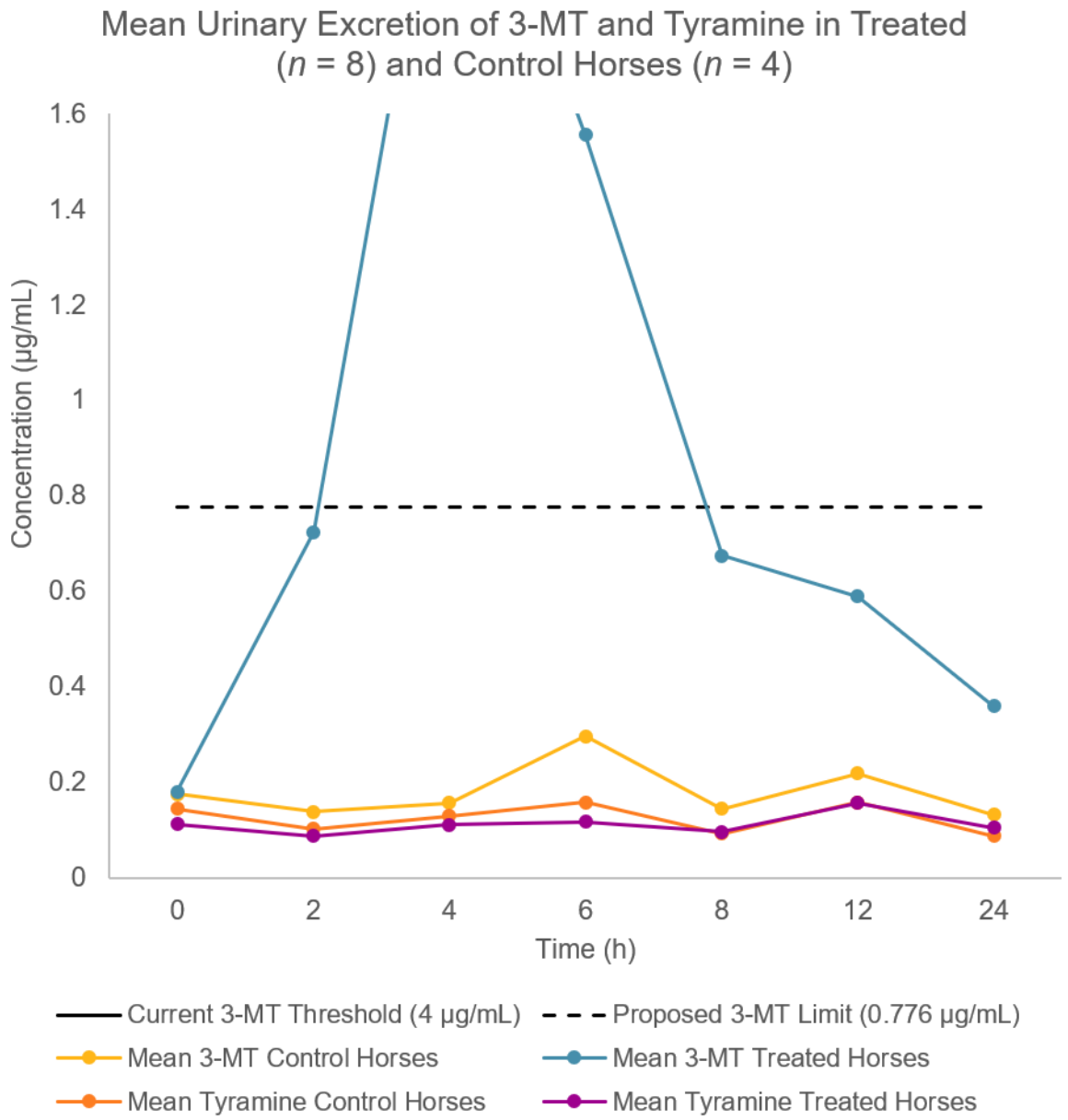


Figure 6-4: A zoomed in view (0 – 24 h) of timewise plotted concentration values for 3-methoxytyramine (3-MT) for individual control (n = 4) and treated (n = 8) horses. The black solid line represents the current 3-MT threshold (4 µg/mL) and the black dashed line indicates the proposed 3-MT intelligence limit (0.776 µg/mL).

Figure 6-5 shows the 3-MT/Tyr ratio data compared to the proposed ratio limit of 5.3, indicated by the dotted black line. The maximum mean ratio result occurred at 4 hours post-administration with a value of 20.7, ranging from 5.0 to 52.0.

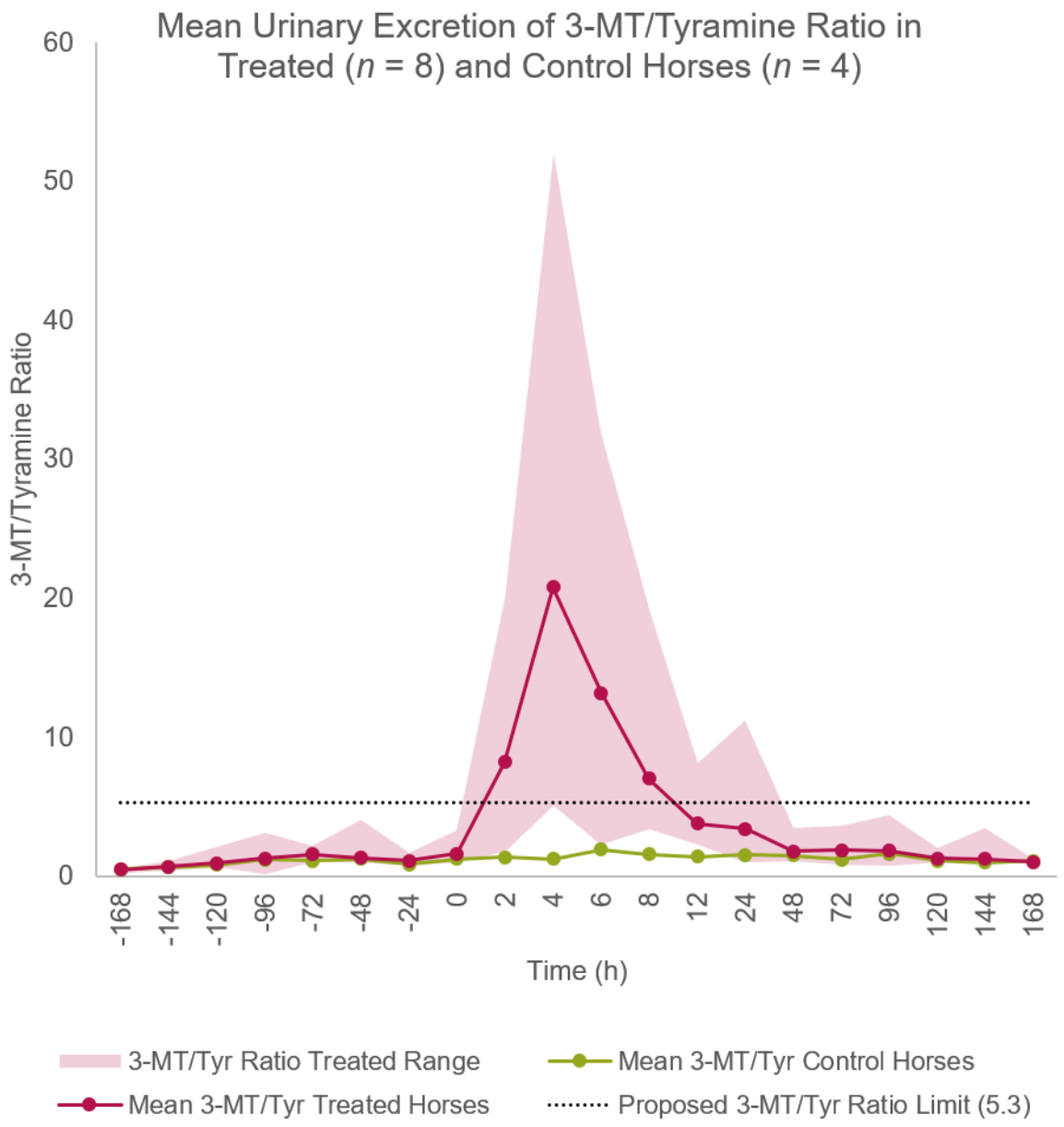


Figure 6-5: Timewise plotted 3-methoxytyramine-to-Tyramine (3-MT/Tyr) ratio values for individual control (n = 4) and treated (n = 8) horses. The black dotted line indicates the proposed 3-MT/Tyr ratio limit (5.3).

Figure 6-6 shows a closer look at where the ratio results fall back below the proposed ratio limit line. The detection window for *Stalevo*® administration extends to at least 8 hours with this ratio limit in place, prolonging the detection period for 2 hours more than the proposed 3-MT limit alone.

Figure 6-6: A zoomed in view (0 – 24 h) of timewise plotted 3-methoxytyramine-to-Tyramine (3-MT/Tyr) ratio values for individual control (n = 4) and treated (n = 8) horses. The black dotted line indicates the proposed 3-MT/Tyr ratio limit (5.3).

6.4 Discussion

Quantitative analysis of 3-MT and Tyr following *Stalevo*® administration revealed the effect that entacapone, a COMT inhibitor, can have when included with a L-DOPA containing medication. 3-MT excretion levels were decreased compared to L-DOPA administrations without a COMT inhibitor present [16].

This influence in 3-MT excretion reduces the efficiency of recognising doping with the current 3-MT threshold, established by Wynne *et al.* (2004), adopted by the IFHA and enforced by

various regulatory authorities across the globe [2, 54, 125]. Of the eight treated horses in this study, only one briefly surpassed the 4 µg/mL threshold, and only 4 hours after administration, before returning to below the reportable limit.

Previous research conducted by Elbourne *et al.* (2022) using two single-horse administration studies involving *Sinemet* showed similar results [16]. *Sinemet* is another L-DOPA containing PD medication which also includes carbidopa, but a COMT inhibitor substance like entacapone, opicapone or tolcapone is absent. The results from this study differ to the earlier findings, in so far as a shorter detection window using the proposed 3-MT limit and 3-MT/Tyr ratio limit has been observed, at least in part due to the introduction of entacapone in the medication. It should also be acknowledged that these administration studies were conducted approximately 5 years apart, on different horses, and analysed using different methodology and analytical instrumentation (GC-MS versus LC-QTOF-MS). Therefore, it is clear that there may have also been other contributing factors to the differences in these results.

This area of research successfully addressed the aim to optimise and validate a fit-for-purpose LC-MS method for the targeted detection of dopaminergic manipulation in equine urine via the quantification of 3-MT and Tyr. This work also reiterated the benefit of implementing the proposed 3-MT intelligence limit and 3-MT/Tyr ratio limit to better detect the sub-threshold doping of L-DOPA.

Overall, these results have demonstrated that the current threshold can be inefficient to detect sub-threshold doping of L-DOPA containing medications, thereby limiting the interpretation of ‘normal’ and ‘manipulated’ dopamine levels in equine urine. Due to this, an untargeted metabolomic search was conducted to investigate more related metabolites that could improve dopaminergic manipulation in equine urine, with the added aim to extend the detection window further from race-day.

Chapter 7: Multivariate Approach for Dopaminergic Manipulation Detection

7 MULTIVARIATE APPROACH FOR DOPAMINERGIC MANIPULATION DETECTION

7.1 INTRODUCTION

Due to the restricted detection window of the 3-MT/Tyr ratio for dopaminergic manipulation observed in **Chapter 6, Section 6.3.2**, a metabolomic approach was investigated with the aim to extend the detection window with more comprehensive biomarkers. This section of research was guided by previous work conducted by B. Keen in the development of a LC-HRMS method for the analysis of dopamine related markers in equine plasma, and the recommendation to implement 3-methoxytyrosine as an indicator for dopaminergic manipulation [169].

Previously acknowledged biomarkers and compounds of interest of L-DOPA, dopamine or related catecholamines, proposed in prior publications have been investigated, in their free and conjugated forms, where possible [12, 26, 54, 56, 57, 59, 63, 64, 169-173]. Most compounds were eliminated due to their lack of consistent presence between multiple administration study batches. This could be considered a limitation of the randomised batch study design implemented in this segment of the thesis.

This area of work also investigated the presence of carbidopa and entacapone in the excreted urine as a potential marker of interest for the detection of *Stalevo*®. However, this was mostly unsuccessful due to the very short half-life of both these compounds, at 2-3 hours and 0.4-0.7 hours for carbidopa and entacapone (identified in humans), respectively [174]. It can also be noted that entacapone is almost completely metabolised in the system prior to excretion, with only approximately 0.2 % excreted unchanged in urine. Due to short elimination half-lives, no true accumulation of levodopa or entacapone occurs when they are administered repeatedly, making their detection from a single administration difficult [174].

This chapter continues the investigation of levodopa misuse using a metabolomics approach to search for more robust urinary biomarkers in their intact sulfate and glucuronide conjugated forms, in addition to the free metabolite. This chapter also details the potential for a machine learning model to extend the detection window for dopaminergic manipulation.

7.2 MATERIALS AND METHODS

7.2.1 Chemicals and Reagents

All chemicals and reagents were used as outlined in **Chapter 6, Section 6.2.1**.

7.2.2 Administration Study

The administration study parameters and details were as outlined in **Chapter 6, Section 6.2.2**.

7.2.3 Randomised Batch Configuration

A single freeze/thaw cycle for the administration study samples was utilised to accomplish batch randomisation and create a pooled QC sample to be analysed alongside each batch. A total of 24 samples were allocated to 10 batches, mimicking the ratio of control to treated horses of the total study (4 control/8 treated horse, Table 7-1). This equated to two timepoint samples from each horse randomly allocated to each group, in addition to a pooled QC sample.

Table 7-1: Randomised batches for metabolomics analysis of Stalevo study samples. Each batch was assigned two samples from each administration horse. Horses are labelled by control (C) and treated (T), mare (M) and gelding (G), and individually numbered. The administration timepoint is denoted in hours below the horse label. Negative (-) values indicate collection occurred before the drug administration, while positive values indicate sampling after the administration was conducted

Sample #	**Batches**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	CM1 2	CM1 -168	CM1 72	CM1 120	CM1 -144	CM1 24	CM1 -48	CM1 12	CM1 8	CM1 4
2	CM1 6	CM1 -96	CM1 144	CM1 -24	CM1 -120	CM1 96	CM1 0	CM1 48	CM1 168	CM1 -72
3	CM6 96	CM6 72	CM6 0	CM6 12	CM6 48	CM6 -168	CM6 168	CM6 -48	CM6 -144	CM6 24
4	CM6 -24	CM6 -120	CM6 6	CM6 144	CM6 2	CM6 120	CM6 -72	CM6 4	CM6 8	CM6 -96
5	CG1 -48	CG1 120	CG1 48	CG1 6	CG1 -120	CG1 8	CG1 -24	CG1 24	CG1 96	CG1 -72
6	CG1 2	CG1 -96	CG1 168	CG1 0	CG1 72	CG1 144	CG1 12	CG1 -168	CG1 -144	CG1 4
7	CG9 96	CG9 -168	CG9 120	CG9 12	CG9 -72	CG9 -96	CG9 2	CG9 -144	CG9 0	CG9 4
8	CG9 168	CG9 72	CG9 144	CG9 -48	CG9 -24	CG9 6	CG9 24	CG9 8	CG9 48	CG9 -120
9	TM2 72	TM2 -72	TM2 144	TM2 -48	TM2 4	TM2 6	TM2 -168	TM2 -96	TM2 -144	TM2 2
10	TM2 -120	TM2 8	TM2 -24	TM2 24	TM2 120	TM2 48	TM2 168	TM2 0	TM2 96	TM2 12
11	TM3 24	TM3 96	TM3 -120	TM3 -48	TM3 8	TM3 -144	TM3 2	TM3 144	TM3 -96	TM3 -72
12	TM3 48	TM3 12	TM3 6	TM3 -168	TM3 120	TM3 0	TM3 4	TM3 168	TM3 72	TM3 -24
13	TM4 -168	TM4 4	TM4 96	TM4 2	TM4 12	TM4 72	TM4 -72	TM4 0	TM4 24	TM4 -48
14	TM4 -144	TM4 8	TM4 120	TM4 48	TM4 -96	TM4 168	TM4 -120	TM4 144	TM4 -24	TM4 6

15	TM90	TM924	TM98	TM9-144	TM9120	TM9-120	TM92	TM96	TM9-48	TM9-168
16	TM9-72	TM94	TM9168	TM972	TM9144	TM9-24	TM996	TM948	TM9-96	TM912
17	TG2-144	TG2-24	TG24	TG2120	TG20	TG2144	TG28	TG2-96	TG22	TG248
18	TG2-120	TG2-168	TG26	TG272	TG212	TG224	TG2168	TG2-72	TG2-48	TG296
19	TG3-168	TG3-72	TG348	TG3168	TG32	TG38	TG312	TG3-48	TG36	TG3-24
20	TG3-120	TG30	TG3120	TG3-96	TG324	TG3144	TG372	TG396	TG34	TG3-144
21	TG48	TG4-144	TG4-120	TG4-24	TG4-168	TG44	TG424	TG4-96	TG42	TG4120
22	TG472	TG4-48	TG496	TG412	TG4144	TG4-72	TG4168	TG40	TG448	TG46
23	TG100	TG1096	TG10168	TG10-168	TG1012	TG102	TG10-96	TG106	TG10-24	TG10-48
24	TG1072	TG1048	TG108	TG1024	TG10-120	TG10-72	TG104	TG10-144	TG10144	TG10120
25										TG101

7.2.4 Sample Preparations and Instrumental Analyses

An aliquot of urine (100 µL) was combined with ACN (300 µL) and internal standard mix (10 µL, 100 ng/mL equivalent concentration, see Table 7-2) in a microcentrifuge tube (650 µL capacity, Cat. No. 3208; Costar; Salt Lake City, UT, USA). The solution was capped and vortexed before being micro-centrifuged using a Beckman Coulter Microfuge 20 (with inserts to hold tubes in place) at 18,000 x g for 5 minutes. The samples were then transferred to small glass test tubes before adding one drop (approximately 20 µL) of methanolic HCl and drying under a gentle flow of nitrogen at 50 °C. The sample was reconstituted in 20 mM ammonium formate in 80% ACN (200 µL), vialed, and stored at 4 °C prior to LC-QTOF-MS analysis.

Table 7-2: Internal standards used in a mix for analysis in positive and negative electrospray ionisation modes. Chemical formula, monoisotopic mass, found retention times and assigned acquisition mode information provided.

Internal Standard Mix	Chemical Formula	Monoisotopic Mass (g/mol)	Retention Time (minutes)	Acquisition Mode (+/- ESI)
Carbidopa–d_5	$C_{10}H_9N_2O_4D_5$	231.12676	1.532	negative
Ethyl glucuronide–d_5	$C_8H_9O_7D_5$	227.10537	3.966	negative
Ethyl sulfate–d_5	$C_2HO_4SD_5$	131.03008	1.192	negative
Gabapentin–d_4	$C_9H_{13}NO_2D_4$	175.15104	3.222	positive
3-Methoxytyramine–d_4	$C_9H_9NO_2D_4$	171.11974	4.325	positive

Table Definition: The D in the chemical formula denotes a deuterated hydrogen atom in the internal standard structures.

LC-QTOF-MS analysis was conducted on the Shimadzu 9030 instrument (Kyoto, Japan) and utilised an IMTAKT (Kyoto, Japan) Intrada Amino Acid column (2 mm x 100 mm, 3 µm) with a Phenomenex (Torrance, CA, USA) C18 SecurityGuard (2 mm x 4 mm). The column oven was operated at 40 °C, with a constant flow rate at 0.3 mL/min. Mobile phases were A: 100 mM ammonium formate in water, and B: 0.3% FA in 100% ACN. The gradient used was: 0 minutes A-B (15:85 %v/v), 0-8 minutes A-B (100:0 %v/v), 8.01 minutes A-B (15:85 %v/v) with a 3-minute post run equilibration time. All injection volumes were 5 µL.

ESI was utilised sequentially in positive and negative modes to acquire free and phase I and II conjugated metabolites. Metabolomic data was acquired through data independent acquisition (DIA) with a precursor ion range of *m/z* 100 – 400 and a product ion range of *m/z* 50 – 400. The method included 20 event windows with *m/z* 15 Q1 transmission width (Table 7-3). The interface temperature was 350 °C with a gas flow of 11.0 L/min. The nebulising gas flow was set to 3.0 L/min. A fixed CE spread was optimised at 30 ± 15 eV.

Table 7-3: Data Independent Acquisition (DIA) instrument method event windows (20, with an initial scan event), featuring a Q1 transmission width of m/z 15. SWATH, another name for DIA, was the term used to identify the data acquisition mode to MS-DIAL.

CycleID	Event Type	*m/z* Begin	*m/z* End
1	SCAN	50	400
2	SWATH	100	115
3	SWATH	115	130
4	SWATH	130	145
5	SWATH	145	160
6	SWATH	160	175
7	SWATH	175	190
8	SWATH	190	205
9	SWATH	205	220
10	SWATH	220	235
11	SWATH	235	250
12	SWATH	250	265
13	SWATH	265	280
14	SWATH	280	295
15	SWATH	295	310
16	SWATH	310	325

17	SWATH	325	340
18	SWATH	340	355
19	SWATH	355	370
20	SWATH	370	385
21	SWATH	385	400

It should be noted that optimisation of analysis was readjusted after batches 1 and 2 were analysed. The changed conditions were flowrate (0.4 mL/min, later adjusted to 0.3 mL/min) and Q1 transmission width (originally *m/z* 10, later extended to *m/z* 15). Unfortunately, these samples were not able to be reanalysed with the updated conditions due to time constraints and instrument availability. As a result, there is a slight difference in retention times, which has been accounted for when comparing results between these batches and the rest of the administration study.

7.2.5 Data Processing Methods

Data was extracted with LabSolutions (version 4.50 SP1) and analysed using Insight Explore (pre-release version). Statistical analysis was performed with MS-DIAL (version 4.9.221218), Microsoft Excel (version 16.42), MetaboAnalyst 6.0 and Python (via Anaconda3, Jupyter Notebook, version 6.4.5). The computer used to conduct the following data processing was a Windows Processor 12th Gen Intel® Core ™ i5-10500, 2667 MHz, running 6 cores, and 16 gigabytes of memory.

MS-DIAL

After file conversion from lcd (LC-MS file) to mzML (MS output file) type, MS-DIAL was used for alignment of the MS data. Thresholds were set for the minimum peak height at 1000, and minimum MS/MS abundance at ten (10). The blank-assisted feature removal function, where likely irrelevant features that appear in samples indicated as blanks are tagged for easy removal, was utilised, in addition to no gap filling by compulsion. All other parameters were unchanged from the default setting. Aligned peak area and peak height matrices were exported in msp (MS/MS readable text file) file format.

Microsoft Excel

Microsoft Excel was used to manually filter and clean the peak area dataset by removal of any features that acquired a signal-to-noise value less than three (*S/N* < 3), if no MS/MS data was found, or a Fill% value less than 1/*the number of samples analysed in the batch* (to check for true triplicate analysis results). The dataset was also rearranged for compatibility with MetaboAnalyst.

Statistical Analysis

MetaboAnalyst's Statistical Analysis (one factor) tool was first used for filtering, normalisation, transformation and statistical analysis of the dataset. The randomised samples were labelled as control horse (Control), pre-administration (Pre) treated horse, post-administration (Post) treated horse, and pooled QC samples (Table 7-4).

Table 7-4: Sample grouping labels for statistical analysis in MetaboAnalyst.

Data Group Label	Explanation
Pre	168 – 0 h prior to administration
Post	1 – 168 h after administration
Control	All timepoints (pre- and post- "administration") collected from control horses
QC	13 replicate injections of the same pooled QC sample

Data filtering removed non-informative variables consisting of features with low repeatability. This was denoted by features present in the pooled QC samples with a RSD greater than 25%, which were removed. Variance filtering was implemented to remove features with near-constant abundance through the experiment denoted by their interquartile range (IQR), with 40% removal. Normalisation by median was used in addition to $\log_{10}$ transformation and no applied scaling. Statistical analysis was conducted on the dataset by way of PCA, volcano plots, and PLS-DA which produced Variable Importance Projection (VIP) scores.

Machine Learning

An investigation of classification models was conducted with MetaboAnalyst's Biomarker Analysis tool, using the different model algorithm functions in multivariate selection, and Python using the XGBoost model package on six features of interest. Optimisation of the model performance was also conducted.

Feature Identification

Feature identification was attempted for a small group of markers identified by statistical analysis and machine learning tools. This was done through manual MS/MS searching and investigation, a custom R script, and MetaboAnalyst's MS Annotation function.

7.3 RESULTS

A total of 24 samples were allocated to ten (10) batches, mimicking the ratio of control to treated horses of the total study (four control and eight treated horse). This equated to two timepoint samples from each horse randomly allocated to each group, in addition to a pooled QC sample. The samples were randomly injected three times throughout analysis, creating technical triplicate datapoints. The pooled QC sample was injected 13 times over the course of the batch analysis, equating to one injection every six (6) samples for optimal alignment results (Table 7-5). Additional information regarding the compounds included in the reference standard mixes can be found in Appendix 28 and Appendix 29.

Table 7-5: Sample injection order for untargeted LC-QTOF-MS analysis of Stalevo® administration study.

Sample Injection Order
Solvent (recon) Blank
Extraction (water) Sample
Solvent (recon) Blank
Urine (matrix) Blank
Solvent (recon) Blank x2

Pooled QC injection 1

Study Samples x6

Pooled QC injection 2

Study Samples x6

Pooled QC injection 3

…

Pooled QC injection 13

Solvent (recon) Blank

Dopamine reference standard mix (urine matrix) x3 (see appendix section 6)

Solvent (recon) Blank

Stalevo reference standard mix (urine matrix) x3 (see appendix section 6)

Solvent (recon) Blank

Column Flush

Table Definitions: Preparations of the Solvent (recon) Blank involved using the methods reconstitution solvent as a blank injection to condition the column prior to sample analysis, the Extraction (water) Sample included purified water that had undergone the same sample extraction process as the urine matrix to identify any unrelated metabolites that could interfere with the biological interpretation of the results, and the Urine (matrix) Blank was conducted using a pooled blank urine spiked with the internal standard (ISTD) mix to also assist in conditioning the column prior to sample analysis.

DIA data was collected for analysed samples, with each batch totalling approximately 100 datafiles for pre-processing. As previously mentioned, each batch contains two (2) randomised samples from each horse in the administration study to minimise bias (Table 7-1). MS-DIAL alignment was first conducted before the data was manually rearranged for compatibility with MetaboAnalyst. Some manual data filtering (*S/N* < 3, no MS/MS acquired, and inadequate Fill%) reduced the dataset from, on average, 20,000 found features, down to roughly 5,000 found features for further investigation. At this point, grouping of the data samples was also applied. The labels, Control, Pre, Post, and QC are explained in **Section 7.2.5**, and presented in Table 7-4.

Metabolomic analysis of the *Stalevo*® administration study samples produced a large amount of data consisting of free and conjugated features that followed data processing and statistical analysis via the pipeline presented below (Figure 7-1).

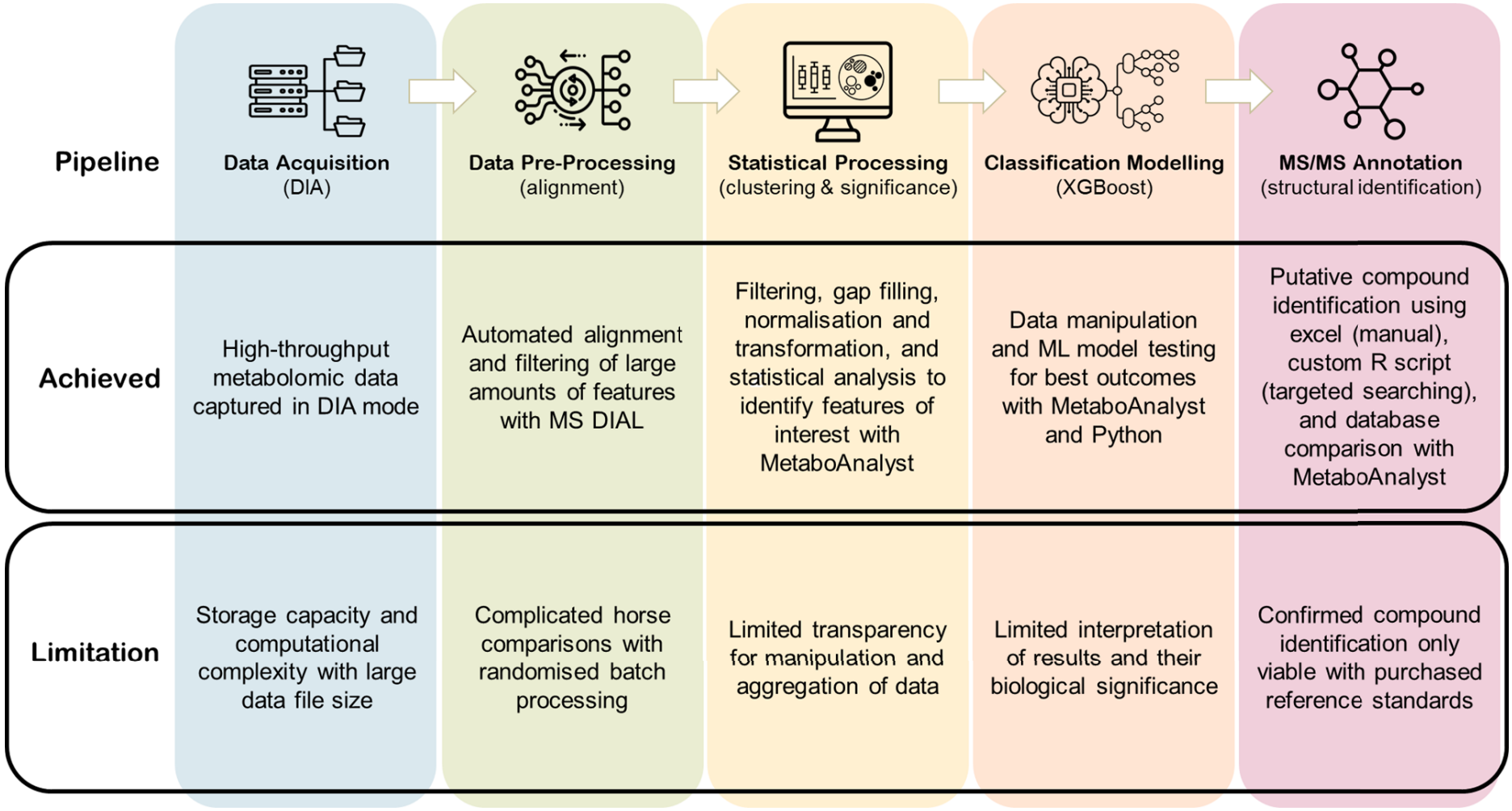


Figure 7-1: Metabolomic data processing pipeline used with what each section was able to achieve and what limitations arose.

7.3.1 Statistical Analysis

PCA plotting was conducted first to ensure quality of the data prior to further investigation. A PCA pairing plot, shown in Figure 7-2, can be used to visualize data structure, class differences, and outliers to find maximal variation between groups of interest. In a PCA plot no class information is given, so its outcome is perceived as an unbiased view of the class separation obtained. In the pairing plot multiple components can be applied; the first principal component (PC1) explains the largest variation in the data, followed by PC2, PC3, etc. [175] Multiple classes can be viewed on the scores plot, in two or three dimensions, and so group separation can be observed, i.e., pre and post-administration.

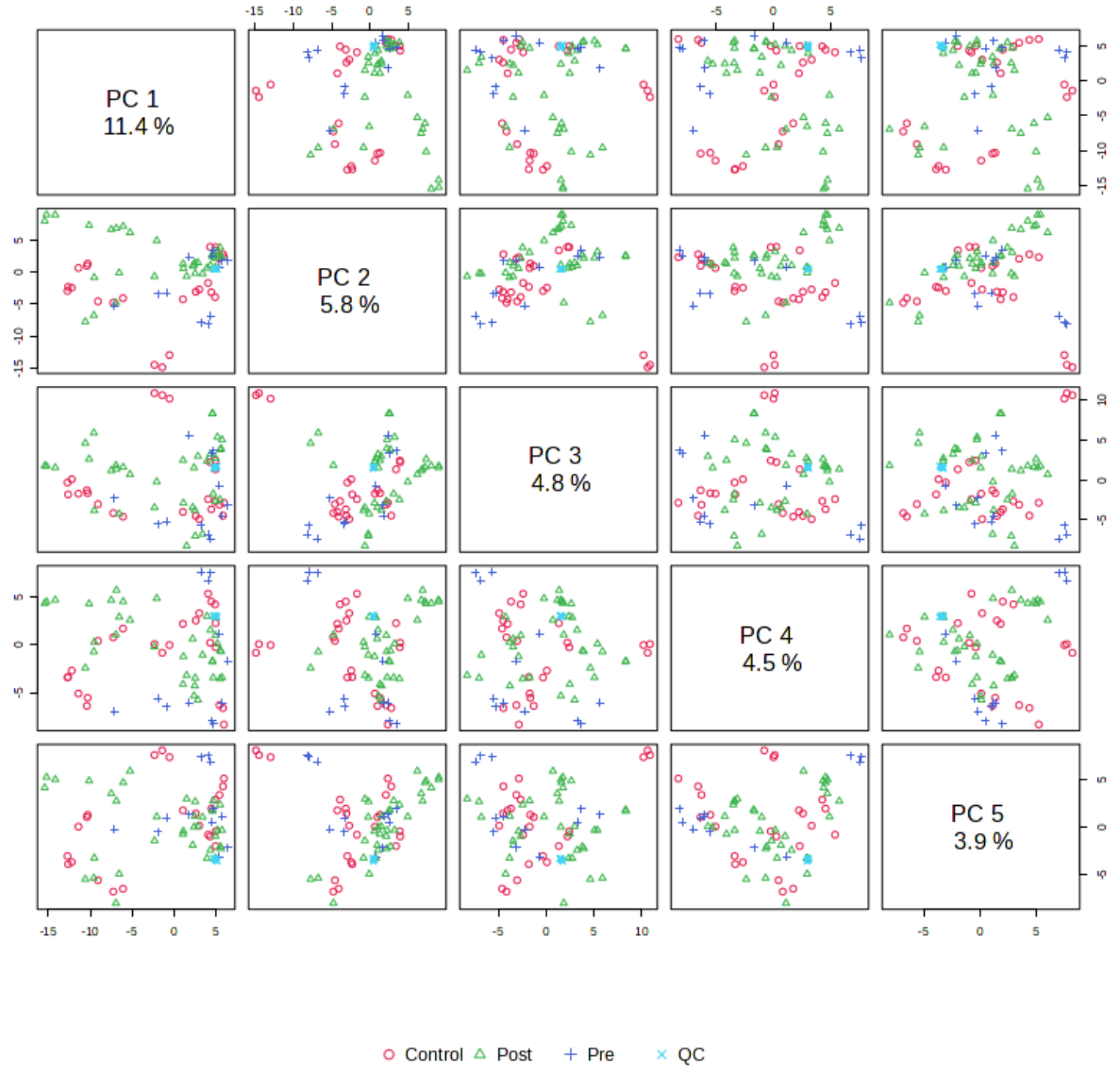


Figure 7-2: Principal component analysis (PCA) pairing plot from batch 7 with data collected in negative electrospray ionisation mode. Red circles (○) indicate control horse samples, green triangles (▲) indicate post-administration treated horse samples, dark blue plus symbols (+) indicate pre-administration treated horse samples, and light blue cross symbols (x) indicate QC samples.

A PCA scores plot containing PC1 and PC2, was used to ensure appropriate visual grouping had occurred with the 13 QC samples (Figure 7-3, coloured light blue dots). This clustering implies consistency in feature data over the course of each batch run length, and was taken as a visual representation of minimal instrument variance.

