THE UNIVERSITY
of EDINBURGH

Female Endocrine Adaptations to Arduous Military Training

Robert Michael Gifford

A Thesis submitted for the degree of Doctor of Philosophy

School of Clinical Sciences

The University of Edinburgh

2019

Declaration

I declare that I have written this Thesis and that the work has not been submitted for any other degree or professional qualification. I confirm that the work submitted is my own, except where work which has formed part of jointly-authored publications has been included. My contribution and the contributions of the other authors have been indicated in the introduction to each chapter. I confirm that appropriate credit has been given within this Thesis where reference has been made to the work of others.

Robert Michael Gifford

28 August 2019

Acknowledgements

I am enormously grateful to the participants who volunteered for the FEAT and E4T studies. If I was completing Sandhurst or planning to ski across Antarctica, I am sure I would decline the offer to participate in an exacting study or give up so much time and so many samples. These amazing Officers, Soldiers and Officer Cadets are the heroes of this work.

Studying these amazing women was a tremendous collaborative effort by experts from many disciplines, to whom I am extremely grateful. Front and centre among them are my supervisors-par-excellence. Prof Rebecca Reynolds has been unfailingly steadfast, long-suffering, approachable and scientifically astute. Despite innumerable other responsibilities, both within Edinburgh and outwith, she has always been generous in time and spirit and never discouraging or impatient (although I gave her every opportunity). Likewise, Col David Woods has been a source of scientific rigour, humour, pragmatism, collateral thinking, military grounding, new perspectives and ideas and, perhaps above all, reassurance. Both have held me back from flights-of-ideas, kept me on the relatively straight and narrow and pushed me towards goals I never remotely thought I could achieve.

I am indebted to Prof Julie Greeves OBE for welcoming me into her team, for her boundless enthusiasm, for her patience and understanding of my clinical background and for allowing her already-stretched team to dedicate so much time to FEAT. Prof Richard Anderson closely advised the project *ab initio,* suggesting the analysis in Chapter 2 and much of the protocol, and was ready to discuss our findings regularly. I am also deeply grateful to Dr Forbes Howie for accommodating our many samples, often at short notice, mentoring me in biochemistry and lab work and for his unfailing generosity and goodwill. I am grateful to Dr Tom O'Leary, Dr Sophie Wardle and Beccy Double for their untiring work in Sandhurst, carrying out much of the urine and saliva sample collection, dual x-ray absorptiometry and high resolution peripheral quantitative scans, sample processing and of course the painstaking energy assessment phases.

The Royal Military Academy Staff, not least Brig Bill Wright, Maj Rich Adamson, Maj Martyn Cook, Maj Bob Whittaker, Maj Jane Ellis and Mr Neil Turner generously volunteered their time, real estate, cadets, tea and toast. Academy staff were always welcoming and professional towards me as a lowly crab. I am enormously grateful.

Jo Singleton, Garry Tucker and her team from the NHS Lothian/ Wellcome Trust CRF were a delight to work with and were key to the success of the field studies. They integrated into the alien environment of Sandhurst with such ease and good humour and did not allow austerity or uncertainty to hamper their research excellence. I am grateful to Finny Paterson and Steve McSwiggan for generously allowing Jo, Garry and the team to be released for such long periods, in particular allowing Jo to travel to Chile (my apologies to those who missed out).

I am grateful to Kirsten Wilson for efficiently and uncomplainingly completing many of our assays (and bearing with my lab skills). I acknowledge the laboratories of Prof Nick Mills, Dr Neil Johnson and Dr Takeshi Fujisawa for the generous use of their Architect analyser, patience with my admin, and supply of many consumables. Dr Natalie Homer, Dr Scott Denham and colleagues stretched their lab to complete one of their largest ever studies in FEAT; this required a monumental effort and I am extremely grateful. Dr Mike Crane performed many additional assays for us and did much to guide the manuscript in Chapter 2, soliciting the help of Angela Ballentyne for Passing Bablok analyses. I acknowledge Prof Clemens Kirschbaum and the Dresden Assay Lab Service, the Glasgow Royal Infirmary Biochemistry Department, NHS Greater Glasgow, led by Dr Susan Johnston for their important contributions. I am grateful to Dr Tommaso Todisco, who was a tireless systematic reviewer, and to Dr Michael Allerhand for performing the meta-analysis.

Being invited to conduct the E4T study was a *coup*, and I am grateful to Air Cdre Rich Withnall for this, and to Prof Adrian Wilson, Dr John Hattersley and Prof Chris Imray. These remarkable scientists opened up a new field of research to me and I am humbled to have been included in their exceptional work. I am grateful to Prof Chris Boos for planning and analysing heartrate variability, and for his generous contribution to the Physiological Reports manuscript.

Dr Fiona Gifford and Giffords Poppy (6) Oliver (4) and Lucy (0) are my inspiration and give me no end of joy (Fiona also helped with data collection). Thank you for allowing me to immerse myself in this job, and for keeping me happy and motivated. I'm indebted to my parents in law Moira and Allistaire for looking after Fi and the kids (and the house) when I've been absent in body and/ or mind, and to my parents for their inspiration and encouragement.

Soli Deo gloria.

Abstract

Women are able to train in arduous combat roles for the first time in the British military, yet data on sex differences in physiological responses to such physically demanding employment are lacking. Military women appear to experience high rates of fractures, menstrual disturbance and adverse psychological outcomes, conditions related to perturbations in metabolic and hormonal function. There is a paucity of studies investigating hypothalamic-pituitary-gonad (HPG) and hypothalamic-pituitary-adrenal (HPA) axis function in military women. Low energy availability (EA) is an important cause of reproductive dysfunction and decreased bone mineral density (BMD) in athletes, while psychological stress and activation of the HPA axis are also associated with HPG axis suppression.

Military training is physically demanding, psychologically stressful and frequently takes place in extreme climates. Therefore, the work in this thesis aimed to characterise female endocrine adaptations to arduous military in two settings: basic military training in the UK and highly arduous exercise in extreme cold. To address female responses to extreme heat, relative rates of heat illness (HI) among men and women were compared in a systematic review and meta-analysis.

Female metabolic and endocrine adaptations to basic military training were explored in a cohort study of women undertaking the 11-month Commissioning Course at the Royal Military Academy, Sandhurst. Changes in body composition were measured by whole-body dual-energy-x-ray absorptiometry (DXA) at baseline and after 3, 7 and 11 months. Energy availability was assessed over 10-day periods using gold-standard techniques during months 3, 6 and 10 of training. Fasting blood sampling at baseline and after 7 and 11 months assessed leptin, insulin and glucose, as well as cortisol, cortisol binding globulin (CBG), oestradiol, anti-Müllerian hormone (AMH), inhibin B and bone turnover markers.

Dynamic assessment of the HPG and HPA was undertaken using Combined Hypothalamic-pituitary-Ovarian and adrenal Cortex (CHOC) tests (LH/FSH and cortisol response to gonadotrophin releasing hormone (GnRH) and adrenocorticotrophic hormone, respectively, at five time points over 1 hour), at baseline and after 9 months. Psychological stress and mood were assessed regularly using questionnaires. Diurnal saliva cortisol was measured approximately every five weeks of training, and 1cm hair samples were assayed for cortisol to give

an average monthly cortisol profile. A urinary progesterone: creatinine ratio cut-off was validated and used to determine if ovulation had taken place in women not using hormonal contraceptives, throughout the Commissioning Course. Changes in BMD were explored using DXA.

Sixty one female Officer Cadets commenced the study (mean age 24.0 ±0.3 years) of whom 52 completed the study and the Course (8 withdrew due to musculoskeletal injuries; one changed career). Fat mass decreased 0.8kg from baseline to month 3 of training, increased 1.8kg to month 7 and reverted to baseline by month 11; fat-free mass did not change. Leptin and insulin resistance increased in association with fat mass. The assessment of EA was hampered by underreporting of energy intake by around 30%, however this measure of EA was weakly associated with physical performance, and paradoxically inversely associated with change in fat mass, suggesting compensatory over-eating after periods of arduous training and low energy intake.

Questionnaires showed lower mood, increased stress and greater sleep deprivation during the course. Diurnal cortisol variability peaked in the first month of training, followed by habituation over the ensuing 10 months. Fasting plasma cortisol decreased, while CBG remained stable during the Course. Hair cortisol increased throughout, possibly reflecting physical exercise. Cortisol response to ACTH was modestly lower at 9 months than baseline. These data suggest a healthy overall response of the HPA axis to demonstrably arduous training.

Of 22 participants (36%) not using hormonal contraceptives, 7 (32%) reported increased menstrual frequency and 7 (32%) reported amenorrhoea or oligomenorrhoea. Of cycles with corresponding urine concentrations, 80% were anovulatory. LH and FSH responses to GnRH were suppressed after 9 months and this effect was apparent in both users and non-users of hormonal contraceptives (regardless of type). Oestradiol, FSH and inhibin B increased during the course, but there was no change in AMH, suggesting hypothalamic suppression of the HPG axis and follicular dysgenesis. Bone turnover was modestly increased, reflected in higher markers of both bone formation and resorption after 8 and 11 months. Total BMD reduced slightly, with modest regional losses from arms, legs and ribs, while trunk, spine and pelvis BMD were preserved.

The female response to arduous exercise in extreme cold was explored in six participants (median (range) 32.7 (28.6–36.1) years) undertaking the first unassisted

all-female team crossing of Antarctica. The measures used above were adapted for use 1 month pre-expedition and 5 and 14 days post-expedition; no through-expedition measurements were possible, and the CHOC test was performed after low-dose dexamethasone suppression. Mean (SD) weight loss was 10.1 (2.3) kg, constituting fat and not fat-free mass. Fasting oestradiol, cortisol and insulin resistance were unaffected, except leptin, which fell during the expedition and recovered partially after 14 days. Hair cortisol was elevated during the expedition. Cortisol, LH and FSH responsiveness were suppressed prior to the expedition, but were unchanged change immediately afterwards. LH responsiveness was higher than pre-expedition at expedition +14 days. Bone turnover was uncoupled immediately after the expedition, with higher resorption markers and lower formation markers, resolving after 14 days. Total BMD was unaffected.

The systematic review and meta-analysis included data containing comparable male and female rates of HI. Incident rate ratios were calculated and adjusted, where possible, for severity, occupation and age. Contrary to expectations based on published laboratory studies of sex-related physiology, the systematic review of found consistently higher rates of HI reported among men than women. The meta-analysis found HI rates were over twice as high in men than women (mean (95% CI) incident risk ratio 2.28 (1.66-3.16)). This pattern was consistent across the age span and all severities of HI, and was greater in occupational HI than non-occupational HI (5.66 (2.53-13.64) vs 2.96 (2.14-4.10).

In conclusion, this thesis demonstrated that women in basic military training do experience suppression of the HPG axis, despite long-term energy sufficiency with evidence of compensatory overeating and resilient HPA axis function. Resilient endocrine function in women who crossed Antarctica provides a proof of concept that with appropriate selection, nutrition and preparation, women can benefit from extremely arduous training. Women are also resilient to extreme heat, as evidenced by lower rates of HI than in men. The mechanism underlying reproductive dysfunction and associated pathology in women requires further elucidation but need not preclude women from entering such roles.

Lay Summary

The United Kingdom has recently allowed women to become soldiers in infantry and tank regiments. However, there are concerns that women doing these jobs are at high risk of illnesses and injuries: stress fractures (bone breakages from running or marching) and psychological problems are commoner among military women than men, while menstrual disturbance and fertility problems may be commoner among military than civilian women. These conditions are linked by hormones, which confer biological differences between men and women and can be altered by arduous training.

The main female hormones, oestradiol and progesterone, alter the body's bone strength, muscle and fat content, temperature control, responses to extremes hot and cold, and ovary function. They are controlled by two hormones, luteinising hormone (LH) and follicle stimulating hormone (FSH), which in turn are controlled by gonadotrophin releasing hormone (GnRH) in the brain. This hormone system is called the hypothalamic-pituitary-gonad (HPG) axis. The HPG axis can be impeded by insufficient energy, when calorie intake from the diet is outweighed by calorie expenditure from exercise (as is often found in professional sportswomen), and by stress. Stress is a complex response, with psychological symptoms and hormonal changes following challenging situations. The main stress hormone is cortisol.

In 68 female Officer Cadets going through 11-month basic military training at the Royal Military Academy, Sandhurst, UK, we measured HPG axis function, cortisol levels and stress, nutrition and exercise, and the risk of stress fracture. To do this we performed regular blood, urine and saliva sampling, scans and assessments of diet and exercise. During training, the HPG axis became suppressed.

Around one third of women not using contraceptive pills found their periods stopped, and almost all stopped ovulating completely. The levels of LH and FSH released in response to GnRH became reduced after 7 months, although morning FSH levels increased. Other blood tests suggested that follicles in the ovaries became larger, but weren't released. Body weight and fat did fluctuate around 2 kg, but were the same at the end as at the beginning of training. Calorie expenditure from exercise was very high, but it was difficult to measure how much the cadets were eating, due to a high number of snacks which weren't recorded. We found that where the ratio of calorie intake to expenditure was low, performance (e.g. 1.5 mile run time or muscle

mass) was worse and gains in fat mass were higher. We think this was because women who ate less during periods of hard exercise tended to overcompensate by eating more and/ or exercising less afterwards. Blood test markers increased which are commonly found with increasing fatness. We found very little evidence of insufficient calorie intake, and the changes we observed in the HPG axis were of a different kind to what has been seen before in athletes.

We found that Officer Cadets found the training very stressful; it dampened their mood and they felt less resilient by the end – which could well be related to tiredness. Different measures of cortisol showed increasing average levels, but decreased rises in cortisol in response to stressful situations, as training progressed. Tests of bone health showed increases in formation of bone during training, but possibly areas of increased risk of stress fracture mid-way through training. As size of bones increased, bone hardening processes (whereby they are strengthened) lagged behind.

We also studied six women before and after the first all-female, unassisted crossing of Antarctica, using many of the same methods as at Sandhurst. While they these women lost significant amounts of fat during the expedition (on average 10.1 kg, or 1.6 stones), their HPG axis function was unaffected. We repeated the tests 2 weeks after the expedition, and found that the HPG axis had improved to near normal levels – indicating that it was suppressed before the expedition. Although average cortisol levels were high during the expedition, there was no change in the cortisol response to stress. Their bone strength was unaffected, and although some blood tests suggested one breakdown straight afterward the expedition these resolved after 2 weeks

Finally, we studied the rates of heat illness (anything from a rash to a coma that is due to exercise and/ or hot weather) from all available records published. Due to known effects of progesterone and oestrogen, we expected to find rates were higher in men than women. In fact we found the opposite – heat illness was on average 2.2 times commoner among men, including when age, occupation and severity of heat illness were accounted for. We think this discrepancy is due to known behavioural differences between the sexes in the heat and, on average, greater appreciation of risk among women.

Military personnel are required to contribute to global operations in unpredictable contexts, and diverse and challenging environments. We studied the biological

responses of women to basic military training, extreme cold and extreme heat. Reproductive problems at Sandhurst were likely due to a combination of psychological challenges, suboptimal nutrition and sleep deprivation. Simple educational measures and psychological and nutritional support could help prevent adverse consequences of training, but long-term follow-up would be needed to determine if there is any effect on fertility. Our studies have also demonstrated the resilience of women to extreme exercise in the cold, and in resilience to heat illness. These findings provide a proof of concept that female sex need not preclude individuals from being considered for any military role and are highly relevant to current military policy.

Publications and Presentations relating to this thesis

Publications

Gifford, R.M., Reynolds, R.M., Greeves, J., Anderson, R.A., Woods, D.R., 2017. Reproductive dysfunction and associated pathology in women undergoing military training. Journal of the Royal Army Medical Corps 163, 301-310. (**Appendix A**)

Gifford, R.M., Reynolds, R.M., 2017. Sex differences in early-life programming of the hypothalamic-pituitary-adrenal axis in humans. Early human development 114, 7-10.

Gifford, R.M., Howie, F., Wilson, K., Johnston, N., Todisco, T., Crane, M., Greeves, J.P., Skorupskaite, K., Woods, D.R., Reynolds, R.M., Anderson, R.A., 2018. Confirmation of ovulation from urinary progesterone analysis: assessment of two automated assay platforms. Scientific reports 8, 17621. (**Section 3.1**)

Gifford, R.M., Wardle, S.L., O'Leary, T., Double, R., Homer, N.Z.M., Kirschbaum, C., Greeves, J., Woods, D.R., Reynolds, R.M., 2019. Positive adaptation of HPA axis function in women during 44 weeks of infantry-based military training. Psychoneuroendocrinology 110, 104432. (**Section 3.3**)

Gifford, R.M., O'Leary, T., Cobb, R., Blackadder-Weinstein, J., Double, R., Wardle, S.L., Anderson, R.A., Thake, C.D., Hattersley, J., Imray, C.H.E., Wilson, A., Greeves, J.P., Reynolds, R.M., Woods, D.R., 2019. Female Reproductive, Adrenal, and Metabolic Changes during an Antarctic Traverse. Medicine and science in sports and exercise 51, 556-567. (**Section 4.1**)

Taylor, N., **Gifford, R.M.,** Cobb, R., Wardle, S.L., Jones, S., Blackadder-Weinstein, J., Hattersley, J., Wilson, A., Imray, C., Greeves, J.P., Reynolds, R., Woods, D.R., 2019. Experience from the selection and nutritional preparation for Expedition ICE MAIDEN: the first successful all-female unassisted Antarctic traverse. Journal of the Royal Army Medical Corps. Published Online First: 15 May 2019, doi: 10.1136/jramc-2019-001175. (**Section 4.2**)

Gifford, R.M., Boos, C.J., Reynolds, R.M., Woods, D.R., 2018. Recovery time and heart rate variability following extreme endurance exercise in healthy women. Physiological reports 6, e13905-e13905.

Hattersley, J., Wilson, A.J., **Gifford, R.M.,** Cobb, R., Thake, C.D., Reynolds, R.M., Woods, D.R., Imray, C.H.E., 2019. Pre- to post-expedition changes in the energy usage of women undertaking sustained expeditionary polar travel. Journal of Applied Physiology 126, 681-690.

O'Leary, T.J., **Gifford, R.M.,** Double, R.L., Reynolds, R.M., Woods, D.R., Wardle, S.L., Greeves, J.P., 2019. Skeletal responses to an all-female unassisted Antarctic traverse. Bone 121, 267-276.

Gifford, R.M., Todisco, T., Stacey, M., Fujisawa, T., Allerhand, M., Woods, D.R., Reynolds, R.M., 2019. Risk of heat illness in men and women: A systematic review and meta-analysis. Environ Res 171, 24-35. (**Chapter 5**)

Gifford, R.M., Allerhand, M., Woods, D.R., Reynolds, R.M., 2019. Response to "Letter to the Editors" regarding the article "Risk of heat illness in men and women: A systematic review and meta-analysis". Environ Res 172, 723.

Oral Presentations

Gifford R. M. Woods D. R., Reynolds R. M., Characterising female endocrine and metabolic changes during the British Army Commissioning Course: the Female Endocrinology in Arduous Training (FEAT) Study. Royal Society of Medicine Colt Foundation Prize (winner), London, UK December 2019

Gifford R. M. Woods D. R., Anderson R. A., Reynolds R. M., Hypothalamic-pituitary-gonad (HPG) axis suppression during basic military training in women despite increased adiposity and insulin resistance. Society for Endocrinology BES (oral poster finalist), Brighton, UK November 2019

Gifford R. M. Woods D. R., Reynolds R. M., Female resilience to extreme climactic exposure and arduous military training: mitigating risks to optimise performance. Military Health System Research Symposium, Orlando, FL, USA, August 2019

Gifford R. M., O'Leary T. J, Greeves J. P. Reynolds R. M., Woods D. Endocrine insights from Expedition ICE MAIDEN. Association of Service Physicians, MJ World

Research Prize (winner), Camberley, UK Nov 2018 *and* Royal Society of Medicine Colt Foundation Prize (second place), London, UK December 2018

Gifford R. M., O'Leary, T Greeves J, Reynolds R, Woods D; Resilient reproductive, bone and adrenal function in Expedition ICE MAIDEN, the first all-female, unassisted Antarctic crossing. Society for Endocrinology BES, November 2018

Gifford R. M., O'Leary T. J, Greeves J. P. Reynolds R. M., Woods D. R., The endocrine effects of arduous military training on women: evidence from Antarctica and Sandhurst. Medical Innovation, Birmingham, UK, October 2018

Gifford R.M., Todisco T., Stacey M., Fujisawa T., Allerhand M., Reynolds R.M., Woods D.R., Risk of heat illness in men and women: a systematic review and meta-analysis. Association of Service Physicians, MJ World Prize (winner), York, UK, November 2017

Poster presentations

Gifford R, Wardle S, O'Leary T, Greeves J, Reynolds R, Woods D, SAT-218 Hypothalamic-Pituitary-Gonadal (HPG) and Hypothalamic-Pituitary-Adrenal (HPA) Axis Responsiveness in Women and Men during 29 Weeks of Basic Military Training, ENDO, New Orleans, USA, March 2019

Gifford R, Woods DR, Reynolds RM; Female military responses to heat, cold and basic military training. Association of Physicians of Great Britain Annual Meeting, Glasgow, UK, April 2019

Gifford R, Woods DR, Reynolds RM; Sex-associated rates of Heat Illness; Planetary Health International Congress, Edinburgh, UK, May 2018

Gifford R, Reynolds R Anderson R Woods D; Is serial urinary progesterone measured via automated chemiluminescent assay a valid alternative to pregnanediol via manual ELISA for the detection of ovulation? Society for Endocrinology BES, Harrogate, November 2017

Gifford RM, Greeves J, Anderson R, Woods DR, Reynolds RM. Characterising key endocrine effects of arduous training on women: a concurrent four-domain approach. Military Health System Research Symposium, Orlando, FL, USA, Aug 2017

List of Abbreviations

ACTH	Adrenocorticotrophic hormone
AMH	Anti-Müllerian hormone
ANOVA	Analysis of variance
AUC	Area under the Curve
BAI	Beck Anxiety Inventory
BDI	Beck Depression Inventory
BEDA-Q	Brief eating disorders in athletes questionnaire
BMD	Bone mineral density
BMI	Body mass index
BSAP	Bone specific isoform of alkaline phosphatase (bone formation marker)
CC	Commissioning Course (suffixed by year of entry and course number, e.g. CC172 was the second Commissioning Course to start in 2017)
COCP	Combined contraceptive pill
CDRISC-10	Connor Davidson Resilience Scale, 10 question version
CHOC	Combined Hypothalamic Ovarian adrenal Cortex test
CMIA	Chemiluminescent microparticle immunoassay
Course	The Commissioning Course, Royal Military Academy, Sandhurst, UK
CTX	C-telopeptide cross-links of type 1 collagen (bone resorption marker)
CV	Coefficient of variation
DE	Disordered eating
DMS	Defence Medical Services
DWHRP	Defense Women's Health Research Program (USA)
DXA	Dual-energy x-ray absorptiometry
E1G	Estrone-3-glucuronide
E2	Oestradiol
E4T	Exploring the Endocrine Effects of Extreme Training
EA	Energy availability
EA_{mvpa}	Energy availability from moderate and vigorous physical activity only
EA_{tpa}	Energy availability from total physical activity
EB	Energy balance
ECG	Electrocardiograph
ECLIA	Electrichemiluminescence immunoassay
ED	Emergency department
EDI	Eating disorders inventory
EDTA	Ethylenediaminetetraacetic acid

EEE	Exercise energy expenditure
EEE_{mvpa}	Exercise energy expenditure from moderate and vigorous physical activity only
EEE_{tpa}	Exercise energy expenditure from total physical activity
EHI	Exertional heat illness
EI	Energy intake
ELISA	Enzyme-linked immunosorbent assay
Ex	Expedition
FAI	Free androgen index
FEAT	Female Endocrinology in Arduous Training
FFM	Fat-free mass
FM	Fat mass
FSH	Follicle-stimulating hormone
FTE	Full time equivalent
GCC	Ground close combat
GnRH	Gonadotrophin releasing hormone
HA	Hypothalamic amenorrhoea
HCC	Hair cortisol concentration
HI	Heat illness
HOMA 2 (-IR)	Homeostatic modelling assessment 2 (of insulin resistance)
HPA	Hypothalamic-pituitary-adrenal
HPG	Hypothalamic-pituitary-gonad
HPO	Hypothalamic-pituitary-ovarian
HRpQCT	High resolution peripheral quantitative computerised tomography
HRV	Heartrate variability
HS	Heat stroke
ICD9(-CM)	International classification of diseases 9 (clinical modification)
ICD10-(CM)	International classification of diseases 10 (clinical modification)
IGF-1	Insulin-like growth factor 1
IOC	International Olympic Committee
IQR	Interquartile range
IRMS	Isotope ratio mass spectrometer
IRR	Incident rate ratio
IUS	Intrauterine system (eluting low dose levonorgestel; either Jaydess® or Mirena®)
KNDy	Kisspeptin-neurokinin B-dynorphin
LARC	Long acting reversible contraceptive
LC	Local classification
LH	Luteinising hormone
LH:HF	Ratio of high and low frequency heartrate variability

LnHF	Fast-Fourier transformed logarithms of high-frequency
LnLF	Fast-Fourier transformed logarithms of low-frequency
LPD	Luteal phase defect
MVPA	Moderate and vigorous physical activity
NEFA	Non-esterified fatty acid
Nil/ IUS	Non contraceptive user and intrauterine system user group
NK	Unknown
NPV	Negative predictive value
NR	Not reported
NS or ns	Not statistically significant; generally p>0.05
OHD3	25, hydroxyl vitamin D3
P1NP	Amino-terminal propeptide of procollagen type 1 (bone formation marker)
P4	Progesterone
PDG	Pregnanediol 3 glucuronide
Phase	10-day energy availability assessment phase
PHQ-9	Patient Health Questionnaire 9 (depression scale)
pNN50	Percentage of successive normal R-R intervals greater than 50 ms
PNS	Parasympathetic nervous system
POMS	Profile of Mood States
PPV	Positive predictive value
Prog	Progestogen-only containing contraceptive
PTSD	Posttraumatic stress disorder
RCDM	Royal Centre of Defence Medicine
RED S	Relative Energy Deficiency in Sports
REE	Resting energy expenditure
RM ANOVA	Repeated measures analysis of variance
RMAS	Royal Military Academy, Sandhurst
RMR	Resting metabolic rate
RMSSD	Root mean square of successive differences
ROC	Receiver-Operator Characteristics
RR	Relative risk
R-R	Duration between two R waves in an electrocardiograph
SD	Standard deviation
SD1	Standard deviation of short term heartrate variability
SD2	Standard deviation of long-term heartrate variability
SHBG	Sex hormone binding globulin
SNS	Sympathetic nervous system
SPSS	Statistics Package for the Social Sciences (IBM, New York, NY, USA)

T3	Triiodothyronine
T4	Thyroxine
TEE	Total energy expenditure
TMB	3,3',5,5'-Tetramethylbenzidine
TPA	Total physical activity
TRIS	Trisaminomethane
TSH	Thyroid stimulating hormone
TSST	Trier psychosocial stress test
TVUS	Transvaginal ultrasound
VO_{2max}	Maximum oxygen uptake
VO_{2peak}	Peak oxygen uptake (during submaximal exercise)
vSLAP	Vienna-Standard Light Antarctic Precipitate
vSMOW	Vienna-Standard Mean Ocean Water
WGCC	Women in ground close combat
ηp^2	Partial eta squared
χ^2	Chi squared

List of figures

List of tables

List of Appendices

Contents

Chapter 1 Introduction

1.1 Context: Women in Ground Close Combat

In July 2016, the Chief of the General Staff, with the support of the other Service Chiefs, recommended to the Secretary of State that women be allowed to serve in infantry and armoured units for the first time in the UK (Ministry of Defence, 2016). Women were subsequently allowed to serve in the Royal Armoured Corps in November 2016, in the Royal Air Force Regiment in September 2017, and in all infantry units from October 2018.

1.1.1 The WGCC Interim Health Report, 2016

The recommendation to the Secretary of State was based on an Interim Health Report (IHR) on Service women serving in the military published by the Women in Ground Close Combat (WGCC) Review (Department of Manning (Army), 2016). The IHR presented recommendations based on three datasets:

1. Routine training injury databases, collected from the Royal Marines (2014-2015), RAF Regiment (2011-2015) and Army Recruiting and Training Division (2011-2015).
2. Defence Statistics (Health) created a study cohort of regular Ground Close Combat (GCC) personnel from three services (men only) plus Royal Artillery and Royal Engineers (mixed sex; Combat Support Arms, Army) and Royal Logistics Corps (mixed sex; Combat Service Support, Army; treated as controls). Military healthcare records were interrogated for this cohort.
3. A survey of personnel from GCC, Royal Artillery, Royal Engineers and Royal Logistics Support, plus Combat Service Support personnel attached to these units, conducted from February – March 2016. A total of 1200 respondents were included of whom 152 (12.7%) were women.

While of central strategic importance to UK research on WGCC, the IHR combined retrospective and cross-sectional data, and carried the usual limitations of such studies. It was likely to have been hampered by incomplete reporting or inaccurate coding on databases (Joint Personnel Administration system or defence primary care database), reporting bias was likely among survey respondents and selection bias may also have been problematic. The proportion of women surveyed was only 53% of the sample size required based on a sample size power calculation. Details

of the survey were not provided and it is not stated how respondents were identified or approached. The IHR was published in July 2016 as an early submission from a five-year research programme commissioned by the Secretary of State. Given the political drive towards a military 'based on ability, not gender' (Ross, 2015) and trends set by militaries from other nations (BBC News, 2016), there was likely an imperative placed on expedient publication of this report, conceivably inducing significant time pressure on its authors. Its data collection was completed in just three months and it is conceivable that the findings were subject to a degree of 'publication bias'. Nevertheless, the IHR struck a circumspect tone and identified three important areas of risk, all of which are relevant to this thesis. These are laid out below in the context of extant literature:

1.1.1.1 Musculoskeletal injury

Musculoskeletal injuries (MSKI) is an umbrella term covering a wide variety of pathology. The IHR categorised MSKI by anatomic location not pathology, other than for stress fractures (fatigue-induced microtraumas). A pooled incidence or incident rate ratio was not given; instead, the IHR contained 97 different rates of MSKI and 41 reports of female: male relative risk (RR). For example, during Army training, male and female rates of overuse injury (for standard Army entrants were 116 and 220 per 1000 trainees during 2011 to 2015, respectively (RR 1.90; 95% CI 1.73, 2.09), while for Officer Cadets these were 780 and 1,112, respectively (RR not given as females demonstrated >1 injury per trainee). The label of 'overuse injury' could apply to a wide array of conditions, among them stress fractures. Rates of all stress fractures (per 1000 trainees during 2011 to 2015) were consistently higher among female than male trainees in the Army ('stress fracture': 4.1 to 7.4 versus 17.8 to 23.8, respectively, $p<0.001$), and in the RAF ('pelvic stress injury': 0.67 versus 32.54, respectively, $p<0.001$.

The IHR consistently identified the greatest sex disparity in MSKI at a hip or pelvis location (higher rates among women), which is likely caused by higher stress fracture rates among women (this interpretation is supported by a report undertaken during the previous US Defence Women's Health Research Program, which identified highest sex disparities in stress factures was at the hip (Friedl, 2005)). A cross-sectional study of over 650,000 US basic military trainees identified women to be at four-fold increased risk of stress fracture than men (Knapik et al., 2012), and a meta-analysis found this sex disparity to be most pronounced at the hip in military

trainees (Wentz et al., 2011). In the IHR data, there appeared to be an inverse relationship between RR and absolute injury rates. For example, female: male RR ranged from 2.95 (95% CI 2.25, 3.86) for all hip injuries in Army Officer Cadets, to RR 48.40 (95% CI 19.3, 121.4) for pelvic stress injuries in RAF trainees. The absolute stress fracture rates for Army Officer Cadets and RAF trainees were 170.5 per 1000 women and 32.5 per 1000 women, respectively.