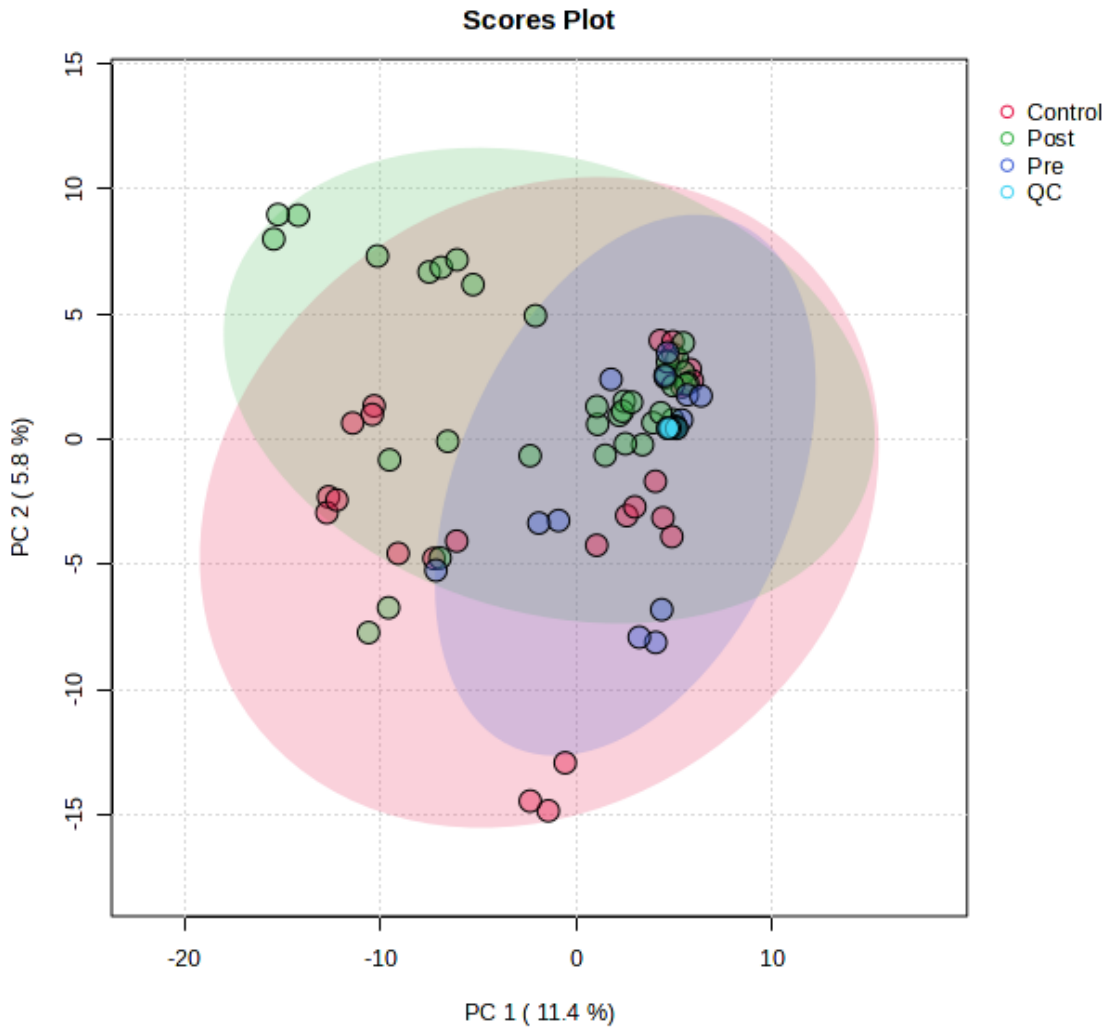


Figure 7-3: Principal component analysis (PCA) scores plot from batch 7 with data collected in negative electrospray ionisation mode. Red dots indicate control horse samples, green dots indicate post-administration treated horse samples, dark blue dots indicate pre-administration treated horse samples, and light blue dots indicate QC samples.

After quality checks were conducted, the control and QC groups were removed from analysis, and pre- and post-administration labelled data proceeded through volcano plotting to identify significantly changed features (Figure 7-4**)**. This change was denoted by high $-\log_{10}$(p-value) and Log_2FC values past preselected cut-offs. The dotted lines indicate the cut-off values selected at an adjusted p-value of 1 (equivalent to an untransformed p-value of 0.1) and a $\log_2$FC value of 2.3 (equivalent to an untransformed FC of 5).

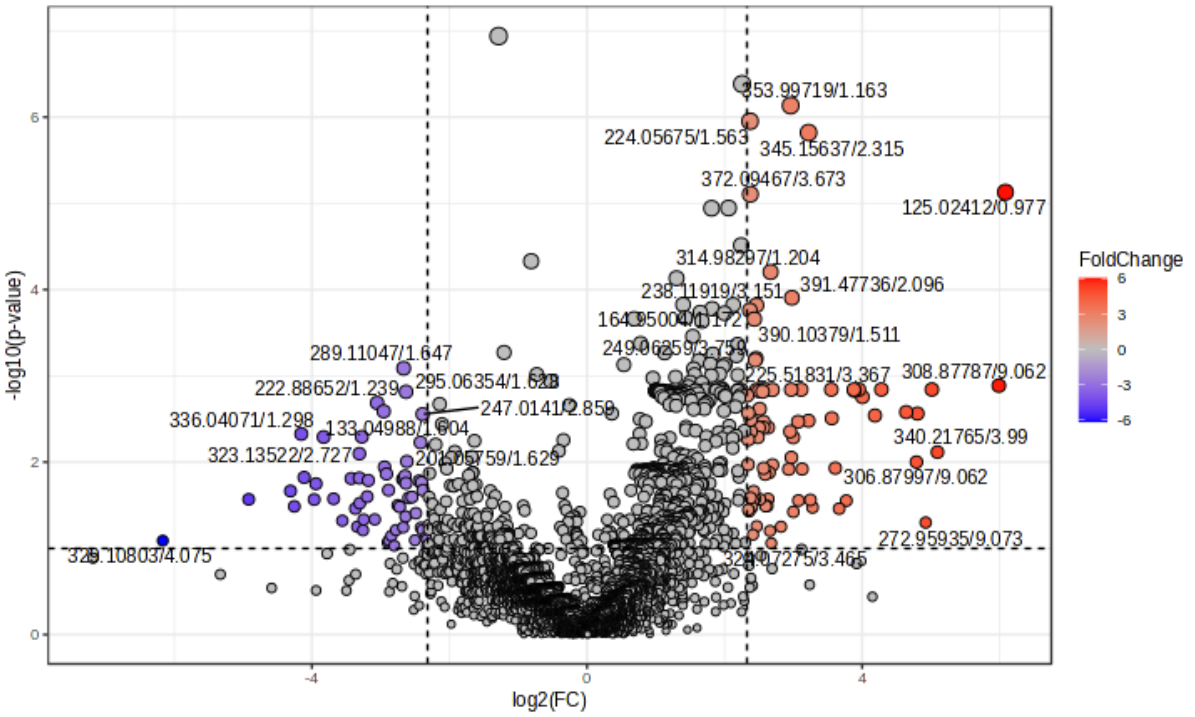


Figure 7-4: Volcano plot results obtained from batch 7 with data collected in negative electrospray ionisation mode, for pre and post administration labelled groups. Significant features are labelled by their deprotonated molecules in m/z / retention time in minutes. Up-regulated significant features are coloured red/brown, and down-regulated significant features are colours blue/purple, non-significant features are coloured grey.

Some variation in the number of significant features extracted per batch was noted and can be seen in Table 7-6. This may be due to the randomised nature of the samples for analysis.

Table 7-6: The number of significant features identified by volcano plotting for each batch, in positive (+) and negative (-) electrospray ionisation (ESI) modes.

	+ESI	**-ESI**
Batch 1	38	27
Batch 2	42	32
Batch 3	90	43
Batch 4	34	51
Batch 5	40	32
Batch 6	50	37
Batch 7	126	139
Batch 8	52	50
Batch 9	39	49
Batch 10	55	37

In addition to volcano plots used to provide significant features of interest, PLS-DA was also conducted and its VIP scores tool was implemented to identify features that contributed heavily to the separation of the pre- and post-administration groups in the pairing and scores plots below (Figure 7-5 and Figure 7-6).

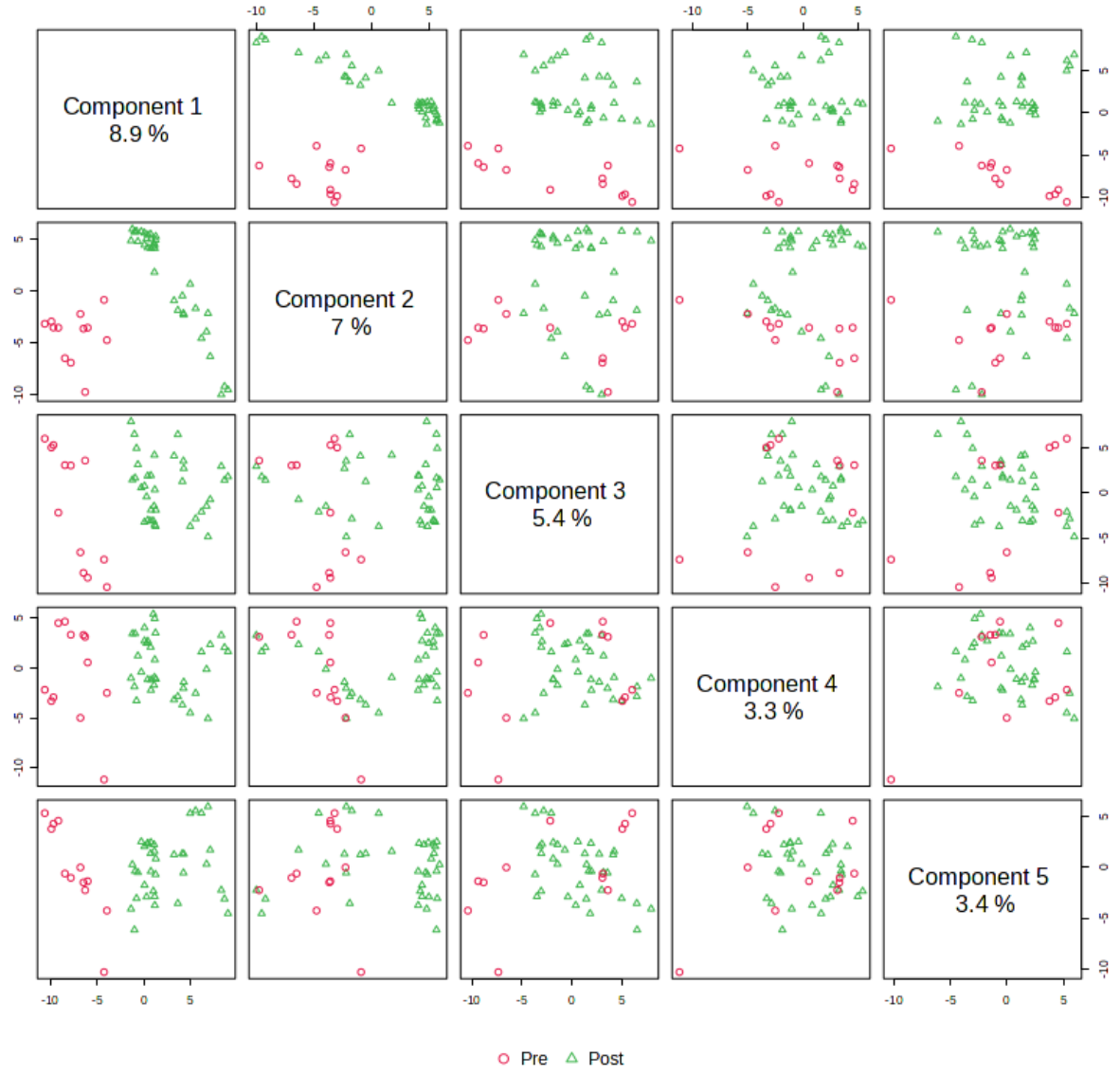


Figure 7-5: Partial least-squares discriminant analysis (PLS-DA) pairing scores from batch 7 with data collected in negative electrospray ionisation mode, pre and post administration labelled groups used. Red circles (○) indicate pre-administration treated horse samples, green triangles (▲) indicate post-administration treated horse samples.

In comparison to PCA plots presented earlier, PLS-DA is a supervised learning tool that incorporates class information and interprets the data given to achieve the maximal separation between the labelled classes.

The PLS-DA scores plots visualise two components on a plane, with the x-axis showing the variation between the labelled groups, and the y-axis showing variation within the groups. Whilst this is a useful aid in visualising class separation, it can suffer from the risk of over-fitting the data, and therefore, must be interpreted with caution. Due to this, the data was used in combination with other statistical tests, to avoid misinterpretation of the results.

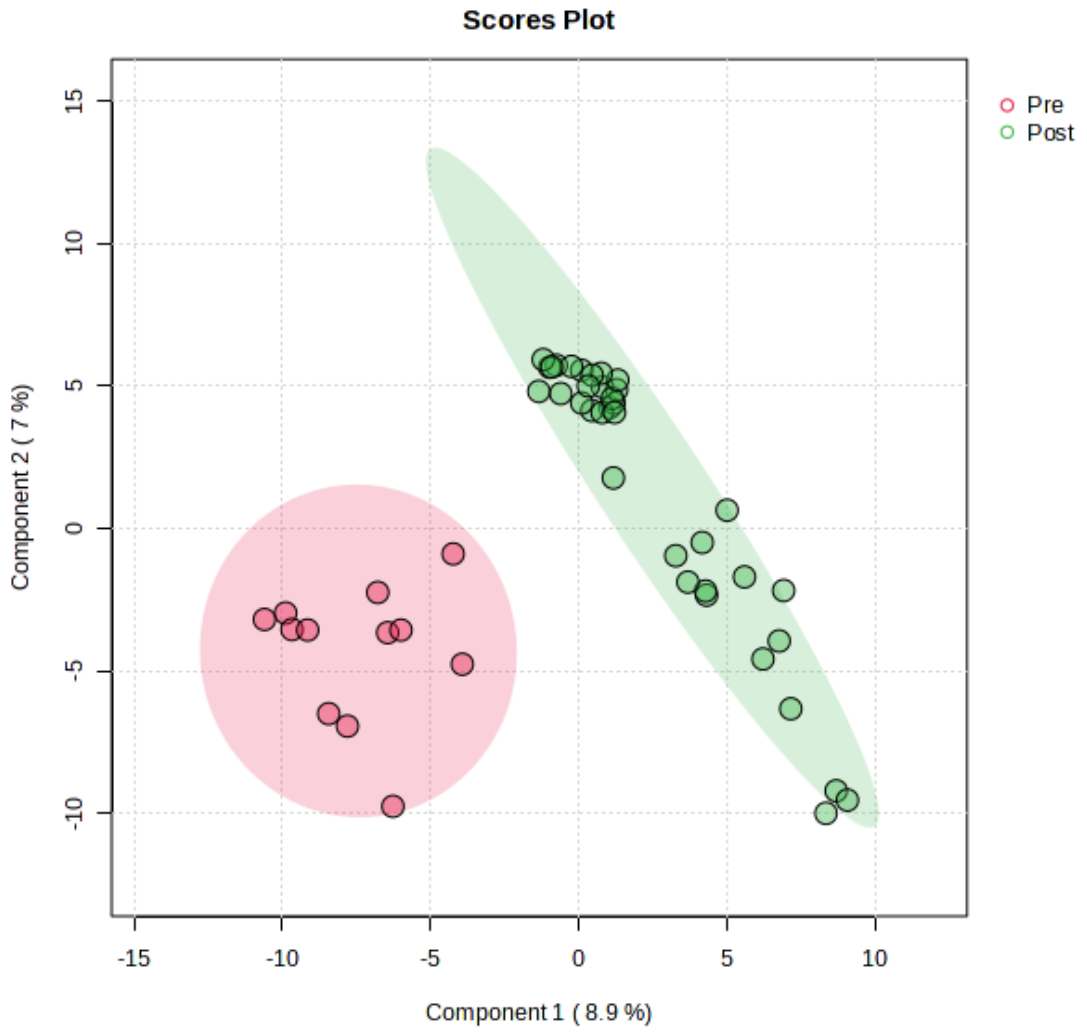


Figure 7-6: Partial least-squares discriminant analysis (PLS-DA) scores plot from batch 7 with data collected in negative electrospray ionisation mode, pre and post-administration groups. the x-axis shows variation between the labelled pre- and post- groups, and the y-axis shows variation within the groups. Red dots indicate pre-administration treated horse samples, green dots indicate post-administration treated horse samples.

The PLS-DA VIP scores plot ranks the top 15 features based on their ability and contribution in separating labelled pre- and post-administration groups for the PLS-DA scores plot (Figure 7-7). VIP scores greater than one (> 1) are suggested to be important, whilst those less than one (< 1) are considered unimportant for the model. The range of VIP scores may vary with each dataset and, in some studies, there may be hundreds of metabolite features with a VIP score around one (1), meaning that the cut-off of one (1) is too low and can be adjusted so that a much higher cut-off is applied [175]. Therefore, in this case the cut-off applied to the dataset was at a higher value of four (4).

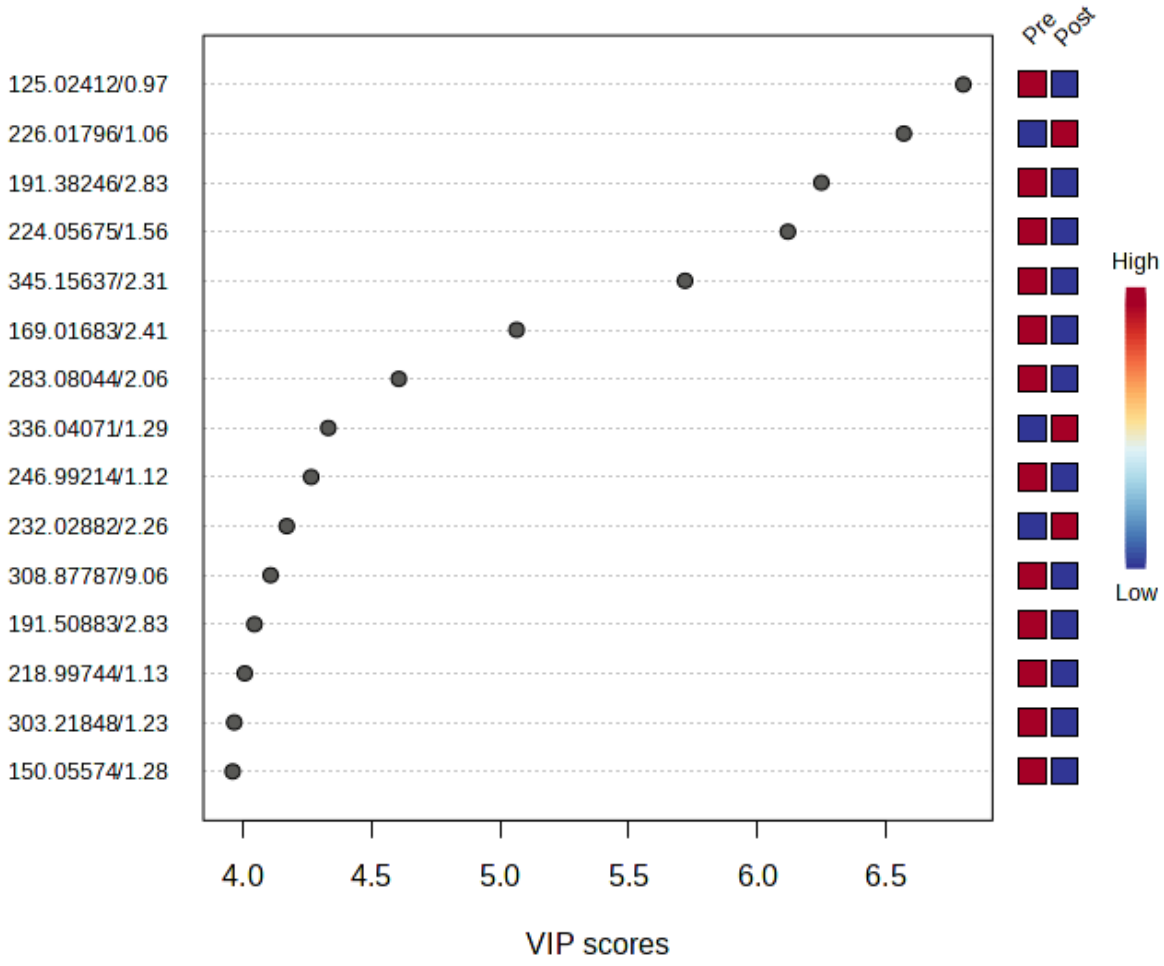


Figure 7-7: Partial least-squares discriminant analysis (PLS-DA) Variable Importance Projection (VIP) scores based on features importance in separating labelled pre and post-administration groups. Data obtained from batch 7 with data collected in negative electrospray ionisation mode. Important features are labelled by their deprotonated molecules in m/z / retention time in minutes. The relative abundance of each feature is indicated by the coloured box, from low in blue to high in red.

7.3.2 Machine Learning

To focus on the timepoints not covered by the proposed ratio in **Chapter 6**, PLS-DA VIP score testing was also conducted on a reduced dataset of the treated horse samples . This dataset included all pre-administration samples, and all samples collected from 12 h to 168 h post-administration, and were labelled as pre and post, respectively (Figure 7-8). The pre-administration timepoints (and associated horse) were -72 h (TM4), -96 h (TG10), -120 h (TM2), and -168 h (TM2). The post-administration timepoints (and associated horse) were 12 h (TG3), 24 h (TG4), 72 h (TG3), 96 h (TM9), and 168 h (TG2, TG4, TM2) (please refer to Table 7-1).

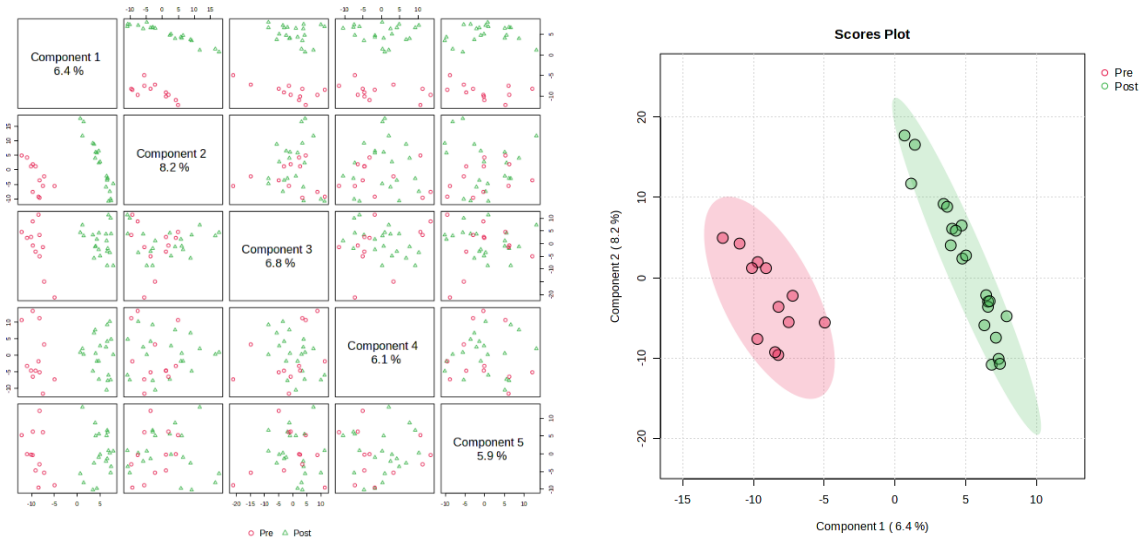


Figure 7-8: Partial least-squares discriminant analysis (PLS-DA) on the reduced timepoint dataset (2 – 8 hour treated horse samples removed) from batch 7 with data collected in negative electrospray ionisation mode. Red indicate pre-administration treated horse samples, green indicate post-administration treated horse samples.

Extracted features were categorised as up-regulated or down-regulated by the VIP plot, comparing pre- and post-administration abundances. VIP score gives 15 features that showed greatest differentiation in PLS-DA pre and post administration groups (Figure 7-9, tabulated in Table 7-7).

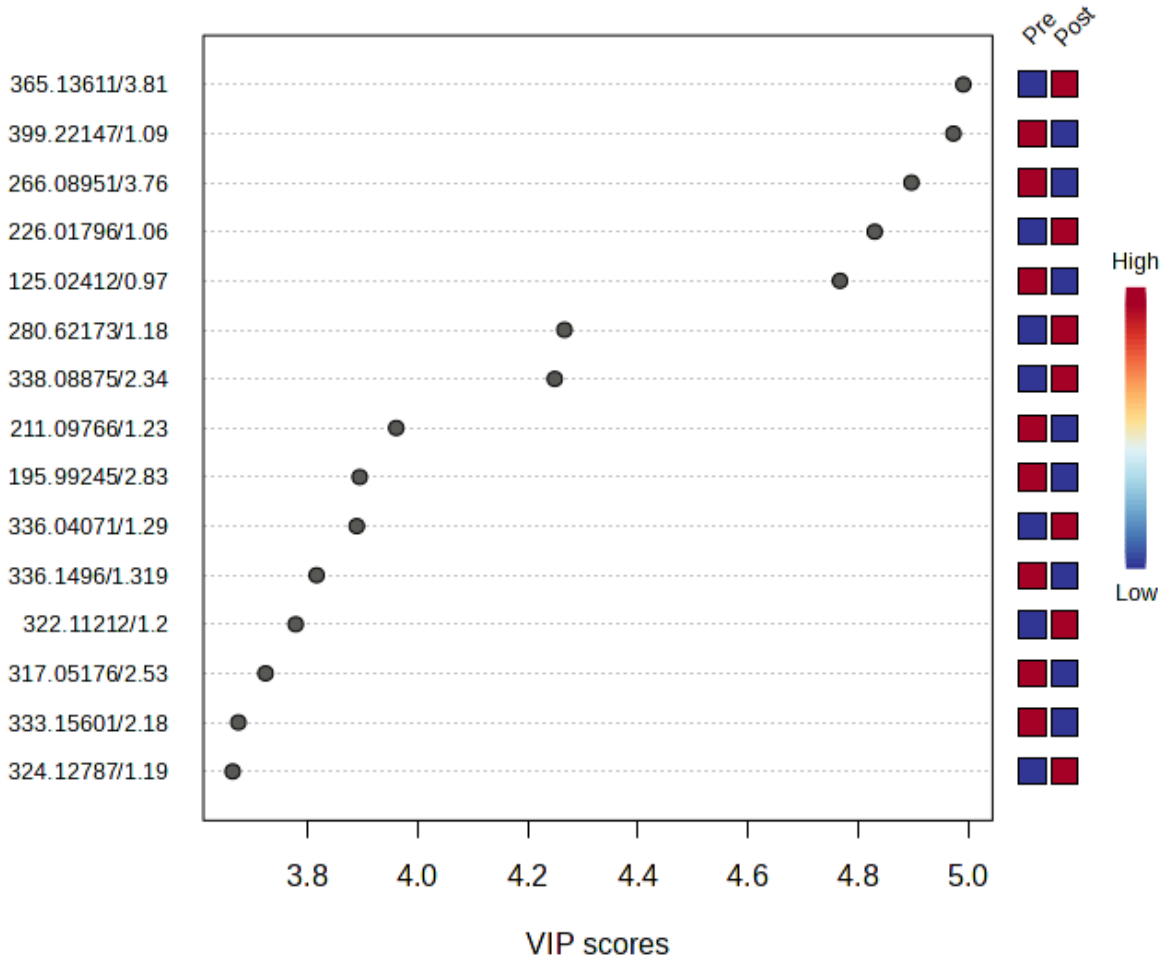


Figure 7-9: Variable Importance Projection (VIP) score results from reduced dataset (2 – 8 hour treated horse samples removed) processing, highlighting top 15 features contributing to pre/post separation from batch 7 with data collected in negative electrospray ionisation mode, focusing on later timepoints. Important features are labelled by their deprotonated molecules in m/z / retention time in minutes. The relative abundance of each feature is indicated by the coloured box, from low in blue to high in red.

Table 7-7: Variable Importance Projection (VIP) score features identified, as shown in Figure 7-9 categorised as up-regulated and down-regulated. Up-regulated features were identified by low abundance in pre-administration samples (blue box for the 'pre' group in Figure 7-9) and higher abundance in the post-administration samples (red box for the 'post' group). Down-regulated features were identified by the inverse, i.e., high (red) abundance in pre-administration samples, and low (blue) abundance in post-administration samples

Direction	Deprotonated Molecule (*m/z*)	Retention Time (RT, min)
Up-regulated	365.13611	3.81
	226.01796	1.06
	280.62173	1.18
	338.08875	2.34
	336.04071	1.29
	322.11212	1.20
	324.12787	1.19
Down-regulated	399.22147	1.09
	266.08951	3.76
	125.02412	0.97
	211.09766	1.23
	195.99245	2.83
	336.14960	1.32
	317.05176	2.53
	333.15601	2.18

Multiple model testing was conducted on these 15 identified features from batch 7 data. Features underwent individual classical univariate ROC plotting, where an AUC score, adjusted p-value, and fold change value were assigned to each (Table 7-8). This assists in identifying their individual ability to classify samples in the dataset.

Table 7-8: Classical univariate ROC plots of 15 extracted features. Area under the curve (AUC) scores, adjusted p-values, and fold change (FC) values were assigned to each feature (presented as deprotonated molecules (m/z) at the corresponding retention time (RT) in minutes.

Deprotonated Molecule / RT (*m/z* / min)	**AUC**	**Adjusted P-value**	**Log_2FC**
338.08875/2.348	0.940476	6.75E-06	2.5522
365.13611/3.81	0.922619	4.09E-06	3.9648
317.05176/2.532	0.914683	1.93E-05	–2.6344
280.62173/1.187	0.902778	9.46E-05	3.2
322.11212/1.20	0.896825	8.56E-05	2.696
324.12787/1.192	0.876984	3.53E-05	2.0675
266.08951/3.765	0.863095	0.000227	–2.2138
333.15601/2.182	0.861111	0.00017	–1.8618
336.1496/1.319	0.857143	2.73E-05	–2.5919
226.01796/1.063	0.849206	0.000102	1.1363
336.04071/1.298	0.835317	0.000336	3.9872
211.09766/1.232	0.801587	0.000798	–3.0519
399.22147/1.091	0.777778	0.001875	–2.3622
195.99245/2.832	0.746032	0.012841	–2.1939

125.02412/0.977	0.732143	0.003469	–5.4271

Model algorithm testing was then conducted with this feature set through MetaboAnalyst's Biomarker Analysis tool, including RF, PLS-DA, Linear SVM, and Logistic Regression (LR). The features were separated and retested with multiple model types to compare the best model performance, based on the AUC result (Table 7-9). The four reduced groups tested were the seven up-regulated features, the eight down-regulated features, the top five features with the most significant adjusted p-values, and finally the top six features with the most significant FC (three with significant positive FC and three with significant negative FC).

Table 7-9: Area under the curve (AUC) results obtained from multiple model testing with a variety of feature groups. RF denotes random forest, PLS-DA indicates partial least-squares discriminant analysis, SVM means linear support vector machines, and FC stands for fold change.

AUC Results	RF	PLS-DA	SVM
All up-regulated features	0.968	0.973	0.954
All down-regulated features	0.963	0.966	0.932
Top 5 significant adjusted p-value features	0.971	0.996	0.988
Top 6 significant FC features	0.963	1.00	0.997

Due to the nature of the desired classification (treated and untreated samples) the top FC features were selected for future model testing. There was also some overlap in the top p-value and FC features groups, namely, *m/z* 317.05176 and *m/z* 365.13611. Whilst good AUC scores were initially obtained with the smaller dataset, the accumulation of all treated horse samples with appropriate timepoints caused a decrease in the models' performance. This was due to an increase in the variability of abundance readings and a lack of consistency in feature availability between batches (particularly regarding *m/z* 336.04071).

Six of those were selected based on their significant FC data collected. These six features were then tested in MetaboAnalyst's Biomarker Analysis tool with a dataset from one example batch, and then advanced to the combined study samples once proven successful. Table 7-10 shows a severe reduction in predictive accuracy using the MetaboAnalyst models once the whole study dataset is implemented (dataset provided in Appendix 30, along with some additional model testing data in Appendix 31).

Table 7-10: Parameters and outcomes of multiple model tests with a variety of dataset sizes.

Dataset used	Model type	Parameters	Outcome
Batch 7 triplicate samples, Pre and Post (12-168 h) treated horses *n* = 33 Pre/post = 12/21 6 features	[a]RF	Built-in classification method and feature ranking	[e]AUC = 0.974 [g]Pred. Acc. = 91.6%
	[b]PLS-DA		AUC = 1.00 Pred. Acc. = 97.5%
	[c]SVM		AUC = 1.00 Pred. Acc. = 95.1%
Combined study averaged dataset (B1-10, Pre and Post (12-168 h) treated horses) *n* = 128 Pre/post = 64/64 6 features	RF	Built-in classification method and feature ranking	AUC = 0.644 Pred. Acc. = 62.1%
	PLS-DA		AUC = 0.733 Pred. Acc. = 66.7%
	SVM		AUC = 0.721 Pred. Acc. = 66.6%
Combined study averaged dataset 6 features	[d]XGB	Train/Test/Val% = 60/30/10 Eval metric = “logloss” Early stopping = 10 Learning rate = 0.3 Min child weight = 6 (tested with max depth = 3 with no change to results)	[e,f]PR-AUC = 0.879 Pred. Acc. = 79.49% [h]Class. Error = 0.2051 F1 Score = 0.7333

	XGB	Train/Test/Val% = **70/20/10**	PR-AUC = 0.860 Pred. Acc. = 73.08% Class. Error = 0.2692 F1 Score = 0.6957
Combined study averaged dataset **4** features (211, 280, 317, 265)	XGB	Train/Test/Val% = 60/30/10 Eval metric = "logloss" Early stopping = 10 Learning rate = 0.3 Min child weight = 6	PR-AUC = 0.879 Pred. Acc. = 79.49% Class. Error = 0.2051 F1 Score = 0.7333
Combined study averaged dataset **3** features (211, 317, 365)	XGB		PR-AUC = 0.853 Pred. Acc. = 61.54% Class. Error = 0.3846 F1 Score = 0.6809

Table Definitions: [a]RF denotes random forest, [b]PLS-DA indicates partial least-squares discriminant analysis, [c]SVM means linear support vector machines, [d]XGB refers to extreme gradient boosting. [e]AUC means area under the curve the addition of [f]PR is for precision recall. [g]Pred.Acc. and [h]Class. Error are the shorthand for predictive accuracy and classification error.

Due to the large decrease in model performance, an XGBoost algorithm (run in Python) was also tested for potential improvement, and greater model suitability. These tests produced higher PR-AUC (Figure 7-10) and classification accuracy scores than previous testing on the same dataset. Using the in-built feature importance function also indicated which features were assisting in the classification process, which were identified in order of importance as *m/z* 211.09766, 365.13611, 317.05176, and 280.62173. Due to this, the two remaining features were dropped from the model as it was evident during testing that they provided no additional improvement to the models' performance, as noted by the unchanged outcomes in Table 7-10, rows 3 and 4.