Accounting for ambiguity introduced by the methods of data capture for the IHR, women appeared to be at consistently greater risk of MSKI than men, markedly more so at the hip and pelvis, due to a higher rate of hip stress fracture. A second important observation was made: men undergoing GCC training experienced higher rates of MSKI, especially stress fractures, than men undergoing standard Army entry training (rates per 1000 trainees from 2011–2015: 10.78 and 1.43, respectively, $p<0.001$; RR 7.56 (95% CI 4.5, 12.69)). The IHR indicated that the risk of stress fracture might be therefore be multiplied to women entering GCC, although the apparent inverse association between absolute injury rates and the female: male RR does not appear to have been investigated.

The databases and survey consistently identified MSKI as the leading cause of medical discharge (being forced to leave a Service due to a medical problem) and medical downgrade (having military duties permanently or temporarily limited due a medical problem) among both men and women in the ‘trained strength’ (all Service Personnel who have completed training). Discharge and downgrade became more common as duration of service increased. The overall rates of medical discharge for women and men, per 1000 trained Service Personnel over two years, were Royal Navy: 75.5 and 51.5, respectively, $p<0.001$ (RR 1.47 (95% CI 1.29, 1.67)), Army: 82.0 and 75.5, respectively $p>0.05$ (RR 1.09 (95% CI 1.01, 1.17) and Royal Air Force: 36.5 and 20.0, respectively, $p<0.05$ (1.83 (95% CI 1.56, 2.14)). The IHR did not give Service-wide rates of medical downgrade, but in a cross-sectional data capture, medical downgrade due to MSKI did not differ between men and women in the Royal Engineers, Royal Artillery or Royal Logistics Corps.

1.1.1.2 Mental health

The IHR cited mental health diagnoses as the second most common reason for medical downgrade, after MSKI. However, women were more likely to be downgraded due to mental health disorders than men. Cross-sectional data from the

Defence Statistics (Health) cohort indicate that, on 1 November 2015, the proportion of Service men and women diagnosed with a behavioural or mental disorder was 2.6% and 5.9%, respectively. Rates of mental health disorders are presented as cross sectional data (on 1 November 2015), and thus will represent prevalence rather than incidence. In the trained strength Royal Logistics Corps and Royal Artillery, more women were downgraded due to mental health disorders than men, although the absolute rates were low (Royal Logistics Corps: 30.1 versus 13.8, $p<0.001$ (RR 2.18 (95% CI 1.54, 3.8)), Royal Artillery: 32.3 versus 14.2 (RR 2.33 (95% CI 1.37, 3.95))). Rates of discharge for mental health reasons were not given.

Rates of mental health diagnoses differed according to occupation: less mental ill health was captured among Royal Marines (men only at that time) than men from other corps, e.g. Royal Logistic Corps (5.4 per 1000 men versus 13.8 per 1000 men, $p<0.001$, RR 0.40 (95% CI 0.27, 0.57). It was unknown if combat exposure increased the relative risk to women of anxiety and depression, although Royal Marines could be expected to have greater exposure to combat than Logisticians, which may account for these differences.

Since the IHR, the number of new mental diagnoses in 2017/2018 has remained stable at 3.1% (the 2015/2016 incidence was 3.2%), however the gender gap appeared wider than reported in the IHR (6.2% for women and 2.7% for men in 2017/2018; **Figure 1-1**) (Defence Statistics (Health), 2018).

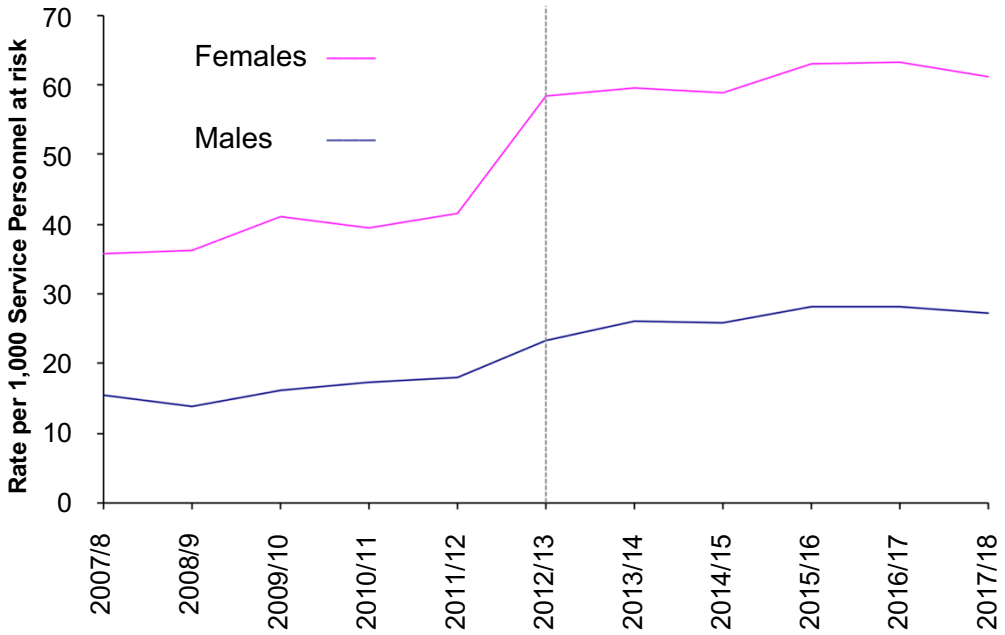


Figure 1-1 Service Personnel assessed to have a mental health disorder at Defence Mental Health Care Services according to gender, adapted from Defence Statistics (Defence Statistics (Health), 2018) and Interim Health Report (Department of Manning (Army), 2016). Dotted line represents revised diagnostic methodology in 2012/2013. Percentages are based on the absolute number at risk. Original data © Crown Copyright

Of particular relevance to this thesis, the prevalence of post-traumatic stress disorder (PTSD) was uncommon across the UK Armed Forces in 2017/2018 (prevalence 0.2%). It is not clear how PTSD was assessed, although these data refer to diagnoses made by Defence Community Mental Health community psychiatric nurses or doctors. The male: female ratio of PTSD diagnoses was not reported.

A range of gender discrepancy has been reported in the incidence of PTSD. A systematic review and meta-analysis by Tolin and Foa (2006) examined the male-to-female odds ratio (OR) of PTSD, making 52 comparisons from 40 non-overlapping studies reporting several different types of traumatic event. An OR of 1.98 (95% CI 1.76, 2.22; $p<0.001$), indicated more PTSD occurred in women than men, after adjustment for duration since the traumatic event, type of traumatic event and age. The authors conducted two further meta-analyses. Pooling 22 outcomes from 19 studies, Tolin and Foa found that women were less likely to experience a

traumatic event than men (OR 0.77 (95% CI 0.65, 0.91; p=0.002)), although the types of traumatic event experienced differed. Data from 64 studies, containing 482 independent comparisons, found men were more likely to experience accidents (OR 0.67, (95% CI 0.57, 0.79; p<0.001)), nonsexual assault (OR 0.62, (95% CI 0.56, 0.69; p<0.001)), combat war or terrorism (OR 0.28, (95% CI 0.18, 0.43; p<.001)), disaster or fire (OR 0.80 (95% CI 0.72, 0.88; p<0.001)), witnessing death or injury (OR 0.80 (95% CI 0.72, 0.90; p<0.001)) or illness or unspecified injury (OR 0.68 (0.68 (95% CI .54, 0.86; p=0.001)). On the other hand, women were more likely to experience adult sexual assault (OR 5.99 (95% CI 4.42, 8.93; p<0.001)) or child sexual abuse (OR 2.66 (95% CI 2.05, 3.44; p<0.001)).

Although these data come predominantly from US civilian data (just four military studies were included, all from USA), they are salient to the risk of PTSD among British women. Whether these observed differences pertain to sex or gender is unclear. Some have noted dimorphic biological sex-associated differences in the hypothalamic-pituitary-adrenal (HPA) axis which putatively predispose women to PTSD. Compared with men, women exhibit lower basal cortisol levels, slower glucocorticoid central negative feedback and basal and stimulated greater corticotrophin releasing hormone (CRH) concentrations (reviewed by Lehrner et al., (2016)). Others have suggested higher rates of PTSD among women are caused by the constraints of gender normative behaviours within society and chronic environmental strain of societal femininity expectations, rather than biological sex differences (Street and Dardis, 2018). Indeed, one study addressed gender differences in stressor exposure closer to what might be expected from WGCC, using in a nationally representative sample of US Iraq and Afghanistan veterans (Vogt et al., 2011). The authors found no clinically significant interactions of gender × stressor (all Cohen's d effect sizes were close to zero).

The influence of sex or gender on mental health outcomes like PTSD is currently moot, as was evident in a recent international roundtable discussion by military experts of resilience for military readiness (Nindl et al., 2018). Resilience is a counterpoint to stress and mental health disorders. Two contributors presented data on the relative psychological (and biological) resilience of men and women, which were diametrically opposed; a consensus could not be reached.

1.1.1.3 Reproductive health

Training and employment in GCC is necessarily arduous, but whether 'arduousness' disrupts reproductive endocrine function, as is seen in athletes during intensive training, is unknown. Rates of infertility among Service women had not been explored until an audit was conducted of all regular, trained Service women aged 15 to 49 years, registered on the Defence primary healthcare system between 2013 and 2015, for inclusion in the IHR. Data before mid-2012 were excluded due to a change in the registration system. Age-stratified data were extracted using Read codes, selected to match a nationwide contemporaneous study of the occurrence of fertility problems among civilians, undertaken by Dhalwani et al., (2013). Only first data entries of fertility-related investigations, interventions, drug prescriptions, referrals or diagnoses were included, to obtain a period prevalence. Prescriptions for infertility medications were not included (they are not on the tri-Service Formulary) and other Read codes of potential relevance were excluded if they did not appear in the study by Dhalwani et al.

The findings of this audit are presented next to data from the study of Dhalwani et al., (2013) in **Table 1-1**. Where 5 cases or fewer were reported, actual numbers were suppressed (in accordance with Defence Statistics regulations), although rates were still reported.

Age (years)	Occurrence of fertility problems (n)	Service personnel (n)	Rate of occurrence of fertility problems among Service women, per 1000 person-years	Comparative rate of fertility problems among civilians, per 1000 person-years, (95% confidence interval)
15-19	~	~	1.23	0.7 (0.7 to 0.8)
20-24	31	7,683	4.03	4.6 (4.5 to 4.7)
25-29	67	11,577	5.79	9.4 (9.3 to 9.6)
30-34	163	10,744	15.17	10.9 (10.8 to 11.2)
35-39	155	6,939	22.34	7.0 (6.9 to 7.1)
40-44	46	4,119	11.17	2.4 (2.3 to 2.5)
45-49	6	1,486	4.04	0.4 (0.3 to 0.4)

Table 1-1 Rate (per 1000 person years) of reported fertility problems in Service women compared with civilian data adapted from Dhalwani et al., (2013) and Department of Manning (Army) (2016). Data source: Audit of UK Defence Medical Information Capability Programme for Read code criteria used in a large, contemporaneous study of infertility rates across the UK (Dhalwani et al., 2013). Data captured for calendar years 2013 to 2015. Occurrences of fertility problems were only counted on the first occasion they were coded for during the reporting period. ~ data suppressed as ≤5 cases.

The rates of presentation for fertility problems among military women were apparently higher than civilians in Dhalwani et al., (2013). However, this finding must be caveated by considering problems inherent to audits. Inaccuracy was likely to be introduced from known inconsistencies in Defence's use of Read codes (Cox et al., 2016). Causality was not assessed, nor were important confounders such as separation of partners (e.g. through Service requirement), chlamydial infection, or policy variations in direct availability of civilian assisted conception services to military personnel in different NHS regions. The IHR suggested that limiting Read codes to those used by Dhalwani et al. might have accounted for a 25% underreporting of infertility. However, the denominator population (taken as the female trained population in June of each year) was only identified through the primary care record system and not corroborated, e.g. by Defence Statistics (Health). If Service women for whom Read codes were reported were not registered on the Defence primary healthcare system at this time, the background population

size could have been underestimated, leading to an overestimate of the occurrence rate.

Overall, these data can only provide a suggestion that infertility may be more common among military women than civilians. However, as will be explored in **sections 1.2 and 1.3,** extant data from other areas make this suggestion highly plausible.

1.1.1.4 Environment

Environmental considerations are also relevant to WGCC. Recent conflicts have taken place in extreme heat (e.g. Southern Iraq, Helmand Province, Afghanistan) and cold (e.g. mountainous areas of Afghanistan) and important sex differences exist in physiological adaptation to extremes of temperature. For example, women are known to vasodilate and sweat less during exercise in heat and humidity due to the effects of oestrogens (Gagnon and Kenny, 2012a; Wong and Hollowed, 2017), while in extreme cold, women shiver less and demonstrate lower average core temperatures than men (Graham, 1988; Wagner and Horvath, 1985). Heat illness and cold injury (especially non-freezing cold injury) represent major health risks for GCC roles (Kuht et al., 2018; Stacey et al., 2015), although they were not discussed in the IHR. While specific risk factors for heat illness remain poorly understood (Moore et al., 2016), female sex could be hypothesised to place women at greater risk of heat illness (Gagnon and Kenny, 2012a). Reassuringly, a recent systematic review found the rate of cold injuries in women to be reduced compared with men by as much as 10-fold, although the sex ratio of background populations was not considered (Heil et al., 2016).

1.1.2 Hormones as the effectors of sex

Hormones are the messengers of biological sex. Where sex differences exist in bone function and MSKI, mental health, reproductive function and thermoregulation, these are influenced, or indeed communicated, by hormones. Hence, appreciating hormonal adaptations to military employment is critical to characterising military employment-related pathology among women. However, there is little evidence from which to quantify the risks of such pathology, because WGCC is such a recent phenomenon. It is conceivable that the pioneers of WGCC may be placing themselves at increased risk of harm and doing so unwittingly. There is therefore a pressing need to develop an evidence base from which to understand sex-

associated adaptations to arduous military training on which to base employment decisions, both for individuals and employers. Understanding adaptations which could contribute to adverse outcomes will enable risk factors of such outcomes to be characterised, allowing individual women to make informed decisions for their careers and policymakers and innovators to mitigate that risk.

1.2 Extant data from Servicewomen

1.2.1 Prevalence of reproductive dysfunction in Service women

Since women have entered GCC roles, and there is a paucity of research examining arduous training-related outcomes in women, generally focussed on reproductive function and its sequelae in basic military training. Only a few studies describe the prevalence of reproductive dysfunction associated with military training, reporting a wide range of results. The earliest, by Colonel Anderson (1979), surveyed all female cadets commencing the United States Military Academy (USMA), West Point, NY in 1976 and 1977 to identify secondary amenorrhoea. Of his 158 respondents (in total across two intakes), those reporting amenorrhoea after 1, 6 and 12 months of training were: 117 (74%), 69 (25%) and 21 (13%), respectively (**Table 1-2**). It could be assumed that “amenorrhoea after 1 month” meant a missed first period (amenorrhoea is normally defined as 3 months or more without a period). The apparent resolution in menstrual regularity was associated with lighter periods and less cramping. Ten years later at the same institution, Colonel Welch (1989) undertook a similar study, broadening the outcome to any menstrual irregularity (including secondary amenorrhoea) and excluding those with previous menstrual irregularity or those using hormonal contraceptives. Of 110 respondents, 77 (70%) reported menstrual irregularity during the first year, of whom 20 (26%) and 60 (78%) reported resumption of normal menses during and after the first year, respectively. These studies did not define menstrual irregularity or specify a cut-off duration for the absence of menses to define amenorrhoea. Both ascribed menstrual disturbance to psychological and exercise stressors rather than energy deficit, undereating or low body fat (Anderson reported the average body fat among female cadets to be 19%). They hypothesised a habituation process: the prevalence of menstrual disturbance decreased as participants became accustomed to the psychological and physical stress of training. Alternatively, stressors could have subsided over time (the ‘shock of capture’ effect).

Study	Design	Setting, participants	Qualitative outcomes	Hormonal contraceptive usage	Prevalence reported
Anderson (1979)	Long-term recall	USMA, freshmen commencing 1977, n=88, age NR	'Secondary amenorrhoea' (duration ND)	NR	Amenorrhoea 75% at 1 month * 45% at 4 months 8% at 12 months
Anderson (1979)	Long- term recall	USMA, freshmen commencing 1976, n=70, age NR	'Secondary amenorrhoea' (duration ND)	NR	Amenorrhoea 73% at 1 month, * 41% at 6 months, 29% at 9 months, 20% at 12 months
Welch (1989)	Long term recall	USMA, sophomore n=45, and freshmen n = 65, age NR. Excluded if prior menstrual irregularity	'Menstrual irregularity, secondary amenorrhoea'(duration ND)	NR (excluded)	Menstrual irregularity: 68% (1990), 72% (1991) Resumption of normal menses: 78% after first year (1990), 26% during first year (1991)
Friedl et al., (1992)	Mail shot survey (focussed on self-reported stress fracture)	US army soldiers, Fort Lewis, Washington, n=1630, median age 24 (IQR 18-52) y	Amenorrhoea – no menses 6 months in the absence of pregnancy	34.9%	Amenorrhoea 14.9%

Lauder (1997)	Questionnaire	USMA reserve officer cadets, n=310, age 21.5 ±1.9 y	'Occasionally skipped periods', 'only a few times per year'	NR	'Occasionally skipping periods' 12%, 'Only a few times per year' 5%
Schneider et al., (1999)	Long term recall questionnaire	USMA, n=158, age 18.4 ±0.8 y	Irregularity 'increased or decreased'	NR (excluded)	Irregularity: 4.2% decreased, 10.1% increased
Lauder et al., (1999)	Interview	USMA, active duty cadets, n=423, age 27.5 ±7.7 y	Oligomenorrhoea - <9 menses in 12 months; amenorrhoea - ≥3 previous cycle equivalents	25% of 34% who met eating disorder criteria	2.1% amenorrhoea, 3.3% oligomenorrhoea
Schneider et al., (2003)	Short term recall by email	USMA, n=116 freshmen, mean age 18.4 ±0.8 y	Menstrual irregularity – percentage of cycles outside 21-45 days (moderate ≤25%, severe 26-50%, extremely severe >50%; mild: >7 day variation in cycle)	NR (excluded)	Regular: 1.7%; irregular - mildly 10.3%, moderately 35.3%, severely 30.2%, extremely 22.4%
Cho et al., (2017)	Written questionnaire after first 8 weeks of training	Korea Third Military Academy, normally menstruating cadets n=40, age range 22 to 28 y	Menstrual irregularity – < 21 days, > 35 days or ≥15 day variability between cycles Amenorrhoea – no menses	29% (excluded)	Menstrual irregularity 60%, amenorrhoea 10%
Pires (2018)	Questionnaires after 4 and 8 weeks	Portuguese Military Academy, Lisbon, cadets entering 1992–2016 n=64, mean (range) age 18.8 (17-25) y	Menstrual dysfunction ND Amenorrhoea on average 5 months	37.3% (excluded)	Menstrual dysfunction 43.8% Amenorrhoea 35.9%

Table 1-2. Comparison of studies assessing menstrual disturbance in military trainees (Adapted and updated from Gifford et al., (2017)) Ages are mean ± standard deviation or unless otherwise stated. * 'Amenorrhoea at 1 month' implies missed first period. NR : not reported; ND: not defined, USMA: United States Military Academy, West Point, New York; IQR, interquartile range; SD, standard deviation.

Subsequent studies used more precise definitions of menstrual disturbance to determine the true prevalence. In a mail-shot survey of serving soldiers in Fort Lewis, Washington DC, Friedl et al., (1992) defined amenorrhoea as 6 months or longer without menses, and identified a prevalence of 14.9% (1,630 respondents, of whom 34.9% were excluded due to hormonal contraceptive usage). Lauder et al., (1997) studied all female cadets attending a reserve officer training camp (n=310, age 21.5 ±1.9 years), using a questionnaire to identify 'occasionally skipped periods' or 'only a few periods per year'. Rates reported were 12% and 5%, respectively (Lauder, 1997). The same authors interviewed full-time USMA cadets were interviewed in detail (n=423, aged 27.5 ±7.7 years), reporting oligomenorrhoea (≤9 menses in 12 months) and amenorrhoea (missed ≥3 consecutive cycles) rates of just 2.1% and 3.3%, respectively (Lauder et al., 1999). Two studies using email surveys at the USMA found significantly higher rates. Using long-term recall questionnaires (n=158, aged 18.4 ±0.81 years) 48.2% of cadets reported decreased menstrual frequency during the past 12 months of training (Schneider et al., 1999). The authors also undertook a study of short-term recall (n=116, aged 18.4 ±7.7 years), in which menstrual disturbance was defined as cycles outside 21-45 days (Schneider et al., 2003). The rates of moderate (≤25%), severe (26 to 50%) and extremely severe (>50%) menstrual disturbance were 35.3%, 30.2% and 22.4%, respectively. More recent studies from South Korea and Portugal suggest even higher rates of amenorrhoea and/or menstrual disturbance in these physically demanding basic training courses, of 70% and 80% (n=40 and n=64), respectively (Cho et al., 2017; Pirez, 2018).

Such differences in reported rates of menstrual disturbance are likely due in part to study designs. The approach of Schnieder et al., (2003), using long term recall via email, was intended to reduce recall bias, but this seems unlikely given the high rates they observed. Similarly, a response to a letter or email mailshot would be made more likely if a participant had a positive finding to report, which might explain the high rate identified by Friedl et al., (1992). The more conservative findings of Lauder et al., (1999) might be more reliable, since these were obtained from interviews during training.

Given the paucity of evidence of the impact of military training on reproductive function in women, meaningful insight may be gained from data in civilian athletes. Military training involves intense physical activity with high pressure to perform, and

infantry have been described as ‘tactical athletes’ (Department of Manning (Army), 2016). A systematic review of reproductive dysfunction and associated pathology in athletes by Gibbs et al*.,* (2013) found the prevalence of secondary amenorrhoea ranged from 1.0 % to 6.0 % (34 studies, n = 5607), while oligomenorrhoea ranged from 0.9 % to 52.5 % (23 studies, n = 4044). Such a wide variation in the prevalence of menstrual disturbance could be accounted for selection and reporting bias, as well as differences in training exposure. However, the prevalence of menstrual disturbance reported among athletes was probably higher than in the general population, where estimates of the prevalence of amenorrhoea range from 2% to 5% (Barrack et al., 2013; Torstveit and Sundgot-Borgen, 2005). The definitions of menstrual disturbance used by Gibbs et al*.,* (2013), and widely accepted elsewhere, are shown in **Table 1-3**.

Eumenorrhoea	Cycle length 26-35 days
Oligomenorrhoea	Cycle length 36-89 days
Amenorrhoea	Cycle length ≥90 days
Ovulatory disturbance, LPD	Short or absent luteal phase (day 14 onwards), usually within a normal-length cycle

Table 1-3 Definitions used in menstrual function. LPD, luteal phase defect

The relationship between the spectrum of reproductive function and menstrual disturbance discussed in this thesis is illustrated in **Figure 1-2.** The athletic literature commonly refers to a shortened latter half of the cycle (luteal phase), or luteal phase defect (LPD). Disturbed cycle function in this way is referred to as ‘ovulatory disturbance’ and the term generally implies the cycle length is normal (Petit and Prior, 2013). Gibbs et al., (2013) presented a few studies which assessed luteal phase defects and anovulation; the prevalence of luteal phase defects ranged from 5.9% to 43.0% (4 studies, n = 118) and anovulation 5.9% to 43.0% (4 studies, n = 101).

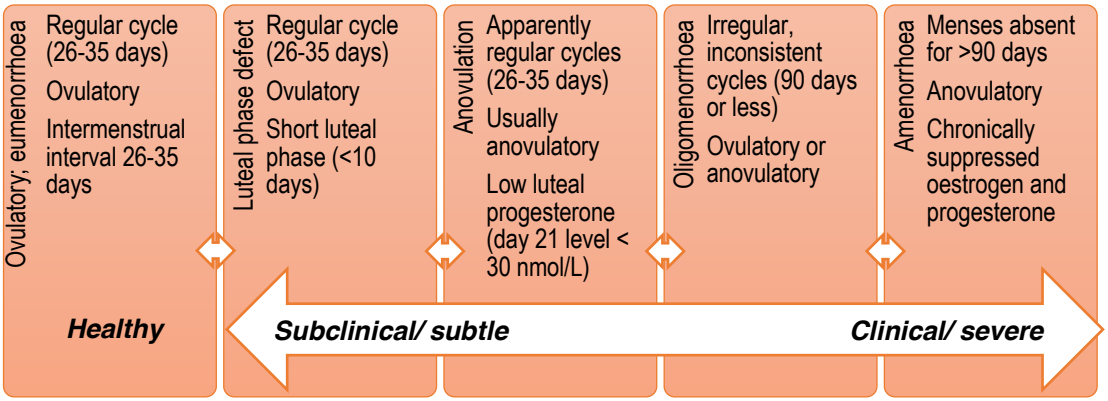


Figure 1-2 Spectrum of reproductive function associated with hypothalamic amenorrhoea, adapted from Gifford et al., (2017)

1.2.2 HPG axis adaptations to exercise and stress

Recent clinical guidance published by the Endocrine Society, written largely by members of the American Society for Reproductive Medicine, acknowledge the association between participation in sports and exercise and hypothalamic amenorrhoea (HA) (Gordon et al., 2017). Other relevant factors which may disrupt the hypothalamic-pituitary-gonadal (HPG) axis include psychological stressors, relative nutritional deficit (low energy availability) and androgen excess, which bring about HA by disrupting gonadotrophic hormone releasing hormone (GnRH) pulsatility at the hypothalamus. Such physiological disruption may reflect disturbances of ovarian, pituitary, hypothalamic and/ or higher cortical function.

The HPG axis encompasses a complex interaction between GnRH, luteinising hormone (LH) and follicle stimulating hormone (FSH), and ovarian oestradiol and progesterone. Kisspeptin, the ligand for G protein-coupled receptor GPR54, is central in the regulation of GnRH release. Released by neurons in the arcuate nucleus together with neurokinin B and dynorphin (hence KNDy neurons), kisspeptin is central to initiating GnRH oscillations for the onset and progression of puberty (Avendano et al., 2017). Throughout reproductive life, GnRH neurons in the median eminence of the inferior hypothalamus are sensitive to downregulation by low levels of kisspeptin, manifesting in reduced LH pulsatility and amenorrhoea (Jayasena et al., 2014). KNDy neurons are receptive to numerous stimuli including leptin and ghrelin, and these hormones appear to be key mediators of HA in several pertinent conditions such as anorexia nervosa, exercise and stress (Hofmann et al., 2017; Scheid et al., 2013; Stieg et al., 2015).

Hormonal profiles of women undertaking exercise emphasising leanness, such as long distance running, cycling or ballet, are often characterised by hypo-oestrogenism and suppression of the normal GnRH pulsatile release (Warren and Perlroth, 2001). As GnRH stimulates pituitary FSH and LH release a suppression of GnRH leads to an attenuation of FSH and particularly LH. This subsequently attenuates production of progesterone and to a lesser extent oestradiol . The final effect of this disturbance is to prolong the follicular phase leading to ineffective, shortened or absent luteal phases (Hakimi and Cameron, 2017).

In contrast to lean sports, in women engaging in non-lean sports, such as contact sports, rowing and handball, the LH:FSH ratio tends to be increased and hyperandrogenism may be observed with high or normal oestrogen (**Figure 1-3**) (Dulac et al., 1986; Frisch et al., 1984). These athletes are also prone to a distinct phenotype of menstrual disturbance, possibly due to high levels of unopposed oestradiol and hyperandrogenism blunting the LH surge (Javed et al., 2015). Gibbs et al., (2013) reviewed studies comparing menstrual disturbance in lean and non-lean training. Overall these identified a higher prevalence of menstrual disturbance in lean than in non-lean sports, (0.7-27.7%, versus 0%-16.5%, 5 studies, n = 1032, no p value given), although this comparison should be treated with caution, since a meta-analysis was not conducted and heterogeneity was not assessed. It is not clear if hyperandrogenism occurs as a consequence of the training or is a result of self-selection; fewer data are available for non-lean than lean sports and conclusions about the relative effects of training type on reproductive function are tentative.

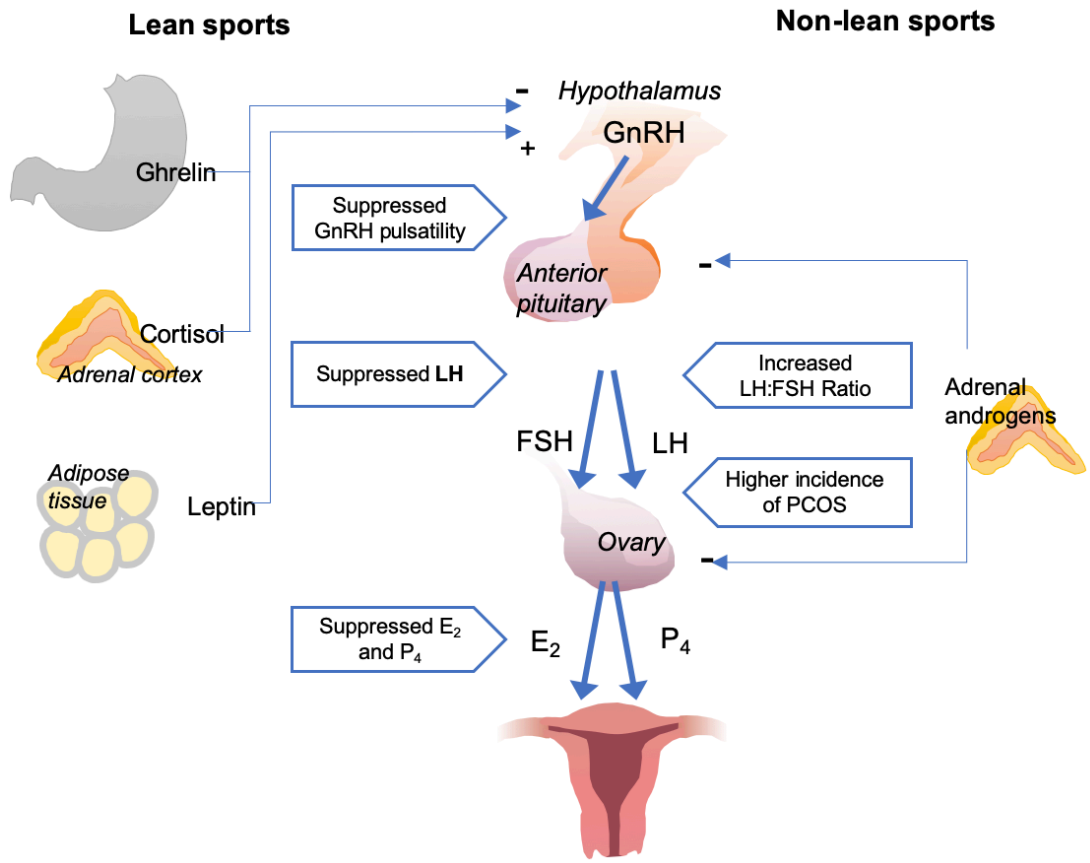


Figure 1-3 Putative causes of hypothalamic amenorrhoea in lean and non-lean sports, adapted from Gifford et al., (2017) *GnRH gonadotrophin releasing hormone, LH luteinising hormone, FSH follicle stimulating hormone, E2 oestradiol, P4, Progesterone, PCOS polycystic ovary syndrome, + stimulatory, – inhibitory.*

1.2.2.1 Competing hypotheses: body composition, stress and energy availability

Three distinct hypotheses have emerged to explain HA observed in athletes. The body composition hypothesis, first described by Frisch and McArthur (1974), held that when adipose tissue falls below a critical threshold, menstrual function is lost, as observed in patients with anorexia nervosa. However, the hypothesis was based largely on correlation rather than experimental evidence, and subsequent studies noted eumenorrhoeic and amenorrhoeic athletes spanned a range of body sizes (Hale et al., 1983; Loucks et al., 1984).

In the 1930s, Selye noted the adrenals of rats were hypertrophied when various stressors interrupted the normal reproductive cycle (Selye, 1939). His experiments led to the ‘exercise stress hypothesis’, whereby activation of the hypothalamic-

pituitary-adrenal (HPA) axis from exercise leads to amenorrhoea. This was thought to explain the observations of Nagata et al., (1986), who documented amenorrhoea in most Japanese nursing students during their highly stressful term, which resolved during the summer holidays. Controlled exercise studies in humans demonstrated blunted stress hormone responses in amenorrhoeic women compared with eumenorrhoeic and non-exercising controls, presumably related to chronically supra-normal levels (Ackerman et al., 2013; Loucks and Horvath, 1984; Loucks et al., 1992; Loucks et al., 1989). Elevated corticotrophin releasing hormone (CRH) levels were thought to cause reproductive dysfunction by interrupting GnRH and LH pulsatility (Kowalski et al., 1993; Petraglia et al., 1987). However, experiments were confounded by the effect of the stressor enforcing the exercise and the energy cost of the exercise, which itself would be associated with a rise in cortisol to stimulate gluconeogenesis and lipolysis (Loucks, 2013).

The observation by Warren (1980) that ballet dancers' energy intake during training could modulate the hypothalamic pituitary axis and induce amenorrhoea led to interest in the idea of brain energy availability (EA). Winterer et al., (1984) developed the hypothesis of brain-to-body caloric ratio: mammals partition energy across 6 hierarchical processes in the following order: cellular maintenance, thermoregulation, locomotion, growth, reproduction and storage. Fuel spent locomotion is unavailable to other less important processes, such as reproduction (Wade and Schneider, 1992).

The hypothesis of inadequate EA has been supported by animal studies, where food restriction in rodents was seen to interrupt the normal reproductive cycle (Petit and Prior, 2013). Loucks (2013) performed a series of short studies comparing sedentary and active eumenorrhoeic women, in whom energy intake was manipulated to limit energy availability. Over 5 consecutive days of exercise, limiting energy intake resulted in suppressed LH pulsatility and reduced total triiodothyronine (TT3) levels, while exercise without dietary restriction had significantly less effect. The question remained: does energy availability exert this effect *per se* or is another component of exercise responsible (such as cortisol or another metabolic marker of exercise stress)? Work by Williams et al*.,* aimed to delineate these factors. Rhesus monkeys were trained to run for increasing periods while food intake was kept constant (Williams et al., 2001). The diet of half of the monkeys was then supplemented without changing the exercise regimen, which led

to restoration of menstrual cycles. The same effect was shown by Williams et al*.,* (1995) in humans undertaking intense running training: initial LH suppression was observed to return to a normal pulsatile pattern upon energy derestriction, supporting the hypothesis that exercise has no effect on the hypothalamus beyond that of decreasing energy availability.

To determine whether energy availability as an entity is independent of the metabolic stress of exercise, it is necessary to control experiments for the metabolic impact of the exercise. One such experiment was carried out on male US Army Rangers over 8 weeks, where soldiers were exposed to arduous physical training with controlled diets (5000 and 2000 kcal per day on alternate days), alongside heat, cold and psychological stressors (soldiers were exposed to 4, 2-week phases in desert, forest, mountain and swamp environments) (Friedl et al., 2000). Anterior pituitary hormones including LH were measured as well as testosterone. Both LH and testosterone were suppressed during training, but were restored to normal during re-feeding, despite continued exposure to such stressors.

It now is widely accepted that the body can adapt to survive an energy deficit by prioritizing exercise over reproductive function (Mountjoy et al., 2018). Reduced EA (defined as dietary energy input minus exercise induced energy expenditure) has become the best explanation of exercise induced reproductive disturbance, especially in lean athletic pursuits (Javed et al., 2015; Reed et al., 2015).

Warren (1980) and Winterer et al., (1984) proposed that dynamic hypothalamic adaptations occur to meet immediate requirements for survival because of stressors, according to the degree the stressor is imposed or withdrawn. Loucks (2013) noted the reduced LH pulsatility and amenorrhoea seen in the setting of low energy availability were associated with suppressed glucose, insulin-like growth factor 1, insulin and total triiodothyronine (TT3), while growth hormone and cortisol levels were raised. These phenomena may be mediated through the anorectic adipokine leptin and orexigenic (appetite stimulating) gut peptide ghrelin (De Souza et al., 2004; Loucks, 2004; Scheid and De Souza, 2010). Leptin is produced proportionately by adipocytes after eating, and concentrations are relatively low in fasting states, while ghrelin is secreted in response to stomach emptying. The primary role of these hormones has traditionally been thought to be hypothalamic signalling of satiety and hunger (Bouassida et al., 2006; De Souza et al., 2004).

Hypothalamic amenorrhoea is associated with disproportionately low levels of leptin and high levels of ghrelin, accompanying blunted LH pulsatility (Ackerman et al., 2012; Scheid et al., 2013)

Welt et al*.,* (2004) demonstrated that recombinant leptin administration restored the menstrual cycle and levels of sex steroids and gonadotrophins in women with hypothalamic amenorrhoea. These results have been replicated such that it seems that there may be a critical leptin threshold below which hypothalamic amenorrhoea occurs (Chou et al., 2011) and leptin therapies promise a new adjunct for treating HA after addressing psychological and social aspects (Kyriakidis et al., 2016).

Despite vigorous refutation of the exercise stress hypothesis by the original proponents of the EA hypothesis (*i.e.* that sole activation of the HPA axis leads to reproductive dysfunction, (Loucks, 2009; Loucks and Redman, 2004)), the observation remains that reduced GnRH drive in HA is associated with elevated central and peripheral levels of cortisol (Ackerman et al., 2013). Vulliemoz et al., (2008) induced negative energy balance and HA in rhesus monkeys (using an infusion of ghrelin) and demonstrated a restoration of LH pulsatility following infusion of astressin B, a CRH antagonist. This suggests that antagonism of the central drive to cortisol production could ameliorate the effect of a negative energy balance on LH. Subsequently, in a cross sectional study of overnight hormone levels in female adolescent amenorrhoeic endurance athletes, Ackerman et al., (2013) demonstrated a rise in ghrelin and fall in leptin and fat free mass (used here as surrogates for chronic energy availability) associated with reduced LH pulsatility, compared with exercising eumenorrhoeic and sedentary controls. Cortisol was shown to be more strongly associated with reduced LH pulsatility, even after correcting for leptin, ghrelin and fat free mass. These studies suggest the action of gut peptides on the hypothalamus in the reduced EA state are closely linked with activation of the HPA axis.

Other means of HPA axis activation might also compound reproductive dysfunction. In a cross sectional study, depression scores were higher in women with HA than eumenorrhoeic controls (Beck Depression Inventory score 7.9 ±8.5 vs 2.9 ±2.9, p=0.02) (Marcus et al., 2001). Cognitive behavioural therapy is now recommended in the treatment of HA (Gordon et al., 2017), although this recommendation is based only on a few small studies. In one unblinded randomised control trial, cognitive

behavioural therapy largely focussed on nutrition, weight and exercise in women with normal BMI restored ovarian activity and ameliorated the characteristic central and peripheral hypercortisolism of HA; leptin increased following treatment without weight gain (Michopoulos et al., 2013).

Functional neuroimaging studies have revealed the brain to be one of the most metabolically active organs in the body. Psychogenic activation could reduce energy availability to the hypothalamus, such that even modest combinations of exercise, low EA (elevated ghrelin) psychological stress (HPA axis activation) and decreased adiposity (low leptin), may result in significant reproductive dysfunction (Berga, 2008). Thus, trends in research intertwine the body composition and stress hypotheses with energy availability, suggesting FA reflects complex interactions between several hormonal axes, nutrition and energy storage.

1.2.2.2 Female athletic triad and Relative Energy Deficiency in Sports

Much research into reproductive disorders has emphasised the female athletic triad (Triad), observed in the setting of low EA and de-prioritisation of reproductive processes to compensate for arduous training (Loucks, 2013). The Triad was originally defined by the presence of an eating disorder, functional amenorrhoea and premature osteoporosis (Otis et al., 1997), although as will be seen, cross-sectional studies in various populations (including military trainees) demonstrated a low prevalence of all three components. The definition was refined by the American College of Sports Medicine in 2007 to encompass spectrums of menstrual disturbance (which may be subclinical), osteopenia and low energy availability (Nattiv et al., 2007). Unfortunately large-scale cross sectional studies aiming to determine the Triad's prevalence have been hampered by stringent inclusion criteria or inconsistent definitions of the components (e.g. self-reported menstrual irregularity rather than serial measurement of hormones) (Coombs et al., 2015; Gibbs et al., 2013). One or two components of the Triad may occur in isolation, and at any point on a spectrum of severity (**Figure 1-4**). In particular, poor bone health may be hard to detect and is frequently overlooked, since imaging is required for screening (Ackerman and Misra, 2018). Arguably, the concept as defined by the American College of Sports Medicine (Nattiv et al., 2007) is not a true 'triad', since it represents three separate conditions which can occur independently (DiPietro and Stachenfeld, 2006).

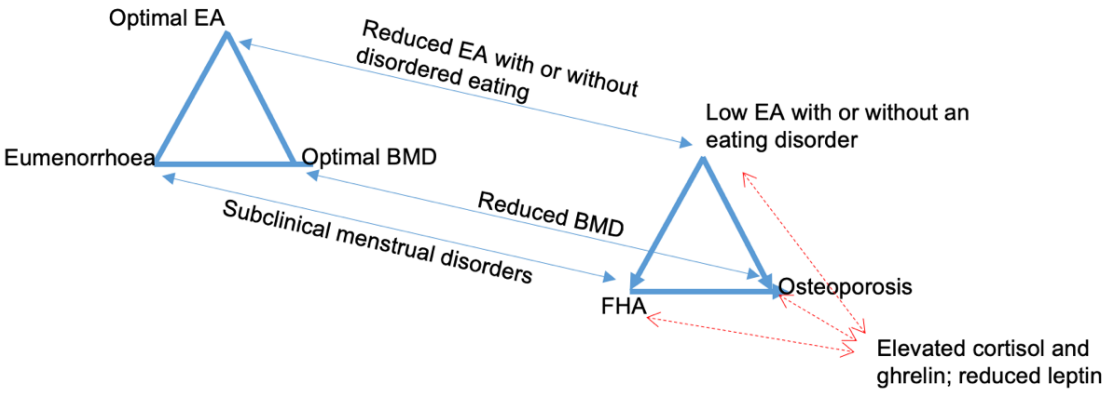


Figure 1-4 Female Athletic triad according to the 2007 American College of Sports Medicine Consensus Statement, adapted from (Gifford et al., 2017) EA, menstrual function and BMD exist in a clinical spectrum, along which athletes are distributed (thin arrows), An athlete moves either left or right along the spectrum according to exercise and diet practice. Energy availability modulates BMD indirectly via effects on menstrual function and 'directly' via changes in metabolic hormones, notably leptin and ghrelin (thick arrows). All three aspects are associated with changes in the HPA axis and (red dashed arrows) BMD: bone mineral density, EA: energy availability, FHA: functional hypothalamic amenorrhoea.

The Triad concept has since been expanded to encompass a wide array of hormonal and metabolic disturbances putatively caused by low EA. This broader syndrome, Relative Energy Deficiency in Sports (RED-S), can be thought of as the consequence of downregulation of hormonal axes in the setting of relative starvation, and may affect men as well as women (Mountjoy et al., 2018). Of particular concern to the military is uncoupled bone turnover, since this is associated with microtrauma and stress fractures (Greeves, 2015). The fact that stress fractures are commoner among female military trainees than males (Knapik et al., 2018) could indicate RED-S is particularly prevalent in the military.

An early study evaluating the prevalence of the Triad undertook a cross-sectional analysis of cadets at USMA, including both regulars and reservists (Lauder et al., 1999). Only 3.3% of 87 cadets had reduced BMD (*T*-score between -1.0 and -2.5), 2.1% had amenorrhoea and none had 'the Triad'. However, the findings were confounded by hormonal contraceptive use in a majority of participants and time constraints on reservist participants limited survey completion. Furthermore, the study evaluated an early definition of the Triad (Otis et al., 1997), and would have

underestimated prevalence by the 2007 ‘spectral’ definition (Nattiv et al., 2007). However, 8% of the cadets studied by Lauder et al*.,* met the criteria for an eating disorder, while 27% were at risk of one (‘disordered eating’). These findings are comparable with studies of civilian athletes (Matzkin et al., 2015). A cross-sectional study of 295 US service personnel (71% male) found 29% of women with normal BMI perceived they were overweight, of whom 81% said they were attempting weight loss (Clark et al., 2017). No men with normal BMI perceived they were overweight. Exposure to military-related stressors can increase the risk of disordered eating in women (Breland et al., 2017), while aspects of military culture may influence preconceptions about body composition and drive for thinness (Clark et al., 2017).

Disordered eating is a psychological attitude in which individuals feel they must limit food intake to achieve or maintain a desired body image (Rumball and Lebrun, 2004). It is perhaps surprising that unhealthy attitudes towards eating were so prevalent in Lauder et al.’s study of West Point cadets (1999). However, such observations have been made across a wide spectrum of sports, including lean and non-lean activities, the prevalence in endurance, technical and ball game sports (24%, 17% and 16%, respectively) being significantly greater than controls (4.6%; $p<0.001$) (Gibbs et al., 2013). In a systematic review of studies of prevalence estimates of diagnosed anorexia nervosa, bulimia nervosa and nonspecified eating disorders among US military men and women, Bartlett and Mitchell (2015) found the prevalence to be 1.1%, 8.1–12.5% and 36–62.8%, and 2.5%, 6.8% and 40.8% among women and men respectively, which were comparable to the general population. The studies included used a range of methods, with probable diagnoses being based on a several different questionnaire scores. However, these authors and others are concerned that DE, which is subclinical and does not infer medical downgrade or prevent entry to the military, might be more common among military than civilian populations. Possible reasons for this were explored in a small study by Breland et al., (2017), who identified the combination of normal practices which promoting binge-eating, psychosocial stressors (not uncommonly sexual trauma), and the need to meet military weight targets. Outside the military, it is widely recommended that coaches and clinicians regularly screen for early evidence of DE in athletes; it may be that the military should be more proactive in this regard (Mountjoy et al., 2018; Williams et al., 2017).

Dietary restriction can be imposed as a component of military training as noted by Hoyt and Friedl in their review of field studies in US and Norwegian Rangers, US Marines and others (Hoyt and Friedl, 2006). Their own study was unusual in that they recruited female as well as male soldiers, who endured 7 days of prolonged physical activity and food and sleep deprivation during Ranger training in Norway (Hoyt et al., 2006). Their gold-standard assessments of total energy expenditure and body composition demonstrated increased usage of fat mass for energy reserve in women compared to men; women lost less weight than men but burned proportionately more fat. The authors did not assess menstrual dysfunction, ghrelin or leptin, and further investigation of sex differences is warranted in advanced military training. Energy deficit was also demonstrated using 24-hour recall in 324 soldiers of 101st Airborne division (21% female) by Beals et al., (2015). They found soldiers significantly under-consumed carbohydrates and micronutrients whilst in barracks, where they exercised for 60-90 min/day. Just 15% of males and 13% of females in barracks ate the recommended amounts of carbohydrate, with no significant differences between sexes. However, in an earlier study examining body fat during an 8-week basic training, where 150 women were estimated to expend a modest 2786 kcal/day, Friedl et al., (2001) reported no increase in self-reported menstrual dysfunction or mean serum oestradiol or progesterone concentrations. Energy intake was not reported.

1.3 Measurement of female endocrine adaptations to military training

Endocrine responses to exercise and arduous military training are more aptly described as functional adaptations than disease processes. Instead of encompassing a simple aetiology and predictable, progressive course, a functional adaptive process involves the complex integration of personal and environmental factors, a labile and reversible course. Physiological effects vary greatly, with the potential for habituation to ongoing training, or reversal when training stops (Mallinson and De Souza, 2014; Petit and Prior, 2013).

1.3.1 Amenorrhoea and reproductive dysfunction

Most literature exploring endocrine adaptations in female athletes and military personnel has focussed on menstrual dysfunction (discussed in **section 1.2.1**). Amenorrhoea and oligomenorrhoea are important symptoms of reproductive

dysfunction, however regular menstrual cycles do not guarantee normal reproductive function. Menstruation is normally caused by a decline in progesterone (P4), but in the absence of ovulation is caused by a decline in oestradiol (E2). Thus, menstruation does not necessarily confirm ovulation or a normal luteal phase, which require a complex interplay between many hormonal and metabolic factors. These are challenging to measure over time in large numbers of women (Petit and Prior, 2013). Therefore, survey-based studies identifying menstrual disturbance are of limited value in the measurement of reproductive function.

The hypothalamic-pituitary-gonadal (HPG) axis is dynamic and undergoes marked fluctuation even in health, and there are several putative endpoints to identify, for example anovulation, LPD or sustained axis suppression, which would be meaningful for downstream end-organ effects such as bone turnover. Normal reproductive function also requires permissible fallopian tubes and a uterine structure to for fertilisation and implantation (Abraham, 1978). For the purposes of this thesis, reproductive dysfunction will be limited to functional disturbance, i.e. hypothalamic amenorrhea (HA), implying physiological disruption with anovulation and/ or a shortened luteal phase (luteal phase defect, LPD) (Gordon et al., 2017). Reproductive disruption arising from anatomical defects will not be considered.

Reproductive dysfunction is defined here as any endocrine imbalance that would impair ovulation. Designing studies to identify reproductive dysfunction in women is challenging, especially since it is essential to conduct such studies in ‘free living’ environments which represent a true exposure to military training. Endocrine pathology associated with reproductive dysfunction includes hormonal changes which could be considered maladaptive and could have downstream effects with negative implications for the female soldier.

1.3.1.1 Ovulation and the luteal phase

Daily transvaginal ultrasound examination is the most accurate means of demonstrating the development and disappearance of a follicle, while regular assessment of endometrial morphology by endometrial biopsy is the gold-standard method to detect a healthy luteal phase (NICE, 2017). Unfortunately both methods are invasive, expensive and have gained limited use as research tools (Petit and Prior, 2013).

A sustained, elevated progesterone (P4) in the luteal phase is often accepted as evidence that ovulation has taken place, most commonly measured 1 week before menstruation (day 21 for a 28 day cycle). A day 21 serum or plasma P4 ≥30 nmol.l^{-1} is considered proof of ovulation (Wathen et al., 1984). However, in exercising women cycle length can vary dramatically, which makes the mid-luteal point difficult to identify. Therefore, serial samples are needed, and this again would be cumbersome in large numbers of women over a long duration and in a field environment. Other indirect methods of determining reproductive function include quantitative basal temperature monitoring and urinary mid-cycle LH surge measurement. Both are subject to user variability (Petit and Prior, 2013).

Daily urinary concentrations of metabolites of P4 and E2 (pregnanediol-3-glucuronide, PDG and estrone-3-glucuronide, E1G) have often been used to demonstrate cyclicity of ovarian activity (Blackwell et al., 2018; De Souza et al., 2010; Ecochard et al., 2017). Expressing these metabolites as a ratio removes the need to adjust either for urinary concentration (e.g. by measuring creatinine or osmolality).

The time-pressured and austere environment of military training precludes detailed measurements such as daily urine or blood sampling over many weeks or months. Study burden is perhaps a greater consideration in military field studies than in clinical trials.

In clinical practice, regular monitoring of the urinary ratio of PDG to creatinine can be used to monitor ovulation and luteal phase length in an outpatient setting, a technique known as tracking. This requires weekly sampling of urine over two or more consecutive cycles and an accurate menstrual diary. A PDG: creatinine threshold of ≥ 5 nmol/mol may be deemed ovulatory and may rise above this level during the luteal phase (Ecochard et al., 2013). Tracking could be a more feasible means of assessing ovulation over long field studies in large numbers of women (Blackwell et al., 2018). Specific gravity (the ratio of density of urine compared with water) represents an alternative for normalising the measurement of urinary hormones. It is measured using refractometry and is independent of muscle activity and creatinine release into the blood stream. However, it requires contemporaneous measurement since storage of samples may precipitate crystals and alter the density (Miller, 2004).

1.3.1.2 Other measures of reproductive function

Assays of ovarian capacity and function have been developed such as Anti-Müllerian hormone (AMH) and inhibin B. These can be performed as one-off tests, are readily available and can be automated, allowing assessment of large numbers of samples. AMH is a glycoprotein produced by granulosa cells of small follicles (≤9 mm) and serves to regulate follicular development. Unlike LH, FSH, E2 and P4, its concentration is stable through the menstrual cycle (Li et al., 2016) making it a reliable and specific measure of ovarian follicular potential, and a useful descriptor of ovarian function. AMH is associated with the number of small follicles, helps differentiate causes of oligo-amenorrhoea such as primary ovarian failure and polycystic ovary syndrome (PCOS), and predicts responsiveness to fertility treatments (Dewailly et al., 2014). Inhibin B is secreted by antral follicles, but is generally produced from larger follicles (peaks at follicle diameter 9-10 mm) and is less discriminatory for causes of secondary amenorrhoea (Li et al., 2011). Inhibin B may be normal or low in women with HA, low in women with premature ovarian failure or menopause, and may be elevated in PCOS (Knauff et al., 2009; Sowers et al., 2008).

While their concentrations vary across the cycle in women, carefully timed basal measurements of plasma E2 and P4 are important, since decreased concentrations are associated with other pathology, such as hypercortisolaemia and impaired bone health (Ackerman et al., 2013; Ackerman et al., 2019b). Hyperandrogenism may contribute to reproductive dysfunction in non-lean sports and PCOS (Hagmar et al., 2009). Elevated adrenal androgens like dehydroepiandrosterone sulphate (DHEA-S) might be beneficial for bone health (Javed et al., 2015). However, elevated DHEA-S may be associated with impaired follicular development, anovulation and amenorrhoea (Lebrun et al., 2013). Evaluation of non-lean sportswomen can be confounded by self-selection of athletes for such sports; higher androgen levels improve muscle mass and power, so women with naturally elevated androgen levels often undertake sports which utilise these traits (Constantini and Warren, 1995). Decreased basal androgen levels following arduous US Army Ranger training have been important in demonstrating the negative effects of stress on the HPG axis in men (Hoyt and Friedl, 2006; Opstad, 2001). Suppression of the HPG axis in men following an arduous 3-month deployment, likely caused by loss of lean mass but not stress, had resolved after 2 weeks of recuperation (Hill et al., 2015). Such

suppression may be mitigated in the short-term by nutritional supplementation (Fortes et al., 2011).

Dynamic function tests measuring the response to stimulation of a hormonal axis are performed when there is concern the axis may be suppressed. Such dynamic function tests can either measure a response at a specific level of an axis, or the whole-axis response to a higher (whole organism) stressor. To isolate the HPG axis, GnRH, or less commonly kisspeptin, can be injected, assessing the responsiveness of anterior pituitary gonadotrophs or hypothalamic KNDy neurones over time, respectively (Jayasena et al., 2014; Morosini et al., 1989).

Recent technological advancements using electrochemical aptamers (single stranded DNA oligonucleotides) have demonstrated rapid, sensitive and specific measurements of LH in real time, using an indwelling catheter in blood or interstitial fluid (Dhillo et al., 2019). Such technology could conceivably be miniaturised into a wearable LH sensor, enhancing the measurement of reproductive function, while allowing detailed longitudinal monitoring in arduous training and extreme environments, as has been demonstrated for continuous glucose monitors (Hill et al., 2018).

1.3.2 The HPA axis

The HPA axis is essential to life in vertebrates, linking central nervous and endocrine systems via the release of glucocorticoids (Kudielka and Kirschbaum, 2005). Cortisol, the primary glucocorticoid in humans, is the main effector of the human HPA axis and carries numerous control and regulatory functions. Cortisol is released in response to external stimuli and plays a central role in mobilising energy stores and increasing energy delivery. Cortisol effects increased vascular tone and cardiac output via the sympathetic nervous system and release of catecholamines from the adrenal medulla, and modulates immune, inflammatory and psychological function (Chen et al., 2016). The adrenal cortex produces cortisol in response to adrenocorticotrophic hormone (ACTH), produced in the anterior pituitary response to corticotrophin releasing hormone (CRH) produced in the hypothalamus. Circulating cortisol is 10–15 % bound to albumin and 80–90% bound to cortisol binding globulin (CBG), so that biologically active cortisol accounts for approximately 5% (Faix, 2013). CBG actively regulates circulating cortisol, serving to buffer concentrations. A reactive central loop is cleaved by proteases at sites of inflammation, resulting in a

10-fold reduction in binding affinity for cortisol and increased cortisol release to tissues (Bae and Kratzsch, 2015). Feedback inhibition from glucocorticoids and a range of afferent neurons from the brainstem, midbrain and limbic system tightly regulates the HPA axis, acting at the level of the pituitary and hypothalamus (Smith and Vale, 2006).

Regulation and sensitivity of the HPA axis evolve from early life to adulthood life in a sex-specific manner (reviewed briefly by Gifford and Reynolds (2017)). From adolescence onwards the differences pertain to sex hormones: oestradiol stimulates CRH and ACTH production, while testosterone has the opposite effect (van der Voorn et al., 2017). Oestrogens also stimulate production of CBG. CBG concentrations are therefore greater, and average free cortisol levels often lower, in women than men, and in COCP-using women than non-users (Bae and Kratzsch, 2015). In normally ovulating women, synthesis of CBG is higher in the luteal phase than in the follicular phase and rises 2 to 3–fold during pregnancy (Bae and Kratzsch, 2015).

Changes in HPA axis function have been observed in tandem with reproductive dysfunction. In patients with HA, there is an inverse association between cortisol concentrations and reproductive function (Ackerman et al., 2013), while downregulation of the HPA axis may restore reproductive function (Vulliemoz et al., 2008).

Since cortisol is released in a pulsatile manner, point measurements should be carefully timed to account for ultradian fluctuation. This is undertaken commonly using the cortisol awakening response (CAR), whereby free cortisol is measured in saliva immediately upon waking, and at predefined intervals thereafter, e.g. 20, 40 and 60 min (Chida and Steptoe, 2009). Changes in the peak and area under the curve of the CAR are informative about the HPA axis response to stress, and have been suppressed during basic military training (Clow et al., 2006). CAR is generally smaller among men than women (Stalder et al., 2016). Cortisol usually reaches a nadir in the evening; the addition of a further measurement before bed allows the assessment of the diurnal slope. This is useful to detect relative evening hypercortisolaemia (flattening of the slope) which may be associated with anxiety or depression (Adam et al., 2017).

Measurement of average cortisol concentrations by continuous collection of urine over several hours is logistically challenging but has been achieved in a military setting, showing a modest increase after 4–week Hellenic Army basic training (Makras et al., 2005). Hair cortisol concentration represents a more convenient alternative. Cortisol is released into hair as it grows and is stable at room temperature (Stalder et al., 2017), hair grows at a consistent rate of 1cm per month, hence hair can yield a retrospective record of average cortisol concentration for the age of a segment. A systematic review and meta-analysis showed concentrations decline by a mean 29% from the first proximal 3 cm segment to second most proximal 3 cm segment (Stalder et al., 2017). Important considerations are sex (men exhibiting 21% higher concentration than women), age, peroxide treatments, which denature cortisol, and COCP, which has a negative effect on hair cortisol concentration (Stalder et al., 2017). Elevated hair cortisol concentrations are consistent with some cortisol measures, e.g. higher 24-hour urine concentrations (Remer et al., 2008) and cortisol reactivity under stress (Kudielka et al., 2009), but not others, e.g. Boesch et al. (2015) found no change in hair cortisol in men during Swiss basic military training (this was attributed to interrupted segments which provided a record from 3–weeks at the beginning and end of training only). Hair cortisol concentrations in this military cohort correlated moderately with education status ($r=0.265$, $p<0.05$), environmental temperature and air humidity ($r=0.425$ and $r=0.509$, respectively, both $p<0.001$).

The HPA axis is commonly assessed clinically at the level of the adrenal measuring response to synthetic adrenocorticotrophic hormone (SynACTHen). In clinical practice in the UK, a supraphysiological dose of 250 µg is used, aiming to detect adrenal insufficiency if the response is below a predefined threshold after 30 or 60 minutes. A lower dose may be used to detect more physiologically relevant adrenal responses over 1 hour and may be of greater value in detecting subtle functional changes before and after training (Reynolds et al., 2001a). Corticotrophin releasing hormone (CRH) is also used to measure HPA axis response at the pituitary level. Cortisol responses to HPA axis stimulation tests, like CRH, are generally of lesser magnitude among men than women (Kudielka et al., 2009). However, perhaps owing to practical considerations, such as cost and availability of SynACTHen or peptide stability, dynamic tests of specific hormonal axes do not appear to be used frequently in studies of exercise or military training. Instead, dynamic responses to

whole-organism stressors, such as hypoglycaemia (insulin tolerance test, ITT) or maximal exercise, have been more commonly used (Cadegiani and Kater, 2017b). Blunted responses of many hormones including cortisol, LH and FSH may be seen in overtraining syndromes (Cadegiani and Kater, 2018b).

Measurement of basal (fasting) cortisol should be caveated, noting that venepuncture represents a stressor and isolated concentrations may not reflect dynamic variations in the HPA axis or be informative of changes in HPA axis function. However, fasting cortisol has been usefully measured before and after arduous training 'interventions'. For example, fasting serum cortisol was elevated in women after 1 week of US Marine Corps basic training, but gradually normalised during the ensuing 12 week course, despite weight loss (albeit fat only; lean mass increased – discussed in section 1.4.2) (Lieberman et al., 2008). This was attributed by the authors to marked improvements in mood scores during the course. Highly arduous training, such as the 8-week US Army Ranger course, is a whole-organism stressor, with psychological stress, energy deficit, weight loss with loss of lean mass, sleep deprivation and submaximal exercise (Friedl et al., 2000). Serial basal hormone measurements immediately after training and over the subsequent days has demonstrated elevated cortisol, with suppressed thyroglobulin, free triiodothyronine (T3) and free thyroxine (T4), gonadotropins, insulin-like growth factor 1 (IGF-1) and testosterone in men (Henning et al., 2014; Opstad, 2001; Opstad et al., 1984). In a series of studies of Norwegian men and women undergoing an intensive 5-day military training course, the combination of physical and mental strain under sleep restriction led to marked suppression of the HPG axis, as well as the HPA and thyroid axes (Opstad, 2001). Four-week survival, evasion, resistance and extraction (SERE) training is associated with elevated cortisol levels compared with non-psychologically stressed controls (Taylor et al., 2007b); cortisol was elevated in response to a variety of simulated interrogation techniques encountered during SERE training (Lieberman et al., 2016a). Adaptations like these are transient and resolve over days to weeks with resumption of normal eating, sleeping and rest (Henning et al., 2014).