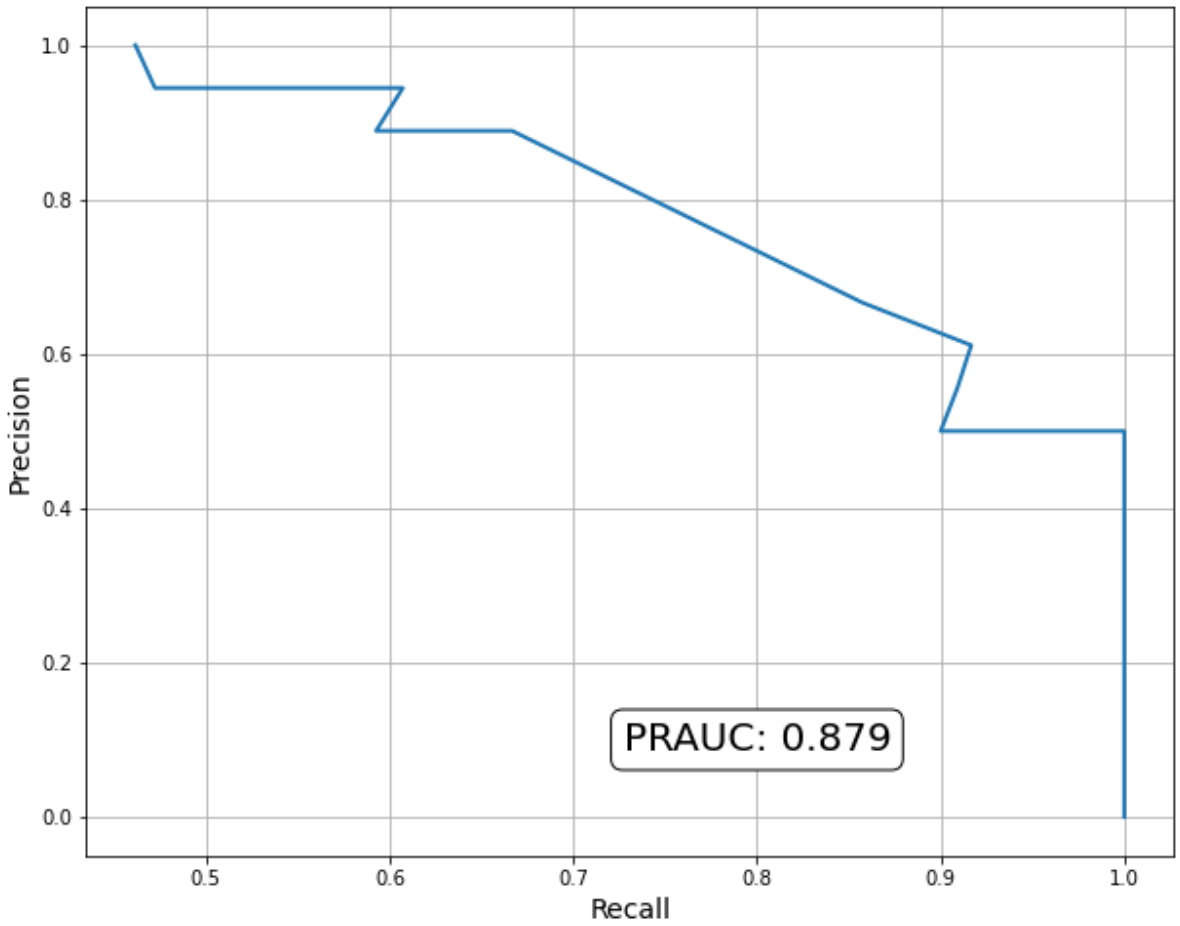


Figure 7-10: Best outcome achieved from XGBoost model with whole study dataset and 4 features included in the profile.

The confusion matrix data printed with the code was used as another model performance testing tool (Table 7-11). These results show a high degree of true negative results, indicated by the value of 20. However, the value of seven (7) in the following row implies a number of false negatives, which translates to positive post-administration samples being incorrectly classified as pre-administration samples. This may be due to the large timepoint range used in this dataset and the variability in excretion of the chosen features over the 7-day time period. Trialling this model with a smaller timeframe, i.e., 12 – 48 hours, may reduce that number of false negatives classified by the model.

Table 7-11: Confusion matrix data regarding best outcome for XGB model with four biomarkers, incorporating the full study dataset. The 0 and 1 indicate pre and post administration classes, respectively.

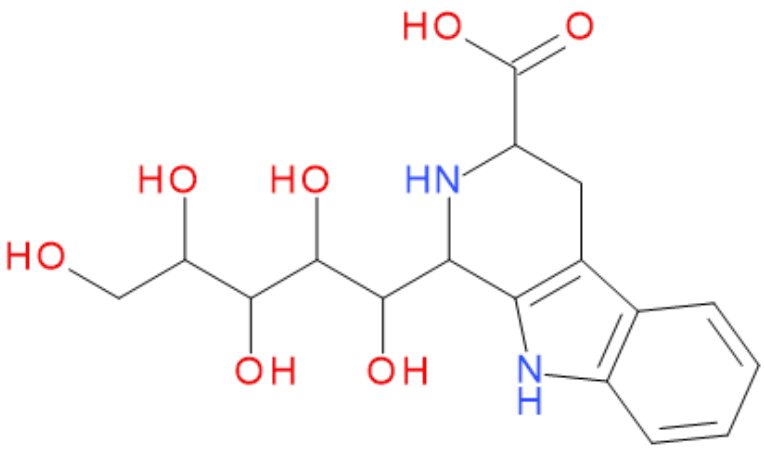
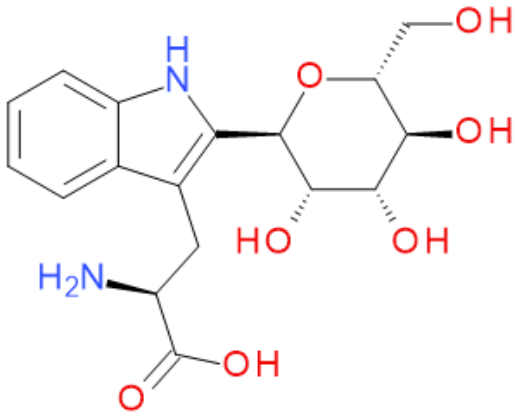
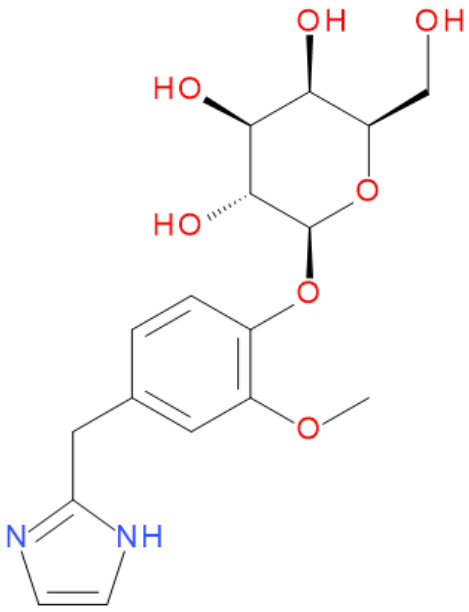
		Predicted	
		Pre-Administration (0)	**Post-Administration (1)**
Actual	**Pre-Administration (0)**	20	1
	Post-Administration (1)	7	11

7.3.3 Feature Identification

Putative identification of these markers was conducted using MetaboAnalyst's Peak Annotation tool for MS/MS DDA and DIA data. MS/MS annotation of *m/z* 211 returned a similarity score with 3,4-methyleneazeliac acid ($C_{11}H_{16}O_4$), a medium-chain fatty acid, and 2-amino-3,4-dimethylimidazo(4,5-f) quinoline ($C_{12}H_{12}N_4$). MS/MS of *m/z* 365 returned a similarity score with 1-(1,2,3,4,5-pentahydroxypent-1-yl)-1,2,3,4-tetrahydro-beta-carboline-3-carboxylate ($C_{17}H_{22}N_2O_7$), a novel carbohydrate-derived β-carboline found in fruit products, tryptophan 2-C-mannoside ($C_{17}H_{22}N_2O_7$), a biomarker of kidney function in blood and urine, and semilepidinoside B ($C_{17}H_{22}N_2O_7$), a phenolic glycoside found in food and plants.

However, the compound matches provided by MetaboAnalyst (Table 7-12 and Table 7-14) do not align with the fragmentation of the named metabolite and the common fragmentation seen across all batches (Table 7-13 and Table 7-15). For example, sulfate like fragmentation was noted for *m/z* 211 ion, however, none of the identified compounds from MetaboAnalyst contained a sulfate conjugate group or similar fragments. This is likely due to the lack of research into these compounds, meaning their presence in a predefined MS/MS library is unlikely.

Table 7-12: MetaboAnalyst identified compounds for MS/MS fragmentation collected for ion m/z 365.

	
1-(1,2,3,4,5-pentahydroxypent-1-yl)-1,2,3,4-tetrahydro-beta-carboline-3-carboxylate	
$C_{17}H_{22}N_2O_7$	
	
tryptophan 2-C-mannoside	semilepidinoside B
$C_{17}H_{22}N_2O_7$	$C_{17}H_{22}N_2O_7$

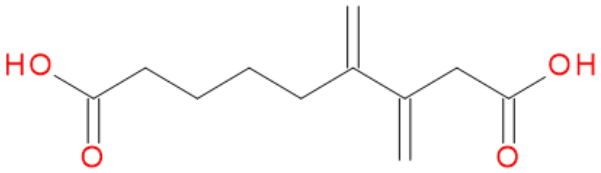
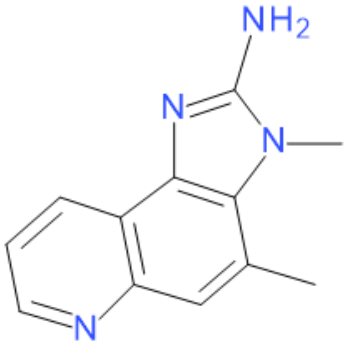
In Table 7-13, the main glucuronide ions that appeared were 113.0236 ($C_5H_5O_3$), 129.0188 ($C_5H_5O_4$), and 175.0257 ($C_6H_7O_6$). The neutral loss of $C_6H_8O_6$ (176.0321 Da) is also evident in the appearance of fragment ion *m/z* 189.1045 (i.e., M–$C_6H_8O_6$).

Table 7-13: Main MS/MS fragment ions found in metabolomic dataset for ion m/z 365.

Fragment Ions (*m/z*)	**Proposed Fragment Structure**
365.0895 (precursor ion)	$C_{17}H_{22}N_2O_7$
189.1045	$C_{11}H_{13}N_2O$
175.0257	$C_6H_7O_6$
145.0977	$C_5H_5O_5$
129.0188	$C_5H_5O_4$
113.0236	$C_5H_5O_3$

Through manual MS/MS searching and fragmentation predictions it is possible for *m/z* 365 to be tryptophan 2-C-mannoside. However, not a lot is known about the metabolism of tryptophan 2-C-mannoside in humans or animals, so very little was found regarding the likely process occurring to cause elevation upon levodopa administration. It is suspected to be produced through the C-mannosylation of tryptophan via the C-mannosyl-transferase enzymatic pathway [176].

Table 7-14: MetaboAnalyst identified compounds for MS/MS fragmentation collected for ion m/z 211.

	
3,4-methyleneazeliac acid	2-amino-3,4-dimethylimidazo (4,5-f) quinoline
$C_{11}H_{16}O_4$	$C_{12}H_{12}N_4$

In Table 7-15, the main sulfated ions that appeared were 79.95628 (SO_3), 80.96407 (HSO_3), and 96.95944 (HSO_4). The neutral loss of SO_3 (79.9568 Da) is also evident in the appearance of fragment ion *m/z* 132.0451 (i.e., M–SO_3). There are still two prominent fragment ions that are unidentified, *m/z* 93.03385 and *m/z* 102.0556. This is mainly due to the still unknown main structure of the compound.

Table 7-15: Main MS/MS fragment ions found in metabolomic dataset for ion m/z 211.

Fragment Ions (*m/z*)	**Proposed Fragment Structure**
211.0977 (precursor ion)	$[M–H]^-$
132.0451	$[M–SO_3]^-$
102.0556	?
96.95944	HSO_4
93.03385	?
80.96407	HSO_3
79.95628	SO_3

7.4 DISCUSSION

Metabolomic analysis of *Stalevo*® administration data returned a group of related features with the potential to extend the detection window for dopaminergic manipulation in racehorses. The classification model, presented above, was designed to differentiate a non-treated horse sample, to a treated horse sample. From the previous direct detection method, the 3-MT/Tyr ratio can detect levodopa misuse up to 8 hours post administration. This model targeted what potential markers might be useful in determining levodopa misuse after 8 hours. To achieve this, the treated samples applied to this model included only timepoints from 12 hours up to 7 days post administration (168 hours).

The area of research aimed to devise a metabolomic multi-tool workflow that could identify metabolites of interest in the detection of dopaminergic manipulation in racehorses. Whilst further work is still necessary, this work was successful in presenting a fit-for-purpose metabolomic pipeline with a focus on phase I and II conjugated compounds.

This pipeline presented above is a practical example of how metabolomic analysis is conducted and the types of results that may be obtained. This work is a cautionary tale of the importance of robust and thorough compound identification prior to the publishing of any novel results, in case the features detected cannot be definitively linked meaningfully to the drug administration that is being targeted. Robust validation of the model with external administration and control samples, preferably from a separate laboratory facility, should be conducted to verify positive correlation of results and reinforce the appropriateness of the identified biomarkers. Without additional computational power (currently unavailable at this time) and some missing datafiles (beyond the author's control) the desired next steps are not possible. However, these steps include the reprocessing of select control and treated horse sample datafiles (mzML) via MS-DIAL for appropriate realignment, and statistical analysis through MetaboAnalyst to acquire features of interest specific to the comparison of post-administration control horse samples and treated horse samples. This time targeting changes post- drug administration that do not include external influences also seen by the control horses.

The results of this work are a testament to the importance of a robust study design with the use of control horses to adjust for external factors that could influence the metabolome of the horse, other than the administered substance. Additionally, the benefit of a randomised study model in minimising the extraction and analysis bias, provided an unforeseen roadblock in the post processing interpretation of the data and results. This proved a challenge and added a time-consuming manual data-mining factor that was not experienced in the ALT administration

study. The introduction of DIA data analysis mode, as opposed to DDA ensured that all potentially relevant information was collected, however, this increased the raw file size ten-fold, and made storage and transfer of the raw data complicated.

7.4.1 Study Limitations

All areas of this presented pipeline were limited by the computational power that was available at the time of this research. As previously mentioned in other chapters (**Chapter 2: The Current Applications and Future Potential Implementations of Machine Learning Strategies in Metabolomic-Based Anti-Doping Research**), high computational power is often required for machine learning and metabolomic data processing due to the large amounts of information collected through HRMS analysis. The computer used for the majority of data processing was a Windows Processor 12th Gen Intel® Core ™ i5-10500, 2667 MHz, running 6 cores, and 16 gigabytes of memory, and in some instances, fell short of what was required for further investigation.

Data Acquisition

DIA mode produces far larger raw MS datafiles in comparison to DDA mode or full scan. Whilst this provides an abundance of information for the benefit of data mining for possible metabolic results, it also adds greater difficulty in MS/MS interpretation. As everything is collected with this method, determining true findings from background noise is time-consuming and a major limitation of this option.

The chosen DIA ion range of *m/z* 50-400 was unfortunately too narrow to detect all potential glucuronide conjugated compounds of interest. One compound in particular is the glucuronide conjugated entacapone, which has a monoisotopic mass of 481.13326 Da. This can, and will be rectified by reanalysing some representative samples with a broader range to scan for these compounds that are anticipated to be present/absent.

Some alteration to the acquisition method was conducted after the first and second batches were analysed due to consultation with an applications specialist from Shimadzu. Due to time restraints the original batches were not reanalysed at the time, however, these changes have been considered when processing the results. The few changes made were a flowrate reduction, from 0.4 mL/min and 0.3 mL/min to improve peak shape, and a smaller number of DIA event windows (reduced from 30 to 20), increasing the Q1 transmission width from *m/z* 10 to 15 to improve datafile size manageability.

Data Pre-Processing

Whilst the randomisation of administration samples for extraction and analysis was beneficial in reducing bias, an unforeseen issue arose in the data pre-processing step, and affected all steps proceeding this. As alignment and filtering was conducted per-batch with MS-DIAL, uncertainty arose when comparing samples from a single horse across multiple batches for timewise plotting of found features. The mix of timepoints per batch meant some features may have been missed due to MS/MS thresholds, or removed for lack of variance in the filtering stage.

Data Processing and Analysis

Using MetaboAnalyst for the filtering, gap filling, normalisation, transformation and statistical analysis of each batch was user-friendly and created a low barrier for entry for more complex processing and analysis techniques. However, this web-based tool was limited by slower processing times due to its remote servers, and lacked transparency on how data manipulation and aggregation was conducted.

Machine Learning

The ML modelling on both MetaboAnalyst and within customised Python scripts were limited in their ability for results interpretation and the biological significance of the model outcomes. Whilst the modelling was successful in identifying features which actively contributed to the classification of pre- and post-drug administration, the role (if any) these features had in the drug metabolism could not be stated. This is the great benefit and limitation of metabolomic research, the ability to capture all metabolomic information, whether it is relevant to the research question or not.

Compound Identification

Features found in the experimental data can only be putatively identified via their MS/MS information without the purchase of true reference standards. However, these standards can often be expensive or difficult to acquire (sometimes entirely unavailable) due to their complex nature. This is especially true for sulfated or glucuronide conjugated compounds.

7.4.2 Future Work

Further optimisation and validation should be conducted on this model prior to implementation for the detection suspected levodopa misuse. The aim would be to implement the proposed

dopamine classification model features for monitoring and verification with raceday samples once validation could be completed.

Further work includes the confirmed identification of the four features presented above with purchased reference standards to align retention time, *m/z*, and MS/MS fragmentation. Further optimisation and validation of this model should be conducted prior to the implementation of these features. Ideally this should be done using an unrelated dataset, where the horses' administration is also known.

These features have been trialled as a ratio with little success, however, the development of a longitudinal profiling system for the abundance of these compounds could assist in providing additional intelligence regarding the detection of dopamine manipulation in racehorses. A system like this could be implemented into the current EBP system Racing NSW has in place to bolster the current threshold used for the detection of dopaminergic manipulation in racehorses.

Chapter 8: Conclusions and Future Recommendations

8 CONCLUSIONS AND FUTURE RECOMMENDATIONS

8.1 CONCLUSIONS

8.1.1 Altrenogest Administration Study

This research aimed to analyse phase I and II metabolites, quantitatively and qualitatively, in mares impacted by differing altrenogest administration routes. A validated method was developed for the targeted quantification of steroid impurities by LC-MS/MS in equine urine. No evidence of elevation in these steroid impurities was observed outside 24 hours of oral altrenogest administration. The quantified steroid impurities in injected altrenogest samples, however, exceeded 1 ng/mL up to two days post-final administration. A progesterone ratio classification model for administration route differentiation was also proposed using an extreme gradient boosting machine learning algorithm, providing 84.21% predictive accuracy for the identification of an intramuscularly injected altrenogest sample.

To improve the detection of intramuscularly injected altrenogest and its differentiation to oral altrenogest, a metabolomic approach to the two equine administration routes of the steroidal progestin altrenogest was also investigated. LC-HRMS analysis of equine urine samples found five sulfated compounds with the potential to differentiate between oral and intramuscularly injected altrenogest administration using a random forest classification model. Using estrone sulfate, testosterone sulfate, 2-methoxyestradiol sulfate, pregnenolone sulfate, and cortisol sulfate, the best model results gave an AUC score of 0.965 with a confidence level of 95% (between 0.931 and 0.995). These compound identifications were confirmed with assistance from the Shimadzu Insight Explore Assign feature, as well as MS/MS spectrum and retention time matching with purchased and synthesised reference standards.

8.1.2 Levodopa Administration Study

This area of research successfully addressed the aim to optimise and validate a fit-for-purpose LC-MS method for the targeted detection of dopaminergic manipulation in equine urine via the quantification of 3-methoxytyramine and tyramine. This work also reiterated the benefit of implementing the proposed 3-methoxytyramine intelligence limit (0.776 µg/mL) and 3-methoxytyramine-to-tyramine ratio limit (5.3) to better detect the sub-threshold doping of levodopa, with a concern that advances in Parkinson's disease medication formulations have negated the effectiveness of the current international threshold for the detection of exogenous levodopa use.

To address these concerns, a metabolomic multi-tool workflow was proposed and implemented to identify metabolites of interest in the detection of dopaminergic manipulation in racehorses. Metabolomic analysis of *Stalevo*® administration data returned a group of related features with the potential to extend the detection window for dopaminergic manipulation in racehorses. The extreme gradient boosting classification model was designed to differentiate a non-treated horse sample, to a treated horse sample. As the 3-methoxytyramine-to-tyramine ratio was proven to detect levodopa misuse up to eight hours post-administration, this model targeted biomarkers potentially useful in determining levodopa misuse after this eight-hour post-administration period. Whilst further work is still necessary, this area of research was successful in presenting a fit-for-purpose metabolomic pipeline with a focus on identifying phase I and II conjugated compounds that may be beneficial in extending the current detection window of dopaminergic manipulation in racehorses.

This thesis successfully addressed its aims to investigate progestogenic and dopaminergic substance misuse in racehorses via the direct quantification and metabolomic analysis in a complementary approach. Additionally, the use of data processing workflows and machine learning tools were implemented successfully to improve urinary biomarker discovery in an equine anti-doping context.

8.1.3 Overall Limitations

There were also some key limitations identified over the duration of this research, both regarding specific areas of the results, as well as some larger overarching issues. An important limitation to address is the ruggedness and durability of high-resolution mass spectrometry equipment. Whilst these analytical instruments have great advantages in their acquisition of high-throughput data and the consequent identification of novel compounds, they may often fall short in their detector longevity (among other internal parts). Detector issues, resulting in a failure to successfully run the instrument, were constant throughout sample analyses of this work, causing delays at highly inconvenient points. Ultimately, this only demonstrated the reality that while high resolution analytical equipment is advantageous when applied in novel research, they are not robust enough to withstand rigorous daily use expected of routine analytical equipment. This means that the translation of some of this research to more rugged liquid chromatography-mass spectrometry instrumentation is still required for appropriate implementation.

As previously stated, high resolution mass spectrometry acquires large amounts of high-throughput data, translating to demands for large volumes of data storage. Unfortunately, during this research the ability to regularly and reliably backup this data was limited (access

restrictions) due to security concerns from the host laboratory organisation. This resulted in the loss of some raw data files relating to the metabolomic analysis of the *Stalevo*® administration study, which was out of the author's control. This negatively affected any plans for reprocessing of the raw data grouped by administration horse, rather than by batch as it was run analytically. In an attempt to obtain some more meaningful information from this data an amalgamation of recovered and previously saved datafiles was done (which amounted to approximately 70% of -ESI mzML files), and a combined realignment and normalisation (of samples from all 12 horses) through MSDIAL was applied. Further analysis and interpretation of this data will be completed outside the completion of this thesis.

Lastly, as this research commenced in 2021, the COVID-19 pandemic created a delay within the method development and validation, resulting in a postponement to sample processing. Due to lockdown restrictions, time lost from the pandemic totalled three and a half months away from the laboratory.

8.2 FUTURE RECOMMENDATIONS

8.2.1 Industry

The work presented in this thesis provides the opportunity for governing racing bodies to consider the implementation of one or many of the following measures for more comprehensive progestogenic and dopaminergic substance doping detection.

The quantification of steroid impurities, trendione, trenbolone and epitrenbolone, in equine urine as opposed to a qualitative screen for their existence. The detection of these compounds cannot be relied upon as illegal altrenogest use as this research has confirmed the detectable presence of these compounds in both oral and injected administration routes. The monitoring of urinary progesterone levels in racehorses with the confirmed presence of altrenogest would also be a simple and low effort implementation that could assist in the differentiation of altrenogest administration routes when assessing multiple single horse sample. Alternatively, implementing the progesterone ratio classification model would be a simple method for an analyst with rudimentary coding skills to provide sound altrenogest route differentiation in race-day urine samples. For more advanced analysts, and for a more robust determination model, the sulfate biomarker classification algorithm should be implemented.

For the detection of dopaminergic manipulation in racehorses, the implementation of the proposed 3-methoxytyramine intelligence limit (0.776 µg/mL) and 3-methoxytyramine-to-

tyramine ratio limit (5.3) is highly recommended to extend the rapidly shortening detection window. A group of levodopa related biomarkers were also presented in this thesis as a model and the monitoring of some of these markers is encouraged. However, further investigation and validation is still necessary prior to any implementation.

8.2.2 Research

This research has responded to imperative questions for Racing NSW regarding the detection of progestogenic and dopaminergic substance misuse in equine urine. In doing so, new areas have been highlighted where there is the potential for research to be expanded. These areas include the application of models proposed in this thesis to longitudinal assessment with a range of horses to ensure repeatability of results and robustness against inter- and intra-metabolic variations. From this testing, the translation of these models into the current EBP biomarker profiling system should be conducted to improve the current detection systems and provide additional intelligence to racing official stewards and veterinarians. Further investigation of the metabolomic analyses results for the identification and verification of appropriate features for the detection of dopaminergic manipulation is still necessary. Lastly, the potential applications of the proposed sulfate biomarker classification model to other doping scenarios could be investigated.

References

References

1. Fragkaki, A. G., Kioukia-Fougia, N., Kiousi, P., Kioussi, M., and Tsivou, M., "Challenges in detecting substances for equine anti-doping," *Drug Testing and Analysis,* vol. 9, no. 9, pp. 1291-1303, 2017, doi: https://doi.org/10.1002/dta.2162.

2. IFHA. "International Agreement: Article 6A - Prohibited Substances." https://www.ifhaonline.org/default.asp?section=IABRWTest&area=2#article6a (accessed 21 May 2020).

3. *World Anti-Doping Code: International Standard*, WADA, www.wada-ama.org, 2025. [Online]. Available: https://www.wada-ama.org/sites/default/files/2024-09/2025list_en_final_clean_12_september_2024.pdf

4. Teale, P., Barton, C., Driver, P. M., and Kay, R. G., "Biomarkers: unrealized potential in sports doping analysis," *Bioanalysis,* vol. 1, no. 6, pp. 1103-1118, 2009, doi: https://doi.org/10.4155/bio.09.87.

5. Zolg, J. W. and Langen, H., "How Industry Is Approaching the Search for New Diagnostic Markers and Biomarkers," *Molecular & Cellular Proteomics,* vol. 3, no. 4, p. 345, 2004, doi: https://doi.org/10.1074/mcp.M400007-MCP200.

6. Scarth, J. P., Teale, P., and Kuuranne, T., "Drug metabolism in the horse: a review," *Drug Testing and Analysis,* vol. 3, no. 1, pp. 19-53, 2011, doi: https://doi.org/10.1002/dta.174.

7. WADA. "Athlete Biological Passport." https://www.wada-ama.org/en/athlete-biological-passport (accessed 21 May 2020).

8. WADA. "Athlete Biological Passport (ABP) Operating Guidelines." WADA AMA. https://www.wada-ama.org/en/resources/athlete-biological-passport/athlete-biological-passport-abp-operating-guidelines (accessed 25 May 2020).

9. Sottas, P.-E., Robinson, N., Rabin, O., and Saugy, M., "The Athlete Biological Passport," *Clinical Chemistry,* vol. 57, no. 7, pp. 969-976, 2011, doi: https://doi.org/10.1373/clinchem.2011.162271.

10. de la Torre, X., Jardines, D., Curcio, D., Colamonici, C., and Botrè, F., "Isotope ratio mass spectrometry in antidoping analysis: The use of endogenous reference compounds," *Rapid Communications in Mass Spectrometry,* vol. 33, no. 6, pp. 579-586, 2019, doi: https://doi.org/10.1002/rcm.8377.

11. Duluard, A., Bailley-Chouriberry, L., Kieken, F., Popot, M.-A., and Bonnaire, Y., "Longitudinal follow-up on racehorses: Veterinary and analytical issues. A one year study of French trotters," presented at the 18th Proc. Int. Conf. Racing Anal. Vets., New Zealand, 2010.

12. Cawley, A. T. and Keledjian, J., "Intelligence-based anti-doping from an equine biological passport," *Drug Testing and Analysis,* vol. 9, no. 9, pp. 1441-1447, 2017, doi: https://doi.org/10.1002/dta.2180.

13. Viljanto, M., Hincks, P., Hillyer, L., Cawley, A., Suann, C., Noble, G., Walker, C. J., Parkin, M. C., Kicman, A. T., and Scarth, J., "Monitoring dehydroepiandrosterone (DHEA) in the urine of Thoroughbred geldings for doping control purposes," (in eng), *Drug Test Anal,* vol. 10, no. 10, pp. 1518-1527, 2018, doi: https://doi.org/10.1002/dta.2411.

14. Viljanto, M., Scarth, J., Hincks, P., Hillyer, L., Cawley, A., Suann, C., Noble, G., Walker, C. J., Kicman, A. T., and Parkin, M. C., "Application of testosterone to epitestosterone ratio to horse urine - a complementary approach to detect the administrations of testosterone and its pro-drugs in Thoroughbred geldings," (in eng), *Drug Test Anal,* vol. 9, no. 9, pp. 1328-1336, 2017, doi: https://doi.org/10.1002/dta.2109.

15. Bonnaire, Y., Dehennin, L., Plou, P., and Toutain, P. L., "Testosterone administration to mares: criteria for detection of testosterone abuse by analysis of metabolites in plasma and urine," (in eng), *J Anal Toxicol,* vol. 19, no. 3, pp. 175-81, 1995, doi: https://doi.org/10.1093/jat/19.3.175.

16. Elbourne, M., Cawley, A., Stanley, S., Bowen, C., and Fu, S., "Intelligence benefit of the 3-methoxytyramine to tyramine ratio in equine urine," *Drug testing and analysis,* vol. 14, no. 5, pp. 936-942, 2022, doi: https://doi.org/10.1002/dta.3264.

17. Klein, D. J., Anthony, T. G., and McKeever, K. H., "Metabolomics in equine sport and exercise," *Journal of Animal Physiology and Animal Nutrition,* vol. 105, no. 1, pp. 140-148, 2021, doi: https://doi.org/10.1111/jpn.13384.

18. Dettmer, K., Aronov, P. A., and Hammock, B. D., "Mass spectrometry - based metabolomics," *Mass Spectrometry Reviews,* vol. 26, no. 1, pp. 51-78, 2007, doi: https://doi.org/10.1002/mas.20108.

19. Keen, B., Cawley, A., Reedy, B., and Fu, S., "Metabolomics in clinical and forensic toxicology, sports anti-doping and veterinary residues," *Drug Testing and Analysis,* vol. 14, no. 5, pp. 794-807, 2022, doi: https://doi.org/10.1002/dta.3245.

20. Cajka, T. and Fiehn, O., "Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics," *Analytical Chemistry,* vol. 88, no. 1, pp. 524-545, 2016, doi: https://doi.org/10.1021/acs.analchem.5b04491.

21. Contrepois, K., Jiang, L., and Snyder, M., "Optimized Analytical Procedures for the Untargeted Metabolomic Profiling of Human Urine and Plasma by Combining Hydrophilic Interaction (HILIC) and Reverse-Phase Liquid Chromatography (RPLC)–Mass Spectrometry*," *Molecular & Cellular Proteomics,* vol. 14, no. 6, pp. 1684-1695, 2015, doi: https://doi.org/10.1074/mcp.m114.046508.

22. Madji Hounoum, B., Blasco, H., Nadal-Desbarats, L., Diémé, B., Montigny, F., Andres, C. R., Emond, P., and Mavel, S., "Analytical methodology for metabolomics study of adherent mammalian cells using NMR, GC-MS and LC-HRMS," *Analytical and Bioanalytical Chemistry,* vol. 407, no. 29, pp. 8861-8872, 2015, doi: https://doi.org/10.1007/s00216-015-9047-x.

23. Stojiljkovic, N., Paris, A., Garcia, P., Popot, M. A., Bonnaire, Y., Tabet, J. C., and Junot, C., "Evaluation of horse urine sample preparation methods for metabolomics using LC coupled to HRMS," (in eng), *Bioanalysis,* vol. 6, no. 6, pp. 785-803, 2014, doi: https://doi.org/10.4155/bio.13.324.

24. Matuszewski, B. K., Constanzer, M. L., and Chavez-Eng, C. M., "Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC−MS/MS," *Analytical Chemistry,* vol. 75, no. 13, pp. 3019-3030, 2003, doi: https://doi.org/10.1021/ac020361s.

25. Courant, F., Antignac, J.-P., Dervilly-Pinel, G., and Le Bizec, B., "Basics of mass spectrometry based metabolomics," *PROTEOMICS,* vol. 14, no. 21-22, pp. 2369-2388, 2014, doi: https://doi.org/10.1002/pmic.201400255.

26. Bicker, J., Fortuna, A., Alves, G., and Falcão, A., "Liquid chromatographic methods for the quantification of catecholamines and their metabolites in several biological samples—A review," *Analytica Chimica Acta,* vol. 768, pp. 12-34, 2013, doi: https://doi.org/10.1016/j.aca.2012.12.030.

27. Houghton, E. and Maynard, S., "Some Aspects of Doping and Medication Control in Equine Sports," Springer Berlin Heidelberg, 2009, pp. 369-409, doi: https://doi.org/10.1007/978-3-540-79088-4_17.

28. Tulipani, S., Llorach, R., Urpi-Sarda, M., and Andres-Lacueva, C., "Comparative Analysis of Sample Preparation Methods To Handle the Complexity of the Blood Fluid Metabolome: When Less Is More," *Analytical Chemistry,* vol. 85, no. 1, pp. 341-348, 2013, doi: https://doi.org/10.1021/ac302919t.

29. Poole, C. F., "Milestones in the Development of Liquid-Phase Extraction Techniques," (Handbooks in Separation Science. United States: Elsevier, 2020, ch. 1, pp. 1-36, doi: https://doi.org/10.1016/B978-0-12-816911-7.00001-3.

30. Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., Van Ommen, B., Pujos-Guillot, E., Verheij, E., Wishart, D., and Wopereis, S., "Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research," *Metabolomics,* vol. 5, no. 4, pp. 435-458, 2009, doi: https://doi.org/10.1007/s11306-009-0168-0.

31. Müller, R. K., "History of Doping and Doping Control," Springer Berlin Heidelberg, 2009, pp. 1-23, doi: https://doi.org/10.1007/978-3-540-79088-4_1.

32. Allen, D. and McWhinney, B., "Quadrupole Time-of-Flight Mass Spectrometry: A Paradigm Shift in Toxicology Screening Applications," *Clinical Biochemist Reviews,* vol. 40, no. 3, pp. 135-146, 2019, doi: https://doi.org/10.33176/aacb-19-00023.

33. Pasikanti, K. K., Ho, P. C., and Chan, E. C. Y., "Development and validation of a gas chromatography/mass spectrometry metabonomic platform for the global profiling of urinary metabolites," *Rapid Communications in Mass Spectrometry,* vol. 22, no. 19, pp. 2984-2992, 2008, doi: https://doi.org/10.1002/rcm.3699.

34. Mooney, M. H., Elliott, C. T., and Le Bizec, B., "Combining biomarker screening and mass-spectrometric analysis to detect hormone abuse in cattle," *TrAC Trends in Analytical Chemistry,* vol. 28, no. 6, pp. 665-675, 2009, doi: https://doi.org/10.1016/j.trac.2009.03.011.

35. Ibáñez, M., Sancho, J. V., Bijlsma, L., van Nuijs, A. L. N., Covaci, A., and Hernández, F., "Comprehensive analytical strategies based on high-resolution time-of-flight mass spectrometry to identify new psychoactive substances," *TrAC Trends in Analytical Chemistry,* vol. 57, pp. 107-117, 2014, doi: https://doi.org/10.1016/j.trac.2014.02.009.

36. Tiller, P. R., Yu, S., Castro-Perez, J., Fillgrove, K. L., and Baillie, T. A., "High-throughput, accurate mass liquid chromatography/tandem mass spectrometry on a quadrupole time-of-flight system as a 'first-line' approach for metabolite identification studies," *Rapid Communications in Mass Spectrometry,* vol. 22, no. 7, pp. 1053-1061, 2008, doi: https://doi.org/10.1002/rcm.3472.