Specific psychosocial stress tests have been developed, for example the Trier Social Stress Test (TSST) (Kirschbaum et al., 1993). The TSST measures saliva cortisol responses to repeatable stressful scenarios (e.g. simulated job interview or solving arithmetic problems in front of an audience of peers). A blunted response

may be seen in anxiety disorders such as PTSD, consistent with enhanced central negative feedback on the hypothalamus (Bandelow et al., 2017). Sex differences in TSST response have been observed in a wide variety of studies (Reviewed by Allen et al., (2017)). The most common disparities are reduced responsiveness in women and greater anticipatory cortisol response in men. These disparities appear to be influenced by sex steroids (e.g. some studies found sex differences to be greater when female progesterone concentrations were lower), while gender behavioural roles appear to frame responses to such 'social' stresses (Allen et al., 2017). Few studies of military social stress exposure have been undertaken in women owing to the relatively low percentage of veterans who are female (Lehrner and Yehuda, 2014). In female veterans Pierce et al., (2016) found past military deployment, but not PTSD, was associated with blunted TSST responses compared to civilian controls. Diurnal cortisol curve was no different in deployed women compared with controls. A group modification of the TSST has been validated in Swiss military training (Boesch et al., 2014), demonstrating particular effectiveness when requiring 'ego involving' effort from the subject, such as defensive speech. This tool could be feasible for a military training environment and is promising as a detailed measure of stress responses for future studies but for logistical reasons this test was not included in my studies.

1.3.3 Bone health

The female athlete triad and RED-S concepts emphasise the importance of bone loss in athletes alongside reproductive dysfunction and concomitant risk of stress fractures (Nattiv et al., 2007; Mountjoy et al., 2018). HA is associated with increased risk of osteopenia and fractures (Gordon et al., 2017), and preventing training related fractures is of particular importance to the military (Greeves, 2015).

Bone is continuously being formed and resorbed throughout the body. The balance of formation and resorption can be assessed using bone turnover markers measured in blood. These include markers of collagen resorption (e.g. procollagen type 1 N-terminal propeptide, P1NP), collagen formation (e.g. β cross linked telopeptide, CTX), and formation and resorption of calcium hypoxyapatite (e.g. osteocalcin and alkaline phosphatase, respectively). Low bone turnover (both low resorption and formation) implies the normal repair of microscopic damage to bone is impaired, which leads to higher risk of overuse fractures. Uncoupling of bone

turnover (an imbalance favouring either resorption or formation) implies a progressive change of bone mineral density and strength (Hughes et al., 2014).

Bone mineral density is most often measured using areal whole-body absorptiometry (dual energy x-ray absorptiometry, DXA), estimating relative composition of tissues and density of bone, based on absorption of x-rays of two different energies. This allows comparison of all skeletal regions, including areas prone to stress fracture (e.g. hip). DXA scores are most commonly expressed as bone mineral density (bone mass divided by bone area) or the number of standard deviations bone mineral density is above or below an age-adjusted normogram (Z score). DXA is limited by its two-dimensional design, which provides an estimate of bone mineral density from regional absorption of x-rays, rather than a direct measurement of density (Mallinson and De Souza, 2014).

Three dimensional, volumetric bone mineral density can be measured using cross-sectional imaging, peripheral quantitative computerised tomography. While limited to the peripheries, this provides a direct measurement of bone volume and yields large amounts of additional data about microscopic architecture, e.g. cortical thickness, trabecular microarchitecture and porosity (Mallinson and De Souza, 2014). High resolution peripheral quantitative computerised tomography (HRpQCT) is a further recent advancement in bone imaging, which can estimate functionally relevant parameters such as fracture threshold and stiffness (De Souza et al., 2017).

1.4 Consideration of stressors in military training

The stressors which contribute to making military training ‘arduous’ are laid out in this section. Military training, which is the context for most of the extant studies measuring aspects of arduousness, will first be briefly defined.

1.4.1 Context

Military training serves as a necessary means for recruits or service personnel to establish or maintain combat readiness. Research into the effects of such training on trainees must not interfere with the training itself and is therefore largely observational. Intervention-based studies would be difficult to perform; interventions would need to be very carefully designed and integrated into training to reduce the likelihood of altering certain training outcomes and affecting the trainee’s career.

Military training can place extreme demands on individuals (Booth et al., 2006). The most strenuous training includes specialist soldier selection and training courses, such as for Royal Marine Commandos, the Parachute Regiment, UK Special Forces or US Army Rangers (collectively termed ‘advanced combat training’ in the USA). Such training places severe physical and mental demands as a means of selecting the strongest individuals (Nindl et al., 2007). Survival, evasion, resistance and extraction training is particularly psychologically demanding, exposing service personnel to the kind of hostility expected in the event of capture (Taylor et al., 2007a).

Basic military training (termed ‘basic combat training’ in the USA) is less objectively demanding, serving to teach core military skills, ‘followership’ and leadership. It can nevertheless be very arduous, since those undertaking it are normally naïve to military life beforehand. Basic military training generally lasts several weeks to months. The numerous psychological stressors involved in adaptation to training include anticipatory stress, performance evaluations, sleep deprivation and conflict (Gold and Friedman, 2000). The fact that most soldiers in basic military training are adolescents presents additional stressors (Hardoff and Halevy, 2006; Schneider et al., 1999).

1.4.2 Energy availability

It has been seen that EA is important to measure in relation to reproductive function during exercise, since low EA causes reproductive dysfunction and relevant associated pathology (section 1.2.2.1). Low EA due to disordered eating may be a significant problem in military training as in athletes, while restriction of energy intake can be included as a component of arduous military training (section 1.2.2.2). Energy availability is challenging to measure in practice and is defined as follows

$$EA = \frac{EI - EEE}{FFM}$$

EA: energy availability, EI: energy intake, EEE: energy expenditure due to exercise and FFM is fat-free mass.

This is distinct from energy balance which is simpler to measure and may be defined as:

$$EB = EI - TEE$$

EB: energy balance, EI: energy intake, TEE: total energy expenditure.

While both are largely determined in healthy individuals by exercise and energy intake, there are important operative differences between EA and EB. Since EA is dependent on energy expenditure due to *exercise,* it may be considered an *input* into physiological systems - the energy left over for metabolic processes after exercise. Conversely, EB represents the overall surplus or deficit after total energy expenditure and may be considered an *output*, either positive, negative or neutral (Loucks et al., 2011).

Change in energy balance is reflected by composition, measured by skinfolds, air displacement plethysmography or DXA. Fat and fat-free mass contain 9.5kCal.g^{-1} and 1.0kCal.g^{-1}, respectively (Dulloo and Jacquet, 1999). Energy balance may be calculated from body composition, allowing for a 10% variation in the thermic effect of food as follows (Compher et al., 2006)

$$EB = 1.0\,\frac{\Delta FFM}{\Delta t} + 9.5\,\frac{\Delta FM}{\Delta t}$$

EB: energy balance in kcal.d^{-1}, ΔFM and ΔFFM: change in fat free mass and fat mass, in g, Δt*:* length of time of the duration over which the change is measured in days.

In the setting of negative EB, women appear to utilise relatively more energy from fat stores than men, to meet the energy demands of exercise (Hoyt et al., 2006). This preferential substrate utilisation may help mitigate against negative effects of energy deficit in women.

Important measurement differences exist between EB and EA. While EB can be determined from changes in body composition over time, EA requires an exact measure of EEE. There is currently some debate as to whether for the purposes of determining EA, EEE represents the energy expenditure due to all physical activity, or the energy expenditure resulting from 'exercise activity', that is, adjusted for sedentary activity (Burke et al., 2018). The advantage of considering all physical activity energy expenditure is that this can be easily calculated if the resting

metabolic rate (RMR) is known, since the difference between TEE and physical activity energy expenditure is RMR (Hopkins et al., 2016):

$$paEE = TEE - RMR$$

paEE: energy expenditure due to all physical activity (including sedentary activity), TEE: total energy expenditure, RMR: resting metabolic rate

RMR can be estimated based on fat free mass (since ~90% of RMR is accounted for by FFM) as follows (Cunningham, 1991):

$$RMR = [370 + (21.6 \times FFM)]$$

RMR: resting metabolic rate, FFM: fat free mass.

However, it is preferable to measure RMR directly rather than predicting it from body composition parameters. In women with anorexia nervosa, the Harris-Benedict equation tends to grossly over-estimate RMR, by as much as two-fold (De Souza et al., 2019a). This is due to both low body fat stores and downregulation of metabolic activity at a tissue level among these women. A study of oligomenorrhoeic and amenorrhoeic athletes with features of RED-S demonstrated 8% lower measured RMR than predicted by DXA-derived body composition, suggesting downregulation of metabolic activity at a tissue level in these women (Koehler et al., 2016).

Considering energy expenditure due to all physical activity is logical, since sedentary activity varies in athletes and may contribute to overall energy deficit. For example, compensatory increases in sedentary activity are expected after single bouts of exercise (Paravidino et al., 2017), or reductions in sedentary activity may occur after increased occupational workload (Gay et al., 2017).

In basic military training, direct measurement of RMR may be challenging, since it requires indirect calorimetry at rest, which is time-consuming and involves significant study burden. Moreover, investigators measuring EA (including the original proponents of the Triad), normally only measured EEE during discrete physical activity ‘interventions’, thus sedentary activity energy expenditure would be excluded (Burke et al., 2018; Loucks, 2006; Loucks et al., 1998). Including sedentary activity in EEE is therefore likely to add a significant overestimation of EA when viewed in the context of extant literature.

Whether or not sedentary activity is included in a working definition of EEE, the EEE measure must include exercise at the level of walking or marching, in addition to intentional 'physical training' activities, since walking and marching can occupy a significant amount time and energy expenditure during military training (O'Leary et al., 2018; Richmond et al., 2010).

Total energy expenditure can be measured using doubly labelled water (DLW), whereby a known, weight-based dose of water labelled with the stable isotopes ^{18}O and ^{2}H is administered. Since exhaled CO_2 does not contain hydrogen, the relative concentrations of ^{18}O and ^{2}H in urine allows derivation of average CO_2 production. The energy equivalence of CO_2 production varies according to the metabolic substrate being oxidised to a much greater extent than O_2 consumption (as measured by calorimetry). It is therefore necessary to assume a mean respiratory quotient value for the duration of the measurement. However, when it is not feasible to undertake indirect calorimetry as in military training, the respiratory quotient can be substituted with food quotient (FQ) as follows:

$$TEE = 4.63\ CO_2 + 16.49\ \left(\frac{CO_2}{FQ}\right) kJ.d^{-1}$$

Where TEE is total energy expenditure and FQ is food quotient.

For omnivorous adults in energy balance, FQ is generally taken as 0.85 and may be adjusted for vegetarians to 0.87 (Ebbeling et al., 2018). FQ is increased in the setting of weight loss (normally assumed to be 88% fat and 22% non-fat, of which 75% is water and 25% protein) (Black et al., 1986).

Exercise energy expenditure can be estimated from activity logs; however, these are prone to recall bias and imprecision from equating predefined metabolic energy equivalents for known activities into those undertaken in the study (Doyle-Lucas et al., 2010; Hoch et al., 2009; Silva et al., 2018; VanHeest et al., 2014). The use of heartrate monitors either continuously or during bouts of exercise has been widely used to estimate EEE, either alone or alongside activity logs (Lagowska et al., 2014; Schaal et al., 2011; Torstveit et al., 2018; Viner et al., 2015). Oxygen uptake (VO_2) can be extrapolated from a known relationship with heartrate, either established under laboratory conditions or estimated from perceived exertion and normograms of heartrate and average energy expenditure (Burke et al., 2018). Wrist or waist-

worn accelerometers have also been used to monitor bodily movements (Brown et al., 2017; Hoch et al., 2009; Woodruff and Meloche, 2013), and carry the advantage of not requiring calibration against calorimetry. However a validation study of several accelerometers found consistent underestimation of total physical activity EEE by 100 to 600 $kCal.d^{-1}$, compared with DLW or calorimetry (Murakami et al., 2016). The precision of accelerometers can be increased by simultaneously carrying out gold-standard measurements of energy expenditure, such as indirect calorimetry or DLW. The latter requires urine sampling only, so is more feasible in military training.

Energy intake is notoriously challenging to measure in humans. The gold standard method is weighed food intake, where researchers weigh individual components of each item eaten at mealtimes (Williamson et al., 2003). This is feasible in some military settings where meals are eaten in a single location (canteen). However military studies have used 24 hour recall (Beals et al., 2015) and food diaries (Castellani et al., 2006; Fallowfield et al., 2014; Schnakenberg et al., 1981; Shay et al., 2009) or a combination (Margolis et al., 2016). These 'traditional' methods are prone to under-reporting and recall bias. Additional challenges military training may present for the assessment of EI result from physical and psychological factors, such as body dissatisfaction, weight consciousness and perfectionism (Martinsen et al., 2014). The assessment of food intake is known to alter habitual intake and reporting, as was quantified in a systematic review and meta-analysis of such methods compared with doubly labelled water: Capling et al. (2016) found consistent under-reporting by mean 2793 ±1134 kJ, or 19%, per day (11 studies).

1.4.2.1 Eating behaviour

Eating behaviours influenced by unhealthy attitudes towards eating may be an important cause of low EA in women undergoing military training (discussed in section 1.2.2.2). Various questionnaires have been devised to determine risk of disordered eating in athletes. Most are derived from the eating disorders inventory. The EAT-26 questionnaire (a derivation of the Eating Attitudes Test (Garner and Garfinkel, 1979)) is a 26-question tool which has been widely used to detect attitudes found in anorexia nervosa (Knapp et al., 2014). However questionnaires like EAT-26 may not be focussed enough on the nuanced eating patterns of athletes (and by extrapolation, military trainees). Martinez et al. (2015) expressed concerns that EAT-26 scores were paradoxically higher in athletes with ad-libitum diet than those with restricted diet. Briefer questionnaires were subsequently developed to

targeting the specific domains of the eating disorders inventory found in female athletes. The Low Energy Availability in Females Questionnaire (Melin et al., 2014) demonstrates excellent test-retest properties, but goes beyond the remit of eating disorders, with several questions on gastrointestinal function (diarrhoea and constipation), contraceptive use, stress fractures and menstrual function. It may be usefully employed where a single tool is required to screen for features of disordered eating and some of its associations, but does not provide great detail and cannot be applied to men. A more focussed measure is the Brief Eating Disorders in Athletes Questionnaire (BEDA-Q), with nine self-rated items which focus on the body dissatisfaction, drive for thinness and perfectionism (Martinsen et al., 2014; Melin et al., 2014), each scored from 1 to 3. Additional binary outcomes (‘are you dieting’, ‘have you ever dieted’ and ‘if so how many times’) is useful and allows greater statistical flexibility for study design. The BEDA-Q has recently received recognition and is a useful tool for time-pressured environments, like military training (Ackerman et al., 2019a; Logue et al., 2018; Peric et al., 2016; Slater et al., 2017).

1.4.3 Type of physical exercise

Physical exercise is a core component of military training, aiming to create physical resilience and maintain health. As discussed in section 1.2.2.1, athletic training may usefully be divided into categories of lean and non-lean, each having specific hormonal associations. The former encompasses physical activities that place an emphasis on endurance, low body weight, lean physique and/ or aestheticism (e.g. marathon running, ballet, gymnastics and figure skating), while the latter encompasses sports which emphasise power, such as rowing, swimming, rugby and hockey (Gibbs et al., 2013).

Military training encompasses both lean and non-lean training and requires endurance, stamina and strength. Most studies published on civilians relate to lean athletes (Gibbs et al., 2013); there is a paucity of literature describing the endocrine effects of training that is more relevant to the military, such as long-distance marches with load bearing, lifting, hauling, digging, or role-specific tasks like section attacks (running while frequently changing direction and position (‘taking cover’) carrying body armour, a weapon and ammunition).

An important difference between professional civilian athletic training and military training is that athletes generally intend to maximize performance for an event for a

specific discipline or disciplines, for example a football match, cycling sportive or marathon. In contrast, the aims of military training are broad, aiming to enable the performance of military tasks at submaximal exertion on operational deployments, which do not have predefined durations (Booth et al., 2006). Military training may involve many different patterns of exertion, for example abrupt onset sprinting carrying a load, without a precisely designed warm-up regimen. Such abrupt exercise onset has been shown to increase propensity to menstrual disturbance in animal models, while gradually-introduced exercise had dramatically less effect on reproduction (Selye's 'general adaptation syndrome') (Petit and Prior, 2013; Rivier and Rivest, 1991; Williams et al., 2001). More modest effects have been observed in humans, but repeated prospective studies demonstrate that the effects of exercise on reproductive function are reduced when the onset of training is more gradual (Bullen et al., 1985; Maïmoun et al., 2013; Williams et al., 1995). Abrupt onset high intensity interval training in male and female Australian Army recruits was associated with blunted CAR compared with extant endurance and strength training. Cortisol diurnal slope was unaffected (Drain et al., 2017).

Adaptation to exercise can be measured using standardised assessments of oxygen consumption, VO_{2peak} and VO_{2max}. Such assessments represent a gold standard of response to training but require relatively specialised equipment and training and are time-consuming. Isometric or dynamic tests of muscular function using dynamometers provide repeatable measures of strength. Dynamic measures include, static lift or 'explosive force' (jumping from a squatting; isometric measurements are those where the strength of single or few muscle groups are tested in isolation. These require participants to be tested while well-rested in a repeatable, controlled setting (Romero-Franco et al., 2017).

Both strength and aerobic fitness assessments are already incorporated into basic military training as standardised fitness assessments. These normally include measures of both strength and aerobic fitness. The best-effort 1.5 mile (2.4km) run has been used for many years by the British Army and correlates closely with VO_{2max} (r = 0.79 (95 % CI 0.73, 0.85), from a meta-analysis of 18 correlations in 873 subjects by Mayorga-Vega et al. (2016))..

Adaptations to exercise may also be manifested in changes in body composition. As was discussed in section 1.4.1, changes in body composition reflect EB. Physical

exercise with negative EB usually causes loss of fat mass, and if energy deficit is extreme will extend to loss in lean mass. In women undertaking basic military training, Lieberman et al. (2008) showed a mean change of–4.7 ±0.2 kg and +2.5 ±0.2 kg in fat and lean mass, respectively, over 13 weeks. In one of several studies of 8-week US Army Ranger training, Nindl et al. (2007) showed men lost 8.5 ±2.3 and 2.6 ±2.0 kg fat mass and lean mass, respectively, which is typical for this kind of endeavour (Hoyt and Friedl, 2006).

Cardiovascular adaptation to exercise is regulated through the autonomic nervous system (ANS). ANS activity can be measured using plasma metabolites of neuroendocrine nephrines (normetanephrine and metanephrine), which are more stable than catecholamines and less prone to fluctuations from the stress of venepuncture or postural changes (Woods et al., 2017). These represent useful markers of acute adaptation, in particular in the context of additional stressors which may be encountered in military training, (Stacey et al., 2018b) demonstrating habituation with adaptation to repeat exercise and stress exposures (Hodge et al., 2013).

Heart rate variability (HRV), an expression of the variation in beat-to-beat (R-R) intervals, is a useful, cheap and non-invasive marker of autonomic adaptation to physical activity. HRV may be measured during (Michael et al., 2016) and/ or after exercise (Armstrong et al., 2012), calculated using R-R interval frequencies, standard deviations of R-R intervals and/ or non-linear methods (R-R interval entropy, reflecting signal regularity and complexity). Increased HRV is caused by relative dominance of parasympathetic activity over sympathetic activity in the ANS. Beneficial adaptations to exercise reflect greater vagal tone, facilitating restorative and vegetative processes (Thayer and Sternberg, 2006). In contrast, sympathetic activation following exercise serves to mobilise energy stores and improve substrate delivery to tissues. Prolonged vagal activation following exercise is maladaptive and is associated with increased plasma metanephrines and heart rate and reduced HRV (Shaffer and Ginsberg, 2017).

Markers of autonomic activity are useful when viewed alongside contextual factors. Differences in ANS activity may emerge at altitude or during heat exposure. A study by Boos et al. (2017) showed HRV to reduce with high altitude exposure, an effect which was greater among women than men, possibly reflecting sex differences in

ANS adaptation to very high altitude or acclimatisation. Physiological strain from heat exposure attenuates ANS benefits of exercise, measured by metanephrines and HRV (Stacey et al., 2018b). Psychological stress may also influence ANS activity (discussed in greater detail in section 1.4.3). A recent study of Olympic rugby players found HRV to be reduced during international but not domestic tournaments, although tournaments were similar in terms of physical exertion (Flatt et al., 2018). Perceived exertion assessed by questionnaires appeared to be more sensitive than HRV in quantifying arduousness of exercise as the players perceived it. These findings suggested that circumstantial stress, including external expectations, chaotic travel and poor sleep quality had a greater effect on ANS function than exercise (Flatt et al., 2018). Since ANS measures are sensitive to the context of measurement, it is important that they are measured in a quiet, controlled environment, at the same time of day (ideally early morning), where the risk of disruption to study participants is minimised (Task Force, 1996).

1.4.4 Psychological stressors

Psychological stress can be defined as the response of an individual to a challenge or threat (stressor) to maintain mental or physical allostasis in that environment (Selye, 1946). A common conception of basic military training is that its objective is to ‘break the individual down then build them up again’ (Bornmann, 2009). In reality, military training seeks to positively equip individuals to function in the military and has specific and measurable objectives (which do not include ‘breaking individuals down’, at least not in the UK). Nevertheless, this axiom illustrates the fact that basic training is psychologically challenging and there are a number of facets to such stress.

Individuals joining the military, especially adolescents, may experience separation stress, whereby current social support structures are exchanged for an austere training environment (Carbone et al., 1998). Adjustment from home life may manifest as homesickness. In 176 US Air Force trainees (61 (35%) female), a homesickness questionnaire score was positively associated with depression and somatic complaints (Banning, 2010). Scores were higher than first year university students (‘freshmen’): means scores 71 versus 64, respectively (no standard deviation, denominator or p value given).

Other stressors include frequent assessment of performance, time pressure and conflict between teamwork and leadership roles (Gold and Friedman, 2000). These are often assessed using mood questionnaire scores as a surrogate. In a study of 8-week US Army basic training, female basic military trainees had higher anxiety scores than men (mean Beck Anxiety Inventory (BAI) scores 14.6 and 10.1, respectively, $F = 7.57$, $p<0.01$) (Alfonso et al., 2006). Regression analyses showed subjects with higher anxiety scores demonstrated greater improvement during training, although a comparison of men and women at the end of training was not given. Scores of Beck Depression Inventory (BDI) and perceived stress scale (PSS, a direct marker of stress) were in the 'moderate' range at the start of training and were normal by the end, with no differences between men and women. Gender differences in anxiety may reflect a self-reporting style among women which is less defensive (Joiner et al., 2000). However, Lieberman et al. (2016b) used the Profile of Mood States (POMS) manual which measures fatigue, anger, tension-anxiety and depression scores. In a study comparing changes in 98 men and 71 women across 9 week basic military training, no gender differences in these traits were demonstrated; both demonstrated improved mood state in all domains by the end of training. The same authors assessed mood in women during more arduous US Marine Corps basic training (Lieberman et al., 2008). All scores were significantly elevated after 1 week of training compared with age-matched norms, but reduced consistently after 5, 8 and 12 weeks, to levels significantly lower than age-matched norms by week 12, possibly reflecting satisfaction with task completion. As discussed in **sections 1.3.3** and **1.4.2**, this study is notable for the pronounced reduction in basal cortisol and positive body composition changes noted during training in these women. Lieberman et al. (2016a) also found that domains of the POMS resolved at different rates after highly stressful captivity training. Depression and tension-anxiety resolved in these men more quickly than fatigue, confusion and total mood disturbance.

It is important to account for stressors experienced by individuals prior to training. In a cross-sectional study by Warner et al. (2007) 1,090 US Army basic trainees (12% female) reported high rates of physical and sexual abuse prior to training (men: 12.3% and 6.6%, women: 33.3% and 24.4%, respectively). These came from a box in a questionnaire, which participants were asked to check if they had experienced such abuse; no further specifics were provided. Depression scores were higher among women than men (mean patient health questionnaire (PHQ-9) scores 6.3

(6.0) and 4.5 (6.3), respectively, no p value given) but not after multivariate adjustment, including for pre-training abuse. Retrospective measures of stress include the perceived stress scale, which asks about general feelings of stress and coping (Cohen et al., 1983). It has been used widely, but only applies to the past month, does not cover specifics like physical and sexual abuse, and the questions may be seen to be vague. The later measure of Rosengren et al. (2004) used in the landmark Interheart study incorporates 'major life stressors' which the participant can define from a list, discriminates between stress at work and at home and applies to the past six months.

Stress-related mood outcomes could impact performance during training. Nakkas et al. (2016) used measures of anxiety, depression and hostility as surrogates of stress. In their study of 675 male Swiss military trainees, all three demonstrated an improvement in stress by week 8. Lower anxiety, depression and hostility scores were associated with increased chance of promotion at the end of training (i.e. better training outcome). In his study of 176 US Air Force cadets, Banning (2010) found low scores of homesickness were associated with better training grades in male cadets (r -0.314, $p<0.001$), but not females (r -0.015, $p>0.05$). Likewise, mood traits consistent with low stress levels may be associated with better physical fitness. In US Army basic trainees (n=300, 22% female), those higher baseline fitness test scores (incorporating strength and aerobic fitness) had lower likelihood of depression at the end of training compared with low fitness test scores, after adjustment for sex and other demographics (odds ratio (OR) 0.40, 95% confidence interval (CI) 0.19–0.84) (Crowley et al., 2015). There was no difference between high and low fitness groups in depression at baseline, although no clear causal relationship between stress and fitness could be made.

Conversely, increased stress is closely associated with training-related injuries. The long-standing framework of Williams and Andersen (1998) has been used to predict sports injuries (particularly overuse injuries) from psychological stress levels in athletes. The framework was validated in a systematic review and two meta-analyses (Ivarsson et al., 2017); anticipatory stress and magnitude of response to stressors were consistently predictive of sports injuries.

An early cross-sectional study of 170 US Air Force women also reiterates the importance of stress for reproductive function. Menstrual disturbance was much

more common among Service personnel affected by stressful life events compared with civilian controls (Gordley et al., 2000). This included amenorrhea (OR, 2.20; 95% CI, 1.08 to 4.50) and abnormal cycle length (OR, 3.42; 95% CI, 1.12 to 10.50). Military women without exposure to a stressful life event were not at significantly altered risk compared with controls. This study did not explore stress or mood ratings, or the working pattern of these women, who had completed training.

Scales for measuring depression and anxiety abound, but studies in basic military training generally choose those that are brief and well validated in young adults, such as POMS (Lieberman et al., 2008; 2016b), patient-health questionnaire 7 (Adler et al., 2015; Warner et al., 2007) and/ or generalised anxiety score 7 (GAD-7) (Adler et al., 2015). Longer questionnaires such as Beck depression and anxiety inventories (BDI and BAI, respectively) are also sometimes used (Alfonso et al., 2006; Carbone et al., 1998)

Autonomic nervous system function represents another marker of stress. Increased sympathetic activity mobilises energy stores and enhances cardiovascular function in response to a stressor (Shaffer and Ginsberg, 2017). Conversely, higher cortical functions like emotional processing and attention are linked to vagal activity (McCraty, 2017). Studies of psychological stress use non-invasive measures of stress, especially HRV and salivary alpha-amylase. Salivary glands are innervated by parasympathetic and sympathetic fibres and alpha-amylase, one of the major protein constituents of saliva, is released in response to stress (Garrett, 1999). Several studies have demonstrated it to be a reliable marker of psychological stress and associated effects on mood (reviewed in Nater and Rohleder (2009)). Sustained cardiovascular responses to stress, with latent resolution after the stressor had been removed have been demonstrated after military captivity training (Lieberman et al., 2016a) and in military sexual abuse victims (Lee and Theus, 2012). Alpha-amylase and HRV were effective at discriminating responses to a TSST among military participants (Boesch et al., 2014), while pre-deployment relaxation training are associated with increased HRV following psychologically stressful training in a predominantly male military cohort (Lewis et al., 2015).

Delineating physical, metabolic, neuroendocrine and psychological effects during military training is difficult due to the immersive nature of training and its multifaceted stressors (Alemany et al., 2008; Hoyt and Friedl, 2006; Opstad, 2001). However,

extant studies consistently demonstrate that stressors encountered during military training have adverse psychiatric, metabolic and training-related outcomes. Some have suggested that low EA, not stress, accounts for reproductive dysfunction and associated pathology during exercise (Loucks et al., 2011; Loucks and Redman, 2004). There is enough evidence to hypothesise that, independently of low EA, psychological stress mediates HA (Michopoulos et al., 2013) impaired bone health (Moran et al., 2013) and increased risk of training injury (Ivarsson et al., 2017).

1.4.4.1 Resilience

Psychological resilience, the ability to overcome the negative effects of setbacks, can be considered a counterpoise to psychological stress and is an important concept for military training (Nindl et al., 2018). A cross sectional study of 696 civilian women identified that psychological resilience (from an 11-point resilience scale, RS-11) was associated with less menstrual disturbance and associated psychological stress (Palm-Fischbacher and Ehlert, 2014). Scales for measuring resilience are manifold. Strong scores from a similar measure, Angela Duckworth's Grit scale, independently predicted retention and better training outcomes in a cross-sectional study of US basic military trainees (Duckworth and Eskreis-Winkler, 2013). The Connor-Davidson Resilience Scale (CDRISC) has been widely used and validated in both men and women in the military demonstrating beneficial associations for a wide variety of mental health and physical performance outcomes (Davidson, 2018). An abbreviated 10-point version (CDRISC-10) demonstrated strong psychometric properties (Davidson, 2018). It is widely held that Service Personnel who are more resilient cope better with military operations and the stress of war, but whether men or women are more psychologically resilient is far from clear (Nindl et al., 2018). A survey of 665 US Armed Forces veterans (298 (45%) female) found higher CDRISC-10 scores among men than women (29.6 ±7.64 vs 27.5 ±7.87, Cohen’s d 0.27) (Portnoy et al., 2018). While p values were not reported, this observation was no longer ‘significant’ after hierarchical regression for trauma type experienced during deployment. Thus, the authors explained that differences in resilience scores were due to different traumatic events experienced by men and women. Men were more likely to score higher than women on a Deployment Support Scale measuring social support received during deployment (44.5 ±11.0 versus 39.0 ±13.1 Cohen’s d =0.46). Measuring psychological traits of

resilience alongside those of stress would therefore be important in studies of basic military training.

1.4.5 Sleep deprivation

Sleep deprivation is a common feature of military training, serves to increase its arduousness, reflecting circumstances encountered in military operations and war (Luxton et al., 2011; Seelig et al., 2010). Using structured discussions in focus groups in US Army basic training, Crowley et al. (2012) found sleep disruption was commonly attributed to noise, work, 'stress' and, perhaps concerningly, hunger. Discussion groups were segregated according to fitness level (high, medium and low). The low fitness group reported greater sleep deprivation than medium or high fitness groups. This might relate to reporting bias or dissatisfaction among cadets doing less well during training, rather than representing a true association, although it is plausible that increased sleep deprivation during training could impair athletic performance (Marcora et al., 2009).

Sleep deprivation is a core component of US Army Ranger training; average sleep duration is 3 to 4 hours per day for a 61 day course (Friedl et al., 2000). Since such training is also associated with severe nutritional deficit and exercise, it is difficult to delineate the importance of sleep (Drain et al., 2017). Norwegian Ranger training lasts 7 days with no organised sleep, but also little or no energy intake (Hoyt et al., 2006). The metabolic and endocrine effects of such training are profound and include loss of lean mass (Friedl et al., 2000; Friedl et al., 2001; Hoyt et al., 2006). Exposing subjects to sleep deprivation in the absence of realistic military training would be unethical, hence ascribing a relative contribution of sleep deprivation to outcomes from Ranger training would be speculative. Spiegel and Leproult et al. conducted studies of male college students whose sleep was restricted to 4h per night without caloric restriction (Leproult and Van Cauter, 2011; Spiegel et al., 1999). Isolated sleep restriction was associated with reduced glucose tolerance, thyroid stimulating hormone and HRV (Spiegel et al., 1999). Sex steroids were suppressed, but significant perturbations to the HPA axis were not seen (Leproult and Van Cauter, 2011). Others have shown that sleep quality but not quantity may affect the HPA axis. In a study of 73 undergaduate students (44 (61%) female), Bassett et al. (2015) performed TSST alongside a questionnaire-based, cross-sectional assessment of sleep. Men reporting lower sleep quality had blunted cortisol response to a TSST, but in women with impaired sleep quality there was no

effect on cortisol response. Sleep quantity did not affect TSST outcome in men or women. The authors proposed that sex-hormone associated differences in HPA axis responsiveness to stress (Kirschbaum et al., 1992) may outweigh the importance of sleep.