37. Girolamo, F., Lante, I., Muraca, M., and Putignani, L., "The Role of Mass Spectrometry in the "Omics" Era," *Current Organic Chemistry,* vol. 17, no. 23, pp. 2891-2905, 2013, doi: https://doi.org/10.2174/13852728178881311181627 25.

38. Narduzzi, L., Dervilly, G., Audran, M., Le Bizec, B., and Buisson, C., "A role for metabolomics in the antidoping toolbox?," *Drug Testing and Analysis,* vol. 12, no. 6, pp. 677-690, 2020, doi: https://doi.org/10.1002/dta.2788.

39. Lu, X., Zhao, X., Bai, C., Zhao, C., Lu, G., and Xu, G., "LC–MS-based metabonomics analysis," *Journal of Chromatography B,* vol. 866, no. 1, pp. 64-76, 2008, doi: https://doi.org/10.1016/j.jchromb.2007.10.022.

40. Hendriks, M. M. W. B., Eeuwijk, F. A. v., Jellema, R. H., Westerhuis, J. A., Reijmers, T. H., Hoefsloot, H. C. J., and Smilde, A. K., "Data-processing strategies for metabolomics studies," *TrAC Trends in Analytical Chemistry,* vol. 30, no. 10, pp. 1685-1698, 2011, doi: https://doi.org/10.1016/j.trac.2011.04.019.

41. Van Den Berg, R. A., Hoefsloot, H. C., Westerhuis, J. A., Smilde, A. K., and Van Der Werf, M. J., "Centering, scaling, and transformations: improving the biological information content of metabolomics data," *BMC Genomics,* vol. 7, no. 1, p. 142, 2006, doi: https://doi.org/10.1186/1471-2164-7-142.

42. Hodgson, D., Howe, S., Jeffcott, L., Reid, S., Mellor, D., and Higgins, A., "Effect of prolonged use of altrenogest on behaviour in mares," *The Veterinary Journal,* vol. 169, no. 3, pp. 322-325, 2005, doi: https://doi.org/10.1016/j.tvjl.2005.03.003.

43. Loy, J., Cawley, A., Sornalingam, K., Scrivener, C. J., Keledjian, J., and Noble, G. K., "Pharmacokinetics of Two Formulations of Altrenogest Administered to Mares," *Drug Testing and Analysis,* 2024, doi: https://doi.org/10.1002/dta.3796.

44. McConaghy, F. F., Green, L. A., Colgan, S., and Morris, L. H., "Studies of the pharmacokinetic profile, in vivo efficacy and safety of injectable altrenogest for the suppression of oestrus in mares," *Australian Veterinary Journal,* vol. 94, no. 7, pp. 248-255, 2016, doi: https://doi.org/10.1111/avj.12459.

45. Machnik, M., Hegger, I., Kietzmann, M., Thevis, M., Guddat, S., and Schänzer, W., "Pharmacokinetics of altrenogest in horses," *Journal of Veterinary Pharmacology and Therapeutics,* vol. 30, no. 1, pp. 86-90, 2007, doi: https://doi.org/10.1111/j.1365-2885.2007.00820.x.

46. NexGenPharmaceuticals. "Altrenogest." NexGen Pharmaceuticals. https://nexgenvetrx.com/blog/horsebreeding/the-administration-of-altrenogest-in-mares/ (accessed).

47. RacingVictoria. "Products Containing Altrenogest - Update." https://www.racingvictoria.com.au/notices/2023-08-22/products-containing-altrenogest-update (accessed 30th September, 2024).

48. Van Gestel, M. F. "Use of Altrenogest in Fillies and Mares." https://www.racingnsw.com.au/news/latest-racing-news/use-of-altrenogest-products-in-fillies-and-mares/ (accessed 12th October, 2021).

49. *Rules of Racing,* R. NSW Racing - General, 2023. Available: https://www.racingnsw.com.au/wp-content/uploads/NSWRules.pdf

50. Railton, S. G. "Warning To Trainers & Veterinarians: Use Of Altrenogest In Fillies And Mares." Racing NSW. https://www.racingnsw.com.au/wp-content/uploads/ALTRENOGEST-WARNING-August-2023.pdf (accessed).

51. Gillon, A., Ho, E. N. M., Chan, G. H. M., Kauff, A., Hughes, G., Lund, R. A., Ashley, Z., Wan, T. S. M., and Heather, A. K., "Unravelling androgens in sport: Altrenogest shows strong activation of the androgen receptor in a mammalian cell bioassay," *Drug Testing and Analysis,* vol. 13, no. 3, pp. 523-528, 2021, doi: https://doi.org/10.1002/dta.2941.

52. Nutt, J. G. and Holford, N. H. G., "The response to levodopa in parkinson's disease: Imposing pharmacological law and order," *Annals of Neurology,* vol. 39, no. 5, pp. pp. 561-573, 1996, doi: https://doi.org/10.1002/ana.410390504.

53. Gonçalves, D., Alves, G., Soares-da-Silva, P., and Falcão, A., "Bioanalytical chromatographic methods for the determination of catechol-O-methyltransferase inhibitors in rodents and human samples: A review," *Analytica Chimica Acta,* vol. 710, pp. 17-32, 2012, doi: https://doi.org/10.1016/j.aca.2011.10.026.

54. Wynne, P. M., Vine, J. H., and Amiet, R. G., "3-Methoxytyramine as an indicator of dopaminergic manipulation in the equine athlete," *Journal of Chromatography B,* vol. 811, no. 1, pp. 93-101, 2004, doi: https://doi.org/10.1016/j.jchromb.2004.03.078.

55. Vine, J. H., Wynne, P. M., and Amiet, R. G., "The quantitative analysis of metabolites of levodopa and dopamine in equine urine," presented at the 12th Proc. Int. Conf. Racing Anal. Vets., 2000.

56. Wynne, P. M., McCaffrey, J. P., Vine, J. H., and Amiet, R. G., "Precursor loading with tyrosine to increase levodopa and dopamine in the horse," presented at the 13th Proc. Int. Conf. Racing Anal. Vets., Cambirdge, UK, 2000.

57. Wynne, P. M., Vine, J. H., and Amiet, R. G., "The quantitative analysis of equine urine for the acid metabolites of levodopa and dopamine," presented at the 14th Proc. Int. Conf. Racing Anal. Vets., 2002.

58. McKinney, A. R., Richards, S. L., Cawley, A. T., Keledjian, J., Wynne, P. M., and Suann, C. J., "Catechol-O-methyltransferase inhibitors and the equine 3-methoxytyramine threshold," presented at the 20th Proc. Int. Conf. Racing Anal. Vets., Mauritius, 2014.

59. Wynne, P. M., Vine, J. H., and Amiet, R. G., "The quantitative analysis of equine urine for metabolites of levodopa and dopamine," presented at the 13th Proc. Int. Conf. Racing Anal. Vets., 2000.

60. Cawley, A., Keen, B., Tou, K., Elbourne, M., and Keledjian, J., "Biomarker ratios," *Drug Testing and Analysis,* vol. 14, no. 5, pp. 983-990, 2022, doi: https://doi.org/10.1002/dta.3250.

61. Cawley, A., "Biomarker analysis," *Drug Testing and Analysis,* vol. 14, no. 5, pp. 791-793, 2022, doi: https://doi.org/10.1002/dta.3268.

62. Richards, S. L., Cawley, A. T., McKinney, A. R., Suann, C. J., Keledjian, J., and Wynne, P. M., "COMT Inhibitors - A potential means to circumvent the 3-methoxytyramine threshold," in *20th Proc. Int. Conf. Racing Anal. Vets.*, Mauritius, 2014, vol. 20: ICRAV.

63. Stanley, S., Van den Berg, K., Foo, H. C., and Deng, D., "Metabolism and elimination of the catechol-o-methyltransferase inhibitor tolcapone in the horse," *Drug Testing and Analysis,* vol. 11, no. 4, pp. 578-585, 2019, doi: https://doi.org/10.1002/dta.2531.

64. Wikberg, T., Vuorela, A., Ottoila, P., and Taskinen, J., "Identification of major metabolites of the catechol-O-methyltransferase inhibitor entacapone in rats and humans," (in eng), *Drug Metab Dispos,* vol. 21, no. 1, pp. 81-92, 1993.

65. Keränen, T., Gordin, A., Karlsson, M., Korpela, K., Pentikäinen, P., Rita, H., Schultz, E., Seppälä, L., and Wikberg, T., "Inhibition of soluble catechol- O -methyltransferase and single-dose pharmacokinetics after oral and intravenous administration of entacapone," *European Journal of Clinical Pharmacology,* vol. 46, no. 2, pp. 151-157, 1994, doi: https://doi.org/10.1007/BF00199880.

66. Männistö, P. T., Ulmanen, I., Lundström, K., Taskinen, J., Tenhunen, J., Tilgmann, C., and Kaakkola, S., "Characteristics of catechol O-methyltransferase (COMT) and properties of selective COMT inhibitors," in *Progress in Drug Research / Fortschritte der Arzneimittelforschung / Progrès des recherches pharmaceutiques*, U. Bachrach *et al.* Eds. Basel: Birkhäuser Basel, 1992, pp. 291-350, doi: https://doi.org/10.1007/978-3-0348-7144-0_9.

67. Stanley, S., Deng, D., Van den Berg, K., and Foo, H. C., "The equine metabolism of the catechol-O-methyltransferase enzyme inhibitor nitecapone," *Drug Testing and Analysis,* vol. 14, no. 5, pp. 929-935, 2022, doi: https://doi.org/10.1002/dta.3163.

68. Wikberg, T., Ottoila, P., and Taskinen, J., "Identification of major urinary metabolites of the catechol-O-methyltransferase inhibitor entacapone in the dog," *European Journal of Drug Metabolism and Pharmacokinetics,* vol. 18, no. 4, pp. 359-367, 1993, doi: https://doi.org/10.1007/bf03190186.

69. Tan, J., Qin, F., and Yuan, J., "Current applications of artificial intelligence combined with urine detection in disease diagnosis and treatment," (in eng), *Transl Androl Urol,* vol. 10, no. 4, pp. 1769-1779, 2021, doi: https://doi.org/10.21037/tau-20-1405.

70. Galal, A., Talal, M., and Moustafa, A., "Applications of machine learning in metabolomics: Disease modeling and classification," *Frontiers in Genetics,* vol. 13, 2022, doi: https://doi.org/10.3389/fgene.2022.1017340.

71. Steuer, A. E., Brockbals, L., and Kraemer, T., "Untargeted metabolomics approaches to improve casework in clinical and forensic toxicology—“Where are we standing and where are we heading?”," *WIREs Forensic Science,* vol. 4, no. 4, p. e1449, 2022, doi: https://doi.org/10.1002/wfs2.1449.

72. Schneider, A., Hommel, G., and Blettner, M., "Linear Regression Analysis," *Deutsches Ärzteblatt international,* 2010, doi: https://doi.org/10.3238/arztebl.2010.0776.

73. Casson, R. J. and Farmer, L. D., "Understanding and checking the assumptions of linear regression: a primer for medical researchers," *Clinical & Experimental Ophthalmology,* vol. 42, no. 6, pp. 590-596, 2014, doi: https://doi.org/10.1111/ceo.12358.

74. Vetter, T. R. and Schober, P., "Regression: The Apple Does Not Fall Far From the Tree," *Anesthesia and analgesia,* vol. 127, no. 1, pp. 277-283, 2018, doi: https://doi.org/10.1213/ANE.0000000000003424.

75. GeeksforGeeks. "Linear Regression in Machine learning." Sanchhaya Education Private Limited. https://www.geeksforgeeks.org/ml-linear-regression/ (accessed January 14th 2025).

76. Stoltzfus, J. C., "Logistic Regression: A Brief Primer," *Academic Emergency Medicine,* vol. 18, no. 10, pp. 1099-1104, 2011, doi: https://doi.org/10.1111/j.1553-2712.2011.01185.x.

77. Shalev-Shwartz, S. and Ben-David, S., "Decision Trees," 2014, pp. 212-218, doi: https://doi.org/10.1017/CBO9781107298019.019.

78. Kamiński, B., Jakubczyk, M., and Szufel, P., "A framework for sensitivity analysis of decision trees," *Central European Journal of Operations Research,* vol. 26, no. 1, pp. 135-159, 2018, doi: https://doi.org/10.1007/s10100-017-0479-6.

79. Hastie, T., Tibshirani, R., and Friedman, J., *Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Second Edition ed. (Springer Series in Statistics). New York: Springer, 2009.

80. Dhall, D., Kaur, R., and Juneja, M., "Machine Learning: A Review of the Algorithms and Its Applications," vol. 597, (Lecture Notes in Electrical Engineering. Switzerland: Springer International Publishing AG, 2019, pp. 47-63, doi: https://doi.org/10.1007/978-3-030-29407-6_5.

81. Saeys, Y., Inza, I., and Larrañaga, P., "A review of feature selection techniques in bioinformatics," *Bioinformatics,* vol. 23, no. 19, pp. 2507-2517, 2007, doi: https://doi.org/10.1093/bioinformatics/btm344.

82. Cunningham, P. and Delany, S. J., "k-Nearest Neighbour Classifiers - A Tutorial," *ACM Computing Surveys,* vol. 54, no. 6, pp. 1-25, 2022, doi: https://doi.org/10.1145/3459665.

83. Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., and Goodacre, R., "A tutorial review: Metabolomics and partial least squares-discriminant analysis – a marriage of convenience or a shotgun wedding," *Analytica Chimica Acta,* vol. 879, pp. 10-23, 2015, doi: https://doi.org/10.1016/j.aca.2015.02.012.

84. GeeksforGeeks. "K means Clustering – Introduction." Sanchhaya Education Private Limited. https://www.geeksforgeeks.org/k-means-clustering-introduction/ (accessed January 14th, 2025).

85. ACC, "Organised Crime and Drugs in Sport," in "New Generation Performance and Image Enhancing Drugs and Organised Criminal Involvement in their use in Professional Sport," Australian Crime Commission, Australian Criminal Intelligence Commission, 2013. [Online]. Available: https://www.ssaa.org.au/assets/news-resources/research/2013-02_acc-organised-crime-and-drugs-in-sport.pdf

86. Szeremeta, M., Pietrowska, K., Niemcunowicz-Janica, A., Kretowski, A., and Ciborowski, M., "Applications of Metabolomics in Forensic Toxicology and Forensic Medicine," *International Journal of Molecular Sciences,* vol. 22, no. 6, p. 3010, 2021, doi: https://doi.org/10.3390/ijms22063010.

87. Wang, M. W. H., Goodman, J. M., and Allen, T. E. H., "Machine Learning in Predictive Toxicology: Recent Applications and Future Directions for Classification Models," *Chemical Research in Toxicology,* vol. 34, no. 2, pp. 217-239, 2021, doi: https://10.1021/acs.chemrestox.0c00316.

88. Jia, X., Wang, T., and Zhu, H., "Advancing Computational Toxicology by Interpretable Machine Learning," *Environmental Science & Technology,* vol. 57, no. 46, pp. 17690-17706, 2023, doi: https://doi.org/10.1021/acs.est.3c00653.

89. Steuer, A. E., Brockbals, L., and Kraemer, T., "Metabolomic Strategies in Biomarker Research-New Approach for Indirect Identification of Drug Consumption and Sample Manipulation in Clinical and Forensic Toxicology?," (in eng), *Front Chem,* vol. 7, p. 319, 2019, doi: https://10.3389/fchem.2019.00319.

90. Chen, X., Shu, W., Zhao, L., and Wan, J., "Advanced mass spectrometric and spectroscopic methods coupled with machine learning for in vitro diagnosis," *VIEW,* vol. 4, no. 1, p. 20220038, 2023, doi: https://doi.org/10.1002/VIW.20220038.

91. Want, E., "Challenges in Applying Chemometrics to LC–MS-Based Global Metabolite Profile Data," *Bioanalysis,* vol. 1, no. 4, pp. 805-819, 2009, doi: https://doi.org/10.4155/bio.09.64.

92. Considine, E. C., Thomas, G., Boulesteix, A. L., Khashan, A. S., and Kenny, L. C., "Critical review of reporting of the data analysis step in metabolomics," *Metabolomics,* vol. 14, no. 1, 2018, doi: https://doi.org/10.1007/s11306-017-1299-3.

93. Lee, E. S. and Durant, T. J. S., "Supervised machine learning in the mass spectrometry laboratory: A tutorial," *Journal of Mass Spectrometry and Advances in the Clinical Lab,* vol. 23, pp. 1-6, 2022, doi: https://doi.org/10.1016/j.jmsacl.2021.12.001.

94. Ren, S., Hinzman, A. A., Kang, E. L., Szczesniak, R. D., and Lu, L. J., "Computational and statistical analysis of metabolomics data," *Metabolomics,* vol. 11, no. 6, pp. 1492-1513, 2015, doi: https://doi.org/10.1007/s11306-015-0823-6.

95. Gray, B., Lubbock, K., Love, C., Ryder, E., Hudson, S., and Scarth, J., "Analytical advances in horseracing medication and doping control from 2018 to 2023," *Drug Testing and Analysis,* vol. n/a, no. n/a, doi: https://doi.org/10.1002/dta.3760.

96. Yang, Q., Xu, W., Sun, X., Chen, Q., and Niu, B., "The Application of Machine Learning in Doping Detection," *Journal of Chemical Information and Modeling,* vol. 64, no. 23, pp. 8673-8683, 2024, doi: https://doi.org/10.1021/acs.jcim.4c01234.

97. Sardela, P. D. d. O., Sardela, V. F., da Silva, A. M. d. S., Pereira, H. M. G., and de Aquino Neto, F. R., "A pilot study of non-targeted screening for stimulant misuse using high-resolution mass spectrometry," *Forensic Toxicology,* vol. 37, no. 2, 2019, doi: https://doi.org/10.1007/s11419-019-00482-1.

98. Rahman, M. R., Bejder, J., Bonne, T. C., Andersen, A. B., Huertas, J. R., Aikin, R., Nordsborg, N. B., and Maass, W., "Detection of Erythropoietin in Blood to Uncover

Doping in Sports using Machine Learning," 2022: IEEE, pp. 193-201, doi: https://doi.org/10.1109/icdh55609.2022.00038.

99. Streun, G. L., Steuer, A. E., Ebert, L. C., Dobay, A., and Kraemer, T., "Interpretable machine learning model to detect chemically adulterated urine samples analyzed by high resolution mass spectrometry," *Clinical Chemistry and Laboratory Medicine (CCLM),* vol. 59, no. 8, pp. 1392-1399, 2021, doi: https://doi.org/10.1515/cclm-2021-0010.

100. Du, Y., Hua, Z., Liu, C., Lv, R., Jia, W., and Su, M., "ATR-FTIR combined with machine learning for the fast non-targeted screening of new psychoactive substances," *Forensic Science International,* vol. 349, p. 111761, 2023, doi: https://doi.org/10.1016/j.forsciint.2023.111761.

101. Chen, M., Bai, H., Zhuo, X., Mo, F., Yao, S., Shi, L., Qin, Y., and He, Y., "Label-Free identification of fentanyl analogues using surface-enhanced Raman scattering and machine learning algorithm," *Vibrational Spectroscopy,* vol. 127, p. 103567, 2023, doi: https://doi.org/10.1016/j.vibspec.2023.103567.

102. Lee, S. Y., Lee, S. T., Suh, S., Ko, B. J., and Oh, H. B., "Revealing Unknown Controlled Substances and New Psychoactive Substances Using High-Resolution LC–MS-MS Machine Learning Models and the Hybrid Similarity Search Algorithm," *Journal of Analytical Toxicology,* vol. 46, no. 7, pp. 732-742, 2022, doi: https://doi.org/10.1093/jat/bkab098.

103. Sheng, W., Sun, R., Zhang, R., Xu, P., Wang, Y., Xu, H., Aa, J., Wang, G., and Xie, Y., "Identification of Biomarkers for Methamphetamine Exposure Time Prediction in Mice Using Metabolomics and Machine Learning Approaches," *Metabolites,* vol. 12, no. 12, p. 1250, 2022, doi: https://doi.org/10.3390/metabo12121250.

104. Krombholz, S., Thomas, A., Piper, T., Lagojda, A., Kühne, D., and Thevis, M., "Urinary phenylethylamine metabolites as potential markers for sports drug testing purposes," *Biomedical Chromatography,* vol. 36, no. 2, 2022, doi: https://doi.org/10.1002/bmc.5274.

105. Rocha, D. G., Lana, M. A. G., De Assis, D. C. S., De Macedo, A. N., Corrêa, J. M. M., Augusti, R., and Faria, A. F., "A novel strategy for the detection of boldenone undecylenate misuse in cattle using ultra - high performance liquid chromatography coupled to high resolution orbitrap mass spectrometry: From non - targeted to targeted," *Drug Testing and Analysis,* vol. 14, no. 4, pp. 667-675, 2022, doi: https://doi.org/10.1002/dta.3208.

106. Suzuki, N. and Yamamoto, M., "Roles of renal erythropoietin-producing (REP) cells in the maintenance of systemic oxygen homeostasis," *Pflügers Archiv - European Journal of Physiology,* vol. 468, no. 1, pp. 3-12, 2016, doi: https://doi.org/10.1007/s00424-015-1740-2.

107. Recny, M. A., Scoble, H. A., and Kim, Y., "Structural characterization of natural human urinary and recombinant DNA-derived erythropoietin. Identification of des-arginine 166 erythropoietin," *Journal of Biological Chemistry,* vol. 262, no. 35, pp. 17156-17163, 1987, doi: https://doi.org/10.1016/S0021-9258(18)45504-4.

108. Weber, A. E., Gallo, M. C., Bolia, I. K., Cleary, E. J., Schroeder, T. E., and Rick Hatch, G. F., 3rd, "Anabolic Androgenic Steroids in Orthopaedic Surgery: Current Concepts

and Clinical Applications," (in eng), *J Am Acad Orthop Surg Glob Res Rev,* vol. 6, no. 1, 2022, doi: https://doi.org/10.5435/JAAOSGlobal-D-21-00156.

109. Evans, N. A., "Current Concepts in Anabolic-Androgenic Steroids," *The American Journal of Sports Medicine,* vol. 32, no. 2, pp. 534-542, 2004, doi: https://doi.org/10.1177/0363546503262202.

110. Moreillon, B., Krumm, B., Saugy, J. J., Saugy, M., Botrè, F., Vesin, J. M., and Faiss, R., "Prediction of plasma volume and total hemoglobin mass with machine learning," *Physiological Reports,* vol. 11, no. 19, 2023, doi: https://doi.org/10.14814/phy2.15834.

111. Chan, G. H. M., Ho, E. N. M., Leung, D. K. K., Wong, K. S., and Wan, T. S. M., "Targeted Metabolomics Approach To Detect the Misuse of Steroidal Aromatase Inhibitors in Equine Sports by Biomarker Profiling," *Analytical Chemistry,* vol. 88, no. 1, pp. 764-772, 2016, doi: https://doi.org/10.1021/acs.analchem.5b03165.

112. Ohnuma, K., Uchida, T., Leung, G. N. W., Ueda, T., Obara, T., and Ishii, H., "Establishment of a post - race biomarkers database and application of pathway analysis to identify potential biomarkers in post-race equine plasma," *Drug Testing and Analysis,* vol. 14, no. 5, pp. 915-928, 2022, doi: https://doi.org/10.1002/dta.3041.

113. Shobana, G. and Bushra, S. N., "Drug Administration Route Classification using Machine Learning Models," in *2020 3rd International Conference on Intelligent Sustainable Systems (ICISS)*, 3-5 Dec. 2020 2020, pp. 654-659, doi: https://doi.org/10.1109/ICISS49785.2020.9315975.

114. Misra, B. B. and van der Hooft, J. J. J., "Updates in metabolomics tools and resources: 2014–2015," *ELECTROPHORESIS,* vol. 37, no. 1, pp. 86-110, 2016, doi: https://doi.org/10.1002/elps.201500417.

115. Misra, B. B., Fahrmann, J. F., and Grapov, D., "Review of emerging metabolomic tools and resources: 2015–2016," *ELECTROPHORESIS,* vol. 38, no. 18, pp. 2257-2274, 2017, doi: https://doi.org/10.1002/elps.201700110.

116. Misra, B. B. and Mohapatra, S., "Tools and resources for metabolomics research community: A 2017–2018 update," *ELECTROPHORESIS,* vol. 40, no. 2, pp. 227-246, 2019, doi: https://doi.org/10.1002/elps.201800428.

117. Misra, B. B., "New software tools, databases, and resources in metabolomics: updates from 2020," *Metabolomics,* vol. 17, no. 5, pp. 49-49, 2021, doi: https://doi.org/10.1007/s11306-021-01796-1.

118. O'Shea, K. and Misra, B. B., "Software tools, databases and resources in metabolomics: updates from 2018 to 2019," *Metabolomics,* vol. 16, no. 3, pp. 36-36, 2020, doi: https://doi.org/10.1007/s11306-020-01657-3.

119. Zhang, J. D., Xue, C., Kolachalama, V. B., and Donald, W. A., "Interpretable Machine Learning on Metabolomics Data Reveals Biomarkers for Parkinson's Disease," *ACS Central Science,* vol. 9, no. 5, pp. 1035-1045, 2023, doi: https://doi.org/10.1021/acscentsci.2c01468.

120. Gramatica, P. and Sangion, A., "A Historical Excursus on the Statistical Validation Parameters for QSAR Models: A Clarification Concerning Metrics and Terminology," *Journal of Chemical Information and Modeling,* vol. 56, no. 6, pp. 1127-1131, 2016, doi: https://doi.org/10.1021/acs.jcim.6b00088.

121. He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., and Zhang, K., "The practical implementation of artificial intelligence technologies in medicine," *Nature Medicine,* vol. 25, no. 1, pp. 30-36, 2019, doi: https://doi.org/10.1038/s41591-018-0307-0.

122. *Performance Specification of the Laboratories,* I. F. o. H. Authorities Epitrenbolone (17α-Trenbolone), 2024. Available: https://ifhaonline.org/default.asp?section=IABRW&area=7

123. Peters, F. T. and Maurer, H. H., "Bioanalytical method validation and its implications for forensic and clinical toxicology - A review," *Accreditation and Quality Assurance,* vol. 7, no. 11, pp. 441-449, 2002, doi: https://doi.org/10.1007/s00769-002-0516-5.

124. Peters, F. T., Drummer, O. H., and Musshoff, F., "Validation of new methods," *Forensic Science International,* vol. 165, no. 2, pp. 216-224, 2007, doi: https://doi.org/10.1016/j.forsciint.2006.05.021.

125. AORC, "AORC Guidelines for the Minimum Criteria for Identification by Chromatography and Mass Spectrometry," Association of Official Racing Chemists, aorc-online, 23 August 2016 2016. [Online]. Available: http://www.aorc-online.org/about/

126. Anthony C Moffat, M. D. O. a. B. W., *Clarke's Analysis of Drugs and Poisons*, Fourth Edition ed. Pharmaceutical Press, 2011.

127. Esquivel, A., Alechaga, É., Monfort, N., and Ventura, R., "Direct quantitation of endogenous steroid sulfates in human urine by liquid chromatography-electrospray tandem mass spectrometry," *Drug Testing and Analysis,* vol. 10, no. 11-12, pp. 1734-1743, 2018, doi: https://doi.org/10.1002/dta.2413.

128. (2014). *1.0, WADA Technical Document – TD2014MRPL*. [Online] Available: https://www.wada-ama.org/sites/default/files/resources/files/WADA-TD2014MRPL-v1-Minimum-Required-Performance-Levels-EN.pdf

129. *Laboratory Accreditation Requirements and Operating Standards*, Consortium, R. M. T., rmtcnet.com, 2020. [Online]. Available: https://rmtcnet.com/wp-content/uploads/LabAccr4_Lab-Code-of-Standards-v4.0.pdf

130. Mottershead, J. "The Mare's Estrous Cycle." equine-reproduction.com. LLC. (accessed September 2023, 2023).

131. Squires, E. L., "Hormonal Manipulation of the Mare: A Review," *Journal of Equine Veterinary Science,* vol. 28, no. 11, pp. 627-634, 2008, doi: https://doi.org/10.1016/j.jevs.2008.10.010.

132. Fedorka, C. E. and Troedsson, M. H. T., "The use of progestins in equine medicine: A review," *Equine Veterinary Education,* vol. 33, no. 9, pp. 494-504, 2021, doi: https://doi.org/10.1111/eve.13332.

133. Crabtree, J. R., "A review of oestrus suppression techniques in mares," *Equine Veterinary Education,* vol. 34, no. 3, pp. 141-151, 2022, doi: https://doi.org/10.1111/eve.13405.

134. Durant, T., "python_paa_profile_classification/walkthrough_python," ed. https://github.com/tjdurant, 2021.

135. Chen, T. and Guestrin, C., "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016. [Online]. Available: https://doi.org/10.1145/2939672.2939785.

136. Boehmke, B. and Greenwell, B., "Gradient Boosting," *Hands-On Machine Learning with R,*1st ed. Taylor & Francis Group: Chapman and Hall/CRC, 2019, pp. 221-245. [Online]. Available: https://doi-org.ezproxy.lib.uts.edu.au/10.1201/9780367816377-12

137. Mohammadi, A., Karimzadeh, S., Banimahd, S. A., Ozsarac, V., and Lourenço, P. B., "The potential of region-specific machine-learning-based ground motion models: Application to Turkey," *Soil Dynamics and Earthquake Engineering,* vol. 172, p. 108008, 2023, doi: https://doi.org/10.1016/j.soildyn.2023.108008.

138. Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E., "Array programming with NumPy," *Nature,* vol. 585, no. 7825, pp. 357-362, 2020, doi: https://doi.org/10.1038/s41586-020-2649-2.

139. *pandas-dev/pandas: Pandas*. (2024). Zenodo. [Online]. Available: https://doi.org/10.5281/zenodo.13819579

140. Hunter, J. D., "Matplotlib: A 2D Graphics Environment," *Computing in Science & Engineering,* vol. 9, no. 3, pp. 90-95, 2007, doi: https://doi.org/10.1109/MCSE.2007.55.

141. Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Duchesnay, E., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research,* vol. 12, no. 85, pp. 2825-2830, 2011. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html.

142. Ginther, O. J., Utt, M. D., and Beg, M. A., "Follicle deviation and diurnal variation in circulating hormone concentrations in mares," *Animal Reproduction Science,* vol. 100, no. 1, pp. 197-203, 2007, doi: https://doi.org/10.1016/j.anireprosci.2006.08.025.

143. Brownlee, J. "XGBoosting: Making You Awesome At XGBoost." https://xgboosting.com/ (accessed December 13th 2024, 2024).

144. Strott, C. A., "Steroid Sulfotransferases," *Endocrine Reviews,* vol. 17, no. 6, pp. 670-697, 1996, doi: https://doi.org/10.1210/edrv-17-6-670.

145. Gomes, R. L., Meredith, W., Snape, C. E., and Sephton, M. A., "Analysis of conjugated steroid androgens: Deconjugation, derivatisation and associated issues," *Journal of Pharmaceutical and Biomedical Analysis,* vol. 49, no. 5, pp. 1133-1140, 2009, doi: https://doi.org/10.1016/j.jpba.2009.01.027.

146. Hintikka, L., Kuuranne, T., Leinonen, A., Thevis, M., Schänzer, W., Halket, J., Cowan, D., Grosse, J., Hemmersbach, P., Nielen, M. W. F., and Kostiainen, R., "Liquid chromatographic–mass spectrometric analysis of glucuronide-conjugated anabolic steroid metabolites: method validation and interlaboratory comparison," *Journal of Mass Spectrometry,* vol. 43, no. 7, pp. 965-973, 2008, doi: https://doi.org/10.1002/jms.1434.

147. KanehisaLaboratories. KEGG PATHWAY: Steroid hormone biosynthesis - Reference pathway [Online] Available: https://www.genome.jp/kegg-bin/show_pathway?hsa00140

148. Fitzgerald, C. C. J., Hedman, R., Uduwela, D. R., Paszerbovics, B., Carroll, A. J., Neeman, T., Cawley, A., Brooker, L., and McLeod, M. D., "Profiling Urinary Sulfate Metabolites With Mass Spectrometry," (in eng), *Front Mol Biosci,* vol. 9, p. 829511, 2022, doi: https://doi.org/10.3389/fmolb.2022.829511.

149. Waller, C. C. and McLeod, M. D., "A simple method for the small scale synthesis and solid-phase extraction purification of steroid sulfates," *Steroids,* vol. 92, pp. 74-80, 2014, doi: https://doi.org/10.1016/j.steroids.2014.09.006.

150. Fitzgerald, C. C. J., Bowen, C., Elbourne, M., Cawley, A., and McLeod, M. D., "Energy-Resolved Fragmentation Aiding the Structure Elucidation of Steroid Biomarkers," *Journal of the American Society for Mass Spectrometry,* vol. 33, no. 7, pp. 1276-1281, 2022, doi: https://doi.org/10.1021/jasms.2c00092.

151. Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., Kanazawa, M., VanderGheynst, J., Fiehn, O., and Arita, M., "MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis," *Nature Methods,* vol. 12, no. 6, pp. 523-526, 2015, doi: https://doi.org/10.1038/nmeth.3393.

152. Hartigan, J. A. and Wong, M. A., "Algorithm AS 136: A K-Means Clustering Algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics),* vol. 28, no. 1, pp. 100-108, 1979, doi: https://doi.org/10.2307/2346830.

153. Picard, R. R. and Cook, R. D., "Cross-Validation of Regression Models," *Journal of the American Statistical Association,* vol. 79, no. 387, pp. 575-583, 1984, doi: https://doi.org/10.2307/2288403.

154. Pang, Z., Chong, J., Li, S., and Xia, J., "MetaboAnalystR 3.0: Toward an Optimized Workflow for Global Metabolomics," *Metabolites,* vol. 10, no. 5, p. 186, 2020, doi: https://doi.org/10.3390/metabo10050186.

155. Ali, J., Khan, R., Ahmad, N., and Maqsood, I., "Random Forests and Decision Trees," (in English), *International Journal of Computer Science Issues (IJCSI),* vol. 9, no. 5, pp. 272-278, 2012. [Online]. Available: http://ezproxy.lib.uts.edu.au/login?url=https://www.proquest.com/scholarly-journals/random-forests-decision-trees/docview/1270319058/se-2?accountid=17095.

156. Breiman, L., "Random Forests," (in English), *Machine Learning,* vol. 45, no. 1, pp. 5-32, 2001, doi: https://doi.org/10.1023/A:1010933404324.

157. Ho, T. K., "Nearest neighbors in random subspaces," Springer Berlin Heidelberg, 1998, pp. 640-648, doi: https://doi.org/10.1007/bfb0033288.

158. Ghosh, T., Zhang, W., Ghosh, D., and Kechris, K., "Predictive Modeling for Metabolomics Data," Springer US, 2020, pp. 313-336, doi: https://doi.org/10.1007/978-1-0716-0239-3_16.

159. Wilde, L. D., Renterghem, P. V., and Eenoo, P. V., "Long-term stability study and evaluation of intact steroid conjugate ratios after the administration of endogenous steroids," *Drug Testing and Analysis,* vol. 14, no. 5, 2022, doi: https://doi.org/10.1002/dta.3096.

160. Dehennin, L., Lafarge, P., Dailly, P., Bailloux, D., and Lafarge, J. P., "Combined profile of androgen glucuro- and sulfoconjugates in post-competition urine of sportsmen: a simple screening procedure using gas chromatography-mass spectrometry," *Journal of Chromatography B: Biomedical Sciences and Applications,* vol. 687, no. 1, pp. 85-91, 1996, doi: https://doi.org/10.1016/s0378-4347(96)00131-4.

161. Mareck, U., Geyer, H., Opfermann, G., Thevis, M., and Schänzer, W., "Factors influencing the steroid profile in doping control analysis," *Journal of Mass Spectrometry,* vol. 43, no. 7, pp. 877-891, 2008, doi: https://doi.org/10.1002/jms.1457.