1.4.6 Extremes of climate

Military operations frequently involve long periods of exposure to heat and/or humidity. Resilience to diverse environmental conditions is therefore an important asset for a fighting force. Heat illness has been a significant problem for the UK military for many years, and may present a significant problem to women in GCC roles (Stacey et al., 2014).

A degree of exercise in hot, humid environments (heat stress) is incorporated into British military training in the UK and overseas. Elevated core temperature arising from environmental conditions and/ or exercise is associated with heat illness. This syndrome represents a spectrum of conditions ranging from the more modest heat cramps or rash, to heat exhaustion (inability to continue to work), heat syncope (brief loss of consciousness) and heat stroke (Raukar et al., 2015). Heat stroke, a catastrophic rise in temperature above 40ºC accompanied by severe neurological manifestations, is further classified as classical or exertional when it arises, respectively, from environmental exposure or exercise (Bouchama and Knochel, 2002).

Assessing heat exposure can be undertaken prospectively in a humidity and temperature-controlled environment, for example to monitor sweating and vasodilation responses alongside changes in core temperature. Detailed assessments of evaporative cooling and alterations in cutaneous circulation have demonstrated subtle sex differences at high heat stress (a combination of heat and exercise), when women sweat less (Gagnon and Kenny, 2012a) and vasodilate less (Wong and Hollowed, 2017) than men. However, differences in thermoeffector responses and core temperature do not equate to heat illness, which arises as an idiosyncratic inflammatory response (Leon and Bouchama, 2015). Sex differences in the occurrence of heat illness should therefore be assessed first from observational or cross-sectional studies, especially those conducted in military training.

Role-specific training for soldiers (e.g. advanced training for Royal Marine Commandos) is undertaken in the Arctic, preparing trained units for operations in cold environments. Sex differences in thermoregulatory responses to extreme cold have been demonstrated; women experience lower core temperatures (Graham, 1988), and have reduced capacity for heat generation through shivering (Wagner and Horvath, 1985) during cold stress than men. The incidence of cold injuries in military training is greater than heat illness (Kuht et al., 2018), however thermoregulatory responses to cold have been studied far less than those to heat (Iyoho et al., 2017). Sex differences in cold thermoregulation have important implications for predicting responses and designing strategies to optimise performance (Iyoho et al., 2017). To measure outcomes of real-world importance such as cold injuries and training success, studies need to be conducted in the context of these extreme training environments. For example, such studies could observe behaviours during cold exposure and might evaluate wearable technologies to measure both environmental conditions and physiological adaptation (Nindl et al., 2018). Such technology should be carefully caveated to ensure the effects of false positives are considered and mitigated, pathological states are differentiated from extremes of normal physiology, and training is not jeopardised by high-fidelity feedback of physiological data in real time (Stacey et al., 2018a).

However, it is not clear if climactic injury (in particular heat illness) is commoner in men or women. No studies have systematically assessed heat illness according to sex, across different countries, occupations (sociocultural gender roles) and climates. Understanding if heat illness is commoner in men or women is an important prerequisite to exploring heat illness in the military.

1.4.7 Other considerations

1.4.7.1 Hormonal contraceptives

Hormonal contraceptives are an important consideration when assessing ‘real world’ hormonal milieu of women of reproductive age. Synthetic oestrogens and to a lesser extent progestogens exert negative feedback on the hypothalamus. These compounds could have important downstream effects, such as modifying circulating cortisol levels with stress-like alterations on the HPA axis (Hertel et al., 2017). Progestogen-only contraceptives may be associated with transient uncoupling of bone turnover (Walsh et al., 2008) and loss of bone mass (Berenson et al., 2004), but this effect appears to be transient (Viola et al., 2011). Oestrogen-containing

contraceptives show a variable effect, dependent on duration, with some evidence of initial benefit on bone mass from short term use (Petitti et al., 2000), which is be more pronounced with higher doses (Tiedeken et al., 2019), but with no apparent benefit from long term use (Beksinska et al., 2009) and with a slight of detriment to bone mass after continued use for 6 to 9 years (Jackowski et al., 2016; Scholes et al., 2010). One recent study in amenorrhoeic athletes demonstrated transdermal E2 replacement can improve bone mineral density in amenorrhoeic athletes, compared with placebo or combined contraceptive pill (COCP) (Ackerman et al., 2019b).

The current prevalence of hormonal contraceptive use among the British military is not known, and data of contraceptive usage more widely are lacking. One cross-sectional database study of over 500,000 women in the UK aged 12-49 years assessed use of contraceptive prescription in 2008 (Cea-Soriano et al., 2014). Hormonal contraceptive use ranged from 16.2% for COCP to 3.9% for progestogen-only implants or injections, although prevalence data were not available at ages pertinent to basic military training (18 to 29 years). Ages of new users were reported and were lower for COCP than progestogen-eluting implant, progestogen-only pill, or levonorgestel releasing intrauterine system (mean (95% CI) age 19.2 (19.1, 19.3) years, 25.2 (24.9, 25.5) years, 30.7 (30.5, 30.9) years, and 38.4 (38.2, 38.6) years, respectively; no statistical comparisons were made).

Rates of contraceptive use reported in the early 1990s by Friedl et al. (1992) and Lauder et al. (1999) among US military women aged 24 (IQR 18, 52) and 27.5 ±7.7 years, were 34.9% and up to 33.6%, respectively (**Table 1-2**). A recent cross-sectional survey of 4,456 civilian women in the UK found hormonal contraceptive use rates for women were significantly higher – aged 16 to 24 years: 52.0% (95% CI 49.1 to 55.0) and aged 25 to 34 years: 43.2% (40.7 to 45.7) (Firman et al., 2018). Recent trends suggest women increasingly favour long-acting reversible contraceptives (LARC) over oral contraceptives. The proportion of women attending Sexual and Reproductive Health Services in England using LARCs doubled from 19% to 41% from 2007 to 2018, while oral contraceptives use fell from 49% to 42% (NHS Digital, 2018). Similar figures have been found in the US military (Batig, 2017)

1.4.7.2 Age

Age is an important consideration to the measurement of endocrine adaptations to training. The gynaecological age is taken as the number of years after menarche

(mean age 12y). Typically women achieve the highest rate (94%) of ovulatory cycles at gynaecologic age 12y, and as female athletes advance beyond this age, the effect on reproductive function may be attenuated (Loucks, 2006; Petit and Prior, 2013). The majority of studies demonstrating luteal phase defects following reduced energy availability describe the effects on gynaecologically mature women to reduce confounders (Roupas and Georgopoulos, 2011). However, most military recruits are younger than this (Schneider et al., 1999). The comprehensive review of Maimoun et al. (2014) describes additional vulnerability to the Triad, especially menstrual disturbance, conferred by younger age. The authors also point to the thyroid and growth hormone axes, which are vulnerable to disruption from reduced energy availability at an earlier age.

Younger age may accentuate the impact of the psychological stress of military training on reproductive function. Sports that expose adolescents to high levels of psychological stress are associated with hypercortisolaemia and menstrual disruption (Ackerman et al., 2013), while the developing emotional maturity of adolescence may also confer susceptibility to anxiety about body image and disordered eating (Hackney, 2008). These observations could have significant ramifications for young women in military training.

1.5 Restraints and limitations to conducting physiological studies in the field

It has been seen that the assessment of endocrine adaptations to military training could be so detailed that they would alter the training itself. This would be self-defeating. Assessing these adaptations together with measures of military training's impact, or arduousness, need to be integrated as subtly as possible into training to avoid altering the careers of participants or the training itself, and to sustain voluntary participation in the study. An observational study framework is required that concurrently assesses patterns of reproductive hormones, corticosteroids, energy balance and bone health, alongside measures of the arduousness of training in military women, without imposing excessive study burden.

The requirement arose to find an appropriate cohort of military women around whom I could to construct such a framework. Infantry training centres, such as at Catterick, Brecon, Lympstone and Honington had promising infrastructure, but unfortunately did not have female candidates. The regular Commissioning Course at the Royal

Military Academy, Sandhurst (RMAS) was a possibility, being well-established, very arduous relative to other basic military training, and mixed sex.

The regular Commissioning Course is undertaken three times per year and consists of three fourteen-week terms interspersed by a week of adventurous training (e.g. mountaineering, water sports or skiing) and 2-3 weeks of leave. Around 220-270 Officer Cadets commence the course during each intake (approximately 5-12% are female and a similar proportion are from overseas militaries). The age of Cadets is 23.7 ±2.2 years at entry, the majority (at least 80%) are university educated and around 40% attended independent schools. Training takes place in mixed platoons of around 30 Cadets. Term 1 (junior term) involves an immersive introduction to military life followed by several rigorous assessment exercises. Officers have recalled the first five weeks as being 'the toughest thing you'll ever do' (BBC News, 2011). Several summative assessments take place during exercises in the second (intermediate) term, including interview-based selection for destination corps or regiment ('choice of arm') in week 12. Success at this highly competitive process is seen as the most crucial assessment of the Course. Thereafter, term 3 (senior term) consists of academic tuition and assessments (war studies, defence and international affairs and communication and applied behavioural science) and applied leadership training. A large exercise in Germany takes place in weeks 9 to 11, before a high-profile drill event for commissioning (Sovereign's parade).

A key restraint imposed is the number of women applying to start the Commissioning Course. Moreover, training Officer Cadets represents a paradigm of arduous training *status quo*, rather than the anticipatedly more arduous WGCC. The Commissioning Course had clear advantages, being *ab initio,* infantry-based and much longer than most extant studies, which generally examined training lasting 6 to 12 weeks. However, because it is already well-established, it is likely to be relevant to current female military employment as well as the future female infantry.

The need therefore arose to compare and contrast the effects observed in RMAS with women undertaking more arduous training. To achieve this, I explored opportunities to study extremely arduous training among military women and was successful in my bid to study Expedition (Ex) ICE MAIDEN. I proposed to transpose the observational model established at RMAS to the study of six women attempting a ski traverse of Antarctica. This expedition was an attempt at becoming the first all-

female team crossing of the continent without the assistance of wind power and aimed to take around 75 days.

Both contexts presented significant restraints. During the Commissioning Course, Officer Cadets frequently undertake programmed activities from 07:30, and large sections of the course are spent on exercises living in tented accommodation in remote parts of the UK or abroad. During Ex ICE MAIDEN, participants pulled everything needed for the expedition in sledges, with two resupply points along the 1,700 km route. However, I worked around these restraints by engaging participants in planning study visits, and performing detailed, repeated measures, mobile technology and state-of-the-art imaging, examining several hormonal and metabolic domains concurrently.

Study designs needed to minimise interruption to the training of these women, who for different reasons were facing significant expectations and pressures. It was necessary to devise measurements which could assess several domains of endocrine adaptation and arduous training impact concurrently. Daily blood tests, while ideal for determining changes in hormone milieu over time, were not feasible in either setting. Urine and saliva samples had to be limited to the minimum number possible. Unfortunately, this meant a cortisol awakening response was not possible. Cadets were often parading within 30-45 minutes of waking, and busy and probably stressed immediately beforehand. It was therefore impossible to measure the stress response to waking in isolation. The sustained, continuous requirements on Cadets' time, imposed restraints which would not apply to most Service personnel. For example, it was not possible to schedule indirect calorimetry or VO_{2max} tests around the Commissioning Course timetable, but these could be conducted before and after Ex ICE MAIDEN.

Whilst the Ex ICE MAIDEN team were 'on the ice' in Antarctica (**Figure 1.5**), biological samples were not possible as participants would be required to carry all consumables and samples in sledges (the team also carried all packaging waste). On-ice measurements were therefore limited to mobile phone-based questionnaires and wearable biosensors (for a pilot study conducted by other investigators).

Figure 1-5 Training for Ex ICE MAIDEN. Participants carried all food, supplies and waste in sledges, which weighed around 80kg. This precluded on-ice biological samples (with kind permission of Ex ICE MAIDEN Higher Management Committee)

The expedition team also had public and diplomatic commitments before and after the expedition, making scheduling of measurements difficult. While it was not possible to conduct measurements immediately before and after the expedition, every effort was made to conduct assessments as soon as was practical.

We used existing structures from the Commissioning Course and Ex ICE MAIDEN to provide data, where these were available. For example, the best effort 1.5 mile run, which is undertaken five times during the Commissioning Course and once beforehand, correlates closely with VO_{2max} (Mayorga-Vega et al., 2016). Retrospective measures such as hair cortisol concentration and psychological stress questionnaires (Rosengren et al., 2004) meant infrequent study visits could provide data pertaining to the intervening weeks (discussed in **Section 1.4.3**).

One important consideration was the requirement for a control group, especially for the study at Sandhurst. This would allow endocrine changes in Officer Cadets to be controlled for the stress of adjusting to a new environment. Controls would need to be well-matched demographically and undertaking a significant change in life circumstance, such as starting a degree course. Unfortunately, it was difficult to find

a sufficiently large population of mainly postgraduate-level civilians undertaking a programme analogous to military training. Professional or high-level sportswomen, or entrants into the police service might provide this opportunity, however no appropriate local setting for this could be identified. Conducting studies in two geographically remote cohorts was unfortunately not logistically feasible.

1.6 Overarching hypothesis and aims

There is an imperative need to understand female endocrine adaptations to arduous training. Neither the female endocrine responses nor the precise stressors have been characterised during military training. This thesis originated from an unprecedented opportunity to measure detailed, relevant adaptations of women during arduous military training, within the necessary constraints of minimising interruption to it.

1.6.1 Hypotheses

1. *Women undergoing basic military training and those undertaking an arduous Antarctic crossing will experience short-term suppression in:*
    a. *HPG axis responsiveness*
    b. *HPA axes responsiveness (measured in early morning cortisol and in response to ACTH)*
    c. *Uncoupling of bone turnover and bone mineral content*
    d. *Evidence of negative energy availability*

The EA hypothesis, while derived predominantly through small studies in civilian athletes training for sports competitions, is ostensibly central to understanding the aetiology of reproductive dysfunction observed in female military trainees (**Section 1.2.2**). Hyperandrogenism observed in a subset of athletes may also be important in some military women (Warren and Perlroth, 2001), while added complexities of HPA axis activation and psychological stressors are likely to make a significant contribution (Selye, 1939; Sharma et al., 2013; Vulliemoz et al., 2008) (**Section 1.2.2.1**). Perturbations in HPG axis function can be prevented by administration of leptin (Chou et al., 2011), which is produced proportionately by adipocytes, while loss in lean mass during Ranger training suppresses the production of sex steroids. Thus, while the EA hypothesis predominates in the literature on female athletes (Mountjoy et al., 2018), the stress and body composition hypotheses retain

credibility when considering the concept of arduousness more broadly (**Section 1.2.2.2**).

Arduous military training is multifaceted, presenting a combination of complex physical and mental challenges. There is no single measure to quantify the arduousness of training. Instead, arduousness may be considered a virtual construct of a number of measurable vectors (the term vector implies both a magnitude and temporal direction or trajectory over the duration of training), such as those described in **Section 1.4**. A combination of individual and environmental factors may therefore lead to endocrine adaptations across a number of axes (Gifford et al., 2017). The cumulative effect of such factors could be to suppress HPG and HPA axes and bone health in women in a manner which could predispose military trainees are to subfertility, adverse response to psychological stressors and low trauma/ stress fractures over time. This may lead to end-organ effects including blunting of the HPA axis and adverse psychological outcomes, impaired bone health and stress fracture risk, and long-term subfertility (**Figure 1-6**). The studies in **Chapters 3 and 4** analyse many of the same variables to test hypotheses 1a to 1d, in basic military training and arduous training in extreme cold, respectively. The environmental influences (green box in **Figure 1-6**) are different between these groups, in particular nutrition, abruptness, psychosocial stress and sleep deprivation. Individual influences which may vary within the populations studied, such as DE and hormonal contraceptive use, were assessed by testing the effect of these on biological outcomes directly.

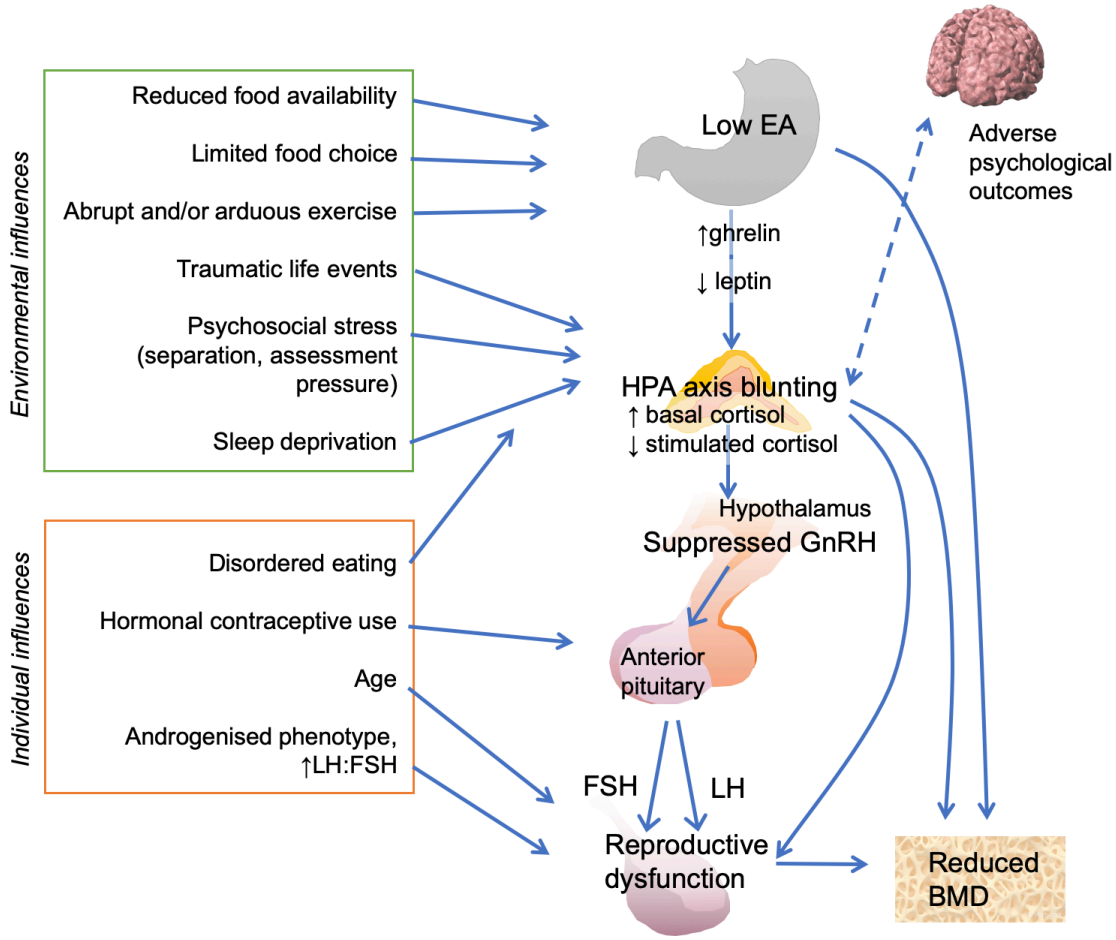


Figure 1-6. Hypothesis: the cumulative, negative effect of individual and environmental factors causes reproductive dysfunction, alterations in the HPA axis, reduced BMD and subfertility. EA energy availability, HPA hypothalamic-pituitary-adrenal, GnRH gonadotrophin releasing hormone, LH luteinising hormone, FSH follicle-stimulating hormone, BMD bone mineral density

2. Women are at greater risk of heat illness than men.

Progesterone causes a rise in core temperature during the latter half of the menstrual cycle, while women sweat and vasodilate less than men at high heat stress, predisposing them to higher core temperatures (Gagnon and Kenny, 2012a; Wong and Hollowed, 2017). However, the relative rates of heat illness in men and women are unknown. Hypothesis 3 will be tested in **Chapter 5**.

Null hypotheses

1. *Arduous military training does not activate the HPA axis in women (or men).*
2. *Arduous military training is not associated with any change to reproductive endocrine function in women (or men).*
3. *Arduous military training does not impair bone health in women (or men).*
4. *Arduous military training is not associated with an energy deficit in women (or men).*
5. *There will be no difference in points 1 to 4 between women and men.*
6. *There will be no difference in rates of heat illness between women and men.*

1.6.2 Aims

This thesis aims to characterise female endocrine adaptations to arduous military training. To achieve this, two cohort studies of military training prospectively evaluated the HPG axis, HPA axis, ANS, EA, EB, metabolic and nutritional markers, body composition, bone mineral density and bone turnover.

Sex hormones are key effectors of differences between the sexes and reproductive function links these endocrine and metabolic domains. Therefore, this thesis aimed to evaluate changes to female reproductive function during prolonged military training. To achieve this, I developed a means of feasibly measuring ovulation during the time-pressured environment of the Commissioning Course, using weekly urine samples accompanied by a menstrual diary over its 44 weeks. Since this process could potentially yield thousands of samples, we sought to determine if the current ELISA for pregnanediol glucuronide could be substituted for a newer, automated assay of progesterone. This would also have wider relevance for detection of ovulation in a clinical setting (**Chapter 3.1**)

This study programme centred around endocrine adaptations during the Commissioning Course. We aimed to determine EA and EB during the Commissioning Course while evaluating a means of measuring EA alongside aerobic capacity and HRV (**Chapter 3.2**). We aimed to measure the psychological stress of training and changes in basal and dynamic function of the HPA axis in women during the course (**Chapter 3.3**). We aimed to detect longitudinal changes in

reproductive function during training, examining the HPG axis and gonadal function, comparing changes in women using hormonal contraceptives since this will reflect real-world hormonal milieu for women of childbearing age. We also compared the reactivity of the HPG axis to men (**Chapter 3.4**).

We aimed to evaluate female endocrine adaptations to arduous exercise in extreme cold, studying the six members of the Ex ICE MAIDEN team before and after the 1,700 km Antarctic traverse undertaken in November 2017 to January 2018. I examined reproductive, adrenal and metabolic changes (**Chapter 4.1**) beforehand as well as 5 and 15 days after, to assess recovery. Given the extraordinary success of the expedition, we aimed to describe the selection, training and nutritional prior to Ex ICE MAIDEN to consider if any findings could be exploited for military training (**Chapter 4.2**).

Lastly, I aimed to consider the female response to thermal stress, by assessing whether heat illness is commoner among women. I aimed to systematically review extant literature reporting sex-associated rates of heat illness and performed a meta-analysis to determine the relative risk of heat illness in women (**Chapter 5**).

In aiming to investigate the multiple stressors and responses, I hope to enhance the ability of Defence to optimise the wellbeing of servicewomen and men (**Chapter 6**). Such work could provide a platform from which to launch long-term cohort studies, assessing the impact of arduous training and GCC on long-term physical and mental health, particularly after events such as operational deployment. It is important that military personnel are aware of the consequences of serving their country, especially when women are likely to do so before they have reached gynaecologic maturity. Since the PhD process would last 3 years, and some of the findings described in **Sections 1.2** and **1.3** might be concerning for military policymakers, my supervisors, other experts in nutrition, bone health and reproductive endocrinology and I published a narrative review describing what was known about reproductive dysfunction and associated pathology in women undergoing military training at the beginning of my PhD studies (abstract included at **Appendix A**). We included recommendations for relevant research (**Table 1-4**). Those addressed by this thesis are highlighted

Topic	What is currently known	Recommendations
Incidence of reproductive dysfunction*	Incidence estimates of menstrual dysfunction in athletes vary greatly (Gibbs et al., 2013). The prevalence in the UK military is not known.	Prospective assessment of changes in ovulation and LPD during arduous training using both self-reporting and biochemical assessment (Ahrens et al., 2014; Reed et al., 2015)
Prevalence of DE and osteopoenia*	Prevalence estimates of osteopenia and DE in civilian athletes vary dramatically (Gibbs et al., 2013). The prevalence of all 3 Triad components combined appears to be low, however the prevalence in the UK military is not known (Lauder et al., 1999; Mallinson and De Souza, 2014).	Prospective cohort studies to determine the prevalence of all Triad components in UK military trainees Bone health and body composition should be assessed using modern imaging modalities. (Mallinson and De Souza, 2014). Validated, robust dietary assessment and eating behaviour tools should be used (Gemming et al., 2015; Martinsen et al., 2014; Melin et al., 2014; Rangan et al., 2016; Thompson et al., 2010)
Influence of androgenisation*	Athletes undertaking non-lean sports are predisposed to reproductive dysfunction to a lesser degree than athletes undertaking lean sports (Gibbs et al., 2013). The mechanism is different, probably involving androgen excess (Hagmar et al., 2009). Androgens might be beneficial to bone health (Javed et al., 2015).	Prospective studies characterising androgenisation in military trainees, associated reproductive changes (FHA and/or PCOS), bone and performance benefits, and whether or not androgenisation changes during training.
Influence of HPA axis and gut peptides*	Cortisol (Ackerman et al., 2013; Taylor et al., 2014) and leptin (Ackerman et al., 2012; Scheid et al., 2013) have been implicated in the pathophysiology of reproductive dysfunction in athletes.	Prospective studies characterising baseline levels of cortisol and gut peptides against controls Observing the changes in these hormones in association with changes in measured reproductive function.
Psychological impact of military training*	Psychological stressors may increase reproductive dysfunction (Gordley et al., 2000; Pauli and Berga, 2010) and reduce energy intake (Ulrich-Lai et al., 2015).	Anxiety, depression (Marcus et al., 2001), PTSD (Lieberman et al., 2016a), and resilience (Palm-Fischbacher and Ehlert, 2014)) should be measured alongside reproductive function.

Long term effects, injury	The effect of exercise associated reproductive dysfunction on long term child-bearing potential is unknown. Stress fracture is associated with reduced BMD and female sex (Knapik et al., 2012; Rauh et al., 2006).	Long-term cohorts of women in various military roles should be undertaken to assess rates of infertility, injury and physical and mental illness.
Training type	Abrupt training onset increases susceptibility to reproductive dysfunction (Roupas and Georgopoulos, 2011).	Controlled studies evaluating interventions to reduce the reproductive function
Highly arduous training*	Highly arduous military training can be associated with marked metabolic and endocrine abnormalities. Such training involves negative energy balance and sleep deprivation (Hoyt et al., 2006; Opstad, 2001)	Characterising the reproductive effects of such training in women, as well as energy availability, and implications on BMD.

Table 1-4. Recommendations for further research on reproductive dysfunction in Service women. Adapted from Gifford et al. (2017). Recommendations addressed in this thesis are marked * LPD, luteal phase defect; DE, disordered eating; Triad, female athletic triad; FHA, functional hypothalamic amenorrhoea; PCOS, polycystic ovary syndrome; BMD, bone mineral density; HPA, hypothalamic-pituitary-gonadal axis; PTSD, posttraumatic stress disorder.

Chapter 2 Clinical and Laboratory Methods

Materials used in this thesis are presented in Appendix B.

2.1 Clinical Studies

The work presented in this Thesis relates to the following studies:

1. The Female Endocrinology in Arduous Training (FEAT) Study (ethical approval 790/MoDREC/16).
2. The Exploring the Endocrine Effects of Extreme Training (E4T) Study (ethical approval 870/MoDREC/16).
3. A systematic review and meta-analysis (**Chapter 5**) conducted in response to a request by WGCC for a Technical Report, "Does female sex confer greater risk of heat illness?".

2.1.1 Study populations

2.1.1.1 FEAT Study

We extended recruitment at RMAS from CC172 and CC173[1], owing to lower numbers of female recruits than anticipated, to include (CC181[2]). Due to low numbers of women initially in CC172, we included a smaller comparator cohort of male Officer Cadets, limited by the ethics committee to n=10. This gave additional benefit of pilot comparison data to compare endocrine and metabolic changes during training between men and women. A comparison of men and women was not the primary aim of the study, which remains within-subjects changes among women (data from men is not included in this thesis).

Officer Cadets undergo medical screening to meet stringent requirements prior to enrolment at RMAS. Cadets were invited by poster to a briefing course at the PCCBC, a 3-day familiarisation visit consisting of physical assessments, exercises and lectures. Inclusion criteria were:

- Entering Commissioning Course at RMAS

[1] The second and third Commissioning Courses, respectively, starting in 2017

[2] The first Commissioning Course starting in 2018

- Aged 18-30 at time of recruitment
- Female (except for 10 men during CC172)
- Voluntary consent

Exclusion criteria:

- Current pregnancy
- Withdrawal of consent
- History of adrenal, ovarian/testicular, ACTH or GnRH insufficiency
- History of pituitary disease or diabetes
- History of osteoporosis or hyperparathyroidism
- History of oral, inhaled or topical steroids use current or in past 3 months

PCCBC took place at RMAS, 2 to 14 weeks prior to the commencement of the Course. For any military study, particular care must be taken to ensure participation is entirely voluntary and not related to subordination by commanders or coercion. At the initial briefing volunteers received an oral presentation by investigators in civilian clothes, not identified as military Officers. This briefing detailed the reasons for the study, its voluntary and confidential nature, what it entails, the right to withdraw without reason at any time, and that participation would not affect the individual's success on the Course or future career. Volunteers were given 24 hours to consider the written information given at the brief and were invited to the start of the study if interested in taking part. **Figure 2-1** illustrates the recruitment and follow-up. **Figure 2-2** illustrates the overlap between three Course intakes, each with three, 14-week terms, and lists the concurrent measurements undertaken.