162. Pozo, O. J., Van Eenoo, P., Van Thuyne, W., Deventer, K., and Delbeke, F. T., "Direct quantification of steroid glucuronides in human urine by liquid chromatography–electrospray tandem mass spectrometry," *Journal of Chromatography A,* vol. 1183, no. 1, pp. 108-118, 2008, doi: https://doi.org/10.1016/j.chroma.2008.01.045.

163. Forsdahl, G., Zanitzer, K., Erceg, D., and Gmeiner, G., "Quantification of endogenous steroid sulfates and glucuronides in human urine after intramuscular administration of testosterone esters," *Steroids,* vol. 157, p. 108614, 2020, doi: https://doi.org/10.1016/j.steroids.2020.108614.

164. Anizan, S., Di Nardo, D., Bichon, E., Monteau, F., Cesbron, N., Antignac, J.-P., and Le Bizec, B., "Targeted phase II metabolites profiling as new screening strategy to investigate natural steroid abuse in animal breeding," *Analytica Chimica Acta,* vol. 700, no. 1-2, pp. 105-113, 2011, doi: https://doi.org/10.1016/j.aca.2010.12.009.

165. Piper, T., Opfermann, G., Thevis, M., and Schänzer, W., "Determination of $^{13}C/^{12}C$ ratios of endogenous urinary steroids excreted as sulpho conjugates," *Rapid Communications in Mass Spectrometry,* vol. 24, no. 21, pp. 3171-3181, 2010, doi: https://doi.org/10.1002/rcm.4762.

166. Piper, T., Putz, M., Schänzer, W., Pop, V., McLeod, M. D., Uduwela, D. R., Stevenson, B. J., and Thevis, M., "Epiandrosterone sulfate prolongs the detectability of testosterone, 4 - androstenedione, and dihydrotestosterone misuse by means of carbon isotope ratio mass spectrometry," *Drug Testing and Analysis,* vol. 9, no. 11-12, pp. 1695-1703, 2017, doi: https://doi.org/10.1002/dta.2291.

167. Gómez, C., Pozo, O. J., Garrostas, L., Segura, J., and Ventura, R., "A new sulphate metabolite as a long-term marker of metandienone misuse," *Steroids,* vol. 78, no. 12-13, pp. 1245-1253, 2013, doi: https://doi.org/10.1016/j.steroids.2013.09.005.

168. Connolly, B. S. M. and Lang, A. E. M., "Pharmacological Treatment of Parkinson Disease : A Review," *JAMA: The Journal of the American Medical Association,* Review vol. 311, no. 16, pp. 1670 – 1683, 2014, doi: https://doi.org/10.1001/jama.2014.3654.

169. Keen, B., Cawley, A., Reedy, B., Noble, G., Loy, J., and Fu, S., "3-Methoxytyrosine as an indicator of dopaminergic manipulation in equine plasma," *Journal of Chromatography B,* vol. 1220, p. 123652, 2023, doi: https://doi.org/10.1016/j.jchromb.2023.123652.

170. Axelrod, J. and Tomchick, R., "Enzymatic O-Methylation of Epinephrine and Other Catechols," *Journal of Biological Chemistry,* vol. 233, no. 3, pp. 702-705, 1958, doi: https://doi.org/10.1016/S0021-9258(18)64731-3.

171. Axelrod, J., Senoh, S., and Witkop, B., "O-Methylation of catechol amines in vivo," *J Biol Chem,* vol. 233, no. 3, pp. 697-701, 1958, doi: https://doi.org/10.1016/S0021-9258(18)64730-1.

172. Oeltmann, T., Carson, R., Shannon, J. R., Ketch, T., and Robertson, D., "Assessment of O-methylated catecholamine levels in plasma and urine for diagnosis of autonomic disorders," *Autonomic Neuroscience,* vol. 116, no. 1-2, pp. 1-10, 2004, doi: https://doi.org/10.1016/j.autneu.2004.08.013.

173. Koupai-Abyazani. M. R., Esaw. B., and Lavoilette. B., "Identification of levodopa and its metabolites in equine biological fluids by liquid chromatography-atmospheric pressure ionisation mass spectrometry," presented at the 10th Proc. Int. Conf. Racing Anal. Vets., Stockholm, Sweden, 1994.

174. ARTG. "Australian Product Information: Stalevo (Levodopa/Carbidopa Monohydrate/Entacapone) Film-Coated Tablets." Australian Register of Therapeutic Goods. https://www.ebs.tga.gov.au/ebs/picmi/picmirepository.nsf/pdf?OpenAgent=&id=CP-2020-PI-02710-1&d=20250131172310101 (accessed).

175. Ivanisevic, J. and Want, E. J., "From Samples to Insights into Metabolism: Uncovering Biologically Relevant Information in LC-HRMS Metabolomics Data," *Metabolites,* vol. 9, no. 12, p. 308, 2019, doi: https://doi.org/10.3390/metabo9120308.

176. Minakata, S., Manabe, S., Inai, Y., Ikezaki, M., Nishitsuji, K., Ito, Y., and Ihara, Y., "Protein C-Mannosylation and C-Mannosyl Tryptophan in Chemical Biology and Medicine," *Molecules,* vol. 26, no. 17, p. 5258, 2021, doi: https://doi.org/10.3390/molecules26175258.

Appendices

Appendices

Appendix 1: Sampling regimen for the oral administration of altrenogest (ALT).

Appendix 2: Sampling regimen for the intramuscular (IM) administration of altrenogest (ALT).

Appendix 3: Concentration of trendione in 5 orally administered ALT horses over a 21-day collection period.

Appendix 4: Concentration of epitrenbolone in 5 orally administered ALT horses over a 21-day collection period.

Appendix 5: Concentration of trenbolone in 5 orally administered ALT horses over a 21-day collection period.

Appendix 6: Mean urinary excretion profile concentrations (ng/mL) of trendione over 21 days in equine urine with oral altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

Appendix 7: Mean urinary excretion profile concentrations (ng/mL) of epitrenbolone over 21 days in equine urine with oral altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

Appendix 8: Concentration of trendione in 5 intramuscularly injected administered ALT horses over a 21-day collection period.

Appendix 9: Concentration of epitrenbolone in 5 intramuscularly injected administered ALT horses over a 21-day collection period.

Appendix 10: Concentration of trenbolone in 5 intramuscularly injected administered ALT horses over a 21-day collection period.

Appendix 11: Mean urinary excretion profile concentrations (ng/mL) of trendione over 21 days in equine urine with intramuscularly injected (IM) altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual

horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

Appendix 12: Mean urinary excretion profile concentrations (ng/mL) of epitrenbolone over 21 days in equine urine with intramuscularly injected (IM) altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

Appendix 13: Peak area ratio of progesterone in 4 orally administered ALT horses over a 21-day collection period.

Appendix 14: Mean fold change of progesterone response in 4 orally administered ALT horses over a 21-day collection period

Appendix 15: Peak area ratio of progesterone in 5 intramuscularly injected administered ALT horses over a 21-day collection period.

Appendix 16: Mean fold change of progesterone response in 5 intramuscularly injected administered ALT horses over a 21-day collection period.

Appendix 17: XGB Classification model script used with progesterone ratio data to differentiate oral and IM ALT administration.

Appendix 18: Progesterone ratio profile classification model dataset (log_{10} transformed datapoints).

Appendix 19: Progesterone ratio profile classification model confusion matrix output from XGBoost model, print out of data presented in-text (Chapter 4, Table 4-9).

Appendix 20: ALT metabolomic analysis reference standards included in the reference mix injected and analysed with each batch.

Appendix 21: ALT sulfate biomarkers relative retention times to the internal standard (ISTD - testosterone sulfate-d3) in the original batch analysis and the later compound verification analysis.

Appendix 22: ALT Sulfates model dataset

Appendix 23: Dosages of Stalevo product currently available with concentrations of levodopa (L-DOPA), carbidopa, and entacapone.

Appendix 24: Method validation results for the quantification of 3-methoxytyramine (3-MT) and tyramine (Tyr) in a water matrix. Validations parameters include sensitivity, linearity (low and high calibration ranges), accuracy, precision, recovery and matrix effects.

Appendix 25: Timewise plotted concentration values for 3-methoxytyramine for individual control (n = 4) and treated (n = 8) horses. The black solid line represents the current 3-methoxytyramine threshold (4 µg/mL) and the black dashed line indicates the proposed 3-methoxytyramine intelligence limit (0.776 µg/mL).

Appendix 26: Timewise plotted concentration values for Tyramine for individual control (n = 4) and treated (n = 8) horses.

Appendix 27: Timewise plotted 3-MT/Tyramine ratio values for individual control (n = 4) and treated (n = 8) horses. The black dotted line indicates the proposed 3-MT/Tyramine ratio limit (5.3)

Appendix 28: Reference standards included in the mix injected alongside the stalevo administration study samples.

Appendix 29: Reference standards included in the mix injected alongside the stalevo administration study samples, specific to compounds found in the stalevo medication.

Appendix 30: Dopamine Biomarker Model Dataset Features are labelled with the deprotonated molecule in m/z and the retention time in minutes. ID denotes the sample name, and the Label indicates the group, pre-administration samples as 0 and post-administration samples as 1.

Appendix 31: Additional dopamine biomarker model tests (RF, random forest model) using MetaboAnalyst and various dataset sizes.

ALTRENOGEST ADMINISTRATION STUDY

ADDITIONAL ADMINISTRATION STUDY INFORMATION

Each administration horse had their heart rate, respiratory rate, rectal temperature and BW manually measured prior to the study commencing to check normal health. Each horse was assigned a stable of yard for the duration of the sampling and fed meals of their regular (unchanged) hay ration of approximately 2 % BW dry matter, and access to clean drinking water as all times. All horses were allowed free exercise in an adjacent paddock to their assigned stable/yard for a minimum of 1 hour each day.

Urine sampling commenced at 7 am on day zero (pre-administration, time zero), and dosing of oral and IM ALT occurred at 8 am. Oral dosing occurred each day at 8 am after first administration, with each 24-hour urine sample being collected at 8 am (Appendix 1 and Appendix 2)

Appendix 1: Sampling regimen for the oral administration of altrenogest (ALT).

Day	Time (approx.)	Oral Admin Urine collection time points
Day 0	from 7am	Pre Administration (T0)
Dose horses with ALT orally	8 am	-
	10 am	2 h
	12 pm	4 h
	2 pm	6 h
	4 pm	8 h
	8 pm	12 h
1	8 am	24 h

2	8 am	48 h
3	8 am	72 h
4	8 am	96 h
5	8 am	120 h
6	8 am	144 h
7	8 am	168 h
8	8 am	192 h
9	8 am	216 h
10	8 am	240 h
11	8 am	288 h
12	8 am	312 h
Day 13 End of dosing elimination profile	from 7am	336 h
Final oral ALT dose	8 am	-
	10 am	338 h
	12 pm	340 h
	2 pm	342 h
	4 pm	344 h
	8 pm	348 h

14	8 am	360 h
15	8 am	384 h
16	8 am	408 h
17	8 am	432 h
18	8 am	456 h
19	8 am	480 h
20	8 am	504 h

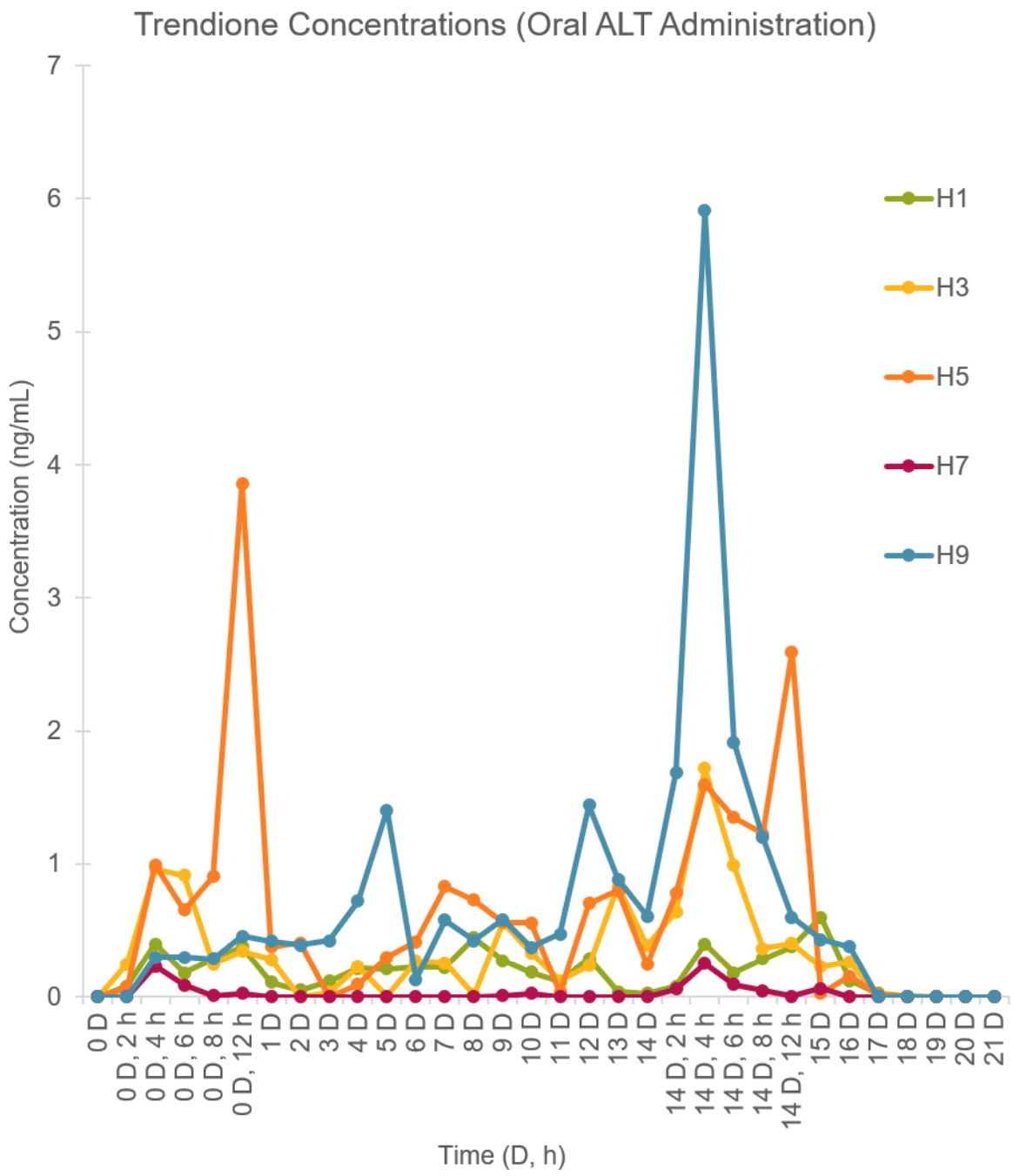
Appendix 2: Sampling regimen for the intramuscular (IM) administration of altrenogest (ALT).

Day	Time (approx.)	IM Admin Urine collection time points
Day 0	from 7am	Pre Administration (T0)
Dose horses with ALT IM	8 am	-
	10 am	2 h
	12 pm	4 h
	2 pm	6 h
	4 pm	8 h
	8 pm	12 h
1	8 am	24 h
2	8 am	48 h
3	8 am	72 h
4	8 am	96 h
5	8 am	120 h
6	8 am	144 h
Day 7 Second dosing elimination profile	from 7am	168 h
Second (final) ALT IM dose	8 am	-

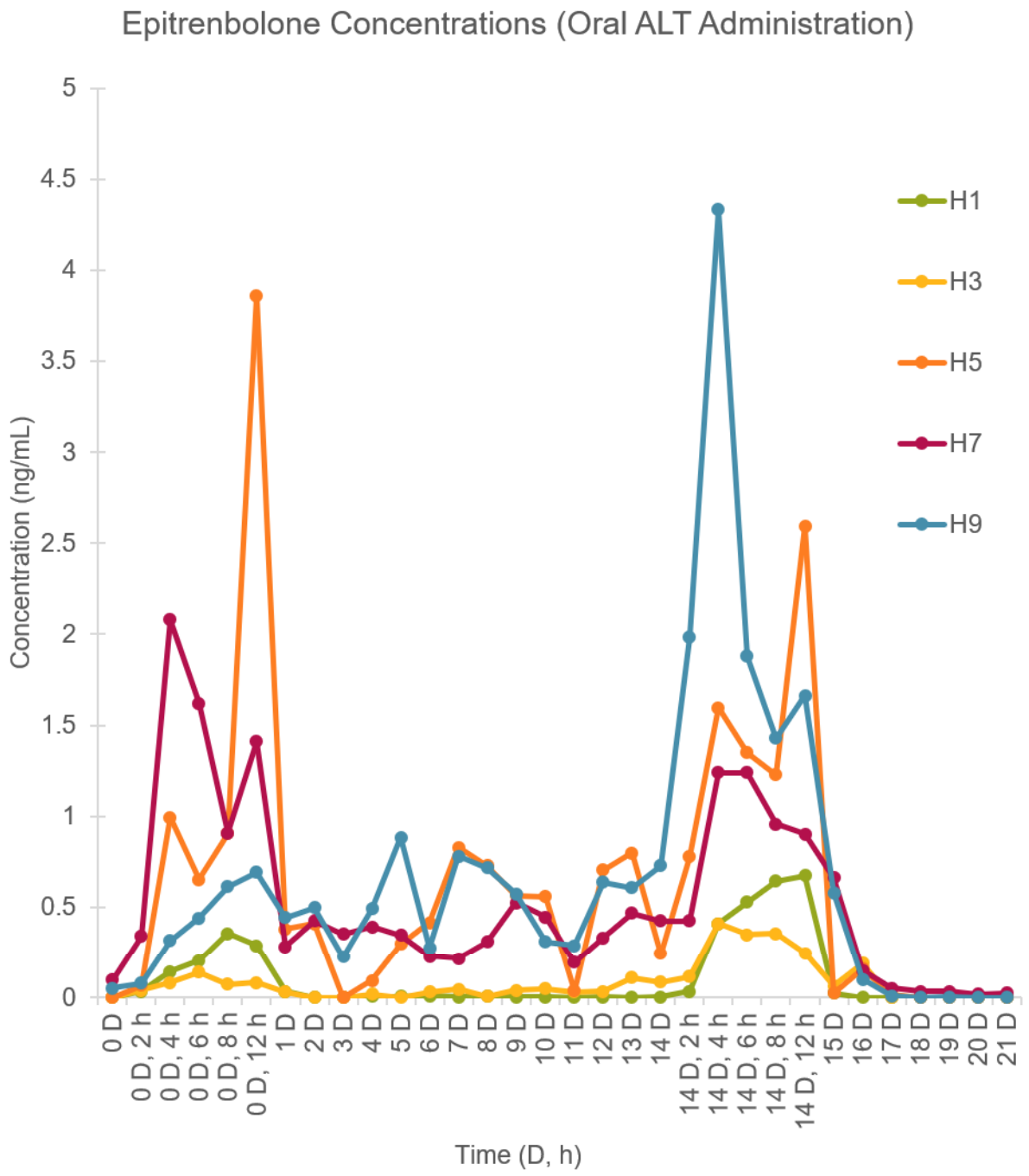
	10 am	170 h
	12 pm	172 h
	2 pm	174 h
	4 pm	176 h
	8 pm	180 h
8	8 am	192 h
9	8 am	216 h
10	8 am	240 h
11	8 am	288 h
12	8 am	312 h
13	8 am	336 h
14	8 am	360 h
15	8 am	384 h
16	8 am	408 h
17	8 am	432 h
18	8 am	456 h
19	8 am	480 h
20	8 am	504 h

QUANTITATIVE RESULTS

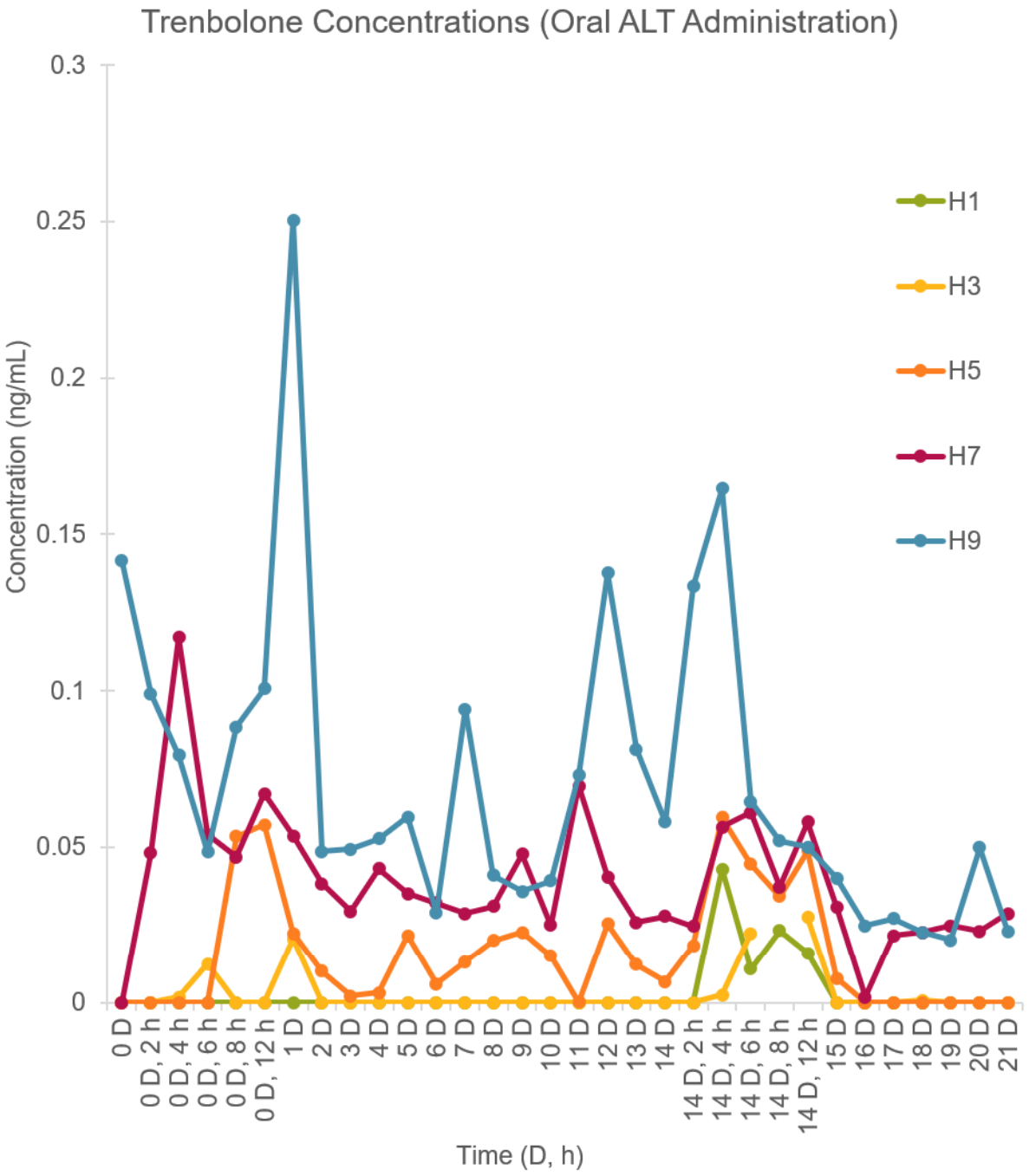
Appendix 3: Concentration of trendione in 5 orally administered altrenogest (ALT) horses over a 21-day collection period.



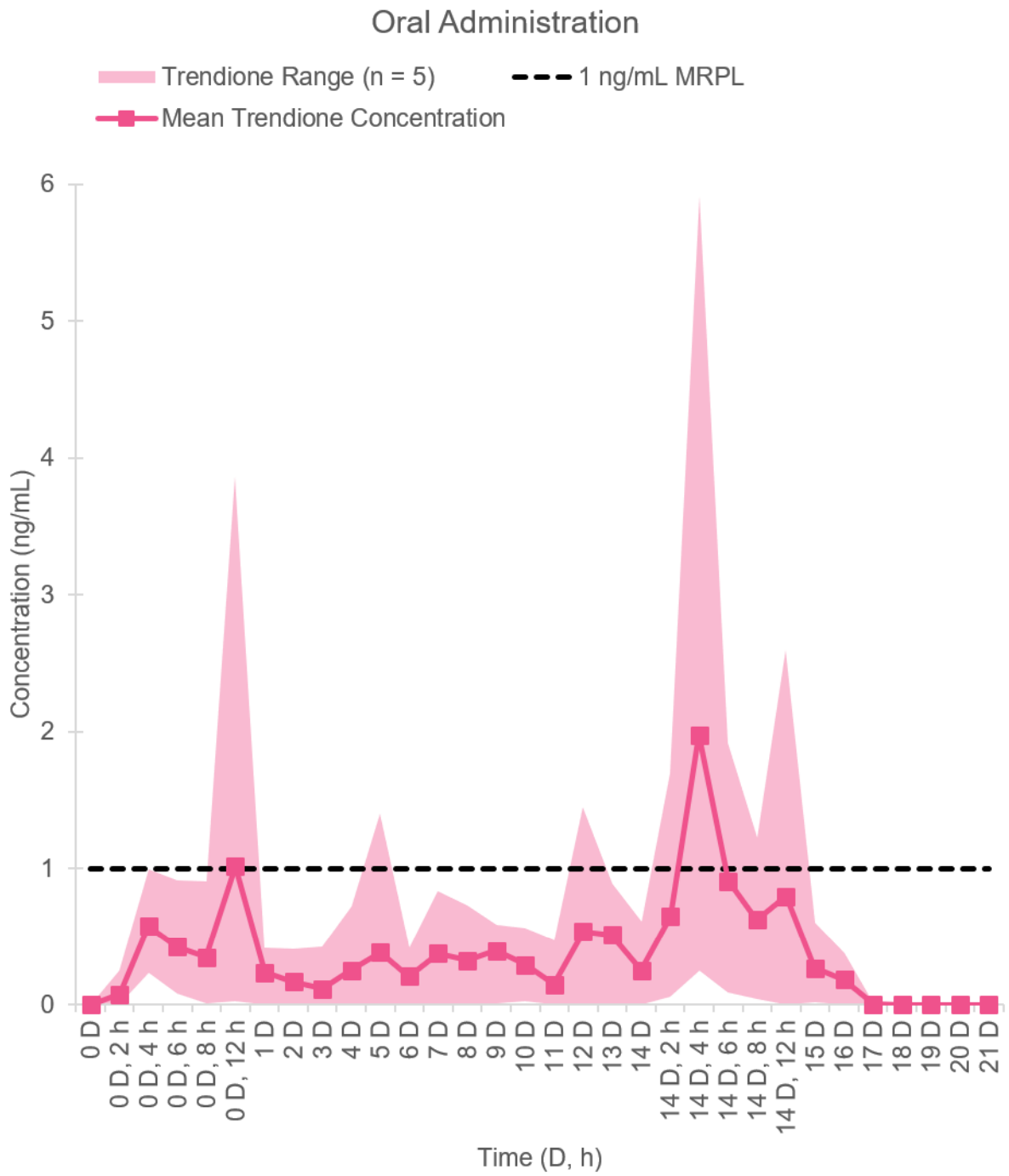
Appendix 4: Concentration of epitrenbolone in 5 orally administered altrenogest (ALT) horses over a 21-day collection period.



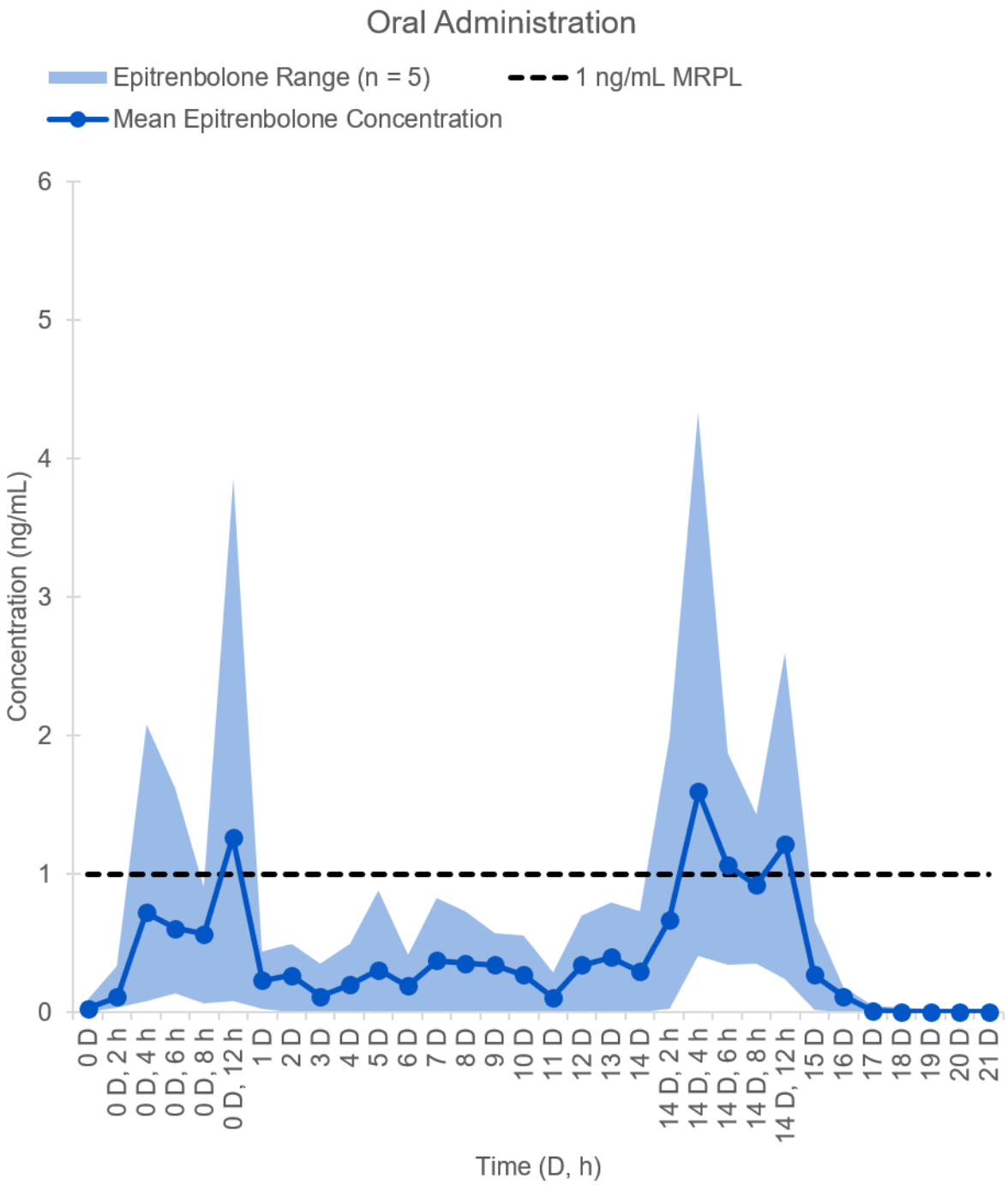
Appendix 5: Concentration of trenbolone in 5 orally administered altrenogest (ALT) horses over a 21-day collection period.



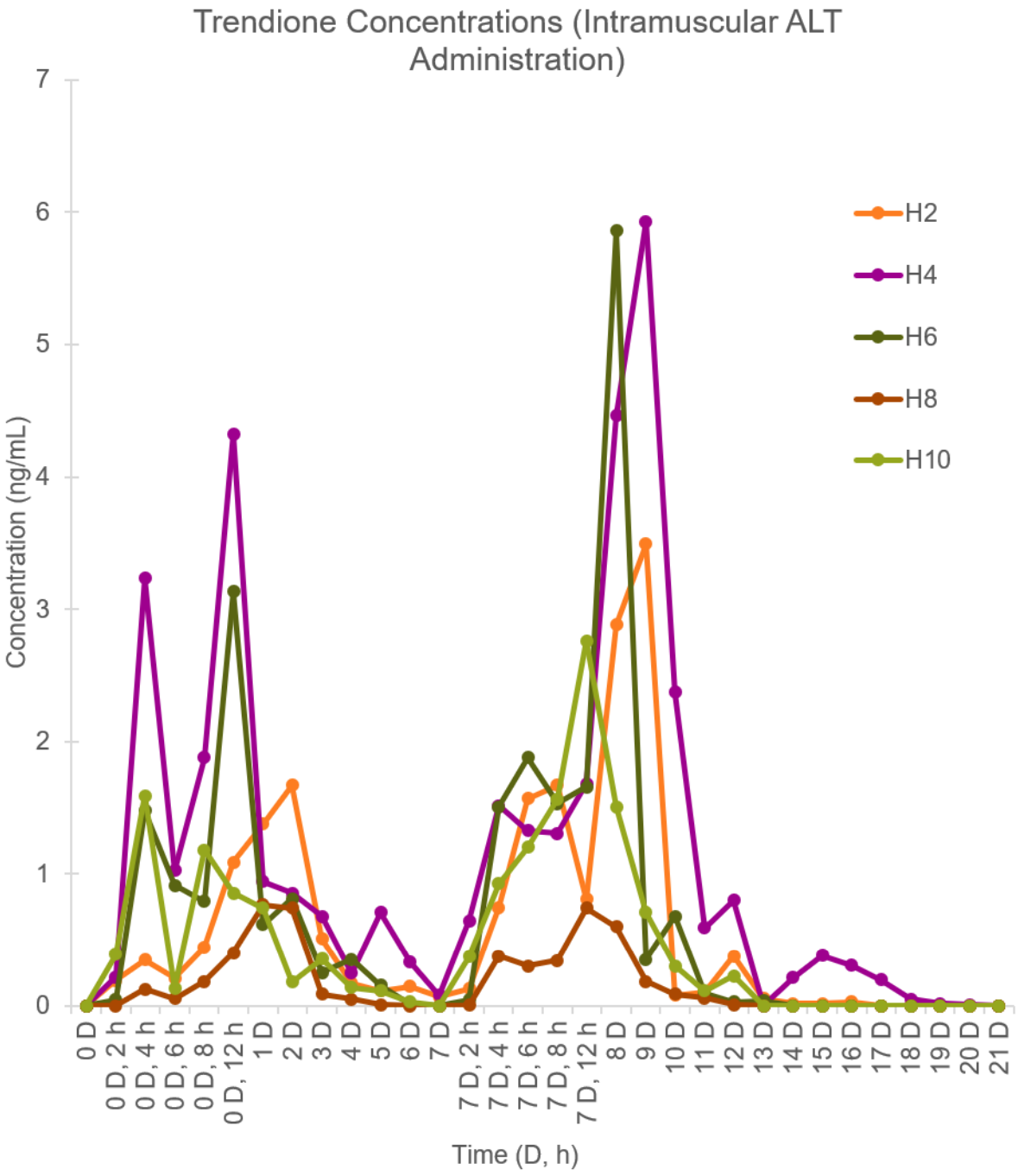
Appendix 6: Mean urinary excretion profile concentrations (ng/mL) of trendione over 21 days in equine urine with oral altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).



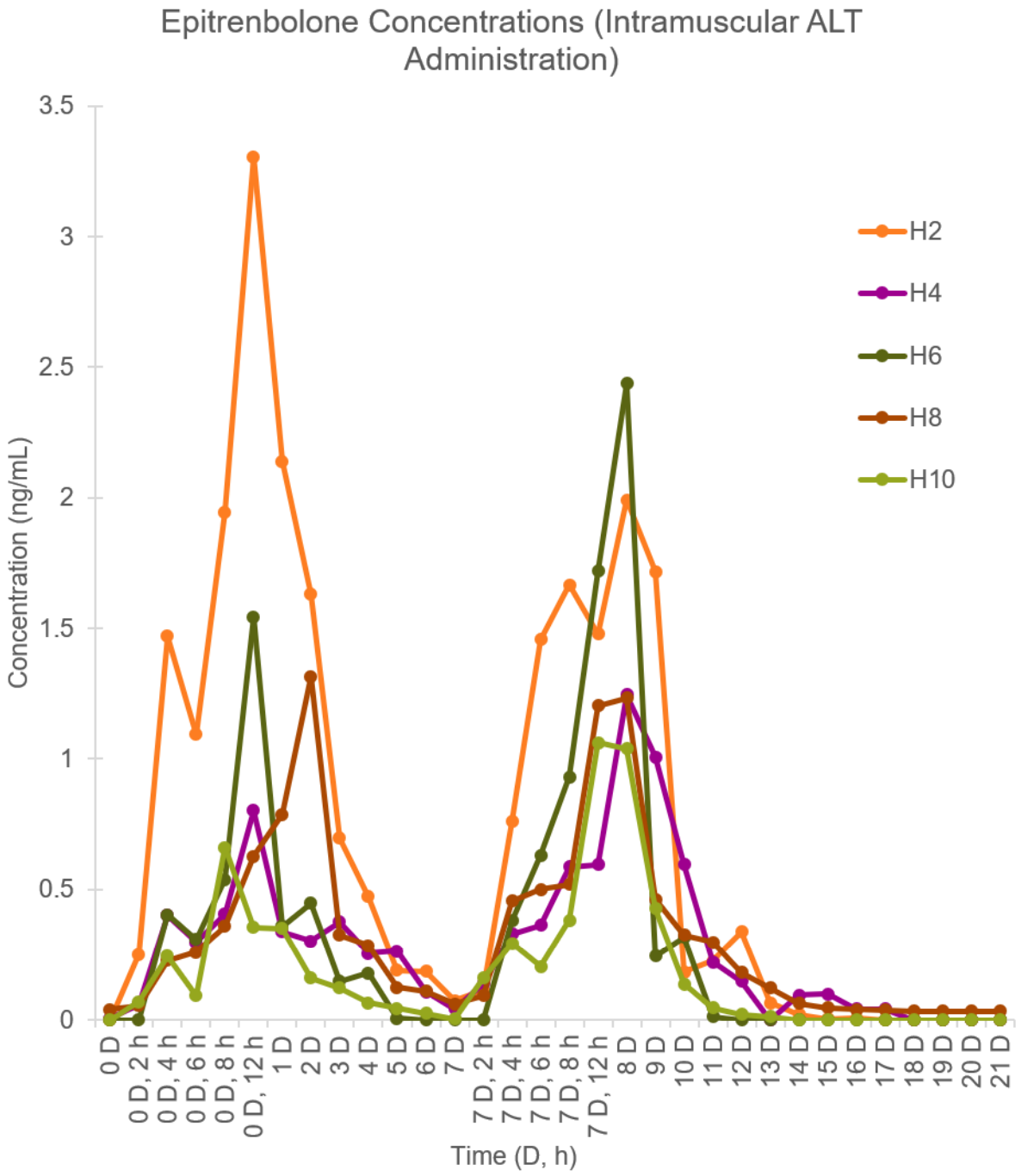
Appendix 7: Mean urinary excretion profile concentrations (ng/mL) of epitrenbolone over 21 days in equine urine with oral altrenogest administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).



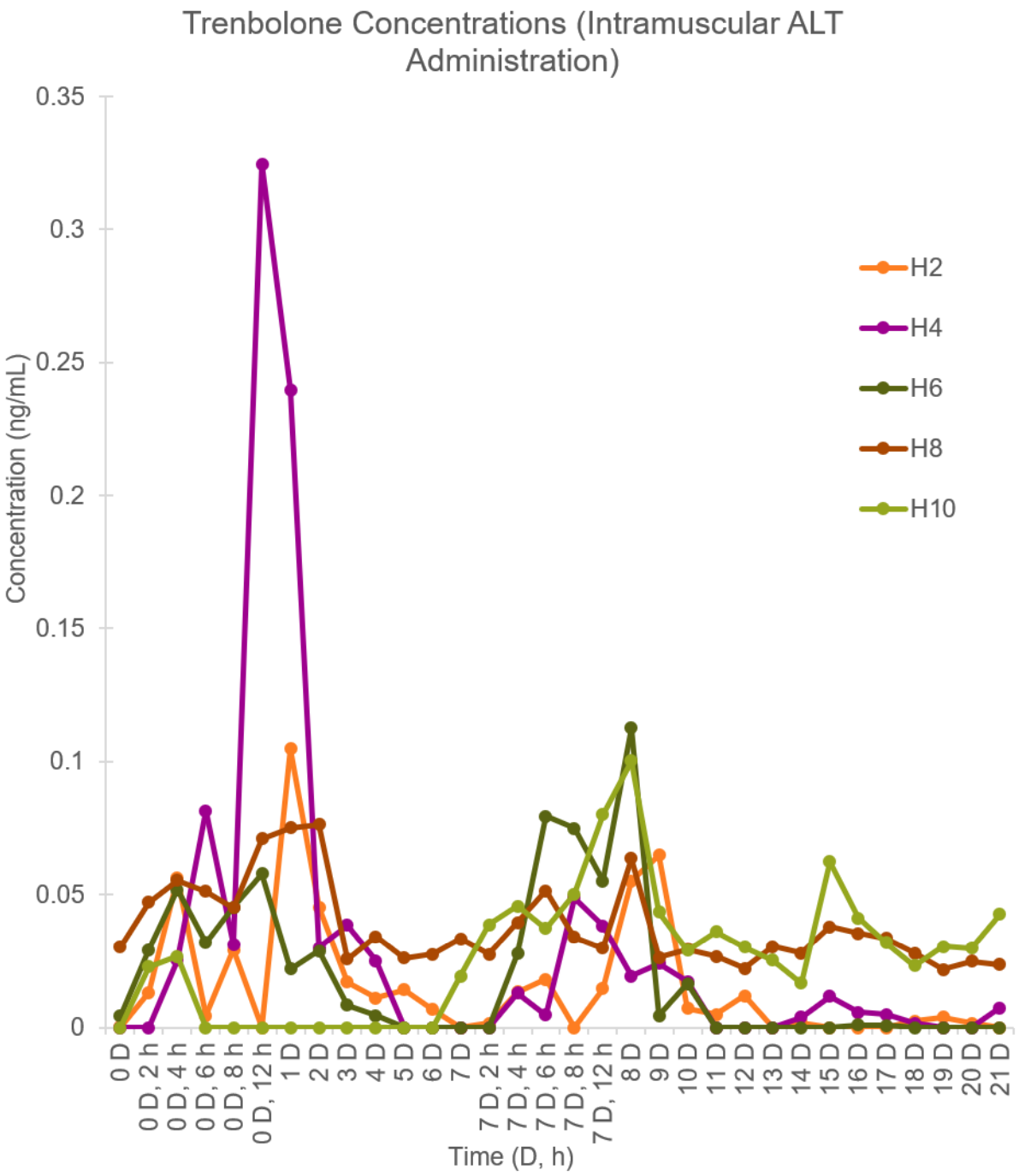
Appendix 8: Concentration of trendione in 5 intramuscularly injected (IM) administered altrenogest (ALT) horses over a 21-day collection period.



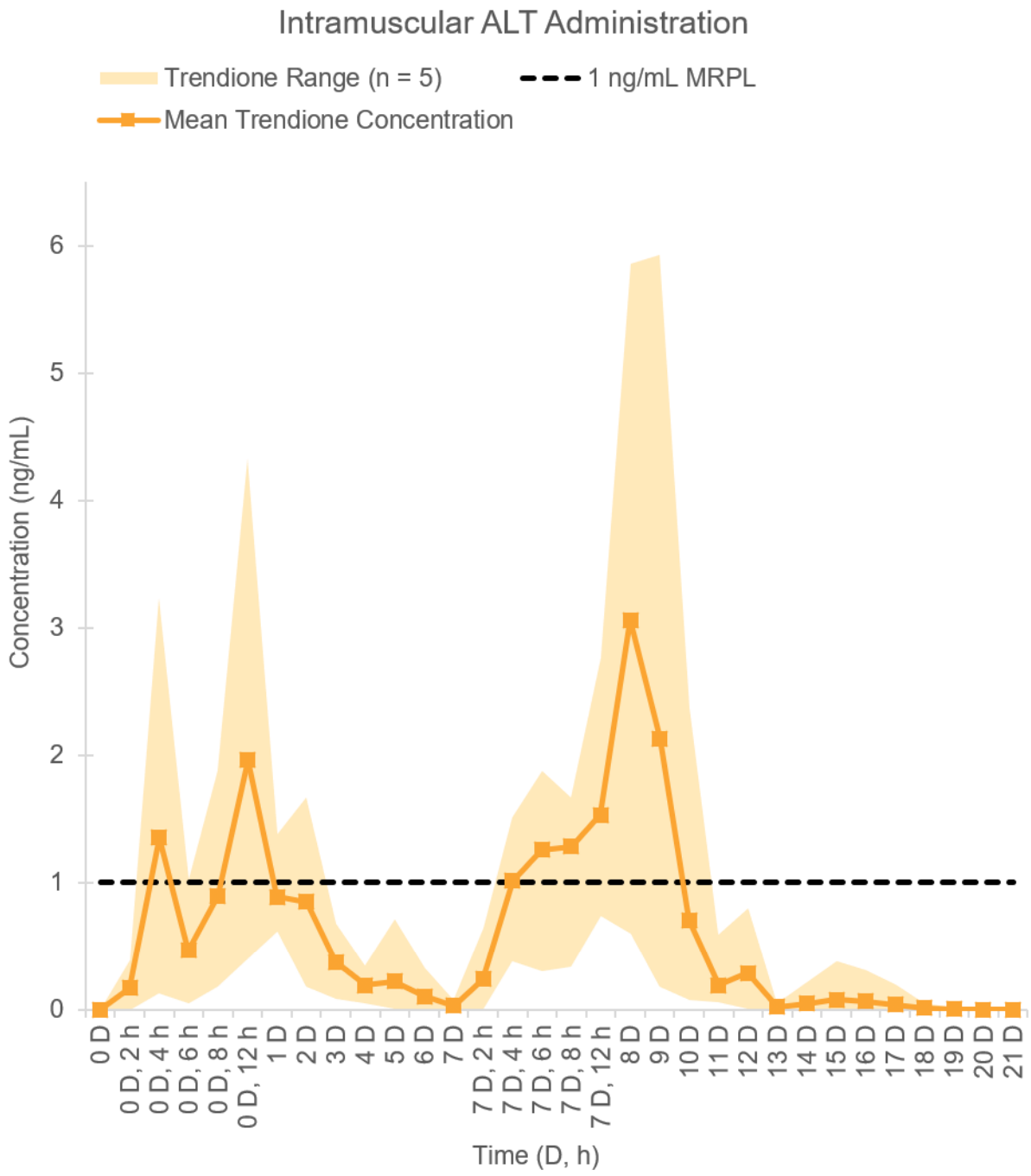
Appendix 9: Concentration of epitrenbolone in 5 intramuscularly injected (IM) administered altrenogest (ALT) horses over a 21-day collection period.



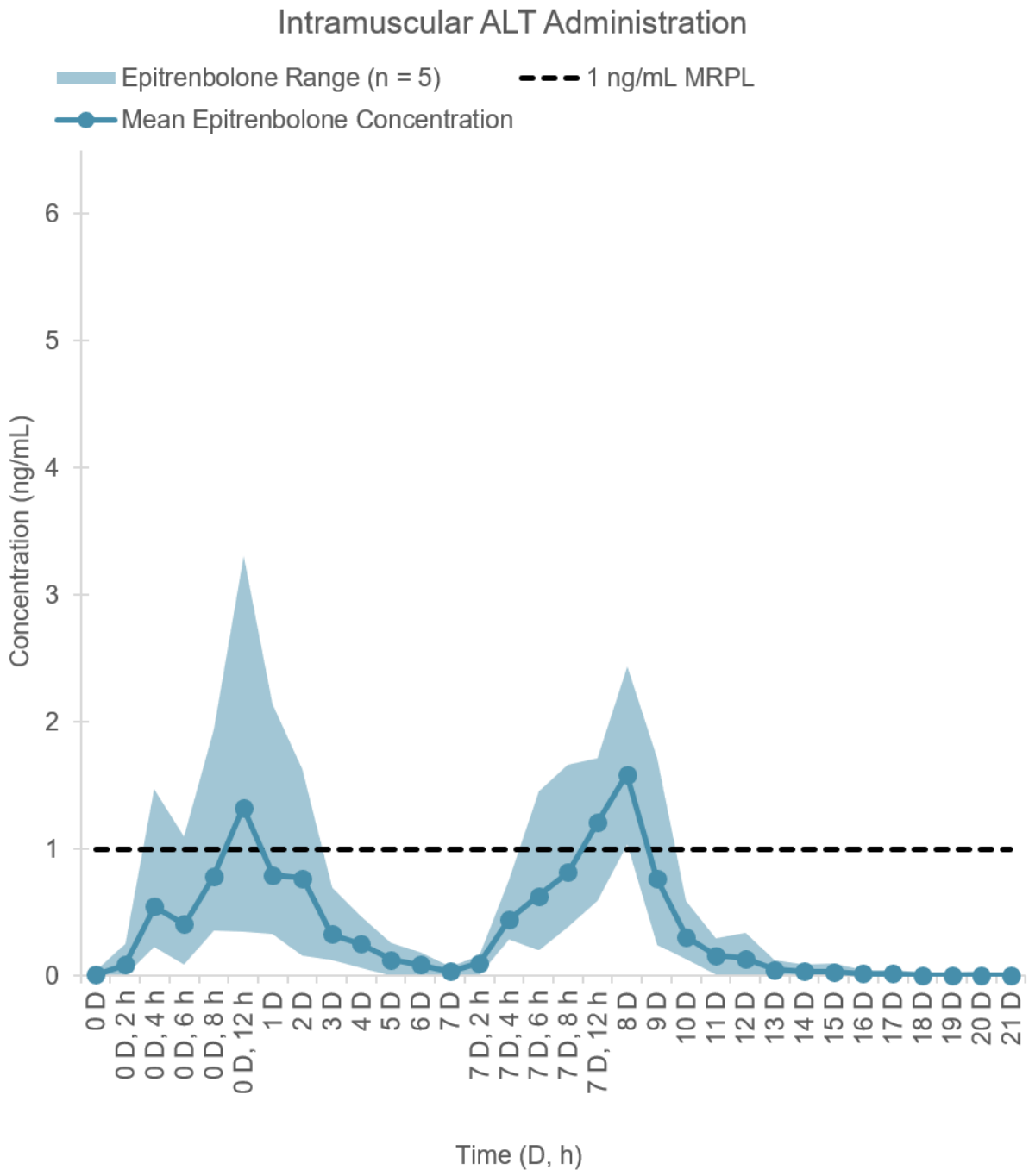
Appendix 10: Concentration of trenbolone in 5 intramuscularly injected (IM) administered altrenogest (ALT) horses over a 21-day collection period.



Appendix 11: Mean urinary excretion profile concentrations (ng/mL) of trendione over 21 days in equine urine with intramuscularly injected (IM) altrenogest (ALT) administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

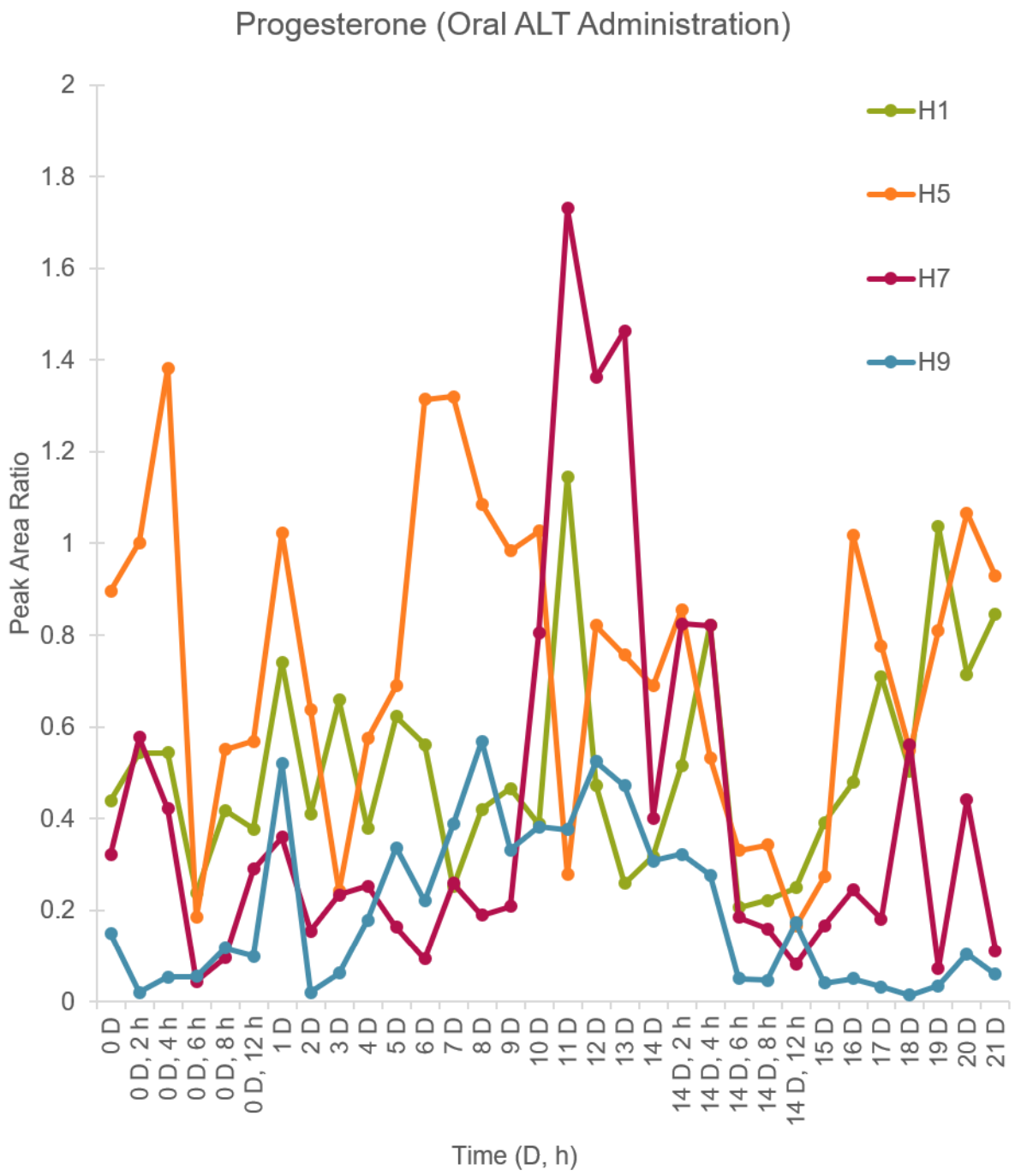


Appendix 12: Mean urinary excretion profile concentrations (ng/mL) of epitrenbolone over 21 days in equine urine with intramuscularly injected (IM) altrenogest (ALT) administration (daily for the initial 14 days), averaged from 5 different horses. Shaded area indicates the range of individual horse samples collected from each timepoint. The black dotted line indicates a 1 ng/mL minimum reporting performance level (MRPL).

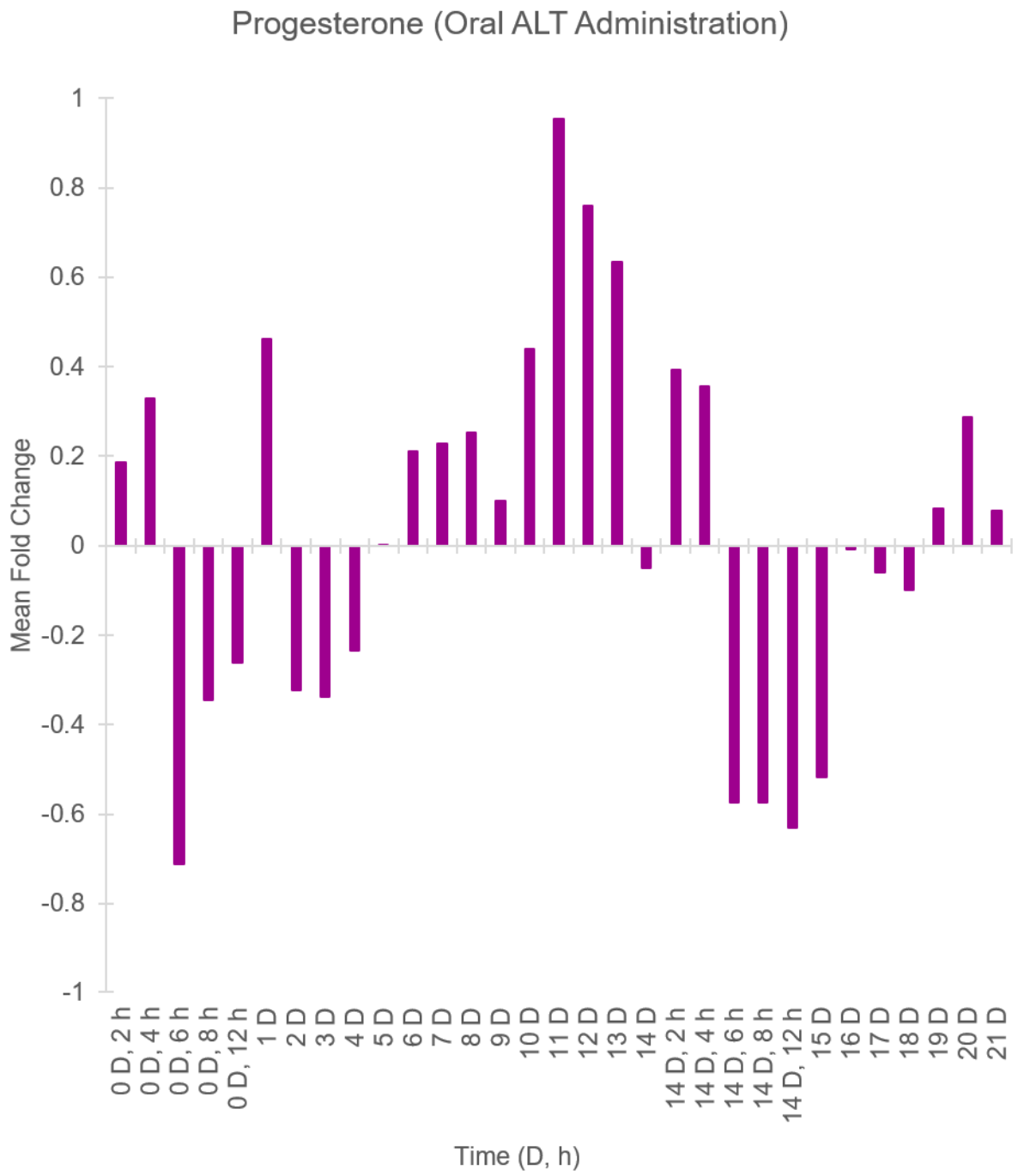


ADDITIONAL PROGESTERONE QUANTITATIVE AND MODEL RESULTS

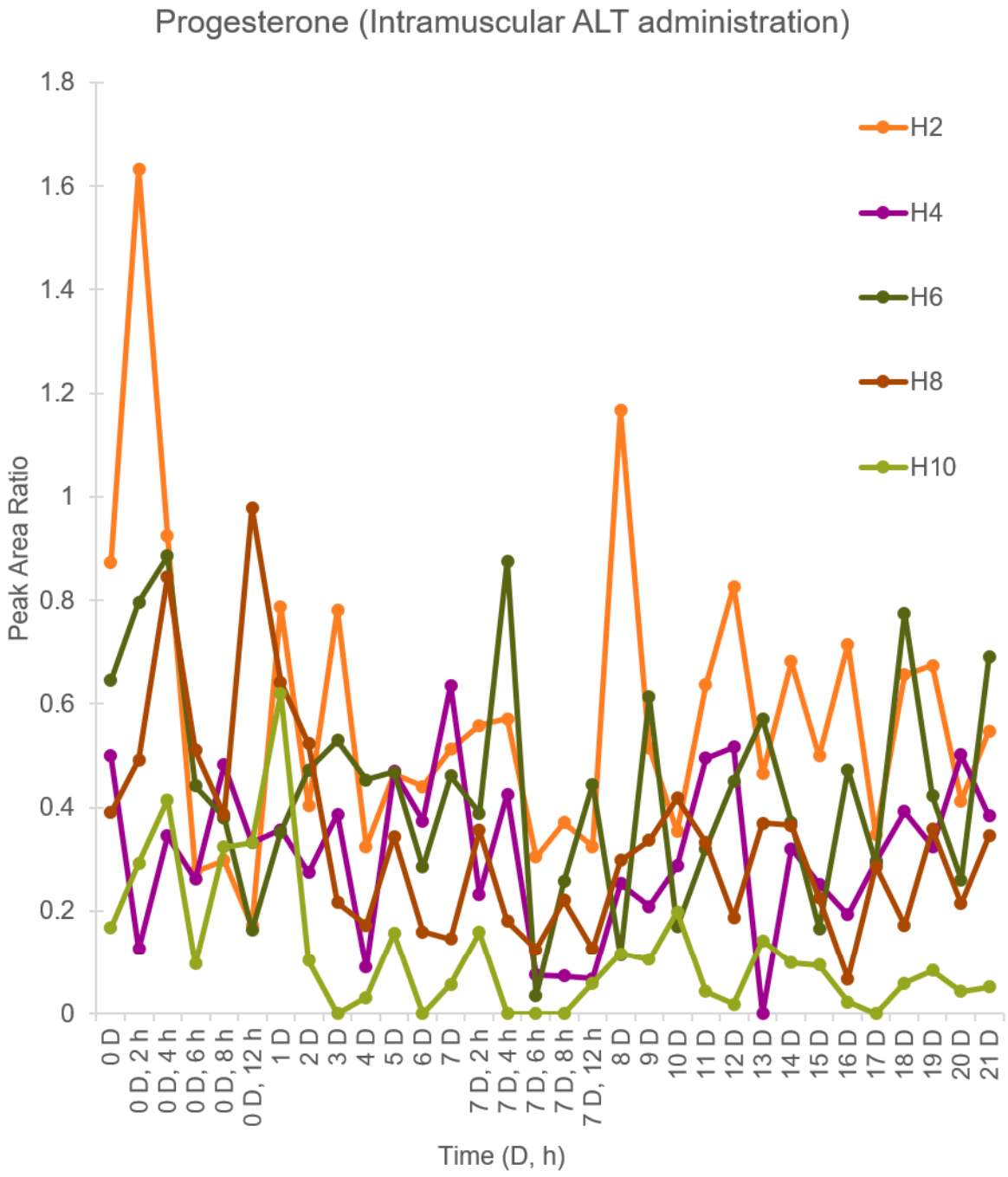
Appendix 13: Peak area ratio of progesterone in 4 orally administered altrenogest (ALT) horses over a 21-day collection period.



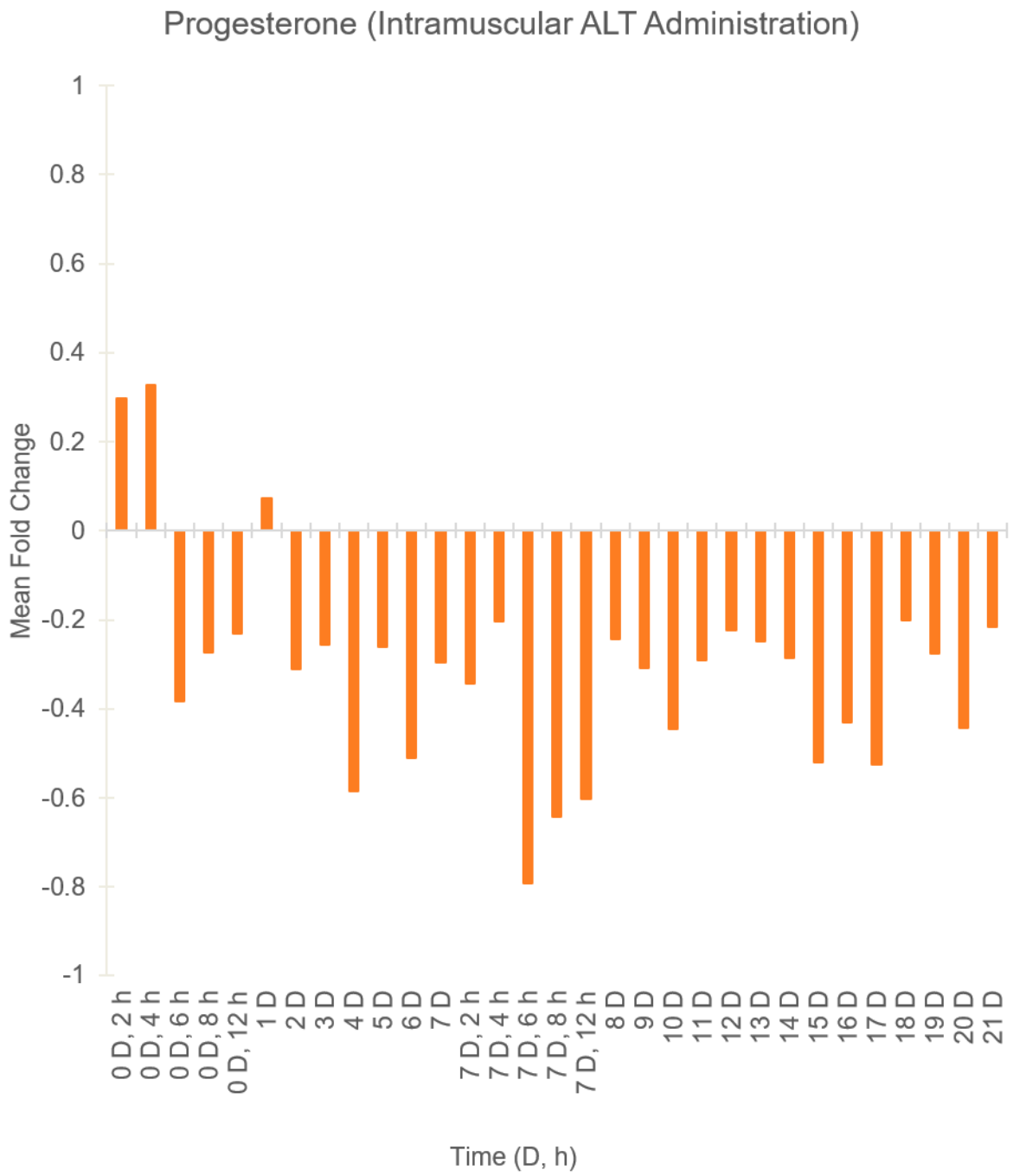
Appendix 14: Mean fold change of progesterone response in 4 orally administered altrenogest (ALT) horses over a 21-day collection period



Appendix 15: Peak area ratio of progesterone in 5 intramuscularly injected (IM) administered altrenogest (ALT) horses over a 21-day collection period.



Appendix 16: Mean fold change of progesterone response in 5 intramuscularly injected (IM) administered altrenogest (ALT) horses over a 21-day collection period.



ML SCRIPT FOR PROGESTERONE RATIO PROFILE CLASSIFICATION MODEL

Appendix 17: XGB Classification model script used with progesterone ratio data to differentiate oral and IM ALT administration.

Introduction

These two vignettes contain walk-throughs of machine learning development in both R and in Python. These tutorials aim to outline the basic steps in training and assessing ML models. The details presented are not meant to discuss the every detail of best practices in machine learning nor do they necessarily show how to develop the best performing model. Instead, the goals are to provide a clear example of how we go from data to predictions in the ML framework and to illustrate general machine learning principles.