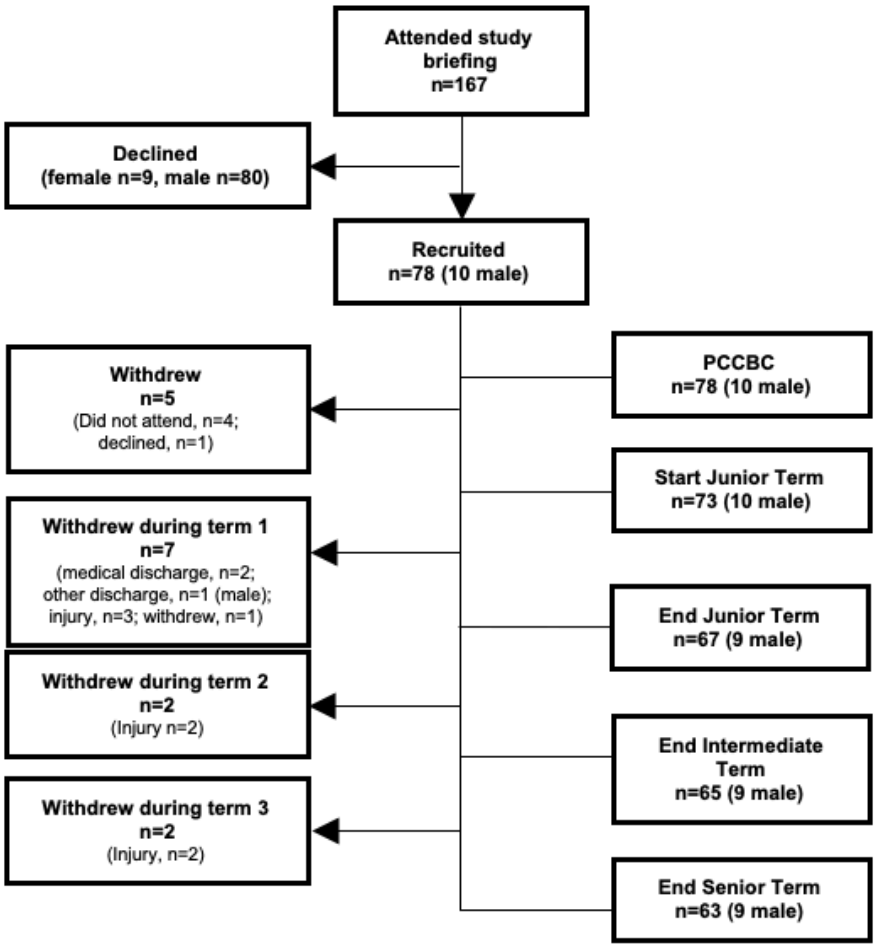
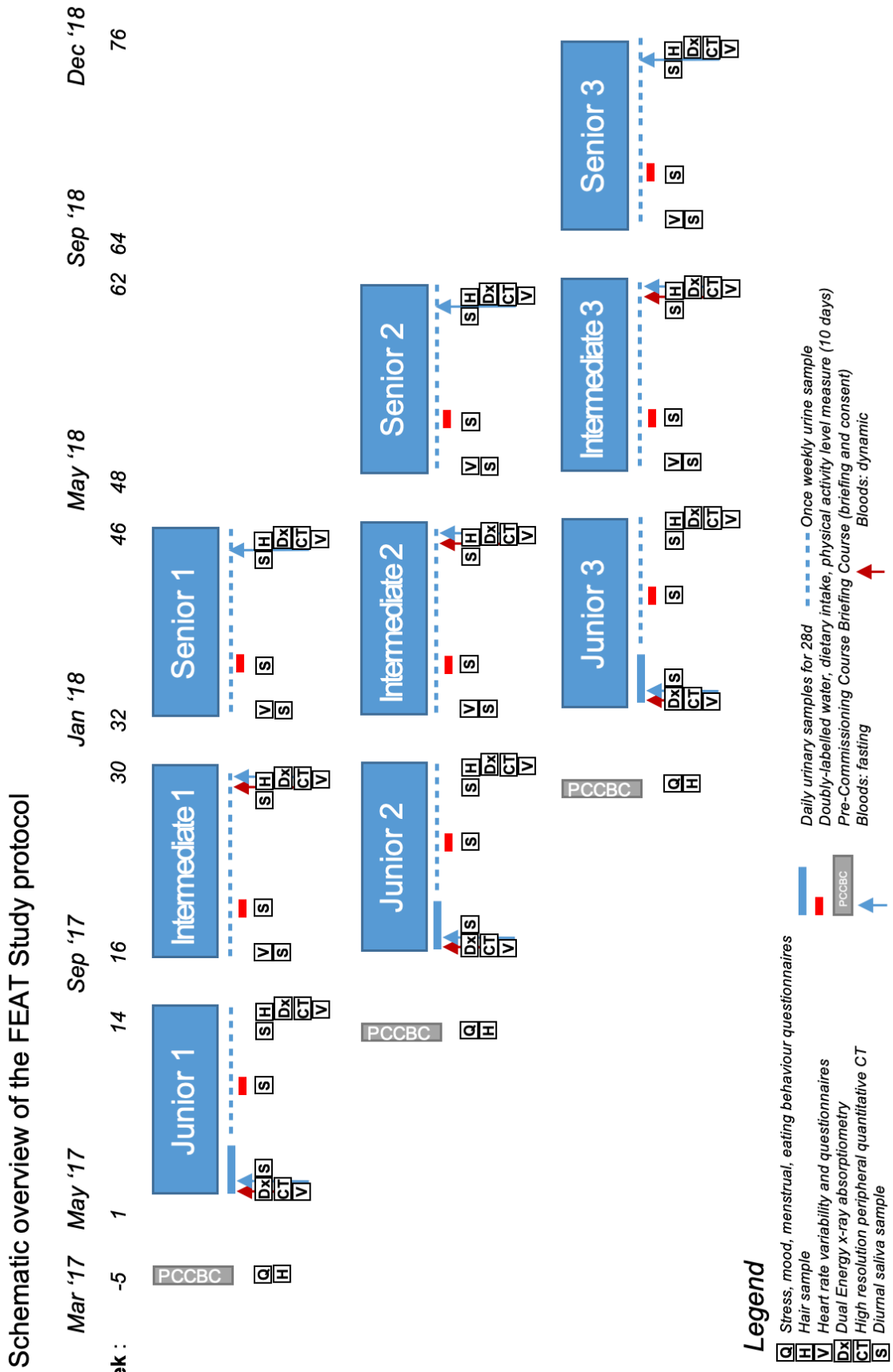


Figure 2-1. FEAT Study Recruitment and Follow up. Terms 1, 2 and 3 and junior, intermediate and senior term are used interchangeably. Study briefing took place during PCCBC (pre-Commissioning Course Briefing Course).

Figure 2-2. Schematic overview of the FEAT Study protocol



2.1.1.2 E4T Study

Participants were invited from Ex ICE MAIDEN (the expedition), a military adventurous training exercise aiming to cross the Antarctic landmass by foot (October 2017– January 2018). The team had already been selected for the expedition and therefore participation, or otherwise, in the research could in no way influence acceptance onto the team.

Physiological research was a key rationale for planning the expedition as stated in expedition publicity (see https://exicemaiden.com/), the team undertook the expedition on a voluntary basis and investigators were invited by the expedition team to undertake this study. Therefore, a high rate of recruitment from the team was anticipated but this did not represent coercion.

2.1.1.2.1 Statistical considerations for the number of participants.

The maximum number of participants was 6 – the number of subjects on the expedition. Below this, only a descriptive analysis of the results would be possible. Small-sample research in medicine and physiology is always problematic from a statistical perspective. However descriptive analysis of measurements, particularly changes in measurements pre- and post-Ex have the potential to provide valuable insights into the physiological changes and adaptations that have occurred in individuals. It should be noted that the breadth of measurements made in this study supported this. Statistical analysis of small samples is possible although the results need to be treated with caution as the power is low. Simple numerical (Monte-Carlo style) modelling suggested that n = 5 was the minimum number of participants required to achieve statistically significant results of difference in resting metabolic rate at $p<0.05$ based on matched-pair comparisons.

Participation in the expedition was contingent on passing a rigorous pre-participation medical examination by an independent medical officer and participants had to be classed as medically fully deployable. The inclusion criterion was therefore only ‘participation on Ex ICE MAIDEN. Exclusion criteria were ‘known allergies to study medication used, withdrawal of consent and steroid use (oral, topical or inhaled) in past 3 months’.

The study comprised three repeated measurements of endocrine and metabolic function, once pre-Ex (in the Research Facility in RMAS) and twice post-Ex (in Punta Arenas, Chile and in RMAS). When the study was being designed, the start

date and duration of the expedition were not known. The protocol is illustrated in **Figure 2-3.**

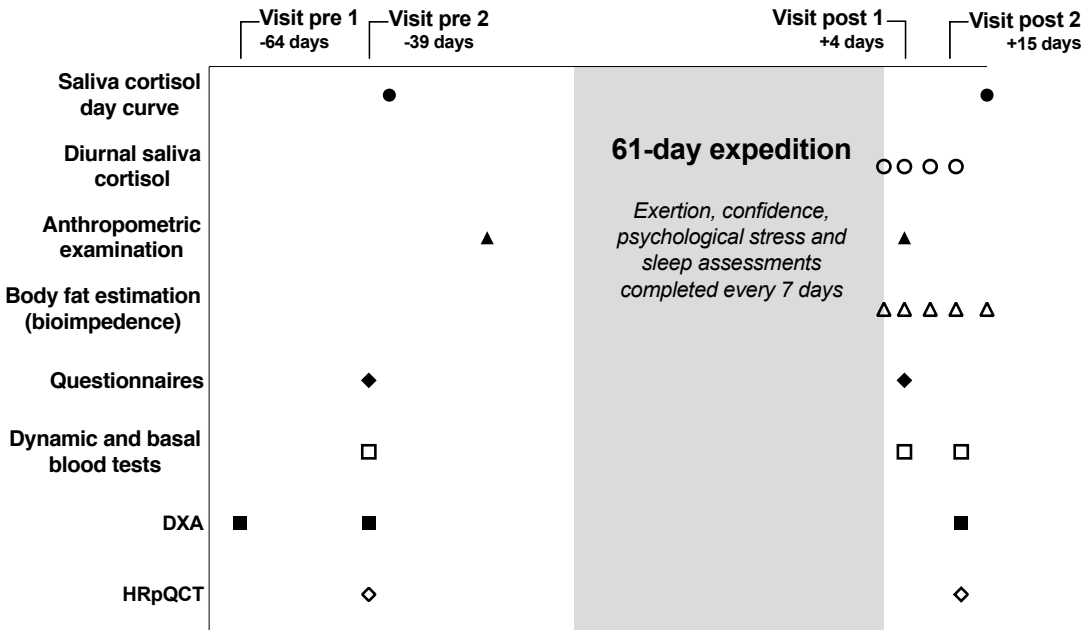
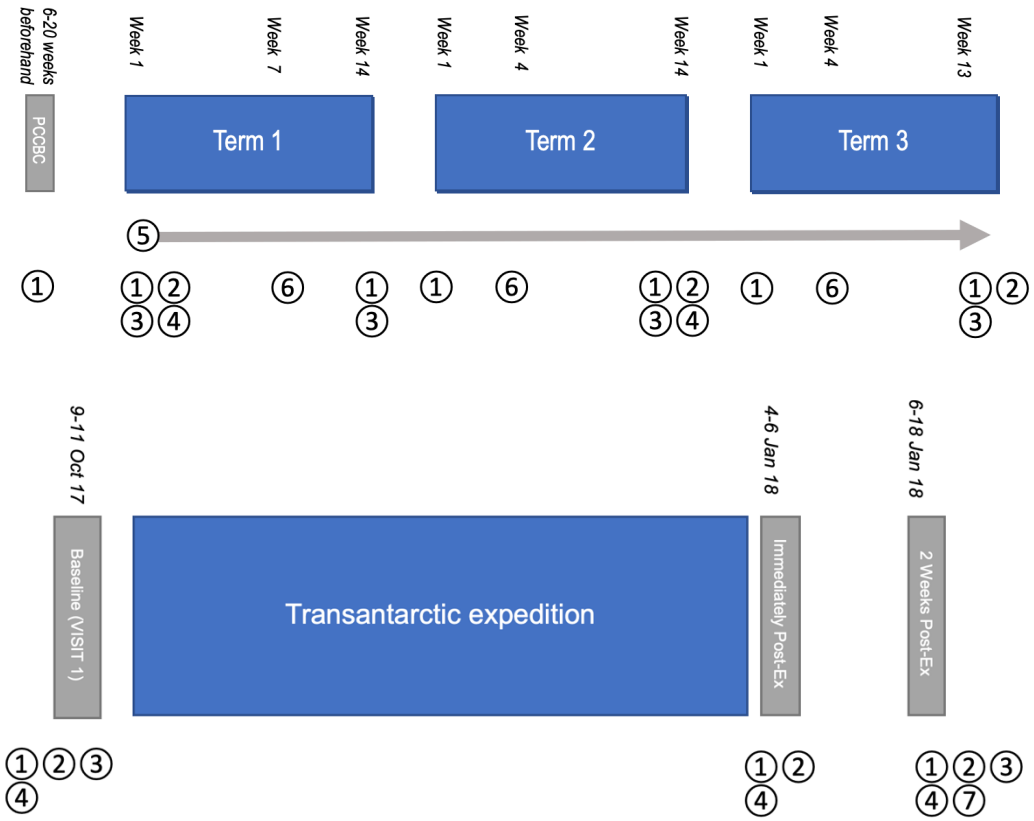


Figure 2-3. E4T Study Protocol. DXA: dual x-ray absorptiometry, HRpQCT: high resolution peripheral quantitative computerised tomography of tibia

2.1.2 Standardised methodology

The FEAT and E4T studies comprised serial measurements of change in 4 domains concurrently: reproductive function, function of the HPA axis, bone health and energy availability. The FEAT and E4T studies aimed to assess the pattern of change in these domains during the Course and an unassisted Antarctic traverse, respectively. The four domains were treated as output/ dependent variables, while the Course or Antarctic training was treated as an input/ independent variable. Both studies utilised discrete basal and dynamic measurements to assess these domains. These were designed to fit around the training. The scheme of investigations is laid out for one FEAT Study cohort in **Figure 2-4**.



	①	②	③	④	⑤	⑥
Reproductive function	Questionnaire	Fasting bloods		Dynamic low dose GnRH test	Serial urinary progesterone and menstrual self-reporting	
HPA axis function and stress	Questionnaire, hair sample, heart rate variability	Fasting bloods		Dynamic low dose ACTH test	Serial diurnal saliva sample, x3 per term	
Bone turnover		Fasting bloods	HRpQCT			
Energy availability	Questionnaire	Fasting bloods	DEXA			10 d energy assessment Phase: EI, EEE and TEE

Figure 2-4. Comparison of domains tested in FEAT and E4T Studies. Tests in FEAT (top) and E4T (bottom) study protocols. RMAS, Royal Military Academy, Sandhurst; HPA Hypothalamic-pituitary-adrenal; GnRH gonadotrophin releasing hormone, ACTH adrenocorticotrophic hormone, HRpQCT high resolution peripheral quantitative computerised tomography, EI energy intake, EEE exercise energy expenditure, TEE total energy expenditure

2.1.2.1 Dynamic testing of the pituitary gonadotroph and adrenal cortex: the "CHOC Test"

Dynamic endocrine testing, although widely used in clinical endocrine practice, has not been used in field studies until recently (Cadegiani and Kater, 2017a; Cadegiani and Kater, 2017b, 2018b; Gifford et al., 2019a). We developed a novel application for existing dynamic endocrine tests, repeating low-dose pituitary and adrenal function testing before and after a training exposure, simultaneously testing two axes (hypothalamic-pituitary-adrenal, HPA and hypothalamic-pituitary-gonad, HPG), termed the Combined Hypothalamic, Ovarian and adrenal Cortex (CHOC) Test, for the purposes of these studies. Provided contraceptive use did not change between visits, the CHOC Test allowed repeated measures of HPG axis function in all participants.

Since the exercise and stress involved in military training would be expected to have a suppressive effect on endocrine axes (for example, as seen in studies of US Army Rangers (Friedl et al., 2000; Nindl et al., 2007)) we chose stimulatory tests, assessing subtle changes in adrenal cortex and pituitary gonadotroph function (Morosini et al., 1989; Reynolds et al., 2001a). Since only small changes would be expected during phase 1 training, and it was important to minimise any potential impact of research participation on training, we also used very low doses of ACTH-$_{1\text{-}24}$ (tetracosactrin acetate as Synacthen®) and gonadotrophin releasing hormone (GnRH, as Gonadorelin hydrochloride), examining subtle changes in luteinising hormone (LH), follicle-stimulating hormone (FSH) and cortisol responsiveness to stimulation before and after training.

2.1.2.1.1 CHOC Test Procedure

Due to constraints imposed by the training schedule, dynamic testing was completed in the late afternoon for the FEAT Study.

For the E4T Study, we were afforded more time for testing, which allowed the CHOC test to be enhanced by performing testing fasted in the morning, under central axis suppression. At 2200h, 10h prior to the CHOC Test, E4T Study participants ingested 0.25 mg dexamethasone. This dose has been used to assess the sensitivity of the HPA axis to a near-physiological level of central negative feedback and to attempt to reduce the baseline variation in morning fasting cortisol prior to the pre-stimulation test cortisol (Kajantie et al., 2003; Reynolds et al., 2001a).

The procedure for both studies was as follows. Participants were allowed to relax before a 21-gauge cannula was inserted into an antecubital or dorsal hand vein and a baseline blood sample was obtained in an EDTA-containing tube. 10 µg Gonadorelin hydrochloride (Intrapharm, Maidenhead UK), followed by 1.0 µg ACTH-(1-24) (tetracosactrin acetate as Synacthen®, Mallinckrodt, Dublin, Ireland), were then injected followed by a 10 mL saline flush. $ACTH_{-1-24}$ was freshly diluted using 249 mL 0.9% NaCl (Baxter, UK), to which Synacthen ® 250 µg in 1 mL had been added, shaken thoroughly and 1 mL of this mixture was injected using a 5 mL syringe to minimise contact with plastic. Venous blood was sampled through the cannula in EDTA-containing tubes 20, 30, 40 and 60 min after drug administration.

2.1.2.2 Reproductive function

2.1.2.2.1 History, contraceptive use

A detailed reproductive history included contraceptive use was taken at baseline. At each study visit thereafter, questionnaires were completed detailing the use of contraceptives or, if no contraceptives were used, menstrual period regularity and heaviness, and first date of the last menstrual period.

2.1.2.2.2 Fasted blood samples

Anti-Müllerian hormone (AMH), inhibin B, prolactin, gonadal steroid hormones, LH and FSH and adrenal androgens were assayed from a single fasted blood sample.

2.1.2.2.3 Continuous assessment of ovulation by urinary sex steroids (FEAT Study female cohort only)

Urinary metabolites of oestradiol and progesterone give an accurate indication of ovulation (Williams et al., 2015). Daily urine samples across the Course would not be feasible, so we measured ovulation over one complete cycle of daily urine sampling at the start. This allowed us to interpret the occurrence of ovulation from weekly measurement of urinary progesterone thereafter (Roupas and Georgopoulos, 2011). The month chosen, the first month at the onset of junior term, was the most pragmatic solution to the requirement for a ‘control cycle’. This period was less physically arduous than subsequent weeks, and while still stressful, changes in menstrual cyclicity were expected to be least at this early stage. Urine containers were labelled by unique identifier number and distributed weekly, and cadets were asked to mark the time and date. Subjects were requested to provide a sample at the same time every day for one month and every Friday thereafter, in the evening. Those taking hormonal contraceptives were not included in analysis. These

data were complemented by a menstrual diary, which all female participants were asked to complete face-to-face and by subsequent text message reminders three times per term.

The confines of the RMAS Commissioning Course necessitated development of methods of detecting ovulation. Serial transvaginal ultrasound (the gold standard) or blood sampling would not have been feasible, while menstrual diaries alone are not sufficiently accurate (Gifford et al., 2017). Pregnanediol glucuronide (PDG), the major metabolite of P4, is 3,000 to 5,000 times more abundant than P4 in urine and assessed by manual ELISA (Stanczyk et al., 1997). We anticipated over 4,500 urine samples from the FEAT Study, which would be onerous to complete by manual ELISA. We therefore performed a study to determine if it is possible to identify ovulation from weekly urine samples using a newer, automated progesterone assay (**Section 3.1**).

2.1.2.3 Adrenal function and stress

2.1.2.3.1 Longitudinal and diurnal cortisol assessment

As cortisol is released in a pulsatile manner, individual samples of blood or urine (unless collected several times per hour) are not a reliable assessment of adrenal function. In clinical practice, assessment of cortisol production is done with 24 hour urine collections, however this not feasible in this setting. Hair sampling has recently been found to provide a reliable long-term measurement of cortisol concentration (Stalder et al., 2017). Hair grows at 1cm per month, and a small sample cut close to the scalp can provide a profile of mean monthly hair cortisol concentration (HCC) for 3-6 months.

Salivary cortisol is complimentary to HCC (Hellhammer et al., 2009). Saliva has been used extensively to assess changes in the diurnal profile of cortisol secretion (highest in the morning, lowest in the evening) in response to various stressors (Adam et al., 2017). We therefore measured salivary cortisol alongside HCC to assess both change in daily profile and average background secretion.

2.1.2.3.2 Socioeconomic, Psychology and Resilience Questionnaires

Changes in HPA activity during psychological stress are influenced by markers of socioeconomic status and education (Cohen et al., 2006). Such changes might be mitigated against by mental resilience (Taylor et al., 2014). A resilient disposition is also likely to moderate the relationship between stress and menstrual disturbance,

such that dispositional resilience might reduce the impact of military training on reproductive function (Palm-Fischbacher and Ehlert, 2014). The questionnaires comprised previously validated self-rating items on a web-based application (SmartSurvey, Tewkesbury, UK)

1 *Demographics and Socioeconomic questionnaires* (Dhalwani et al., 2013; Turrell et al., 2007)

2 *Psychosocial stress questionnaire* (stressful life events) (Rosengren et al., 2004)

3 *Impact of events-revised* (IES-R; symptoms of post-traumatic stress disorder relating to any stressful life events) (Marmar et al., 1996)

4 *The Beck Anxiety Inventory* (BAI; anxiety) (Beck et al., 1988)

5 *Patient Health Questionnaire – 9* (PHQ-9; low mood) (Du Preez et al., 2017; Kocalevent et al., 2013)

6 *Connor Davidson Resilience Scale-10* (CDRISC-10; Resilience) (Campbell-Sills and Stein, 2007)

Psychosocial stress was defined as feeling irritable, filled with anxiety, or as having sleeping difficulties because of conditions at work or at home, with the following response options: ‘never’, ‘some periods’, ‘several periods’ or ‘permanent stress’. We altered the original questionnaire of Rosengren et al. (2004), by combining stress at work and at home into a single question, when participants spent very little time at home during the periods assessed. In asking about the level of financial stress, three options were given: ‘little or none’, ‘moderate’ or ‘high or severe’, while major life events such as major family conflict, divorce or separation were categorised into ‘none’ or ‘one or more’. Participants were then asked to complete the IES-R with reference to any major life event(s) identified (Weiss, 1997).

The BAI assesses how much each of 21 anxiety symptoms has bothered participants in the past month on a 4-point Likert scale from (0 = “not at all” to 3 = “severely – it bothered me a lot”). The BAI has demonstrated high internal consistency in a military population (α coefficient .91), adequate test-retest reliability

(r = 0.75) and correlates highly with other measures of anxiety (Beck et al., 1988; Nathan et al., 2012).

The PHQ-9 consists of nine criteria, scored from 0 to 3 (0 = not at all, 3 = nearly every day). It has demonstrated good performance as a diagnostic and severity measure in military and general populations (Martin et al., 2006; Wells et al., 2013). We analysed scores on a continuous scale of 0 to 27, rather than diagnostic cut-off, to detect subtle differences in a low number of participants.

The CDRISC 10 is a measure of the ability to respond to adversity and comprises 10 items, scored from 0 (“not true at all”) to 4 (“true nearly all the time”), abridged from the 25-point CDRISC, on the basis of factor analysis (Campbell-Sills and Stein, 2007). It demonstrates excellent psychometric properties in young adults (Campbell-Sills et al., 2009) and military populations (Green et al., 2014; Johnson et al., 2011). In 328 military personnel, Maguen et al. found the CDRISC correlates with the PTSD Checklist-Military (PCL-M) measure of PTSD (r=-0.22, p<0.01), negative and positive effect on the Positive and Negative Affectivity Schedule (PANAS) (r=-0.32, r=0.62, respectively, both p<0.01) (Maguen et al., 2008). Permission was granted from the author to use the CDRISC 10 in the present study.

2.1.2.3.3 Heart Rate Variability (Sympathetic-adrenal medulla axis)

Heart rate variability (HRV) is an index of sympathetic and parasympathetic autonomic activity, which may be associated with overtraining and psychological stress (Kiviniemi et al., 2014). Participants are asked to relax in a quiet environment in the early morning, having not consumed caffeine for 3 hours beforehand. Two standard ECG electrodes are placed over the wrists for 5 minutes and a trace is recorded using a portable HRV device (CheckMyHeart, DailyCare Biomedical, Taiwan). R-R intervals were identified using proprietary software (CheckMyHeart for PC, version 3.0, DailyCare Biomedical) and confirmed visually. Specialised software (Kubios HRV Premium version 3.2.0 for mac, http://www.kubios.com) was used to obtain time domain, frequency domain and non-linear HRV analyses.

2.1.2.4 Bone mineral content, microarchitecture and turnover

2.1.2.4.1 Dual-energy x-ray absorptiometry (DXA)

Whole body areal bone mineral density (aBMD) was assessed using DXA (Lunar iDXA, GE Healthcare, UK), with participants wearing shorts and a t-shirt, before and

after junior term, and at the ends of intermediate and senior term (FEAT Study) and before and the Ex ICE MAIDEN (E4T Study). Arms, legs, trunk, ribs, pelvis and spine aBMD were obtained from the whole- body scan using Lunar iDXA enCORE software according to manufacturer's instructions (GE Healthcare, Chalfront St Giles, UK). During the same DXA scan, measurements of body composition were also obtained (**Sections 3-2, 3-4 and 4-1**).

2.1.2.4.2 High resolution, peripheral, quantitative computerised tomography (HR-pQCT)

To assess volumetric BMD (vBMD), bone microarchitecture, bone strength and cortical porosity, a HR-pQCT system (Xtreme eCT II, Scanco Medical AG, Switzerland) was performed at the ultradistal (4 %) and diaphyseal (30 %) sites of the right tibia (unless there was history of fracture, in which case the left tibia was used). These measures are the gold standard means of assessing bone turnover radiographically (Cheung et al., 2013). HR-pQCT uses a simultaneous acquisition of 110 parallel CT slices at 0.62μm resolution, providing a 3D representation of a 9.02 mm section of the tibia in the axial direction. The standard evaluation provided by the manufacturer was used to provide the following measurements: total vBMD, trabecular vBMD, cortical vBMD, trabecular bone volume fraction, trabecular area, cortical area, cortical thickness, trabecular thickness, trabecular number (per mm), trabecular separation, % cortical porosity and cortical pore diameter. Micro-finite element analysis was performed to calculate the biomechanical properties under uniaxial compression, specifically stiffness ($kN.mm^{-1}$) and failure load (kN) (Vilayphiou et al., 2011). HR-pQCT is included in the methods for completeness; the abstract of a publication containing HR-pQCT data from the E4T study is included at **Appendix C**. A manuscript containing HR-pQCT data from the FEAT study is in preparation.

2.1.2.4.3 Bone markers

Total 20(OH) vitamin D3, adjusted calcium and phosphate were measured from each fasted blood sample. Two markers of bone formation and resorption (procollagen type 1 N-terminal propeptide (P1NP) and sclerostin, and bone specific alkaline phosphatase (BSAP) and beta cross-linked c-telopeptide, respectively) were also measured.

2.1.2.5 Energy availability

Assessment of energy availability (EA) was putatively central to understanding endocrine adaptations to military training. In the FEAT Study this was achieved by obtaining 4 DXA scans for body composition, three fasting blood samples including measures of nutritional status, and with one energy assessment phase (EAP) per term. The EAP represented a 'snapshot' of total energy expenditure, energy intake and exercise energy expenditure, during 10 day periods chosen each term as representative of the entire CC.

The practicalities of Ex ICE MAIDEN did not permit participants to undergo continuous measurements during the E4T Study. Instead, the E4T Study subjects underwent a more detailed assessment of energy metabolism before and after the expedition by collaborators in the University Hospital of Coventry and Warwick Human Metabolic Research Unit (HMRU), followed by fitness testing (combined graded exercise test to volitional exhaustion and maximal fat oxidation test). These elements were designed by collaborators from HMRU and are beyond the scope of this thesis.

2.1.2.5.1 Energy availability assessment in FEAT Study

Doubly labelled water (DLW) is the gold standard means of assessing total energy expenditure (TEE) (Speakman, 1997). One, 10-day monitoring periods of EA was undertaken during each term. Participants were given a dose of DLW (10 atom percent excess ^{18}O and 5 atom percent excess ^{2}H) the night before the start of the observation period, and after collection of a baseline urine sample. Subsequent urine samples were collected from the second void the following morning, from a further void on the same day and from two voids in the middle and at the end of the 10-day period. Urine samples were analysed by isotope radio mass spectrometry with an analytic precision of 0.2 ppm for ^{2}H and 0.4ppnm for ^{18}O at the MRC Human Nutrition Centre, Cambridge. This technique has been extensively used to measure EE in previous military training studies (Hoyt et al., 2006; O'Leary et al., 2018; Richmond et al., 2012; Siddall et al., 2019).

Throughout the 10-day EAP, subjects wore a wrist-worn 3-way accelerometer set to a measurement frequency of 60 Hz (GENEActiv™ Original, Activinsights Ltd., Cambs). Acceleration magnitude was plotted against accumulated time spent at each magnitude. After adjustment for very low accelerations and non-wear time, the

area under the curve was integrated to provide relative the exercise exposures at different intensities during each EAP.

During the 10 day DLW and accelerometry periods, subjects recorded all food consumed using a standardised diary, checked at the end of each day in a one-to-one meeting with a researcher, who prompted recall of any incompletely recorded meals or snacks. Researchers also measured food intake for the first cohort (starting May 2017) by weighed dietary analysis at each scheduled meal and from all empty snack wrappers, collected at the end of each day. Dietary intake was analysed using a dietary analysis software package (Nutritics for PC, Nutritics Ltd, Dublin, Ireland).

2.1.2.5.2 Eating behaviour

Disordered eating behaviours are ostensibly significant contributors to the RED-S in civilian athletes. The Brief Eating Disorders in Athletes Questionnaire (BEDA-Q) was measured at each study visit.

The BEDA-Q is a concise assessment of disordered eating risk, designed to focus on nine items from the drive for thinness, body dissatisfaction and perfectionism domains of the eating disorders inventory (EDI). It demonstrated 82% sensitivity and 85% specificity discriminating eating disorders in 221 adolescent female athletes. These nine items are scored 0-3 (one is reverse scored). In a small study of ballerinas aged 17-32, BEDA-Q score did not correlate with age ($r = -0.08$, $p>0.05$). Two additional polar questions are asked, which are beneficial for our repeated measures design: “are you trying to lose weight?”, “have you ever tried to lose weight?”, and “If so, have you tried 1-2, 3-5 or >5 times?”. Since the response categories in this latter question were designed to discriminate excessive weight loss attempts in a younger population than that studied here, we asked respondents to estimate the actual number of times they had attempted to lose weight.

Although it has not been widely validated, BEDA-Q has been recognised increasingly (Ackerman et al., 2019a; Knapp et al., 2014; Mountjoy et al., 2014; Pope et al., 2015) and offered the present study advantages in its brevity compared with the EDI or Eating Disorders Examination Questionnaire (EDE-Q), and indirect approach of questioning, for example compared with the SCOFF questionnaire and its derivatives (Cooper et al., 1989; Morgan et al., 1999). Briefer alternatives developed for screening in the general population, such as the Eating Attitudes

Test-26, have been criticised for inconsistent scoring among athletes (Martinez Rodriguez et al., 2015; Pope et al., 2015).

2.1.2.5.3 Body mass index (BMI)

At the start of each study, height was measured using and Seca 217 stadiometer (Seca Scientific, Birmingham, UK). At every study visit, semi-nude (underwear, shorts and t-shirt) weight was measured using Seca 874 digital scales. Weight was also measured before and after EAPs, to attempt to adjust TEE for the substrate metabolised (where weight is lost, in most circumstances fat, which has a higher energy value quotient per mole CO_2, or food quotient).

2.1.2.5.4 Body composition – Dual-energy X-Ray Absorptiometry

During the DXA scans, regional fat-free mass and body fat distribution were calculated using the same scanning software as for bone composition (section 2.4.2.4.1).

2.1.2.5.5 Metabolic markers of nutrition

Markers of metabolic status were assayed from fasting blood tests, including glucose, insulin, c-peptide, non-esterified fatty acids (NEFA), zinc, magnesium, ferritin, iron, transferrin, IGF-1, thyroid stimulating hormone (TSH), free T4 and total T3.

2.2 Study context

2.2.1 The Royal Military Academy, Sandhurst

Staff at the Royal Military Academy, Sandhurst (RMAS, Camberley, UK) led by the Commander, Sandhurst Group, Brigadier Bill Wright OBE, permitted the clinical studies to take place following ethical approval (**Chapters 3 and 4**). RMAS Staff authorised recruitment and data collection for the FEAT Study (**Chapter 3**), during nine Pre-Commissioning Course Briefing Courses (PCCBCs). I was invited to brief around 300 staff involved in training the regular Commissioning Course (Course), prior to the starts of each term. The RMAS Facilities Manager also provided four rooms within the historic Old College building for conversion into a Research Facility for the study in **Chapter 3**, which was also used for the study in **Chapter 4**: an 8-bedded clinical laboratory, rest area and two rooms for imaging. A benchtop centrifuge,–80ºC freezer, dual x-ray absorptiometry (DXA) scanner and high resolution peripheral quantitative scanner (HRpQCT) scanner were then installed following necessary engineering works and approval by the MoD Radiation Protection Adviser. Twelve refrigerators were placed in the accommodation and office areas of RMAS for daily and weekly samples to be placed for the study in **Chapter 3**. Researchers also accessed field exercises (Exercise ALLENBY'S ADVANCE, Brecon Beacons, UK) on three occasions (**Chapter 3**).

2.2.2 Ex ICE MAIDEN

The Exploring the Endocrine Effects of Extreme Training (E4T) Study (**Chapter 4**) was conducted in response to an open request made by the Ex ICE MAIDEN leaders and the Royal Centre for Defence Medicine (RCDM) Director of Research and Clinical Innovation. Travel and subsistence for study visits in Chile was funded by the ICE MAIDEN Higher Management Committee.

2.2.3 Edinburgh Clinical Research Facility

The Edinburgh Clinical Research Facility is jointly operated by NHS Lothian and the University of Edinburgh. Research Nurses from the Facility normally work in Edinburgh, but a team led by Sister Jo Singleton travelled to RMAS on 7 occasions to conduct visits for the study in **Chapter 3**. Nurses also attended the RMAS Research Facility to conduct study measurements for the study in **Chapter 4** on two occasions and Sister Singleton travelled to Punta Arenas, South Chile to conduct clinical measurements for this study immediately after the Ex ICE MAIDEN.

2.3 Laboratory materials and Methods

2.3.1 ELISA

Commercially available enzyme linked immunosorbent assay (ELISA) kits for inhibin B, sclerostin, BSAP, beta CTX and leptin are specified in results chapters. The ELISAs used in this thesis follows the same principles as the competitive in-house PDG ELISA (**chapter 3 section 3.1)**: analyte-bound horseradish peroxidase (HRP) competes with the analyte of interest from the sample (PDG) for rabbit anti-analyte antibody. The residual HRP bound to rabbit antibody on the surface of the plate, after washing off the sample-HRP mixture, produces a colour reaction when 3,3',5,5'-Tetramethylbenzidine (TMB) is added followed by H_2SO4 'stop solution'. The strength of this reaction (optical density) is dependent on the concentration of analyte.

The PDG ELISA required optimisation prior to use. A DARS concentration of 1/500 was found to generate a superior coefficient of variation (CV) to 1/250 (**Figure 2-5**). Similar CV improvements were also made with an anti-PDG antibody concentration of 1 in 40,000 (superior to 1 in 80,000, or 1 in 32,000 or 1 in 16,000) and sample: HRP ratio of 1: 5 compared with in 1: 10 or 1: 20.

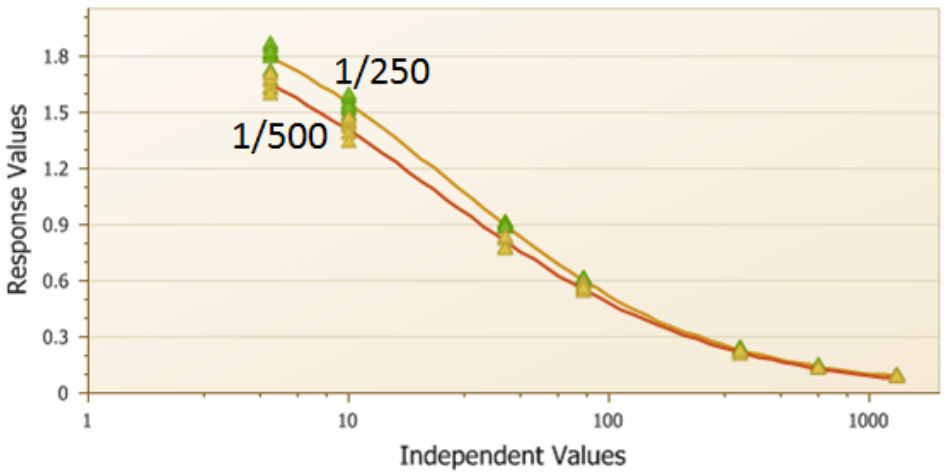


Figure 2-5 Pregnanediol Glucuronide ELISA Standard Curves for 1/250 and 1/500 DARS

Standard curves during optimisation process of pregnanediol glucuronide (PDG) ELISA. 1/250 and 1/500 refer to concentrations of donkey antirabbit serum (DARS) concentration in coating buffer. Mean coefficients of variation (CVs) were slightly better at 1/500 than 1/250 (6.4% and 9.2%, respectively).

The finalised PDG ELISA was as follows. A 96 well plate was coated with pre-precipitated DARS (0.2 µg anti rabbit IgG in 100uL: 1/500) in ELISA coating buffer for 14 hours at 4℃. The coated plate was washed twice with 50 mmol/L TRIS buffer containing 137 mmol/L NaCl and 0.05% tween 20 (Wash Buffer). The plate was blocked with 220µL 10 mmol/L phosphate buffer and 0.5% w/v bovine serum albumin (BSA) and washed twice with wash buffer. The sameple (20µL) was added with 80µL PDG-horseradish peroxidase 1 in 200,000 in PBS 0.1% BSA (Assay Buffer) and shaken for 2 minutes. Rabbit anti-PDG Antibody (in house reagent) 1 in 40,000 assay buffer (50µL) was then added and incubated in a shaker at 30℃ for 2h. The plates were washed 5 times with wash buffer and 120µL 3,3',5,5'-Tetramethylbenzidine (TMB) was added. After 12-15 minutes 80µL 2M H_2SO_4 stop solution was added and the plate read on a plate reader at 450nm.

The following ELISAs were used in **Chapters 3 and 4:**

- Inhibin B (Beckman Coulter, High Wycombe, UK)
- Sclerostin (Biomedica Medizinprodukte GmbH, Vienna, Austria)
- Bone specific alkaline phosphatase (BSAP) (Quidel, Athens, OH, USA)
- Leptin (Quantikine ®, R&D Systems, Minneapolis, MN, USA)
- Total cortisol binding globulin (CBG) according to the methods of Lewis and Elder (2013).

2.3.2 Centrifugal analyser

Commercial kits for creatinine, lactate, NEFA, albumin, transferrin, calcium, phosphate, magnesium and zinc were adapted for use on a Cobas Fara centrifugal analyser (Roche Diagnostics Ltd, Welwyn Garden City, UK). The principle of a centrifugal analyser is to create complexes with analytes of interest and propel the mixtures using centrifugal forces into separate cuvettes for measurement. A reaction in tip of the cuvette is detected by a photometer and read by the analyser. Details of kits and precision are specified in relevant chapters and **Appendix B**.

2.3.3 Roche e411 ® and Abbott Architect ® analysers

An automated Roche e411 ® electrochemiluminescence assay was used to measure serum insulin, C-peptide, AMH, ferritin, IGF1, SHBG, P1NP, CTX and urine progesterone.

An automated Abbott Architect ® chemiluminescent microparticle immunoassay was used to measure serum LH, FSH, oestradiol, progesterone and urine progesterone and creatinine.

The details of kits and precision of e411 ® and Architect ® assays are specified **sections 3.1**, **3.4** and **4.1**.

2.3.4 Liquid Chromatography Mass Spectrometry

Mass spectrometry (MS) is widely regarded as the 'gold standard' for identification and quantification of steroids, since it directly detects the molecule, rather than an epitope, and is not liable to inaccuracies caused by cross reactivity with binding proteins.

Liquid chromatography (LC) is undertaken prior to MS, where molecules within a liquid sample are forced through a column at high pressure, and are separated into ranges of mass based on the duration of elution through the column. MS then occurs when the sample is ionised by bombarding it with electrons, causing it to fragment into product ions. The ions are separated according to their mass-to-charge ratio and detected. Mass-to-charge ratios from the analyte of interest are identified by correlation with known molecular masses through a characteristic pattern of fragmentation.

Prior to measurement, analytes must undergo extraction. Extraction and LC/MS were undertaken for plasma cortisol, hair cortisol and plasma total vitamin D. Details of extraction from plasma or hair are specified in **Sections 3.2, 3.4 and 4.1**.

2.3.5 Statistical analyses

Statistical analysis was performed using SPSS (IBM ® SPSS ®, version 24, Chicago, Illinois). Area under the Curve (AUC) and peak response for cortisol, FSH and LH during the 1-hour dynamic tests were calculated using the trapezoidal rule. Homeostatic modelling of insulin resistance was calculated according to the methods of Levy et al. (1998).

Data were assessed for normality using the Shapiro-Wilk test. Any non-normally distributed data were transformed using natural logarithm or square root transformation prior to analysis with parametric tests.

Chapter 3.1. Correlation of automated P4 assays was performed by Pearson's correlation analysis. Assays were compared using Passing Bablok regression and a Bland-Altman plot was used to check for bias and heterogeneity. Luteal and follicular change measured by PDG and two P4 assays was calculated by one-way ANOVA. Receiver operator characteristics curves were constructed and compared for the best automated P4 Assay and PDG.

Clinical Studies (Sections 3.1, 3.2. 3.3, 4.1 and 4.2). In general, for clinical studies, descriptive statistics were used to describe baseline characteristics and those who completed studies were compared with those who withdrew using independent samples t tests or Chi squared tests. Repeated measures analyses of variance (RM ANOVA) was used to compare the change in continuous normally distributed data over time for participants who completed each study, since there was little or no missing data, with post hoc paired tests where any significant changes were identified. Linear mixed models (with covariance structures to account for missing data) were not required. Specific statistical methods for each study are described in each section.

Data are presented as mean ±SD unless otherwise stated. Statistical significance was set at $p<0.05$ except where Bonferroni adjustment for multiple comparisons was used (applied to specific tests in **Sections 3.1, 3.2, 3.3, 4.1**). Figures were prepared using GraphPad Prism v8.0 (GraphPad Software Inc, California, USA)

Chapter 3 Female endocrinology adaptation during basic military training

This chapter summarises the findings of the Female Endocrinology in Arduous Training (FEAT) Study. Methodological developments which were prerequisites for the characterisation of hormonal adaptation to military training (as identified in **Chapter 1**) are addressed in **Sections 3.1** and **3.2**: an efficient method to identify ovulation from weekly urine samples using an automated assay, and the measurement of energy availability (EA) in military training, respectively. In **Section 3.2** the association between EA measures and training-related adaptations to military training are also assessed. **Section 3.3** describes HPA axis and psychological adaptation to basic military training. **Section 3.4** describes biomarkers of energy status, and HPG axis and skeletal adaptations to basic military training.

3.1 Confirmation of ovulation from urinary progesterone: assessment of two automated assay platforms.

This Section was published in Scientific Reports under the same title by Robert M Gifford (RG), Dr Forbes Howie (FH), Miss Kirsten Wilson (KW), Dr Neil Johnston (NJ), Dr Tommaso Todisco (TT), Dr Mike Crane (MC), Dr Karolina Skorupskaite (KS), Prof David Woods (DW), Prof Rebecca Reynolds (RR), and Prof Richard Anderson (RA) (Gifford et al., 2018b). KS, MC and RA sought ethical approval, recruited the participants and collected the samples for the daily ovulation cohort, for a separate study investigating a neurokinin B antagonist contraceptive (Skorupskaite et al., 2017). RG designed the experiment with supervisory input from RA, FH, MC and RR. RG developed and conducted the pregnanediol 3 glucuronide (PDG) in-house assay with supervisory input from FH and KW. RG collated and analysed the data (except Passing-Bablok analyses, which were conducted by Angela Ballantyne in the NHS Lothian department of biochemistry), wrote the first draft of the manuscript and responded to reviewers' comments. FH, RG, KW, TT, MC and NJ undertook automated analyser assays. All authors provided final editorial input to the manuscript, tables and figures.

In summary, using urine samples from two cohorts (healthy women who provided daily samples, and women attending a fertility service who provided weekly samples), this work demonstrated that an automated progesterone immunoassay

(Abbott Architect ®, marketed for serum) provides comparable sensitivity and specificity to the current gold-standard, a pregananediol-3-glucoronide (PDG) enzyme-linked immunosorbent assay (ELISA). This finding had the potential to create efficiency savings for clinical laboratories and the FEAT Study, where a large number of urine samples for ovulation detection was anticipated. The automated assay platform validated in section 3.1 was used to identify ovulation for the study in section 3.4.

3.1.1 Abstract

Urinary concentrations of the major progesterone (P4) metabolite pregnanediol-3-glucuronide (PDG) are used to confirm ovulation. We aimed to determine whether automated immunoassay of urinary P4 was as efficacious as PDG to confirm ovulation. Daily urine samples from 20 cycles in 14 healthy women in whom ovulation was dated by ultrasound, and serial weekly samples from 21 women in whom ovulation was unknown were analysed. Daily samples were assayed by two automated P4 immunoassays (Roche Cobas and Abbott Architect) and PDG ELISA. Serial samples were assayed for P4 by Architect and PDG by ELISA. In women with detailed monitoring of ovulation, median (95% CI) luteal phase increase was greatest for PDG, 427% (261–661), 278% (187–354) for P4 Architect and least for P4 Cobas, 146% (130–191), p<0.0001. Cobas P4 also showed marked inaccuracy in serial dilution. Similar ROC AUCs were observed for individual threshold values and two-sample percent rise analyses for P4 Architect and PDG (both >0.92). In serial samples classified as (an)ovulatory by PDG, P4 Architect gave ROC AUC 0.95 (95% CI 0.89 to 1.01), with sensitivity and specificity for confirmation of ovulation of 0.90 and 0.91 at a cutoff of 1.67 µmol/mol. Automated P4 may potentially be as efficacious as PDG ELISA but research from a range of clinical settings is required.

3.1.2 Background

The confirmation of ovulation is important for the investigation of infertility, for women planning conception and for researchers understanding the impact of interventions on ovarian function. Transvaginal ultrasound (TVUS) detection of the growth and disappearance of a follicle is the gold standard technique, but it is invasive and repeated measures are often undesirable or not feasible in an outpatient setting (Guermandi et al., 2001; Lynch et al., 2014). The assessment of

cyclical hormonal concentrations represents an objective alternative. Clinical guidelines suggest blood measurement of progesterone (P4) during the luteal phase (Kamel, 2010; Lindsay and Vitrikas, 2015; NICE, 2017). Elevated serum P4 has high specificity, but may require repeated venepuncture and is invasive (Su et al., 2017). The use of assays of urinary metabolites of P4 (pregnanediol 3 glucuronide, PDG) and oestradiol (e.g. oestrone-3-glucuronide) and/or luteinising hormone (LH), correcting for urinary creatinine to adjust for an individual's fluid status, has been widely used for many decades and allows convenient repeated sample collection (Baird et al., 1991; Blackwell et al., 2018). P4 is inactivated to pregnanediol by reduction at the C5, C3 and C20 position, and glucuronic acid is attached via a glycosidic bond, forming PDG.

A rise in urine PDG above a certain threshold (commonly 5 µg/ ml) is required to confirm ovulation (Lynch et al., 2014). Urine PDG demonstrates excellent agreement with P4 in both serum (O'Connor et al., 2003; Roos et al., 2015) and urine (Stanczyk et al., 1997). In practice, PDG assays often require a relatively onerous manual competitive ELISA, and the related expense limits the availability of testing in a clinical setting. Unlike PDG, assays for P4 are readily available for automated analysers commonly used in hospital laboratories, and therefore have potential advantages in expediency, cost-effectiveness and availability.

We therefore aimed to determine whether daily measurement of creatinine-corrected urinary progesterone using an automated progesterone assay could be used for reliable confirmation of ovulation in a cohort in whom ovulation had already been reliably identified (confirmatory cohort). We also aimed to explore whether a P4 threshold value for one sample, or a percent rise between two samples (one follicular and one luteal) was the more discriminatory in the confirmatory cohort. Furthermore, in weekly samples from an additional cohort, in whom no further information on ovulation status was known (exploratory cohort), we aimed to explore the sensitivity and specificity of weekly P4 in confirming ovulation (threshold value and two sample percent rise) with PDG as the referent.

3.1.3 Materials and Methods

3.1.3.1 Human subject recruitment

Ethical approval was obtained from South East Scotland Research Ethics Committee (Ref: 09/S1101/67). The study conformed to the principles outlined in the

Declaration of Helsinki. The study consisted of two cohorts: a confirmatory cohort, evaluating the ability of urine P4 to confirm ovulation diagnosed by ultrasound, relative to PDG, and an exploratory cohort, comparing P4 and PDG in a likely real-world setting, where true ovulation status is unknown, daily testing is not possible and weekly urinary PDG is used clinically.

The confirmatory cohort comprised 14 healthy women aged 27 to 43 years with self-reported regular menses, who were controls in a study in which timing of ovulation was characterised using gold-standard techniques (Skorupskaite et al., 2017). All provided informed consent. Inclusion criteria were: reproductive age (between menarche and menopause), no steroidal contraception or other hormonal medication, intrauterine device use or fertility treatment, normal physical examination, and renal and liver function and electrolytes within normal limits. Participants were excluded if they could not complete the required sampling regimen or became pregnant.

The exploratory cohort comprised 21 women attending a reproductive endocrinology service in Edinburgh, undergoing a standard clinical assessment. None were taking hormonal contraceptives or fertility treatment, all were of reproductive age and all had completed the required sampling regimen. Informed consent was not required for the exploratory cohort, since the investigation was part of their routine investigations. Investigators were blinded to the findings of history, examination and other investigations. Seven healthy men aged 28 to 61 years provided single urine aliquots for assessment of linearity and dilution recovery.

3.1.3.2 Sample size

Since confirmatory and exploratory cohorts were assessed as part of other research or clinical activities, they may be considered convenience samples. Although sample sizes (confirmatory: 20 cycles from 14 women, exploratory: 42 cycles from 21 women) were smaller than previous studies, (Abdulla et al., 2018; Ecochard et al., 2017; Leiva et al., 2015; Lynch et al., 2014; O'Connor et al., 2003), assessments were in greater detail (e.g. daily and weekly urine sampling, respectively, see '*Capability in confirming ovulation'*), hence we anticipated the sample size would be sufficient to confirm the ability of urinary P4 to identify ovulation and explore its diagnostic utility versus PDG.

3.1.3.3 Creatinine and LH assays

LH was measured in serum and urine by in-house ELISA using two different anti-human LH beta subunit mouse monoclonal antibodies (Medix Biochemica, Kauniainen, Finland), as described elsewhere (George et al., 2011). While LH may be unstable in urine at −20 °C (Robinson et al., 2007), a measured peak in urine LH would be intended to be supportive of a serum measurement, and TVUS, both within 2–3 days. Urine creatinine was determined using the creatininase/creatinase specific enzymatic method utilizing a commercial kit (Alpha Laboratories Ltd. Eastleigh, UK) adapted for use on a Cobas Fara centrifugal analyser (Roche Diagnostics Ltd, Welwyn Garden City, UK) (Borner et al., 1979).

3.1.3.4 Urine steroids assays

Urine samples were stored at −20 °C until steroid analysis. Measurement of PDG and two automated P4 immunoassays were undertaken on each sample.

PDG was measured in duplicate by competitive PDG ELISA. A 96-well plate was coated with 100 µL of pre-precipitated donkey 0.2 µg anti rabbit IgG serum per well (Scottish Antibody Production Unit, Carluke, UK) in ELISA coating buffer for 14 hours at 4 °C and washed twice with 50 mmol/L TRIS buffer containing 137 mmol/L NaCl and 0.05% tween 20 (wash buffer). The plate was blocked with 220 µL 10 mmol/L phosphate buffer and 0.5% w/v bovine serum albumin (BSA), washed twice with wash buffer, 20 µL of sample was added with 80 µL PDG-HRP (in house reagent) 1 in 200,000 in PBS 0.1%BSA (assay buffer) and shaken for 2 minutes. 50 µL Rabbit anti-PDG Ab (in house reagent) 1 in 40,000 assay buffer were then added and incubated in a shaker at 30 °C for 2 h. The plates were washed 5 times with wash buffer and 120 µL 3,3′,5,5′-Tetramethylbenzidine (TMB) was added. After 12–15 minutes 80 µL 2 M H2SO4 stop solution was added and the plate read on a plate reader at 450 nm.

Automated P4 chemiluminescent microparticle immunoassay (Abbott Laboratories, Lake Bluff, Illinois, USA): P4 was measured on the Abbott Architect c8000 automated analyser, using a proprietary serum assay kit (Architect System Progesterone, Abbott Ireland Diagnostics Division, Longford, Ireland) according to the manufacturer's instructions. The analytical sensitivity was quoted as ≤0.3 nmol/L. No significant cross-reactants are quoted by the manufacturer.

Automated P4 electrochemiluminescence immunoassay (Roche Diagnostics Ltd, Welwyn Garden City, UK): P4 was measured on the Roche Cobas automated immunoanalyser, using a proprietary serum assay kit (Elecsys® Progesterone II, Roche Diagnostics, Indianapolis, Indiana, USA) according to the manufacturer's instructions. The analytical sensitivity quoted as 1.0 nmol/L. The only significant cross-reactant quoted by the manufacturer is 5 β-dihydroprogesterone at 20.7%.

3.1.3.5 Comparison and correlation between assay methods

A total of 536 daily, early morning urine samples were assayed using PDG ELISA, Architect and Cobas P4.

Since neither P4 assay quoted cross-reactivity with PDG, we measured P4 using both assays in three spiked male urine samples (100 nmol/L PDG) and three unspiked samples. Mean progesterone concentrations were compared between spiked and unspiked samples to give percent cross-reactivity.

3.1.3.6 Freeze thaw stability

There was one freeze-thaw cycle between each assay, in the order PDG, Cobas, Architect. PDG is known to be stable following up to 10 freeze-thaw cycles (O'Connor et al., 2003), but the stability of urinary progesterone measured by automated assay has not previously been demonstrated. We examined the effect of up to five freeze-thaw cycles on hormone concentration. An aliquot of male urine was spiked with 240 nmol/L P4 and divided into 18 aliquots. Three aliquots were each subjected to 0, 1, 2, 3, 4 or 5 freeze-thaw cycles. P4 was then measured using both assays and the percentage decrease from the index samples (0) calculated.

3.1.3.7 Assay precision

Standard samples supplied by the manufacturers spanning the low, middle and high range were measured with both P4 methods at four different time points within each run (Architect 4 runs, Cobas 5 runs). Runs were carried out on separate days and the reagent lot was varied to simulate normal operating procedures.

Within-assay precision was determined by repeating the assays in four replicates simultaneously for creatinine, LH and PDG. The intra-assay CVs were <3%, <5% and <10% while inter-assay CVs were <5%, <10% and <10% for creatinine, LH and PDG, respectively.

3.1.3.8 Linearity and dilution recovery

Seven male samples were spiked with 12.5 mmol/L P4 and were diluted serially with unspiked male urine 2, 4 and 8-fold. These samples were tested on both P4 methods, to assess if any disparity diminished with subsequent dilutions.

3.1.3.9 Capability in confirming ovulation

Daily early morning urine samples were collected across 20 menstrual cycles from the confirmatory cohort. Ovulation was confirmed by the appearance and disappearance of a dominant follicle on transvaginal ultrasound (TVUS), performed every 2–3 days. The precise day of ovulation was determined from the surge in daily urinary LH, corroborated by serum LH (measured every 2–3 days). Samples from ovulation day −10 to −3 were categorised as follicular, and ovulation day +3 to +10 as luteal.

In order to compare the rates of confirmation of ovulation for P4 Architect with PDG ELISA, a further eight, weekly samples were assayed from the confirmatory cohort (in whom the presence or absence of ovulation was otherwise undetermined). These women provided one urine sample every seven days for eight weeks, starting on a random day of the cycle.

3.1.3.10 Statistical analysis

Non-normally distributed data were log transformed prior to analysis. Correlation of P4 assays with PDG was performed by Pearson's correlation analysis. Values obtained by Cobas and Architect were also compared using Passing-Bablok regression. A Bland-Altman plot was used to check graphically for systematic bias and heterogeneity across the range of values. For assay precision, a coefficient of variance (CV) within or between assays of 10% or less was considered acceptable. We estimated within and between series imprecision using analysis of variance (ANOVA). For linearity and dilution recovery, the correlation between observed and expected values was compared using Pearson's test.

To assess the performance in confirming ovulation, all three assays were compared graphically by plotting the median, 10th and 90th percentile concentrations by day of ovulation. Follicular and luteal P4 concentrations by each assay were compared using paired samples t tests. Percent luteal change for all three assays was calculated using the median ratio of all combinations of follicular and luteal creatinine-corrected concentrations by one-way ANOVA. The sensitivity and

specificity of PDG was compared with the closest-correlating P4 assay using 1; daily urine samples (ovulation confirmed above a threshold concentration) and 2; a percent-rise between pairs of samples (ovulation confirmed above a certain percent rise) to confirm ovulation. Receiver-operator characteristics (ROC) curves were constructed and compared.

To compare the ovulation detection rate of P4 with PDG in the exploratory cohort, the diagnostic threshold concentrations and ratios calculated by the ROC curves in the confirmatory cohort were applied to these weekly samples. Where PDG exceeded the threshold concentration, this cycle was deemed ovulatory. The rise in PDG or P4 was assigned as week 3 for graphical purposes. Other cycles were deemed anovulatory. Luteal percentage changes were calculated as the difference between a sample and each of the other seven weekly samples (56 combinations for each woman). The sensitivity and specificity of P4 relative to PDG (the referent) were calculated for anovulatory and ovulatory cycles. The peak samples from each 4-sample consecutive series were analysed by ROC curve.

A p-value < 0.05 was considered statistically significant. Statistical analyses were undertaken using Analyse-It version 2.2 (Leeds, UK) and SPSS Statistics for Mac version 23.0 (IBM, New York, USA).

3.1.4 Results

3.1.4.1 Comparison and correlation between assay methods

Bland-Altman analysis demonstrated positive bias towards Cobas values with a trend towards greater discrepancy at lower concentrations of P4 than by Architect (**Figure 3-1-1**). Passing Bablok plots similarly demonstrated higher values by Cobas at lower concentrations, with higher values by Architect at higher concentrations. The correlations between PDG and P4 by Cobas and P4 by Architect were r = 0.454, r = 0.708, respectively, both p<0.001.

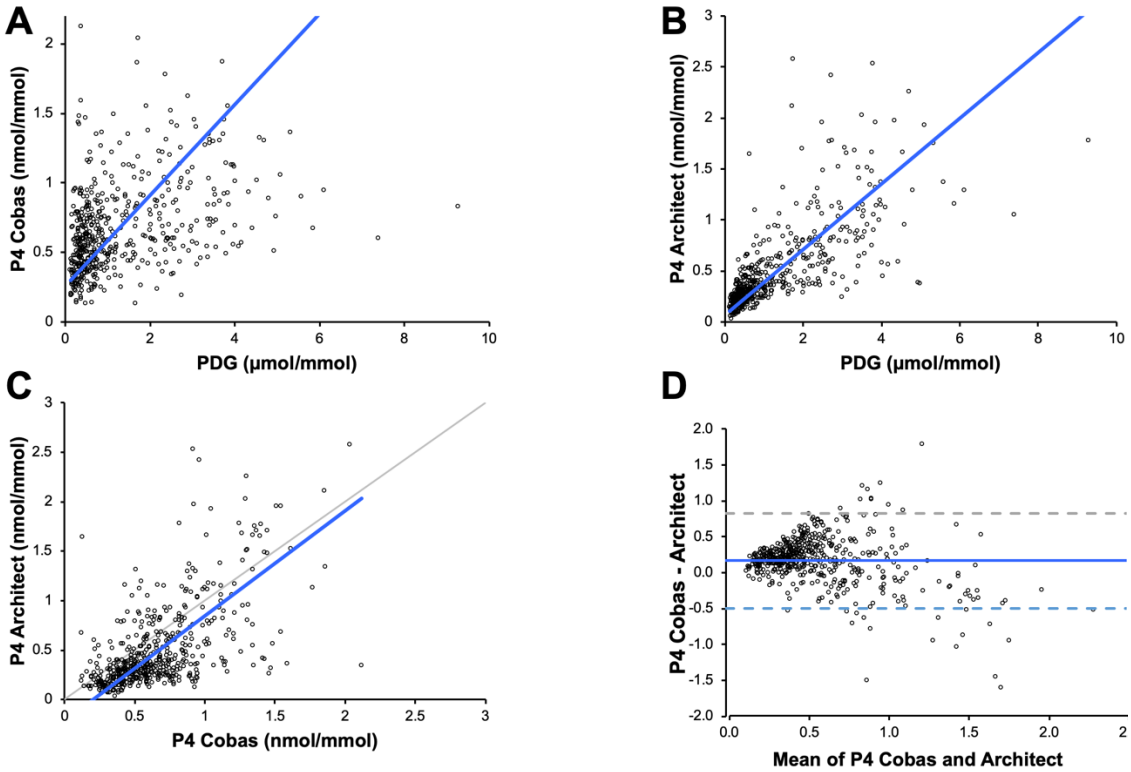


Figure 3-1-1. Passing Bablok and Bland Altman Plots. Passing Bablok plot of creatinine-corrected P4 Cobas and PDG ELISA (A), P4 Architect and PDG ELISA (B), P4 Architect and P4 Cobas (C) and Bland Altman Plot of the two P4 assays (D). P4: Progesterone, PDG: pregnanediol glucuronide, Cobas: Roche Cobas e411 electrochemiluminescence assay, Architect: Abbott Architect chemiluminescent microparticle immunoassay.

3.1.4.2 Cross-reactivity of P4 assays for PDG

The cross-reactivity for PDG in urine was 0.59% for Architect and 0.54% for Cobas.

3.1.4.3 Freeze-thaw stability of P4

The mean (SD) concentration of the index samples was Architect: 247 (3.52) nmol/L, Cobas: 249 (6.3) nmol/L. The percent change in P4 following freeze thaw cycles is shown in **Table 3-1-1**. After 1 freeze-thaw cycle, the mean (SD) change in concentration for Architect and Cobas, respectively was −2.61 (0.93) % and −2.21 (6.01) %, after 3 freeze thaw cycles −4.85 (2.99) % and +3.30 (7.12) %, and after 5 freeze-thaw cycles −7.60 (3.23) % and −7.02 (2.99) %.

Freeze-thaw cycles (n)	P4, Architect (mean (SD) % of index sample)	P4, Architect (mean (SD) % of index sample)
1	97.4 (0.93)	97.8 (6.01)
2	98.1 (2.83)	100.0 (6.70)
3	95.1 (2.99)	100.0 (7.12)
4	96.4 (0.66)	100.0 (2.43)
5	92.4 (3.23)	93.0 (2.99)

Table 3-1-1. Measured P4 concentrations by freeze-thaw cycle. Male urine was spiked with 240 nmol/L P4 and underwent up to five freeze-thaw cycles. Given as percentage of index sample (zero freeze-thaw cycles).

3.1.4.4 Assay precision

Coefficient variations for Architect and Cobas were within-run: < 2.5% and < 0.8%, and between-run: <3.8% and <1.4%, respectively.