A machine learning walkthrough in Python

Loading packages and data

In [3]:
```
%pip install xgboost

import os
import pandas as pd
import numpy as np
import xgboost as xgb

from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from sklearn.metrics import roc_curve
from sklearn.metrics import precision_recall_curve
from sklearn.metrics import average_precision_score
import sklearn.metrics
```

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: xgboost in c:\users\98109731\appdata\roaming\python\py
thon39\site-packages (2.0.3)
Requirement already satisfied: numpy in c:\software\anaconda3\lib\site-packages (from
xgboost) (1.20.3)
Requirement already satisfied: scipy in c:\software\anaconda3\lib\site-packages (from
xgboost) (1.7.1)
Note: you may need to restart the kernel to use updated packages.
```

Data Source

The data for this walkthrough is available in the supplemental material here:
https://academic.oup.com/clinchem/article/66/9/1210/5900235

Data processing

We proceed by loading our data into Python.

The variable `path_to_data` is the location of the csv file with the data. If you don't know how to set environment variables on your computer, just replace everything to the right of the equal (=) sign with the full path to the csv file like so:

path_to_data = "/your/path/to/the/data/clc317479-file001.csv"

Please note that if you are on Windows, you'll need to change the backslashes to forward slashes. And don't forget the double quotes!

The function `pd.read_csv()` will import the data into a data frame.

A data frame is a rectangular representation of data where each row represents an observation and each column represents a variable.

Each row represents a patient, and the column contains the data for each patient. The data frame contains an identifying `SID` for each patient, the patient's `SEX`, concentrations of different amino acids, alloisoleucine in `Allo`, homocysteine in `Hcys`, and argininosuccinic acid lyase deficiency in `ASA`. The `Class` column contains the "labels" normal and abnormal for each PAA profile.

In [4]:
```
pwd
```

Out[4]:
```
'C:\\Users\\98109731\\OneDrive - UTS\\Desktop\\Code\\ML Profile Classification Model
\\Python'
```

In [5]:
```
path_to_data = os.path.join('C:\\Users\\98109731\\OneDrive - UTS\\Desktop\\Code\\ML
df = pd.read_csv(path_to_data)
print(df.shape)
df.head()   # print a preview of the data frame
```

```
(190, 9)
```

Out[5]:

	ID	Breed	Trendione	Epitrenbolone	Trenbolone	Progesterone	Td/Pr	Etr/Pr	Admir
0	H1P1_24	0	9027	20793	0	24431	0.369490	0.851091	0.0
1	H1P1_48	0	4751	6364	0	20626	0.230340	0.308543	0.0
2	H1P1_72	0	12181	9996	133	37619	0.323799	0.265717	0.0
3	H1P1_96	0	18079	11431	0	15119	1.195780	0.756069	0.0
4	H1P1_120	0	17755	13572	0	22342	0.794692	0.607466	0.0

In order to prepare our data frame for machine learning purposes, we need to prepare our data into a form that the machine learning algorithms will accept.

1. We need to get rid any features that we will not input into the algorithm. In this case, we will need to remove the patient identifiers (the `SID` column). We do this by using the `drop()` function to remove the `SID` column

2. We need to convert any categorical variables into numerical codes. For example, the `SEX` column has values of `F` for female, `M` for male, and `U` for unidentified. We use the `loc()` function to find the rows with a specific value (e.g. `F`) and then replace it was the appropriate numerical code (e.g. `0`). We do this for all columns with categorical data.

3. We separate the features (e.g. all of the concentrations in our data) and the labels (e.g. normal or abnormal) and then save them as a Numpy matrix. The data is saved into a matrix `X`, and the labels are saved into a matrix `Y`.

In [6]:
```
df = df.drop(['ID'], axis = 1)

# save data as X and labels as Y
X = df.iloc[:188,:-1].to_numpy()
Y = df.iloc[:188,-1:].to_numpy()
```

In [7]:
```
print(X.shape)
#print(X)
```

```
(188, 7)
```

In [8]:
```
print(Y.shape)
#print(Y)
```

```
(188, 1)
```

Splitting our data into a training, test, and validation set

In [9]:
```
SEED = 7
TEST_SIZE = 0.30
VAL_SIZE = 0.10

# Split dataset into Train and Test datasets
x_train, x_test, y_train, y_test = train_test_split(X,
                                                    Y,
                                                    test_size = TEST_SIZE,
                                                    random_state = SEED)
x_train, x_val, y_train, y_val   = train_test_split(x_train,
                                                    y_train,
                                                    test_size = VAL_SIZE,
                                                    random_state = SEED)
```

ML Training Protocol

In [10]:
```
model_xgb = xgb.XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None,  eval_metric=["logloss"], early_st
              enable_categorical=False, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=0.3, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=3, max_leaves=None,
              min_child_weight=1, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None, nthread=4,
              num_parallel_tree=None)

model_xgb.fit(
  X=x_train,
  y=y_train.flatten(),
  eval_set = [(x_train, y_train), (x_val, y_val)],
)
```

```
[0]     validation_0-logloss:0.60089    validation_1-logloss:0.68446
[1]     validation_0-logloss:0.54092    validation_1-logloss:0.59456
[2]     validation_0-logloss:0.47985    validation_1-logloss:0.49274
```

```
[3]    validation_0-logloss:0.44947    validation_1-logloss:0.47914
[4]    validation_0-logloss:0.42018    validation_1-logloss:0.43191
[5]    validation_0-logloss:0.39853    validation_1-logloss:0.42177
[6]    validation_0-logloss:0.35646    validation_1-logloss:0.35486
[7]    validation_0-logloss:0.33624    validation_1-logloss:0.33106
[8]    validation_0-logloss:0.31944    validation_1-logloss:0.31121
[9]    validation_0-logloss:0.29942    validation_1-logloss:0.28492
[10]   validation_0-logloss:0.28283    validation_1-logloss:0.28232
[11]   validation_0-logloss:0.27251    validation_1-logloss:0.27373
[12]   validation_0-logloss:0.25468    validation_1-logloss:0.23402
[13]   validation_0-logloss:0.24598    validation_1-logloss:0.22201
[14]   validation_0-logloss:0.23809    validation_1-logloss:0.21904
[15]   validation_0-logloss:0.22341    validation_1-logloss:0.21290
[16]   validation_0-logloss:0.21514    validation_1-logloss:0.19930
[17]   validation_0-logloss:0.20801    validation_1-logloss:0.19658
[18]   validation_0-logloss:0.20148    validation_1-logloss:0.18967
[19]   validation_0-logloss:0.19758    validation_1-logloss:0.18539
[20]   validation_0-logloss:0.18984    validation_1-logloss:0.17711
[21]   validation_0-logloss:0.18462    validation_1-logloss:0.16901
[22]   validation_0-logloss:0.17960    validation_1-logloss:0.16001
[23]   validation_0-logloss:0.17608    validation_1-logloss:0.15206
[24]   validation_0-logloss:0.16796    validation_1-logloss:0.15548
[25]   validation_0-logloss:0.16167    validation_1-logloss:0.15997
[26]   validation_0-logloss:0.15835    validation_1-logloss:0.16213
[27]   validation_0-logloss:0.15165    validation_1-logloss:0.16562
[28]   validation_0-logloss:0.14824    validation_1-logloss:0.16350
[29]   validation_0-logloss:0.14304    validation_1-logloss:0.15749
[30]   validation_0-logloss:0.13996    validation_1-logloss:0.15172
[31]   validation_0-logloss:0.13781    validation_1-logloss:0.14517
[32]   validation_0-logloss:0.12861    validation_1-logloss:0.14403
[33]   validation_0-logloss:0.12617    validation_1-logloss:0.14144
[34]   validation_0-logloss:0.12376    validation_1-logloss:0.13975
[35]   validation_0-logloss:0.12150    validation_1-logloss:0.13608
[36]   validation_0-logloss:0.11968    validation_1-logloss:0.13769
[37]   validation_0-logloss:0.11730    validation_1-logloss:0.13817
[38]   validation_0-logloss:0.11440    validation_1-logloss:0.13345
[39]   validation_0-logloss:0.11259    validation_1-logloss:0.13138
[40]   validation_0-logloss:0.11078    validation_1-logloss:0.13483
[41]   validation_0-logloss:0.10911    validation_1-logloss:0.13301
[42]   validation_0-logloss:0.10749    validation_1-logloss:0.13378
[43]   validation_0-logloss:0.10608    validation_1-logloss:0.13189
[44]   validation_0-logloss:0.10425    validation_1-logloss:0.12397
[45]   validation_0-logloss:0.10311    validation_1-logloss:0.12220
[46]   validation_0-logloss:0.09947    validation_1-logloss:0.11790
[47]   validation_0-logloss:0.09808    validation_1-logloss:0.11619
[48]   validation_0-logloss:0.09543    validation_1-logloss:0.11678
[49]   validation_0-logloss:0.09425    validation_1-logloss:0.11116
[50]   validation_0-logloss:0.09212    validation_1-logloss:0.11689
[51]   validation_0-logloss:0.09002    validation_1-logloss:0.11352
[52]   validation_0-logloss:0.08808    validation_1-logloss:0.11405
[53]   validation_0-logloss:0.08673    validation_1-logloss:0.11729
[54]   validation_0-logloss:0.08553    validation_1-logloss:0.11845
[55]   validation_0-logloss:0.08387    validation_1-logloss:0.11791
[56]   validation_0-logloss:0.08267    validation_1-logloss:0.11370
[57]   validation_0-logloss:0.08130    validation_1-logloss:0.11823
[58]   validation_0-logloss:0.07866    validation_1-logloss:0.11425
```

Out[10]:
```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=10,
              enable_categorical=False, eval_metric=['logloss'],
              feature_types=None, gamma=None, grow_policy=None,
              importance_type=None, interaction_constraints=None,
              learning_rate=0.3, max_bin=None, max_cat_threshold=None,
              max_cat_to_onehot=None, max_delta_step=None, max_depth=3,
```

```
                    max_leaves=None, min_child_weight=1, missing=nan,
                    monotone_constraints=None, multi_strategy=None, n_estimators=None,
                    n_jobs=None, nthread=4, num_parallel_tree=None, ...)
```

In [11]:
```
from xgboost import XGBClassifier
from sklearn.model_selection import GridSearchCV
```

In [12]:
```
estimator = XGBClassifier(
    objective= 'binary:logistic',
    nthread=4,
    seed=42
)
```

In [13]:
```
parameters = {
    'max_depth': range (3, 10, 2),
    'min_child_weight':range(1,6,2)
}
```

In [14]:
```
grid_search = GridSearchCV(
    estimator = estimator,
    param_grid = parameters,
    scoring = 'roc_auc',
    n_jobs = 4,
    cv = 5,
    verbose=True
)
```

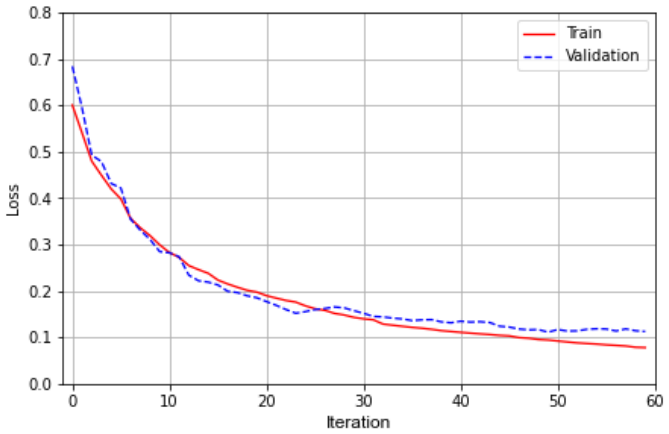
In [15]:
```
grid_search.fit(X, Y)
```

Out[15]:
```
Fitting 5 folds for each of 12 candidates, totalling 60 fits
GridSearchCV(cv=5,
             estimator=XGBClassifier(base_score=None, booster=None,
                                     callbacks=None, colsample_bylevel=None,
                                     colsample_bynode=None,
                                     colsample_bytree=None, device=None,
                                     early_stopping_rounds=None,
                                     enable_categorical=False, eval_metric=None,
                                     feature_types=None, gamma=None,
                                     grow_policy=None, importance_type=None,
                                     interaction_constraints=None,
                                     learning_rate=None,...
                                     max_cat_threshold=None,
                                     max_cat_to_onehot=None,
                                     max_delta_step=None, max_depth=None,
                                     max_leaves=None, min_child_weight=None,
                                     missing=nan, monotone_constraints=None,
                                     multi_strategy=None, n_estimators=None,
                                     n_jobs=None, nthread=4,
                                     num_parallel_tree=None, ...),
             n_jobs=4,
             param_grid={'max_depth': range(3, 10, 2),
                         'min_child_weight': range(1, 6, 2)},
             scoring='roc_auc', verbose=True)
```

In [16]:
```
grid_search.best_estimator_
```

Out[16]:
```
XGBClassifier(base_score=None, booster=None, callbacks=None,
              colsample_bylevel=None, colsample_bynode=None,
              colsample_bytree=None, device=None, early_stopping_rounds=None,
              enable_categorical=False, eval_metric=None, feature_types=None,
              gamma=None, grow_policy=None, importance_type=None,
              interaction_constraints=None, learning_rate=None, max_bin=None,
              max_cat_threshold=None, max_cat_to_onehot=None,
              max_delta_step=None, max_depth=5, max_leaves=None,
              min_child_weight=5, missing=nan, monotone_constraints=None,
              multi_strategy=None, n_estimators=None, n_jobs=None, nthread=4,
              num_parallel_tree=None, ...)
```

In [18]:
```
import matplotlib.pyplot as plt
from matplotlib.font_manager import FontProperties


results = model_xgb.evals_result()
epochs = len(results['validation_0']['logloss'])
x_axis = range(0, epochs)

fig, axs = plt.subplots(figsize=(7, 4.5), facecolor='w', edgecolor='k')

font = FontProperties()
font.set_name('Arial')

plt.plot(x_axis, results['validation_0']['logloss'],
         label='Train', color='red', linewidth=1.3)

plt.plot(x_axis, results['validation_1']['logloss'],
         label='Validation', color='blue',linestyle='--', linewidth=1.3)

#plt.axvline(x=16, color='black', linestyle='--', linewidth=1)

plt.legend()
plt.xlim(-1, 60)
plt.ylim(0, 0.8)
plt.ylabel('Loss')
plt.xlabel('Iteration', fontproperties=font, fontsize=12)

plt.grid(True)

plt.show()
```



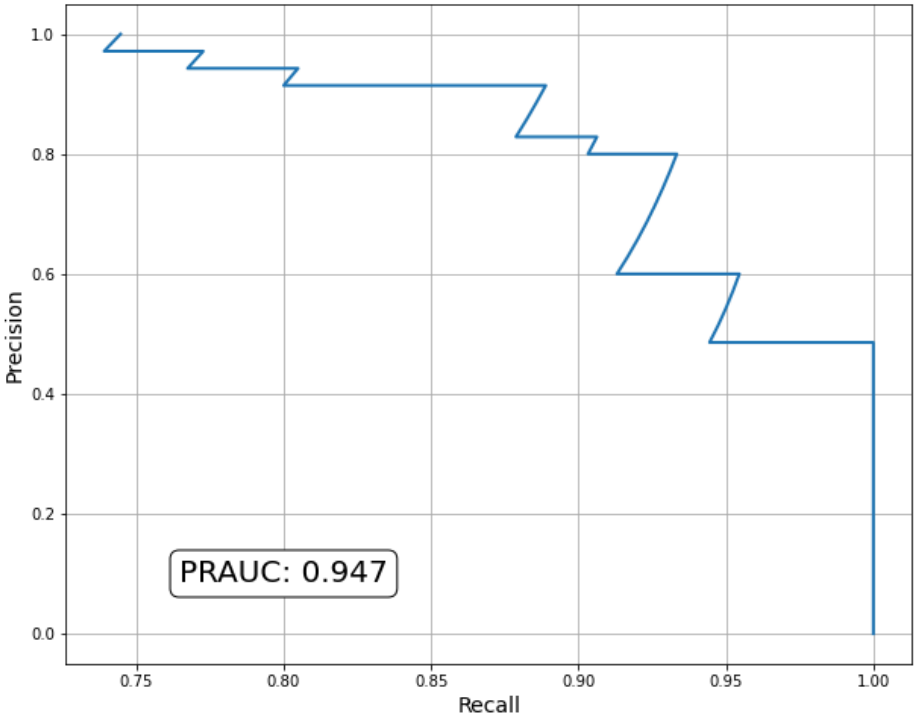
In [19]:

```
y_preds = model_xgb.predict(x_test).astype(int)

# get prediction probabilities for test data
y_probs = model_xgb.predict_proba(x_test).astype(float)

y_test = y_test.astype(int)

# evaluate predictions
accuracy = accuracy_score(y_test, y_preds)
print("Binomial Classification Accuracy: %.2f%%" % (accuracy * 100.0))
```

```
Binomial Classification Accuracy: 82.46%
```

In [37]:

```
precision, recall, _ = precision_recall_curve(y_test, y_probs[:,1])

auprc = round(sklearn.metrics.auc(recall, precision), 3)

fig, axs = plt.subplots(1,1, figsize=(10, 8), facecolor='w', edgecolor='k')

axs.annotate(
    "PRAUC: {0}".format(auprc),
    xy=(0.8,0.1),
    xycoords="data",
    va="center",
    ha="center",
    fontsize=20,
    bbox=dict(boxstyle="round",fc="w")
)

plt.plot(precision, recall, label="Test", linewidth=2, linestyle='-')
axs.set_xlabel('Recall', fontsize=14)
axs.set_ylabel('Precision', fontsize=14)

plt.grid(True)
plt.show()
```



In [28]:
```
# Make predictions on new data
UNK1 = [[0, 1981036, 151485, 70977, 14753, 134.2802142, 10.26888107]]
prediction_UNK1 = model_xgb.predict(UNK1)
print(prediction_UNK1)
```

```
[1]
```

In [33]:
```
print(model_xgb.predict_proba(UNK1))
```

```
[[0.01514578 0.9848542 ]]
```

In [30]:
```
UNK2 = [[0, 1301245, 144239, 49629, 13944, 93.31934882, 10.34416236]]
prediction_UNK2 = model_xgb.predict(UNK2)
print(prediction_UNK2)
```

```
[1]
```

In [32]:
```
print(model_xgb.predict_proba(UNK2))
```

```
[[0.01514578 0.9848542 ]]
```

PROGESTERONE RATIO PROFILE CLASSIFICATION MODEL DATASET

Appendix 18: Progesterone ratio profile classification model dataset (log_{10} transformed datapoints).

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H1P1_24	0	3.96	4.32	N/A	4.39	-0.43	-0.07	0
H1P1_48	0	3.68	3.80	N/A	4.31	-0.64	-0.51	0
H1P1_72	0	4.09	4.00	2.12	4.58	-0.49	-0.58	0
H1P1_96	0	4.26	4.06	N/A	4.18	0.08	-0.12	0
H1P1_120	0	4.25	4.13	N/A	4.35	-0.10	-0.22	0
H1P1_144	0	4.25	4.08	N/A	4.36	-0.12	-0.28	0
H1P1_168	0	4.29	4.05	N/A	4.08	0.21	-0.03	0
H1P1_192	0	4.62	4.15	2.91	4.21	0.41	-0.06	0
H1P1_216	0	4.40	4.09	N/A	4.28	0.12	-0.19	0
H1P1_240	0	4.23	4.17	2.76	4.49	-0.25	-0.32	0
H1P1_264	0	4.16	4.07	2.95	4.70	-0.54	-0.63	0
H1P1_288	0	4.57	4.19	3.04	4.41	0.16	-0.22	0
H1P1_312	0	3.85	4.21	3.08	4.42	-0.57	-0.22	0
H1P1_336	0	3.59	4.26	N/A	4.48	-0.88	-0.21	0
H1P1_360	0	4.55	4.15	N/A	4.21	0.34	-0.06	0
H1P1_384	0	4.00	3.65	3.05	4.28	-0.29	-0.63	0
H1P1_408	0	3.32	3.70	N/A	4.34	-1.02	-0.64	0
H1P1_432	0	2.79	3.80	3.25	4.45	-1.65	-0.65	0
H1P1_456	0	2.76	3.89	3.06	4.66	-1.90	-0.78	0
H1P1_480	0	2.23	3.97	2.75	4.48	-2.25	-0.51	0
H1P1_504	0	2.39	3.90	2.49	4.48	-2.08	-0.57	0
H2P1_24	0	5.10	5.84	4.10	4.61	0.49	1.23	1
H2P1_48	0	4.98	5.50	3.59	4.38	0.61	1.13	1
H2P1_72	0	4.63	5.33	3.38	4.66	-0.03	0.66	1

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H2P1_96	0	4.17	5.09	3.25	4.30	-0.13	0.79	1
H2P1_120	0	3.91	4.67	3.22	4.50	-0.58	0.17	1
H2P1_144	0	4.09	4.71	3.09	4.41	-0.32	0.30	1
H2P1_168	0	3.77	4.38	N/A	4.31	-0.54	0.07	1
H2P1_192	0	5.31	5.69	3.81	4.47	0.84	1.22	1
H2P1_216	0	5.22	5.47	3.68	4.34	0.87	1.12	1
H2P1_240	0	4.19	5.10	3.36	4.64	-0.45	0.45	1
H2P1_264	0	4.19	5.10	3.23	4.65	-0.46	0.45	1
H2P1_288	0	4.48	5.00	3.31	4.50	-0.03	0.49	1
H2P1_312	0	3.67	4.38	2.95	4.40	-0.73	-0.02	1
H2P1_336	0	3.49	4.20	2.83	4.53	-1.04	-0.32	1
H2P1_360	0	3.34	3.82	2.79	4.47	-1.13	-0.65	1
H2P1_384	0	3.38	3.65	1.95	4.30	-0.91	-0.64	1
H2P1_408	0	2.92	3.60	2.65	4.23	-1.31	-0.63	1
H2P1_432	0	3.12	3.41	3.09	4.48	-1.35	-1.07	1
H2P1_456	0	2.78	3.59	2.94	4.52	-1.74	-0.93	1
H2P1_480	0	2.61	3.69	2.96	4.45	-1.84	-0.77	1
H2P1_504	0	2.54	N/A	2.38	4.49	-1.95	N/A	1
H4P1_24	0	4.38	4.62	4.03	4.15	0.23	0.46	1
H4P1_48	0	4.74	5.03	3.78	4.11	0.63	0.92	1
H4P1_72	0	4.58	5.01	3.67	4.12	0.47	0.89	1
H4P1_96	0	4.24	4.88	3.57	3.81	0.44	1.07	1
H4P1_120	0	4.54	4.82	3.28	4.27	0.27	0.55	1
H4P1_144	0	4.25	4.49	3.00	4.21	0.04	0.28	1
H4P1_168	0	3.96	4.35	3.43	4.53	-0.58	-0.18	1
H4P1_192	0	5.24	5.36	3.68	3.96	1.28	1.40	1

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H4P1_216	0	5.53	5.45	3.81	3.95	1.57	1.50	1
H4P1_240	0	5.19	5.29	3.79	4.13	1.06	1.16	1
H4P1_264	0	4.61	4.87	3.55	4.38	0.23	0.50	1
H4P1_288	0	4.77	4.81	3.47	4.41	0.36	0.40	1
H4P1_336	0	4.18	4.55	3.65	4.13	0.05	0.41	1
H4P1_360	0	4.28	4.47	3.58	4.04	0.24	0.43	1
H4P1_384	0	4.20	4.22	3.47	4.04	0.15	0.17	1
H4P1_408	0	4.16	4.31	3.47	4.22	-0.06	0.09	1
H4P1_432	0	3.90	4.07	3.48	4.12	-0.22	-0.04	1
H4P1_456	0	3.62	3.56	3.05	4.23	-0.61	-0.67	1
H4P1_480	0	3.67	3.48	3.27	4.25	-0.59	-0.77	1
H4P1_504	0	3.49	3.44	3.48	4.17	-0.69	-0.73	1
H5P1_24	0	4.56	5.00	3.81	4.58	-0.02	0.43	0
H5P1_48	0	4.57	4.66	3.50	4.48	0.09	0.18	0
H5P1_72	0	3.64	4.41	3.43	4.30	-0.67	0.11	0
H5P1_96	0	4.28	4.91	3.55	4.53	-0.25	0.39	0
H5P1_120	0	4.40	4.92	3.71	4.41	-0.01	0.51	0
H5P1_144	0	4.68	4.81	3.62	4.73	-0.05	0.08	0
H5P1_168	0	4.86	4.79	3.65	4.78	0.08	0.00	0
H5P1_192	0	4.81	5.08	3.79	4.75	0.06	0.32	0
H5P1_216	0	4.68	4.88	3.76	4.77	-0.09	0.11	0
H5P1_240	0	4.83	5.04	3.69	4.81	0.02	0.23	0
H5P1_264	0	3.95	4.73	3.48	3.79	0.16	0.94	0
H5P1_288	0	4.91	5.08	3.93	4.61	0.30	0.47	0
H5P1_312	0	4.95	5.00	3.73	4.60	0.35	0.40	0
H5P1_336	0	4.46	4.72	3.53	4.39	0.07	0.33	0

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H5P1_360	0	3.88	4.98	3.55	4.25	-0.37	0.73	0
H5P1_384	0	4.22	4.36	3.03	4.68	-0.46	-0.32	0
H5P1_408	0	3.80	3.81	2.94	4.51	-0.71	-0.70	0
H5P1_432	0	2.99	3.74	2.32	4.49	-1.51	-0.75	0
H5P1_456	0	2.57	3.60	N/A	4.60	-2.03	-1.01	0
H5P1_480	0	2.70	3.10	3.14	4.60	-1.90	-1.50	0
H5P1_504	0	2.65	3.49	N/A	4.54	-1.89	-1.05	0
H6P1_24	0	5.03	5.69	4.17	4.16	0.87	1.53	1
H6P1_48	0	4.85	5.46	3.89	3.97	0.88	1.49	1
H6P1_72	0	4.44	5.14	3.66	3.80	0.64	1.34	1
H6P1_96	0	4.46	5.07	3.56	3.90	0.56	1.17	1
H6P1_120	0	4.27	4.65	3.36	4.10	0.17	0.55	1
H6P1_144	0	3.75	4.38	3.27	3.95	-0.21	0.42	1
H6P1_168	0	2.87	3.72	2.82	4.28	-1.41	-0.56	1
H6P1_192	0	5.47	5.95	4.17	3.44	2.03	2.50	1
H6P1_216	0	4.53	5.26	3.50	4.20	0.33	1.06	1
H6P1_240	0	4.77	5.32	3.73	3.73	1.04	1.59	1
H6P1_264	0	4.16	4.73	3.17	4.03	0.13	0.70	1
H6P1_288	0	3.93	4.40	3.29	3.99	-0.07	0.40	1
H6P1_312	0	3.71	4.12	3.20	3.96	-0.25	0.16	1
H6P1_336	0	3.02	3.67	3.28	4.01	-0.99	-0.34	1
H6P1_360	0	2.80	3.38	3.07	3.65	-0.85	-0.27	1
H6P1_384	0	2.76	1.67	3.20	4.01	-1.25	-2.34	1
H6P1_408	0	2.75	3.12	3.29	3.74	-0.99	-0.62	1
H6P1_432	0	2.34	2.78	3.07	3.94	-1.60	-1.16	1
H6P1_456	0	2.16	2.79	3.23	3.98	-1.83	-1.19	1

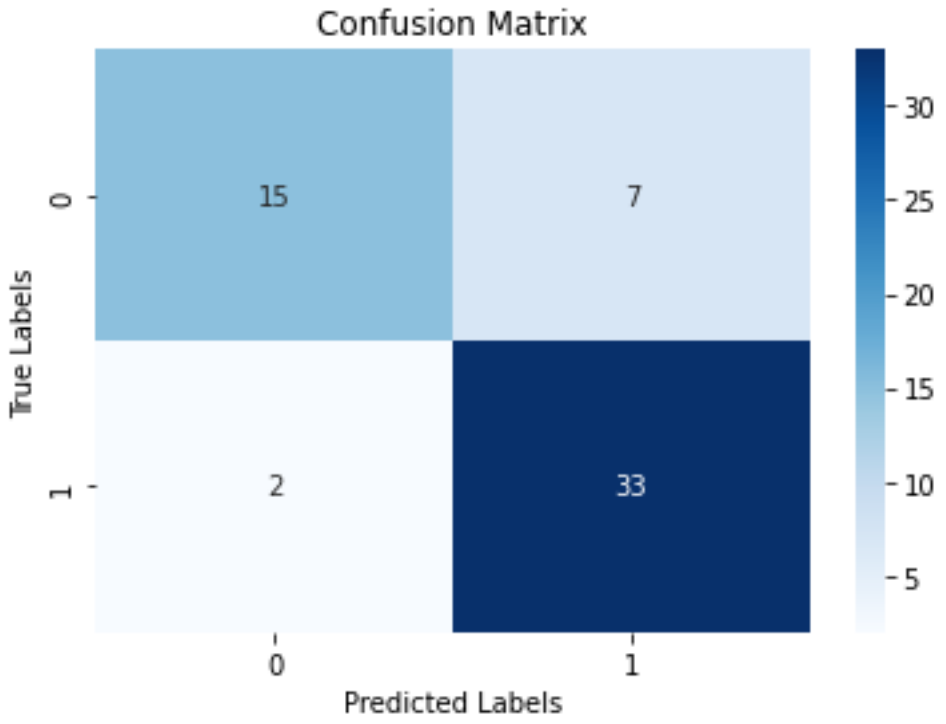
ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H6P1_480	0	2.67	3.36	3.41	3.92	-1.25	-0.56	1
H6P1_504	0	2.34	3.14	3.21	4.03	-1.68	-0.89	1
H7P1_24	1	2.93	4.79	3.33	4.10	-1.18	0.69	0
H7P1_48	1	2.82	4.77	3.29	3.79	-0.97	0.97	0
H7P1_72	1	3.20	4.79	3.16	3.53	-0.33	1.26	0
H7P1_96	1	2.96	4.79	3.35	3.62	-0.66	1.17	0
H7P1_120	1	3.21	4.88	3.19	3.80	-0.59	1.08	0
H7P1_144	1	3.11	4.61	3.07	3.31	-0.20	1.30	0
H7P1_168	1	2.66	4.65	3.01	3.79	-1.13	0.86	0
H7P1_192	1	3.31	4.80	3.10	3.62	-0.31	1.18	0
H7P1_216	1	3.35	4.92	3.17	3.75	-0.40	1.18	0
H7P1_240	1	3.50	4.93	3.01	4.25	-0.74	0.68	0
H7P1_264	1	3.23	4.61	3.78	4.53	-1.30	0.08	0
H7P1_288	1	3.40	4.87	3.42	4.51	-1.11	0.36	0
H7P1_312	1	3.26	4.96	N/A	4.59	-1.33	0.38	0
H7P1_336	1	3.03	4.93	2.94	4.05	-1.02	0.88	0
H7P1_360	1	3.61	5.06	3.09	3.73	-0.12	1.33	0
H7P1_384	1	3.12	4.31	2.06	3.67	-0.55	0.64	0
H7P1_408	1	2.48	3.68	2.05	3.62	-1.15	0.05	0
H7P1_432	1	2.18	3.38	2.42	3.92	-1.75	-0.55	0
H7P1_456	1	2.29	3.12	3.03	3.40	-1.10	-0.28	0
H7P1_480	1	1.94	3.07	2.74	3.93	-1.99	-0.85	0
H7P1_504	1	2.13	2.79	3.07	3.52	-1.38	-0.73	0
H8P1_24	1	4.46	5.13	3.57	3.83	0.63	1.30	1
H8P1_48	1	4.33	5.21	3.52	3.52	0.80	1.69	1
H8P1_72	1	3.70	4.69	3.14	3.68	0.01	1.01	1

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H8P1_96	1	3.59	4.70	3.12	3.45	0.13	1.25	1
H8P1_120	1	3.47	4.32	2.96	3.57	-0.10	0.75	1
H8P1_144	1	3.41	4.30	3.13	3.62	-0.21	0.68	1
H8P1_168	1	2.75	3.82	3.31	3.58	-0.83	0.24	1
H8P1_192	1	4.32	5.26	3.44	3.40	0.91	1.86	1
H8P1_216	1	3.98	4.95	3.07	3.41	0.57	1.54	1
H8P1_240	1	3.78	4.79	3.25	3.65	0.13	1.14	1
H8P1_264	1	3.68	4.73	3.18	3.68	0.00	1.04	1
H8P1_288	1	3.46	4.45	3.03	3.64	-0.18	0.81	1
H8P1_312	1	3.13	4.19	3.08	3.60	-0.48	0.59	1
H8P1_336	1	2.84	3.87	3.11	3.38	-0.55	0.49	1
H8P1_360	1	2.82	3.62	3.34	3.56	-0.74	0.06	1
H8P1_384	1	2.86	3.13	3.32	3.30	-0.44	-0.17	1
H8P1_408	1	2.44	3.27	3.05	3.47	-1.03	-0.20	1
H8P1_432	1	2.23	2.97	2.95	3.61	-1.38	-0.64	1
H8P1_456	1	2.20	2.80	2.92	3.61	-1.40	-0.80	1
H8P1_480	1	2.28	3.05	3.13	3.64	-1.36	-0.59	1
H8P1_504	1	2.42	2.91	3.09	3.59	-1.17	-0.68	1
H9P1_24	1	5.08	5.69	4.89	5.44	-0.36	0.25	0
H9P1_48	1	4.94	5.71	4.04	4.24	0.70	1.47	0
H9P1_72	1	5.30	5.79	3.95	4.60	0.70	1.19	0
H9P1_96	1	5.22	5.82	4.19	4.91	0.32	0.92	0
H9P1_120	1	5.56	6.06	4.08	5.11	0.45	0.95	0
H9P1_144	1	4.95	5.89	3.91	5.09	-0.14	0.80	0
H9P1_168	1	5.20	6.02	4.12	5.33	-0.12	0.70	0
H9P1_192	1	5.21	6.15	4.07	5.43	-0.21	0.72	0

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H9P1_216	1	5.20	5.87	3.82	5.22	-0.03	0.65	0
H9P1_240	1	5.34	5.97	4.13	5.29	0.05	0.68	0
H9P1_264	1	5.38	5.86	4.36	5.30	0.08	0.56	0
H9P1_288	1	5.73	6.11	4.56	5.35	0.38	0.76	0
H9P1_312	1	5.52	6.04	4.50	5.30	0.22	0.75	0
H9P1_336	1	5.41	6.16	4.31	5.06	0.35	1.09	0
H9P1_360	1	5.28	6.11	4.15	4.22	1.05	1.88	0
H9P1_384	1	4.91	5.22	2.89	4.44	0.47	0.78	0
H9P1_408	1	4.01	4.77	3.38	4.48	-0.47	0.29	0
H9P1_432	1	3.26	4.47	3.42	4.08	-0.82	0.39	0
H9P1_456	1	3.34	4.75	3.43	4.47	-1.13	0.29	0
H9P1_480	1	N/A	N/A	3.93	4.79	N/A	N/A	0
H9P1_504	1	3.46	N/A	3.41	4.61	-1.15	N/A	0
H10P1_24	0	5.53	5.94	N/A	4.72	0.81	1.22	1
H10P1_48	0	4.87	5.55	N/A	4.23	0.65	1.32	1
H10P1_72	0	5.20	5.54	N/A	3.77	1.43	1.77	1
H10P1_96	0	4.84	5.27	N/A	3.82	1.02	1.45	1
H10P1_120	0	4.82	5.15	N/A	3.98	0.84	1.18	1
H10P1_144	0	4.47	4.87	N/A	N/A	N/A	N/A	1
H10P1_168	0	4.03	4.58	3.38	4.32	-0.29	0.26	1
H10P1_192	0	5.33	5.99	4.26	4.07	1.25	1.91	1
H10P1_216	0	5.35	5.93	4.16	4.45	0.90	1.48	1
H10P1_240	0	5.21	5.60	4.01	4.64	0.57	0.96	1
H10P1_264	0	4.87	5.26	3.71	4.18	0.68	1.08	1
H10P1_288	0	5.00	4.88	3.82	4.06	0.94	0.82	1
H10P1_312	0	4.27	4.78	3.86	4.46	-0.19	0.32	1

ID	Breed	Td	Etr	Tr	Pr	Td/Pr	Etr/Pr	Admin
H10P1_336	0	3.00	3.85	3.17	4.74	-1.75	-0.89	1
H10P1_360	0	3.34	N/A	4.14	4.04	-0.70	N/A	1
H10P1_384	0	3.62	N/A	3.99	4.03	-0.41	N/A	1
H10P1_408	0	N/A	N/A	3.73	N/A	N/A	N/A	1
H10P1_432	0	N/A	N/A	3.67	3.93	N/A	N/A	1
H10P1_456	0	3.50	N/A	3.87	4.44	-0.94	N/A	1
H10P1_480	0	N/A	N/A	3.94	3.92	N/A	N/A	1
H10P1_504	0	N/A	N/A	3.92	4.09	N/A	N/A	1
UNK1	0	6.30	5.18	4.85	4.17	2.13	1.01	N/A
UNK2	0	6.11	5.16	4.70	4.14	1.97	1.01	N/A

Appendix 19: Progesterone ratio profile classification model confusion matrix output from XGBoost model, print out of data presented in-text (Chapter 4, Table 4-9).



ALTRENOGEST METABOLOMIC MODELLING ADDITIONAL INFORMATION

Appendix 20: Altrenogest (ALT) metabolomic analysis reference standards included in the reference mix injected and analysed with each batch.

ALT Reference Standards Mix	Structural Formula	Theoretical Molecular Ion *(m/z)*	Acquisition Mode (+/-ESI)
Testosterone	$C_{19}H_{28}O_2$	289.21621	positive
Epitestosterone	$C_{19}H_{28}O_2$	289.21621	positive
Altrenogest	$C_{21}H_{26}O_2$	311.20056	positive
Pregnenolone	$C_{21}H_{32}O_2$	317.24751	positive
Progesterone	$C_{21}H_{30}O_2$	315.23186	positive
Dehydroepiandrosterone (DHEA)	$C_{19}H_{28}O_2$	289.21621	positive
Testosterone Sulfate	$C_{19}H_{28}O_5S$	367.15792	negative
Nandrolone Sulfate	$C_{18}H_{26}O_5S$	353.14280	negative
Altrenogest Sulfate	$C_{21}H_{26}O_5S$	389.14266	negative
Pregnenolone Sulfate	$C_{21}H_{32}O_5S$	395.18922	negative
DHEA Sulfate	$C_{19}H_{28}O_5S$	367.15792	negative
5α-androstane-3β,17α-diol, 3-sulfate (ABAS)	$C_{19}H_{32}O_5S$	371.18922	negative

Appendix 21:Altrenogest (ALT) sulfate biomarkers relative retention times (RRT) to the internal standard (ISTD - testosterone sulfate-d3) in the original batch analysis and the later compound verification analysis.

Analyte	Old RT	Old ISTD RT	Old RRT	New RT	New ISTD RT	New RRT	RRT %Diff
E1S	9.333	10.174	0.917338	9.8558	10.65	0.925427	0.991259
TS	10.108	10.174	0.993513	10.6643	10.65	1.001343	0.992181
2-ME2S	8.462	10.174	0.831728	8.9821	10.65	0.84339	0.986173
PregS	12.454	10.174	1.224101	13.1031	10.65	1.230338	0.99493
CS	4.963	10.174	0.487812	5.5476	10.65	0.520901	0.936477

ALT SULFATES RANDOM FOREST (RF) MODEL DATASET

Appendix 22: Altrenogest (ALT) Sulfates model dataset

Samples	Label	E1S	TS	ME2S	PregS	CS
H5_2_a	0	82	674	2859	0	1227
H5_2_b	0	91	811	2892	0	1582
H5_2_c	0	43	426	2794	0	1112
H5_4_a	0	90	307	4267	0	2934
H5_4_b	0	54	431	4469	0	2825
H5_4_c	0	112	355	4076	0	3178
H5_6_a	0	83	240	2376	0	416
H5_6_b	0	55	167	2106	0	423
H5_6_c	0	81	143	1454	0	380
H5_8_a	0	0	184	1768	0	943
H5_8_b	0	0	203	1884	0	1088
H5_8_c	0	0	273	2265	0	1017
H5_12_a	0	0	37	1726	0	2000
H5_12_b	0	0	181	2125	0	2514
H5_12_c	0	32	252	1776	0	2544
H5_24_a	0	0	229	5338	0	4833
H5_24_b	0	0	110	5024	0	4919
H5_24_c	0	0	242	5084	0	5200
H5_48_a	0	30	303	2614	0	1670

Samples	Label	E1S	TS	ME2S	PregS	CS
H5_48_b	0	0	78	2394	0	1707
H5_48_c	0	0	118	2570	0	1775
H5_72_a	0	0	0	594	0	546
H5_72_b	0	0	0	557	0	694
H5_72_c	0	0	58	546	0	552
H5_96_a	0	0	50	1141	0	1633
H5_96_b	0	0	109	1548	0	1335
H5_96_c	0	0	32	1398	0	1330
H5_120_a	0	0	90	1985	0	1632
H5_120_b	0	0	156	2077	0	1459
H5_120_c	0	0	241	1845	0	2611
H5_144_a	0	0	100	3294	0	1468
H5_144_b	0	0	119	3068	0	1429
H5_144_c	0	0	107	2900	0	1973
H5_168_a	0	0	28	1586	0	1939
H5_168_b	0	0	147	2992	0	3587
H5_168_c	0	0	112	2883	0	3852
H5_192_a	0	0	177	4322	0	2269
H5_192_b	0	32	201	4101	0	2515
H5_192_c	0	0	109	4262	0	2946
H5_216_a	0	0	83	2906	0	1610

Samples	Label	E1S	TS	ME2S	PregS	CS
H5_216_b	0	0	80	2352	0	1825
H5_216_c	0	0	45	2697	0	2120
H5_240_a	0	38	169	3635	0	1531
H5_240_b	0	0	114	3453	0	2016
H5_240_c	0	0	195	3442	0	2049
H5_264_a	0	0	0	262	0	298
H5_264_b	0	0	97	278	0	150
H5_264_c	0	0	59	372	0	111
H5_288_a	0	0	39	1972	0	1378
H5_288_b	0	0	187	2091	0	1591
H5_288_c	0	0	55	1678	0	1463
H5_312_a	0	32	57	2335	6	418
H5_312_b	0	42	106	2094	0	386
H5_312_c	0	0	86	2579	0	449
H5_336_a	0	0	153	1267	0	1989
H5_336_b	0	0	0	1563	0	2078
H5_336_c	0	0	139	1540	0	2422
H5_338_a	0	54	53	2797	0	1086
H5_338_b	0	0	0	774	0	803
H5_338_c	0	0	0	0	0	0
H5_340_a	0	158	96	4019	0	2742

Samples	Label	E1S	TS	ME2S	PregS	CS
H5_340_b	0	150	0	4266	0	2247
H5_340_c	0	220	228	4225	0	2542
H5_342_a	0	105	169	1928	0	1387
H5_342_b	0	54	0	2178	0	1296
H5_342_c	0	111	75	1735	0	948
H5_344_a	0	76	247	1289	0	1367
H5_344_b	0	104	49	1333	0	1293
H5_344_c	0	114	98	1392	0	1533
H5_348_a	0	77	47	1185	0	829
H5_348_b	0	91	47	1163	0	562
H5_348_c	0	98	96	1103	0	593
H5_360_a	0	0	102	1542	0	507
H5_360_b	0	0	165	1478	0	473
H5_360_c	0	0	0	1474	0	431
H5_384_a	0	824	127	3467	0	1835
H5_384_b	0	882	115	3445	0	1843
H5_384_c	0	888	181	3163	0	1672
H5_408_a	0	4166	430	2818	14	1149
H5_408_b	0	4129	320	2973	0	1113
H5_408_c	0	3881	323	2998	0	1160
H5_432_a	0	1454	199	2541	0	875

Samples	Label	E1S	TS	ME2S	PregS	CS
H5_432_b	0	1619	141	2759	0	793
H5_432_c	0	1310	0	2978	0	730
H5_456_a	0	790	188	2176	0	316
H5_456_b	0	799	209	2514	0	85
H5_456_c	0	859	128	2432	0	113
H5_504_a	0	389	95	4132	0	1261
H5_504_b	0	420	159	4511	0	1208
H5_504_c	0	255	231	4247	0	1112
H6_2_a	1	0	397	1366	0	3103
H6_2_b	1	0	680	1591	0	3701
H6_2_c	1	0	174	1000	0	1526
H6_4_a	1	0	416	3143	0	6355
H6_4_b	1	0	501	2561	0	4220
H6_4_c	1	0	528	2634	0	6402
H6_6_a	1	0	256	1268	0	1369
H6_6_b	1	0	337	1235	0	956
H6_6_c	1	0	288	1048	0	622
H6_8_a	1	0	172	373	0	1276
H6_8_b	1	0	84	568	0	1057
H6_8_c	1	0	208	468	0	1176
H6_12_a	1	0	174	1225	0	2939

Samples	Label	E1S	TS	ME2S	PregS	CS
H6_12_b	1	0	25	1162	0	2628
H6_12_c	1	0	105	1115	0	1948
H6_24_a	1	0	378	962	0	930
H6_24_b	1	0	34	872	0	956
H6_24_c	1	0	0	1042	0	1213
H6_48_a	1	0	160	1206	0	649
H6_48_b	1	0	199	1143	0	617
H6_48_c	1	0	68	1164	0	0
H6_72_a	1	0	94	1317	0	180
H6_72_b	1	0	352	1401	0	331
H6_72_c	1	0	699	1308	0	599
H6_96_a	1	0	36	1659	23	3703
H6_96_b	1	41	217	1513	0	2368
H6_96_c	1	34	433	1856	0	2768
H6_120_a	1	0	42	973	0	1128
H6_120_b	1	0	319	1356	0	1085
H6_120_c	1	38	44	1098	0	1460
H6_144_a	1	0	200	620	0	326
H6_144_b	1	0	270	514	0	484
H6_144_c	1	0	213	637	0	158
H6_168_a	1	183	0	85	0	1322

Samples	Label	E1S	TS	ME2S	PregS	CS
H6_168_b	1	224	123	0	0	1318
H6_168_c	1	0	29	0	0	1083
H6_170_a	1	522	361	1037	0	1169
H6_170_b	1	444	290	861	0	1289
H6_170_c	1	499	313	890	0	1355
H6_172_a	1	1339	95	1930	0	2428
H6_172_b	1	944	168	1100	0	2421
H6_172_c	1	1650	279	1675	0	2724
H6_174_a	1	1421	89	804	57	645
H6_174_b	1	1123	0	1174	230	1246
H6_174_c	1	612	109	1199	28	1496
H6_176_a	1	779	134	161	65	278
H6_176_b	1	977	116	277	0	163
H6_176_c	1	809	70	221	0	65
H6_180_a	1	744	0	265	0	36
H6_180_b	1	762	123	197	0	17
H6_180_c	1	859	17	297	0	49
H6_192_a	1	559	159	1701	0	2584
H6_192_b	1	507	182	1953	0	2442
H6_192_c	1	497	116	1885	0	2914
H6_216_a	1	0	110	207	0	302

Samples	Label	E1S	TS	ME2S	PregS	CS
H6_216_b	1	0	261	306	0	1573
H6_216_c	1	0	0	298	0	1863
H6_240_a	1	0	14	857	0	320
H6_240_b	1	0	223	802	0	529
H6_240_c	1	0	57	804	0	97
H6_264_a	1	0	217	620	0	1337
H6_264_b	1	0	253	497	0	1482
H6_264_c	1	35	139	619	0	1220
H6_288_a	1	119	288	817	30	1119
H6_288_b	1	0	0	357	0	981
H6_288_c	1	0	239	837	0	1697
H6_312_a	1	669	41	1034	489	1277
H6_312_b	1	798	211	1213	0	747
H6_312_c	1	876	39	1263	0	2006
H6_336_a	1	788	189	417	0	2261
H6_336_b	1	559	269	348	0	1826
H6_336_c	1	780	10	479	0	2132
H6_360_a	1	3967	381	1489	0	911
H6_360_b	1	3814	244	1543	0	934
H6_360_c	1	4180	408	1527	0	633
H6_384_a	1	9320	825	2725	0	1632

Samples	Label	E1S	TS	ME2S	PregS	CS
H6_384_b	1	10157	740	2138	0	2324
H6_384_c	1	8778	921	2401	0	1209
H6_408_a	1	9487	746	1699	0	2262
H6_408_b	1	8911	340	1504	0	2148
H6_408_c	1	8575	708	1687	0	1843
H6_432_a	1	8314	586	1076	0	824
H6_432_b	1	8618	533	1156	51	1186
H6_432_c	1	8416	320	1198	0	831
H6_456_a	1	5518	212	279	30	90
H6_456_b	1	4030	367	494	0	240
H6_456_c	1	5529	342	439	0	188
H6_480_a	1	18725	1192	1464	0	219
H6_480_b	1	19419	958	1426	0	0
H6_480_c	1	18309	843	1287	0	93
H6_504_a	1	29891	1103	2079	0	2042
H6_504_b	1	24560	1119	1859	0	2583
H6_504_c	1	27573	545	2183	0	1818
UNK1_019	N/A	49997	4009	18471	106794	467
UNK2_020	N/A	52716	3547	17766	112264	421
UNK3_021	N/A	51326	3853	18639	105501	431

LEVODOPA ADMINISTRATION STUDY

ADDITIONAL STALEVO® 100 DOSAGE INFORMATION [174]

Appendix 23: Dosages of Stalevo product currently available with concentrations of levodopa (L-DOPA), carbidopa, and entacapone.

Label	L-DOPA (mg)	Carbidopa monohydrate (mg)	Entacapone (mg)
Stalevo 50	50	12.5	200
Stalevo 75	75	18.75	200
Stalevo 100	100	25	200
Stalevo 125	125	31.25	200
Stalevo 150	150	37.5	200
Stalevo 200	200	50	200

Stalevo® 100 is a schedule 2 (S4) prescription only medication containing L-DOPA, carbidopa and entacapone. This orally administered medication is formulated in six strengths, all containing a 4:1 ratio of L-DOPA and carbidopa, and all with 200 mg of entacapone.

The recommended daily dosage of Stalevo for the intended treatment of human PD is no more than 200 mg carbidopa monohydrate (limited benefits and greater risk of adverse side effects), no more than 2000 mg entacapone, and no more than 1500 mg L-DOPA. Therefore, a maximum of ten tablets per day for the lower strengths and a limit of seven tablets per day for the higher strengths is advised.

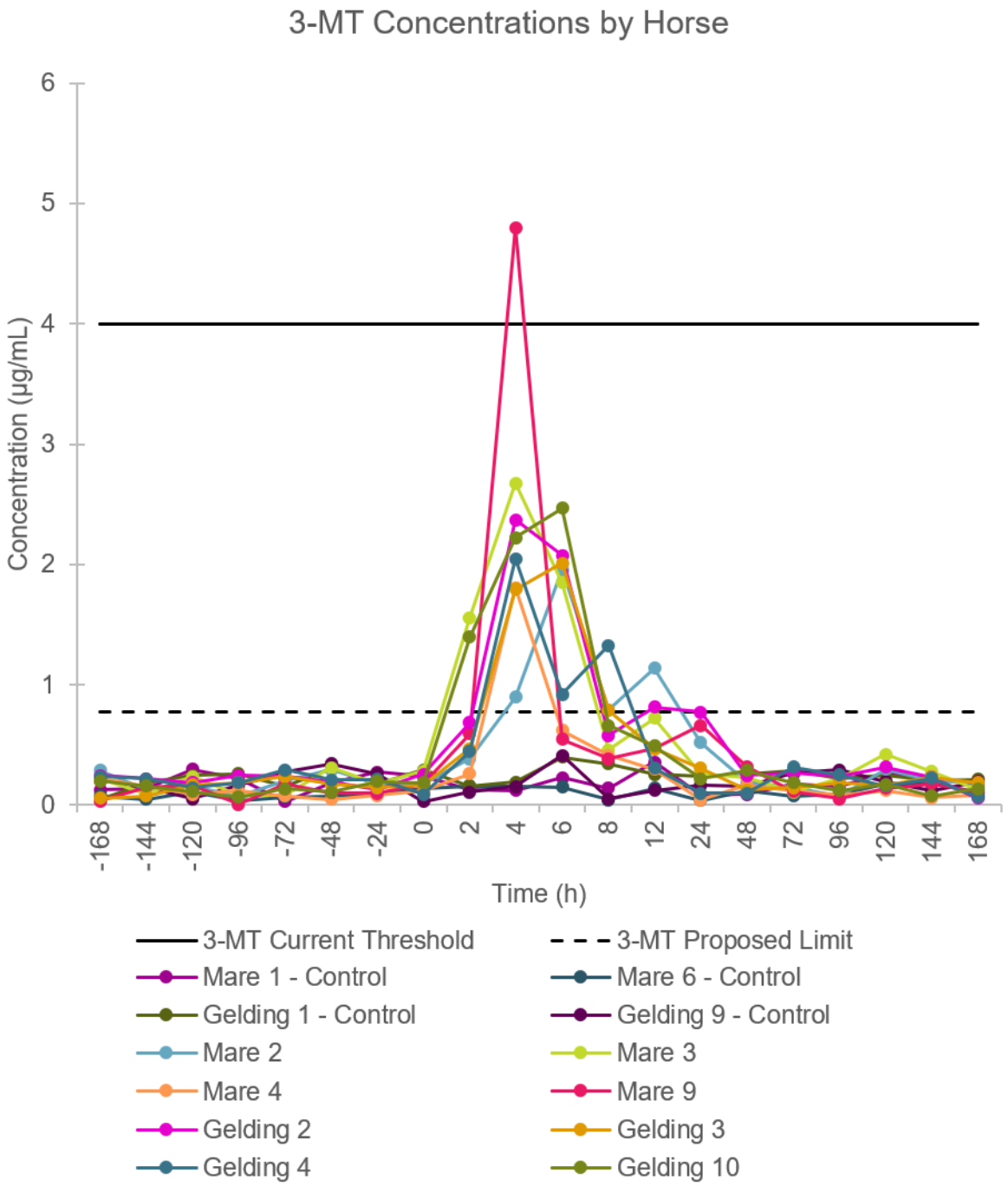
Appendix 24: Method validation results for the quantification of 3-methoxytyramine (3-MT) and tyramine (Tyr) in a water matrix. Validations parameters include sensitivity, linearity (low and high calibration ranges), accuracy, precision, recovery and matrix effects.

		3-MT	**Tyr**
Sensitivity	LOD	< 50 ng/mL	< 50 ng/mL
	LLOQ	< 100 ng/mL	< 100 ng/mL
Linearity (R^2)	[a]Low Range	N/A	N/A
	[b]High Range	0.8945	0.9807
Accuracy (%RE)	LQC	2.6	0.9
	MQC	4.1	-22.9
	HQC	-4.4	-31.2
Precision (%RSD)	LQC	2.7	14.0
	MQC	2.5	2.5
	HQC	3.9	3.9

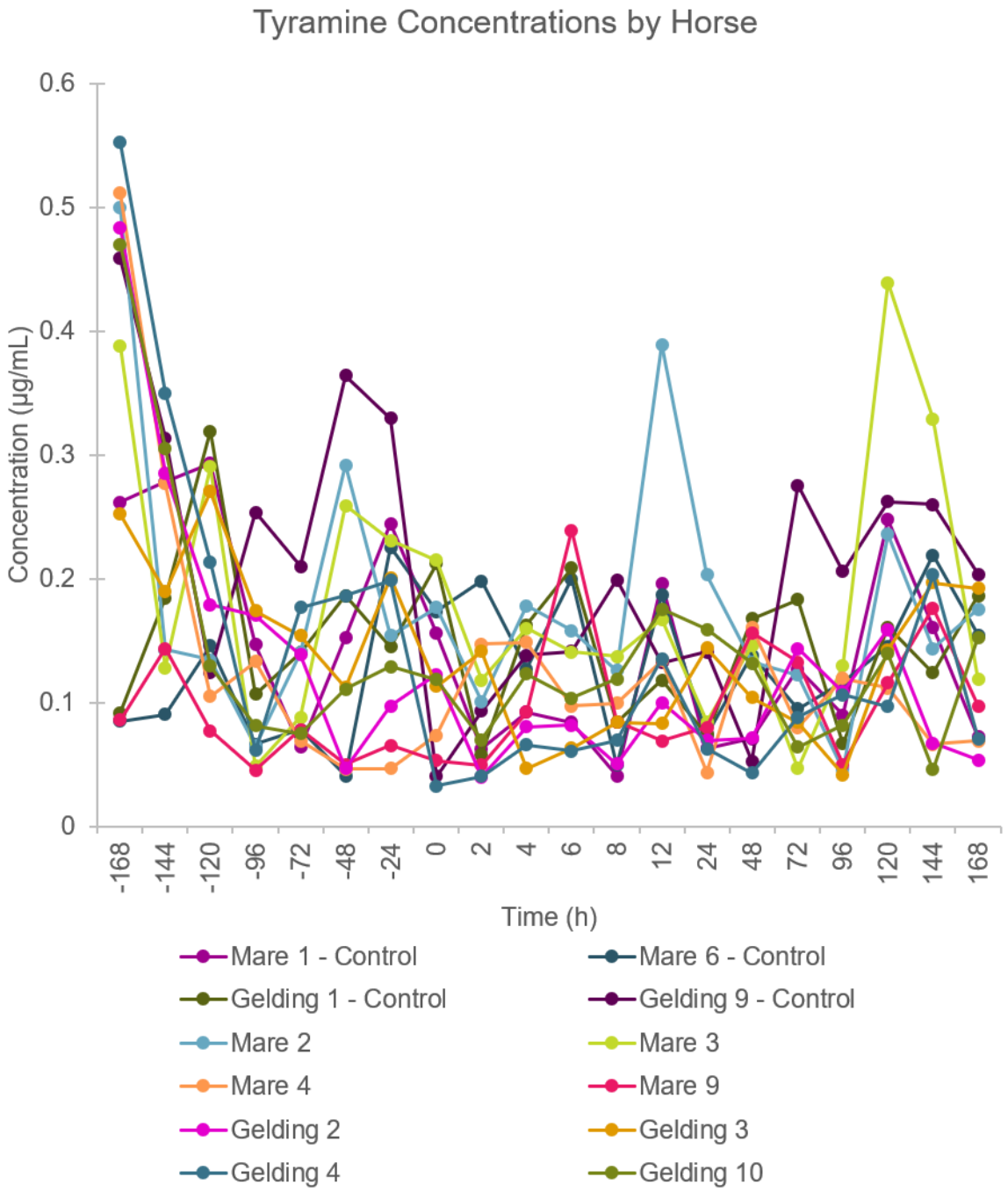
Linearity Definitions: [a]Low Range refers to a calibration range of 0.05 – 1 µg/mL, whereas, [b]High Range refers to a calibration range of 2 – 8 µg/mL. LQC, MQC, and HQC spiked concentrations were 2, 4, and 6 µg/mL, respectively.

ADDITIONAL QUANTITATIVE STALEVO ADMINISTRATION STUDY RESULTS

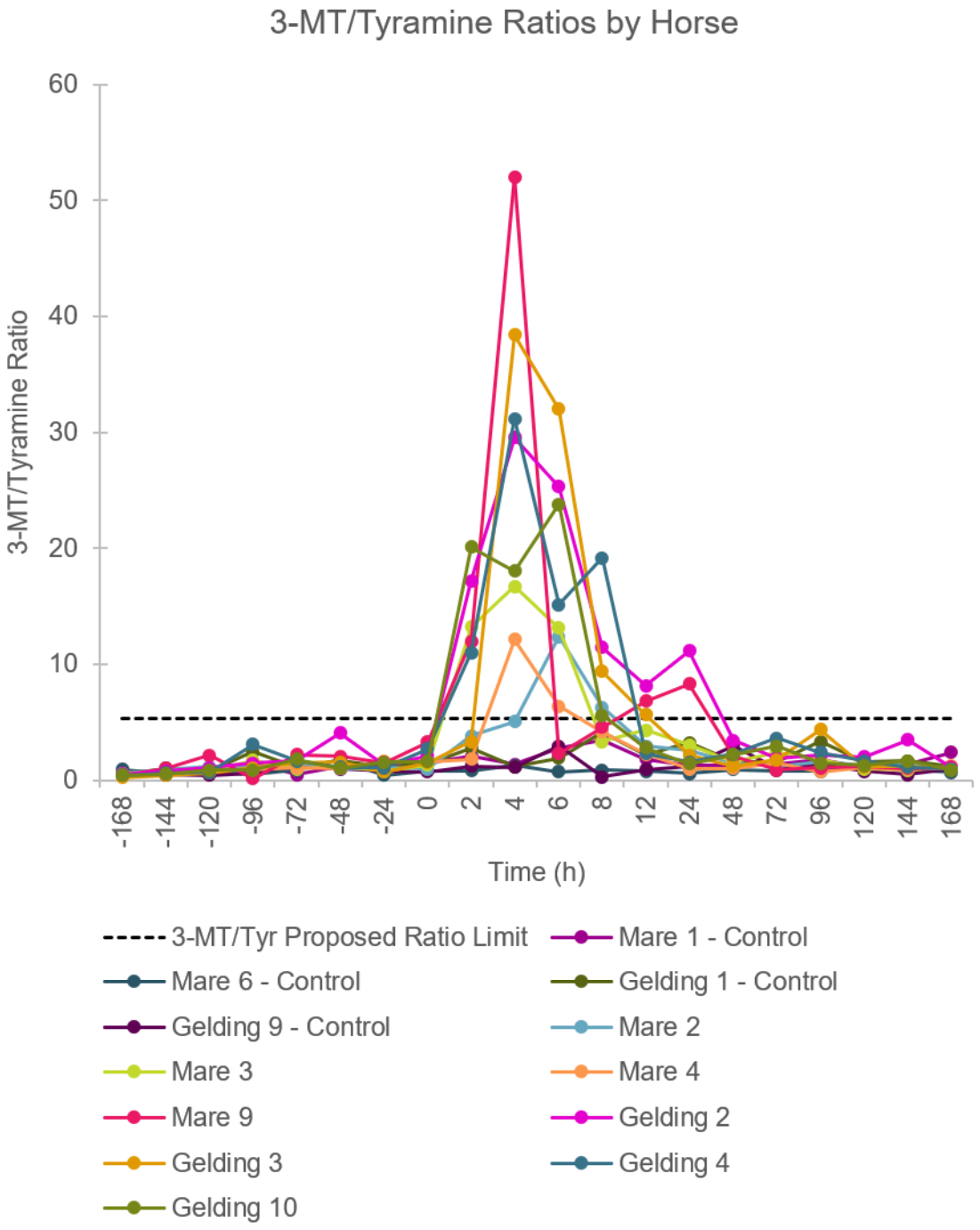
Appendix 25: Timewise plotted concentration values for 3-methoxytyramine (3-MT) for individual control (n = 4) and treated (n = 8) horses. The black solid line represents the current 3-MT threshold (4 µg/mL) and the black dashed line indicates the proposed 3-MT intelligence limit (0.776 µg/mL).



Appendix 26: Timewise plotted concentration values for tyramine for individual control (n = 4) and treated (n = 8) horses.



Appendix 27: Timewise plotted 3-methoxytyramine-to-tyramine (3-MT/Tyr) ratio values for individual control (n = 4) and treated (n = 8) horses. The black dotted line indicates the proposed 3-MT/Tyr ratio limit (5.3)



LEVODOPA METABOLOMIC ANALYSIS INFORMATION

Appendix 28: Reference standards included in the mix injected alongside the Stalevo administration study samples.

Dopamine Reference Standards Mix	**Structural Formula**	**Theoretical Molecular Ion *(m/z)***	**Acquisition Mode (+/-ESI)**
3-Methoxytyramine	$C_9H_{13}NO_2$	168.1019	positive
3-Methoxytyramine glucuronide	$C_{15}H_{21}NO_8$	342.1198	negative
3-Methoxytyramine sulfate	$C_9H_{13}NO_5S$	246.0444	negative
3-*O*-Methyl-carbidopa	$C_{11}H_{16}N_2O_4$	241.1183	positive
DL-4-Hydroxy-3-methoxymandelic acid	$C_9H_{10}O_5$	197.0456	negative
DL-Metanephrine	$C_{10}H_{15}NO_3$	198.1125	positive
DL-Normetanephrine	$C_9H_{13}NO_3$	184.0968	positive
Dopamine	$C_8H_{11}NO_2$	154.0863	positive
Dopamine-3-*O*-sulfate	$C_8H_{11}NO_5S$	232.0285	negative
Glutamic acid	$C_5H_9NO_4$	148.0604	positive
Homovanillic acid	$C_9H_{10}O_4$	181.0506	negative
Hordenine	$C_{10}H_{15}NO$	166.1226	positive
L-Phenylalanine	$C_9H_{11}NO_2$	166.0863	positive
L-Tyrosine	$C_9H_{11}NO_3$	182.0812	positive

N-Methylphenethylamine	$C_9H_{13}N$	136.1121	positive
N-Methyl-*p*-tyramine	$C_9H_{13}NO$	152.1070	positive
Phenethylamine	$C_8H_{11}N$	122.0964	positive
p-Tyramine	$C_8H_{11}NO$	138.0920	positive

Appendix 29: Reference standards included in the mix injected alongside the Stalevo administration study samples, specific to compounds found in the Stalevo medication.

Stalevo Reference Standards Mix	**Structural Formula**	**Theoretical Molecular Ion *(m/z)***	**Acquisition Mode (+/-ESI)**
Carbidopa	$C_{10}H_{14}N_2O_4$	225.0881	negative
Entacapone	$C_{14}H_{15}N_3O_5$	304.0939	negative
Levodopa	$C_9H_{11}NO_4$	198.0761	positive

DOPAMINE BIOMARKER MODEL DATASET:

Appendix 30: Dopamine Biomarker Model Dataset Features are labelled with the deprotonated molecule in m/z and the retention time in minutes. ID denotes the sample name, and the Label indicates the group, pre-administration samples as 0 and post-administration samples as 1.

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B1_TG10 72	0	531161	0	0	0	N/A	1
B1_TG4 72	0	327288	0	0	0	N/A	1
B1_TM2 72	0	435522	39466	0	0	N/A	1
B1_TM3 24	0	60701	0	0	0	N/A	1
B1_TM3 48	108061	43301	0	44658	0	N/A	1
B2_TG10 48	0	547916	11430	0	0	N/A	1
B2_TG10 96	0	251040	0	0	0	N/A	1
B2_TM3 12	0	129476	10426	298614	32912	N/A	1
B2_TM3 96	0	333737	0	0	0	N/A	1
B2_TM9 24	120940	42287	0	127704	34337	N/A	1
B3_TG10 168	0	43416	N/A	60437	0	N/A	1
B3_TG3 120	0	0	N/A	0	0	N/A	1
B3_TG3 48	0	0	N/A	0	8876	N/A	1
B3_TG4 96	0	0	N/A	0	5453	N/A	1
B3_TM2 144	0	0	N/A	128733	51219	N/A	1
B3_TM4 120	9517	0	N/A	0	0	N/A	1
B3_TM4 96	0	0	N/A	0	0	N/A	1
B3_TM9 168	0	15229	N/A	0	495187	N/A	1

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B4_TG10 24	0	84071	3703	49231	51633	N/A	1
B4_TG2 120	0	103583	0	160740	18611	N/A	1
B4_TG2 72	35935	0	4719	229920	0	N/A	1
B4_TG3 168	32508	0	32151	0	140537	N/A	1
B4_TG4 12	0	195690	19990	0	79401	N/A	1
B4_TM2 24	0	0	16851	96703	629130	N/A	1
B4_TM4 48	66585	0	0	334799	43750	N/A	1
B4_TM9 72	44475	0	0	0	50153	N/A	1
B5_TG10 12	0	0	N/A	0	0	N/A	1
B5_TG2 12	0	125991	N/A	0	22709	N/A	1
B5_TG3 24	0	0	N/A	0	118284	N/A	1
B5_TG4 144	0	173683	N/A	0	29053	N/A	1
B5_TM2 120	0	158493	N/A	0	29227	N/A	1
B5_TM3 120	0	761151	N/A	0	0	N/A	1
B5_TM4 12	0	0	N/A	175158	46177	N/A	1
B5_TM9 120	15743	0	N/A	0	285487	N/A	1
B5_TM9 144	0	473881	N/A	0	132115	N/A	1
B6_TG2 144	0	652265	7847	0	0	N/A	1
B6_TG2 24	0	0	11413	0	966091	N/A	1
B6_TG3 144	0	119164	0	0	124878	N/A	1

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B6_TM2 48	0	346977	0	0	365168	N/A	1
B6_TM4 168	0	0	0	0	314543	N/A	1
B6_TM4 72	58254	1053751	0	0	0	N/A	1
B7_TG2 168	0	95406	53067	35113	9451	187942	1
B7_TG3 12	0	0	53695	0	65765	0	1
B7_TG3 72	0	0	0	0	20057	0	1
B7_TG4 168	0	61271	22557	0	99454	35178	1
B7_TG4 24	0	0	55374	0	96370	111850	1
B7_TM2 168	18376	0	84364	84466	151166	0	1
B7_TM9 96	0	18918	5296	0	169951	151935	1
B8_TG3 96	0	636797	N/A	114881	0	N/A	1
B8_TM3 144	622531	29390	N/A	15972	76744	N/A	1
B8_TM3 168	0	16751	N/A	0	402095	N/A	1
B8_TM4 144	136609	0	N/A	0	83529	N/A	1
B8_TM9 48	0	83727	N/A	0	0	N/A	1
B9_TG10 144	0	251018	N/A	0	0	N/A	1
B9_TG4 48	0	399371	N/A	182099	84429	N/A	1
B9_TM2 96	0	477459	N/A	0	0	N/A	1
B9_TM3 72	5731	427333	N/A	340267	0	N/A	1
B9_TM4 24	0	45229	N/A	0	142621	N/A	1

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B10_TG10 120	0	758938	2137	75651	0	0	1
B10_TG2 48	0	741469	21047	262797	0	0	1
B10_TG2 96	0	786991	10937	95421	0	0	1
B10_TG4 120	0	646182	6864	106128	0	0	1
B10_TM2 12	6153	307802	70270	224751	131833	0	1
B10_TM9 12	0	292915	2520	247855	34281	0	1
B1_TG10 0	0	131960	4217	50960	0	N/A	0
B1_TG2 -120	0	186950	0	0	0	N/A	0
B1_TG2 -144	0	290332	0	0	0	N/A	0
B1_TG3 -120	0	309097	0	67217	0	N/A	0
B1_TG3 -168	75447	0	0	0	0	N/A	0
B1_TM2 -120	0	47822	14627	0	0	N/A	0
B1_TM4 -144	21949	322964	0	0	0	N/A	0
B1_TM4 -168	1021612	0	0	0	0	N/A	0
B1_TM9 0	0	28122	0	0	22864	N/A	0
B1_TM9 -72	1063738	0	11195	0	0	N/A	0
B2_TG2 -168	428757	0	0	0	0	N/A	0
B2_TG2 -24	0	93607	0	0	0	N/A	0
B2_TG3 0	0	0	36100	0	11740	N/A	0

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B2_TG3 -72	0	183446	9437	711659	0	N/A	0
B2_TG4 -144	39549	300827	0	397230	0	N/A	0
B2_TG4 -48	0	178784	0	372918	0	N/A	0
B2_TM2 -72	0	576653	8046	266153	0	N/A	0
B3_TG4 -120	0	0	N/A	0	0	N/A	0
B3_TM2 -24	0	0	N/A	0	26090	N/A	0
B3_TM3 -120	51645	0	N/A	57192	0	N/A	0
B4_TG10 -168	63597	0	0	0	0	N/A	0
B4_TG3 -96	0	0	3049	404799	0	N/A	0
B4_TG4 -24	0	347477	22932	138404	0	N/A	0
B4_TM2 -48	0	0	19667	0	0	N/A	0
B4_TM3 -168	0	0	0	61715	15556	N/A	0
B4_TM3 -48	0	97379	0	118253	0	N/A	0
B4_TM9 -144	5559	175624	3797	325847	0	N/A	0
B5_TG10 -120	0	0	N/A	0	0	N/A	0
B5_TG2 0	0	39565	N/A	0	0	N/A	0
B5_TG4 -168	527835	0	N/A	57691	4078	N/A	0
B5_TM4 -96	56376	385025	N/A	0	0	N/A	0
B6_TG10 -72	0	0	0	0	0	N/A	0
B6_TG4 -72	0	0	0	0	36055	N/A	0

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B6_TM3 0	42432	156361	0	0	97538	N/A	0
B6_TM3 -144	0	756026	0	0	0	N/A	0
B6_TM9 -120	58360	0	0	0	134776	N/A	0
B6_TM9 -24	0	509364	0	0	0	N/A	0
B7_TG10 -96	0	425420	3369	262626	0	0	0
B7_TM2 -168	707260	0	4368	68741	0	0	0
B7_TM4 -120	236132	370121	3469	361896	0	0	0
B7_TM4 -72	0	340424	0	115784	0	0	0
B8_TG10 -144	0	23046	N/A	18505	0	N/A	0
B8_TG2 -72	0	643763	N/A	24636	0	N/A	0
B8_TG2 -96	0	668232	N/A	174945	0	N/A	0
B8_TG3 -48	0	788086	N/A	117227	0	N/A	0
B8_TG4 0	0	114118	N/A	0	0	N/A	0
B8_TG4 -96	6746	760901	N/A	157034	11091	N/A	0
B8_TM2 0	4719	0	N/A	0	3968	N/A	0
B8_TM2 -96	0	192834	N/A	0	0	N/A	0
B8_TM4 0	0	60976	N/A	0	0	N/A	0
B9_TG10 -24	0	153005	N/A	47097	0	N/A	0
B9_TG2 -48	0	881694	N/A	91699	0	N/A	0
B9_TM2 -144	0	307635	N/A	0	27044	N/A	0

ID	125.023 74/0.954	211.097 44/1.209	280.622 44/1.232	317.051 67/2.509	365.135 8/3.743	336.040 71/1.29	Label
B9_TM3 -96	0	667971	N/A	116954	0	N/A	0
B9_TM4 -24	71412	69586	N/A	0	0	N/A	0
B9_TM9 -48	1335906	0	N/A	0	115665	N/A	0
B9_TM9 -96	0	0	N/A	0	28117	N/A	0
B10_TG10 -48	0	576787	0	61141	0	0	0
B10_TG3 -144	0	489897	6025	255789	0	0	0
B10_TG3 -24	15382	287477	24914	0	58904	0	0
B10_TM3 -24	15455	77071	36426	658320	0	0	0
B10_TM3 -72	0	494781	56601	55407	0	0	0
B10_TM4 -48	0	322972	2735	30803	0	0	0
B10_TM9 -168	0	0	0	0	0	25610	0

Appendix 31: Additional dopamine biomarker model tests (RF, random forest model) using MetaboAnalyst and various dataset sizes.

Model type	Dataset used	Parameters	Outcome
RF	Combined study averaged dataset (B1-10, Pre and Post (**2-72 h**) treated horses) n = 129 Pre/post split = 64/65		AUC = 0.643 (best score given to 3 feature ROC)
RF	Combined study averaged dataset (B1-10, Pre and Post (2-72 h) treated horses) n = 129 Pre/post split = 64/65	Manual Model Builder: 211, 280, 317, 365	AUC = 0.618 Pred Acc = 60.3% with 100 CV
RF	Combined study averaged dataset (B1-10, Pre and Post (2-72 h) treated horses) n = 129 Pre/post split = 64/65	Manual Model Builder: 211, 317, 365	AUC = 0.635 Pred Acc = 59.8% with 100 CV
RF	Combined study averaged dataset (B1-10, Pre and Post (2-72 h) treated horses) n = 129 Pre/post split = 64/65	Manual Model Builder: 317, 365	AUC = 0.639 Pred Acc = 60.5% with 100 CV
RF	Combined study averaged dataset (B1-10, Pre and Post (2-72 h) treated horses) n = 129 Pre/post split = 64/65	Manual Model Builder: 211, 365	AUC = 0.652 Pred Acc = 64.7% with 100 CV