3.1.4.5 Linearity and dilution recovery

P4 recovery measured by Architect demonstrated high linearity and close to 100% recovery, whereas Cobas demonstrated excess recovery (**Table 3-1-2**). The disparity between observed and expected values by Cobas diminished following serial dilution (170% spiked undiluted and 143% following 8x dilution, versus 113% and 112%, for Architect respectively), indicating a likely matrix effect. Pearson's correlations with expected values were r = 0.987 for Architect and r = 0.947 Cobas.

P4 added, nmol/L	% recovery, mean (SD)	
	Cobas	Architect
Male		
12.5	170 (41.7)	113 (13.3)
6.25 (x2 dilution)	160 (30.0)	106 (16.4)
3.13 (x4 dilution)	153 (26.6)	109 (16.7)
1.56 (x8 dilution)	143 (22.0)	112 (67.0)

Table 3-1-2. Analytical recovery of P4 from urine in the 2 automated assays. SD; standard deviation. Architect: Abbott Architect progesterone chemiluminescent microparticle immunoassay. Cobas: Roche Cobas electrochemiluminescence immunoassay. Male samples, n=7; P4, progesterone

3.1.4.6 Capability in confirming ovulation

In the confirmatory cohort, the median (range) cycle length was 28 (25–38) days. Day of ovulation was confirmed by transvaginal ultrasound and serum and urinary LH peak (median day 14 range day 12 – day 20). The median, 10th and 90th centiles of corrected PDG and P4 measured by Cobas and Architect by cycle day are illustrated in **Figure 3-1-2**. Follicular P4 was higher when measured by Cobas than by Architect (median (IQR) creatinine-corrected follicular concentration 1.41 (1.14–1.85) nmol/mol versus 0.66 (0.50–0.89) nmol/mol, respectively, $p<0.001$), while luteal concentrations did not significantly differ (2.30 (1.63–3.13) nmol/mol versus 2.19 (1.52–3.24) nmol/mmol, respectively, $p = 0.7$) (**Figure 3-1-3**). The median (IQR) percent luteal change for Cobas P4, Architect P4 and PDG were 146 (130–191)%, 278 (187–354)%, and 427 (261–661)%, respectively, $p<0.001$.

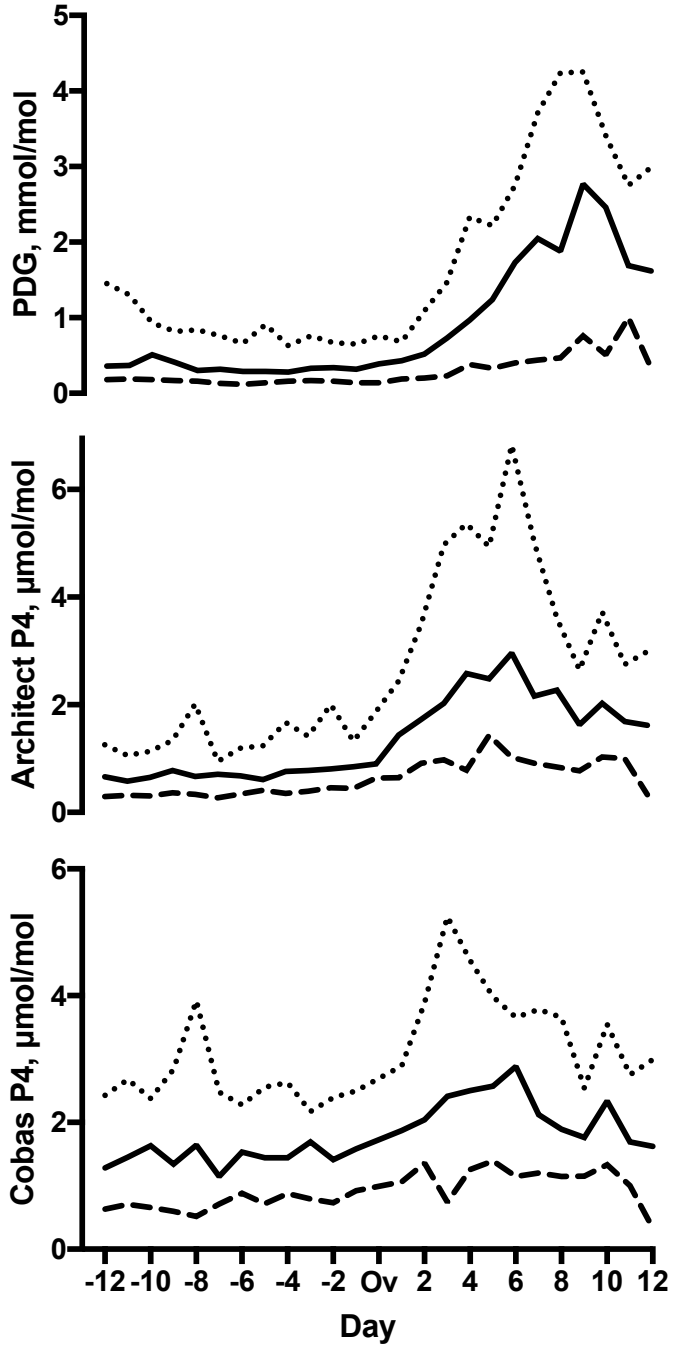


Figure 3-1-2. Analysis of pregnanediol glucuronide (PDG) and progesterone (P4) in daily urine samples in the confirmatory cohort (ovulation identified by ultrasound). Solid line – median. Dashed line – 10th centile. Dotted line – 90th centile. Ov: day of ovulation. Day represents cycle day relative to day of ovulation. PDG: pregnanediol-3-glucuronide, P4: Progesterone, Architect: Abbott Architect chemiluminescent microparticle immunoassay. Cobas: Roche Cobas e411 electrochemiluminescence immunoassay.

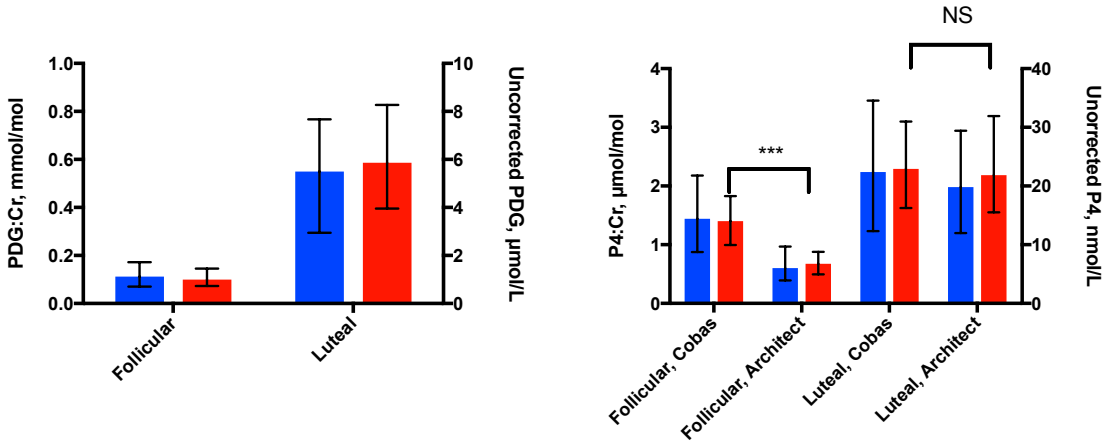


Figure 3-1-3. Median luteal and follicular values, uncorrected and corrected for creatinine. Blue: uncorrected; red: corrected for creatinine. Bars are median values, whiskers represent interquartile ranges. PDG:Cr, pregnanediol glucuronide corrected for creatinine, P4:Cr progesterone, corrected for creatinine; Architect: Abbott Architect progesterone chemiluminescent microparticle immunoassay. Cobas: Roche Cobas progesterone electrochemiluminescence immunoassay ***p<0.001 NS p>0.05.

The Architect P4 assay was therefore chosen for comparison with PDG for efficacy in confirming ovulation. Assay ROC areas under curve (AUCs) for single sample threshold were 0.951 (95% CI 0.923 to 0.978) and 0.944 (95% CI 0.916 to 0.973), for PDG and P4 respectively (p=0.7). The ROC AUCs for luteal percent rise were 0.927 (95% CI (0.915 to 0.940) and 0.950 (95% CI (0.940 to 0.961), respectively (p=0.003; **Figure 3-1-4**). The optimal individual sample threshold identified to confirm ovulation for P4 and PDG were 1.14 µmol/mmol and 0.208 mmol/mmol, respectively, yielding sensitivity and specificity of 0.88 to 0.99 with no meaningful differences between single threshold concentration or luteal percent rise, or between assays (**Table 3-1-3**). There was no significant difference between P4 and PDG ROC AUCs for threshold concentration (**Table 3-1-4**). The optimal percent luteal value rise for P4 and PDG were 165% and 195%, respectively.

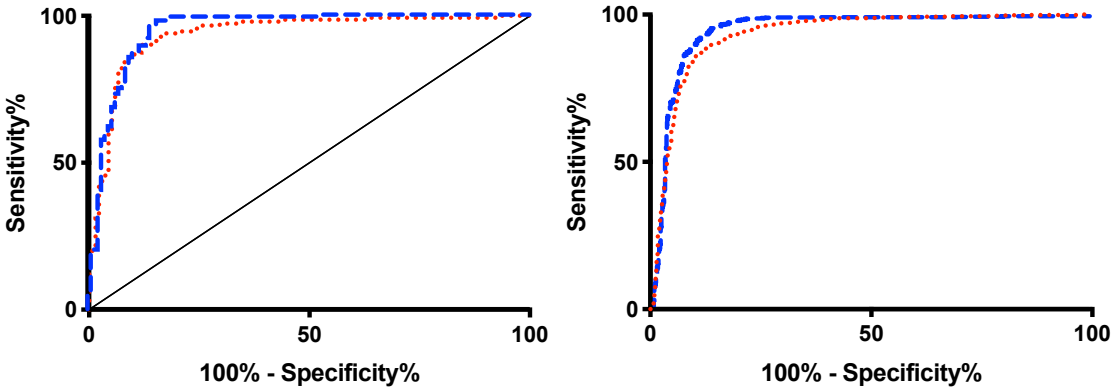


Figure 3-1-4. Receiver-Operator Characteristics creatinine-corrected PDG ELISA versus creatinine-corrected Architect P4 for single-sample threshold concentration (top panel) and two-sample follicular to luteal rise (bottom panel). Left panel: single sample threshold concentration to confirm ovulation; area under curve (95% CI) for P4 threshold (red dotted line) was 0.944 (0.916–0.973) and for PDG threshold (blue dashed line) was 0.951 (95% CI 0.923-0.978), p = 0.7. Right panel: two-sample luteal rise to confirm ovulation. Area under the curve (95% CI) for P4 two-sample difference (red dotted line) was 0.927 (0.915–0.940) and for PDG two-sample difference (blue dashed line) 0.950 (0.940–0.961), p = 0.003. Solid line: identity.

Indicator of ovulation	Value	Sensitivity	Specificity	PPV	NPV
PDG follicular to luteal percent-rise, %	195	0.90	0.90	0.90	0.90
P4 follicular to luteal percent-rise, %	165	0.88	0.88	0.87	0.88
PDG threshold value, mmol/mol	0.208	0.89	0.90	0.88	0.91
P4 luteal threshold value, µmol/mol	1.14	0.88	0.90	0.89	0.90

Table 3-1-3. Indicators of ovulation: Sensitivity specificity and predictive values of corrected pregnanediol and P4 by Architect assay single values and two-sample luteal percent increase (confirmatory cohort - ovulation identified by ultrasound). PDG, pregnanediol glucuronide. P4: progesterone, PPV: positive predictive value, NPV: negative predictive value.

	P4 single value 0.944 (95% CI .916-.973	P4 luteal percent increase 0.927 (95% CI .915 - .940)	PDG luteal 0.951 (95% CI .923-.978)	PDG luteal percent increase 0.950 (95% CI .940-.961)
P4 single value 0.944 (95% CI .916-.973		Difference 0.017 P = 0.278	Difference 0.007 P = 0.730	Difference 0.006 P =0.696
P4 luteal percent increase 0.927 (95% CI .915 - .940)	Difference -0.017 P = 0.278		Difference 0.0151 P= 0.112	Difference 0.023 P = 0.003
PDG single value 0.951(95% CI .923-.978)	Difference -0.007 P = 0.730	Difference -0.0151 P = 0.112		Difference 0.001 P = 0.946
PDG luteal percent increase 0.950 (95% CI .940-.961)	Difference -0.006 P = 0.696	Difference -0.023 P = 0.003	Difference -0.001 P = 0.946	

Table 3-1-4. Differences between areas under receiver operator characteristics curves between progesterone (P4) and pregnanediol (PDG) single-sample threshold and two-sample percent rise to confirm ovulation (difference and two-tailed p-value), in cohort of women in whom ovulation had been confirmed. PDG, pregnanediol glucuronide; P4 progesterone; CI, confidence interval.

In the exploratory cohort, 20 ovulatory and 22 anovulatory cycles were identified (**Figure 3-1-5**). For ovulatory cycles (n = 20, from 13 women), sensitivity and specificity of P4 compared with PDG were threshold: 0.89 and 0.95, percent change: 0.95 and 0.91, respectively. For anovulatory cycles (n = 22, from 14 women), sensitivity and specificity of P4 compared with PDG were threshold: 1.00 and 0.98, percent change: 0.87 and 0.92 (**Table 3-1-5**). Receiver-operator characteristic AUC for peak values identified from ovulatory and anovulatory women was 0.95 (95% CI 0.89 to 1.01). At a cut-off P4 of 1.67 µmol/mol, sensitivity was 0.90 and specificity 0.91.

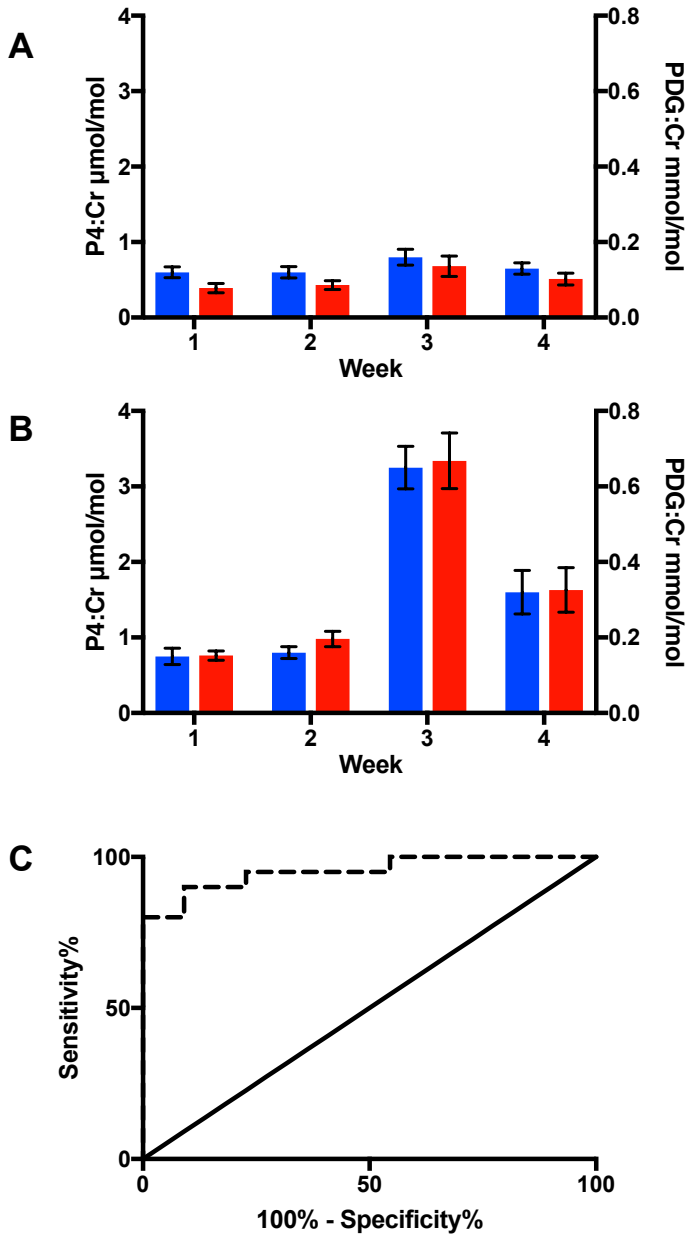


Figure 3-1-5. Weekly creatinine-corrected PDG and P4 in the exploratory cohort. (A) anovulatory cycles (n=22, 14 women), (B) ovulatory cycles (n=20, 13 women). Values are mean±SEM. Ovulation was determined by reaching a threshold cutoff in PDG and week 3 designated by the rise in PDG/P4. C:Receiver operator characteristics (ROC) for peak P4:Cr from each 4-sample consecutive series. Area under the curve 0.952 (95% CI 0.89 to 1.01). Blue bars, PDG:Cr: urinary creatinine-corrected pregnanediol glucuronide, red bars, P4:Cr: urinary creatinine-corrected progesterone. Dashed line: ROC, solid line: identity.

	PDG above luteal threshold concentration	PDG below luteal threshold concentration	
Ovulatory cycles (n=20, from 13 women)			
P4 single sample cut-off confirmed ovulation	25	2	Sensitivity of single sample cut-off P4 relative to PDG: 0.89
P4 single sample cut-off did not confirm ovulation	3	41	Specificity of single sample cut-off P4 relative to PDG: 0.95
P4 two sample percent rise confirmed ovulation	139	38	Sensitivity of P4 two sample percent rise relative to PDG: 0.95
P4 two sample percent rise did not confirm ovulation	8	375	Specificity of P4 two sample percent rise relative to PDG: 0.91
Anovulatory cycles (n=22, from 14 women)			
P4 single sample cut-off confirmed ovulation	0	2	Sensitivity of single sample cut-off P4 relative to PDG: 1.00
P4 single sample cut-off did not confirm ovulation	0	86	Specificity of single sample cut-off P4 relative to PDG: 0.98
P4 two sample percent rise confirmed ovulation	26	47	Sensitivity of P4 two sample percent rise relative to PDG: 0.87
P4 two sample percent rise did not confirm ovulation	4	539	Specificity of P4 two sample percent rise relative to PDG: 0.92

Table 3-1-5. Comparison of progesterone (P4) sensitivity and specificity relative to pregnanediol glucuronide (PDG) in the exploratory cohort (ovulation had not been otherwise identified). P4, progesterone; PDG, pregnanediol glucuronide

3.1.5 Discussion

In a sample of women in whom ovulation was carefully characterised (confirmatory cohort) we found that measurement of urinary P4 using an automated assay reproducibly demonstrated comparable relative concentration changes to PDG. Compared with PDG, the Architect P4 assay demonstrated satisfactory receiver operator characteristics and positive and negative predictive values.

In women in whom ovulation was otherwise undefined (exploratory cohort), P4 Architect was closely comparable to PDG concentration as the referent, with AUC CIs crossing unity and sensitivity and specificity of 90% and 91% respectively. The luteal percent change method estimated a marginally lower sensitivity than single threshold concentration, however this was likely due to the impartial analytical approach (each weekly sample was compared with seven other samples). We were unable to demonstrate any difference between a single threshold value or percent luteal rise in confirmation of likely ovulation.

The Architect method demonstrated a greater luteal rise for P4 than reported by Stanczyk *et al*. (1997), however this was still significantly less than was seen for PDG. A specific ELISA for PDG remains superior in confirming ovulation using urine samples to either automated P4 assay. In the exploratory cohort, the sensitivity and specificity of P4 were calculated relative to PDG, as a gold-standard technique such as TVUS had not been undertaken. This confirmed the potential of urinary P4 analysis using the Architect system for confirmation of ovulation in a clinical setting. The manual PDG assay however requires several hours to perform with overnight plate coating and/ or antibody incubation (Munro et al., 1991; O'Connor et al., 2003). Such a time investment will carry cost implications. Recent studies advocating the use of a PDG threshold concentration to confirm ovulation utilised a time-resolved fluorimetric immunosorbent assay, but details of the assay were not described (Ecochard et al., 2017; Ecochard et al., 2013; Leiva et al., 2015; Roos et al., 2015). Time-resolved fluorimetry requires more specialised equipment than the competitive TMB-based ELISA, hence it is likely this technique will be predominantly utilised by specialised reproductive laboratories. Liquid chromatography and tandem mass spectroscopy represent an accurate alternative (Sinreih et al., 2015), however the cost is likely to be prohibitive in the general laboratory. Autoanalyzers such as those tested here are in widespread use for plasma/serum in general biochemistry laboratories and using them for confirmation of ovulation, where available, would be of great practical value, reducing direct and indirect costs and improving efficiency. The message is autoanalyser use improves efficacy and is practical so is less resource intensive than a manual ELISA. They are also less likely to be associated with human error. The Abbott and Cobas P4 assays are not developed or marketed for urine but this analysis suggests that the Abbott assay shows good characteristics and may potentially be of clinical value in this context, after further validation in larger cohorts. While Architect P4 demonstrated a correlation with PDG of r = 0.71,

it shows potential for clinical application, since the identification of change from follicular to luteal concentrations is robust (90% sensitive and 91% specific in this exploratory cohort of 21 women). It may also prove a useful tool for population-based research studies, where large numbers of samples need to be analysed.

The Cobas demonstrated a matrix effect for measurement of P4 in urine, overestimating concentrations thus limiting the assay's ability to differentiate between follicular and luteal samples. Cobas also demonstrated a poor percentage recovery, an effect which was reduced by serial dilutions with phosphate buffered saline. Architect by comparison was unaffected by matrix effect in urine and showed good recovery, and thus was chosen for further comparisons.

As far as we are aware this is the first time PDG ELISA has been compared with an automated assay of P4 in urine for the confirmation of ovulation. Strengths of our study include the detailed assessment of ovulation and excellent adherence to a daily urine sampling regimen. Our study has several weaknesses. An important limitation is that the confirmatory population was relatively small, although statistical significance was achieved for the key comparison of fold increase in luteal versus follicular P4. Ultrasound, blood and urine hormone measurements in this cohort provide more detail than previous studies and represent a gold standard of ovulation determination, hence we feel this sample size was sufficient to confirm the ability of daily urinary Architect P4 to identify ovulation. Nevertheless, these data should therefore be interpreted with caution and substantially larger sample sizes are required for determination of reference ranges for threshold or cutoff values. Our assays were not contemporaneous, with 1 freeze-thaw cycle between each of them. While significant degradation of steroids was not detected, in future researchers should aim to run the assays concurrently (O'Connor et al., 2003; Reyna et al., 2001).

Larger studies including TVUS in both ovulatory and anovulatory women are required to determine the best sampling strategy to confirm ovulation. It would also be necessary to determine the efficacy of automated P4 assessment versus PDG in a range of ovulatory patterns before recommending this test for widespread clinical use, for which sufficient reliability has not yet been demonstrated.

3.2 Energy availability and physical adaptation during basic military training

Section 3.2 is in submission to the American Journal of Endocrinology and Metabolism under the title “higher energy availability is associated with favourable adaptations to infantry-based military training in women,” by Dr Robert M Gifford (RG), Prof Julie P Greeves (JG), DR SL Wardle (SW), Dr TJ O’Leary, (TO) Miss RL Double (RD), Dr Michelle Venables (MV), Prof Chris Boos (CB), Mr J Langford (JL), Prof DR Woods (DW) and Prof RM Reynolds (RR). The study was planned, ethical approval was obtained, and participants recruited by RG with supervisory input from RR, DW, JG and SW with expertise on heartrate variability provided by CB. SW, RD and TO conducted energy availability assessment phases and measurement of body composition with very occasional assistance from RG (planning, facilitating start-up and study visits, and liaison with Royal Military Academy staff) and supervisory input from JG. Heartrate variability, weight and questionnaires were measured by RG with assistance from RD, SW and the nurses from the Wellcome Trust Clinical Research Facility, Edinburgh, led by Jo Singleton. Fitness test data (1.5 mile run time) were kindly provided by Royal Military Academy Physical Training Instructors. Measurement of ^{2}H and ^{18}O and calculation of total energy expenditure by the doubly labelled water method was completed by MV. Calculation of exercise energy expenditure from accelerometry was completed by JL. Calculation of energy intake was completed by SW, TO and RD. Analyses were conducted by RG with supervisory input from DW, RR and CB. RG wrote the first draft of the manuscript, tables and figures and all authors provided final editorial oversight to the final version.

In summary, during three 10-d phases of basic military training, this study found a measure of exercise energy expenditure using a wrist-worn accelerometer demonstrated robust characteristics compared with a gold-standard free living measurement of total energy expenditure. We found total energy expenditure during basic military training was of a similar magnitude to studies of professional athletes. Assessment of energy availability was hampered by underestimation of energy intake by around 32%. However, despite this, energy availability did correlate with some adaptations to training (fat mass loss, fat-free mass gain, faster 1.5 mile run time) but not others (heartrate variability). Energy availability also correlated with eating disorder score. This study underlines the importance of energy availability for

positive training adaptation, demonstrates the potential utility of wrist-worn accelerometry to support measurement of energy availability in a variety of military training environments, but found that real-world measurement of energy availability is hampered by the challenge of accurately measuring energy intake.

3.2.1 Abstract

Energy availability (EA) underpins positive adaptation to exercise, however its measurement is hampered by inconsistent definitions of exercise energy expenditure (EEE) in field studies. In women undertaking three heterogeneous 10-d assessment phases (Phases) during basic military training, we studied an accelerometry-based measure of EA based on moderate and vigorous physical activity (EA_{mvpa}) alongside EA based on a doubly-labelled water-based estimate of total physical activity (EA_{tpa}); energy intake (EI) was measured using food diaries and 24-hour recall. We then tested the association of EA with physical adaptations (body composition by dual-energy x-ray absorptiometry, 1.5 mile best-effort run and heartrate variability) and eating behaviour.

Fifty-nine women enrolled in the study and 47 completed three Phases (mean (SD) age 23.9 (2.6) years). EA_{mvpa} correlated strongly with EA_{tpa} ($r=0.85$, $p<0.0001$) and was consistently 10.0 kcal/kg/d lower than EA_{tpa} across the range of measurement, but both were low (mean (SD) 19.8 (16.4) and 9.81 (19.8) kcal/kg/d, respectively, $p<0.001$) despite no overall difference in body mass during Phases; low EA was likely driven by under-reporting of EI. Training was associated with no overall change in total, fat-free or fat mass despite modest fluctuations throughout (partial Eta squared 0.074, 0.242 and 0.079, respectively, $p<0.001$). 1.5 mile run time improved –0:15 min (SD 0:40) and time domain, nonlinear and autonomic nervous system indices of heartrate variability demonstrated greater parasympathetic activity at rest (all $p<0.01$). EA_{mvpa} demonstrated modest correlations with improvement in 1.5 mile run time, fat mass loss, fat-free mass gain, and inversely with disordered eating risk ($r = 0.208$, 0.340, 0.187, –0.206, all $p<0.01$). There was no correlation between EA and heartrate variability. An accelerometry–based measure of EA demonstrates potential to support field studies, despite underestimation of EI. These findings underline the importance of EA for positive adaptation to training.

3.2.2 Introduction

The International Olympic Committee has highlighted the wide array of physiological and psychological ramifications in sports, underpinned by low energy availability (EA), when dietary intake and energy expended during exercise are mismatched (Mountjoy et al., 2018). Organisms apportion energy hierarchically for survival (Wade and Schneider, 1992). When exercise energy expenditure (EEE) is too high and/ or energy intake (EI) too low, locomotion is prioritised over reproductive function or bone turnover, and insufficient energy is available to maintain these processes. Most evidence of the clinical sequelae of low EA demonstrate its importance in female athletes, in whom it leads to menstrual dysfunction and osteopoenia (the female athlete triad) (De Souza et al., 2014a). Men appear to be less prone to adverse effects from low EA (De Souza et al., 2019b).

The measurement of EA requires subtraction of EEE from EI, expressed as kcal per kg fat-free mass (FFM) per day. It is necessary to measure EEE rather than simply total energy expenditure (TEE), since the latter determines energy balance (EI minus TEE). While energy balance is an important 'net output' of a metabolic environment, determining EA might be of greater clinical importance for athletes because it represents the energy 'input' to cellular process after the cost of exercise (Loucks et al., 2011).

Definitions of EEE vary between studies, which may make interpretation challenging. Exercise energy expenditure has been estimated as that due to total physical activity (EEE_{tpa}, e.g. (Brown et al., 2017; Silva et al., 2017)), which would include non-exercise activity (Levine, 2002), although the seminal studies demonstrating the importance of low EA for reproductive dysfunction (Loucks and Thuma, 2003; Loucks et al., 1998), and many others since (e.g. (Cialdella-Kam et al., 2014; Silva et al., 2018; Torstveit et al., 2018; VanHeest et al., 2014)), have measured EEE from moderate or vigorous physical activity only (EEE_{mvpa}). Investigators have measured EEE_{mvpa} using activity logs during free living (often validated by heart rate monitors or accelerometers (Cialdella-Kam et al., 2014; Hoch et al., 2009; Hoch et al., 2011; Reed et al., 2015; Schaal et al., 2011; Torstveit et al., 2018; VanHeest et al., 2014)), measuring EEE only during programmed periods of exercise (Loucks and Thuma, 2003; Loucks et al., 1998; Scheid et al., 2013) or by calculating non-exercise physical activity energy expenditure which can be deducted

from EEE_{tpa} (Lieberman et al., 2018; Williams et al., 2015). Measuring EEE_{mvpa} carries advantages over EEE_{tpa}, since the former excludes non-exercise physical activity below 1.6 metabolic equivalent (METs) tasks, which is in itself an important determination of health (Kerrigan et al., 2018), but is distinct from the intentional exercise activity addressed by the RED-S or female athlete triad paradigms. Compared with EEE_{tpa}, EEE_{mvpa} is more closely aligned with EA theory, is generalisable to the extant literature on EA, and is more relevant clinically for trainers seeking to assess and restore low EA in the field (Burke et al., 2018).

Energy intake recorded from food diaries, questionnaires and 24 h recall is prone to bias from mis-reporting (Magkos and Yannakoulia, 2003), which may be exacerbated by lapses in motivation and fatigue if reporting is prolonged (Hill and Davies, 2001; Magkos and Yannakoulia, 2003). Capling *et al.* reviewed studies reporting EI in athletes alongside measured TEE and measured, or estimated, resting metabolic rate (RMR), finding systematic mean under-reporting of 20% (Capling et al., 2017).

Chronically low EA may impair performance through impaired muscle mass, recovery and cardiovascular function, although few studies have measured physical training adaptation alongside EA (Mountjoy et al., 2018). The autonomic nervous system (ANS) plays a crucial link between cardiovascular and physical adaptation to exercise. It can be assessed non-invasively by heartrate variability, which reflects the careful balance between parasympathetic and sympathetic nervous system (PNS and SNS) activity (Shaffer and Ginsberg, 2017). Increased resting PNS activity accompanies beneficial physical training adaptation (Bellenger et al., 2016), while SNS activity can accompany overtraining, psychological stress and reduced sleep (Booth et al., 2006; Kim et al., 2018).

Basic military training, characterised by prolonged working days, restricted sleep, and high physical demands and interspersed by periods of low or moderate level activity (e.g. transiting around training areas) is a relevant, free-living setting to assess the impact of EA on adaptations to training. Military training is also highly relevant for women: a ban on women joining the infantry has recently been lifted in several nations, so that women will now be required to train more arduously. This may increase the risk of low EA-associated conditions, such as osteopoenia and stress fracture (reviewed by Friedl (2005)) or reproductive dysfunction (reviewed by Gifford et al. (2017)).

The primary aim of this study was to compare EA_{mvpa} measurement using accelerometry with EA_{tpa} based on DLW, widely considered the gold-standard measure of free-living TEE, in women during three, 10-day periods of a 44-week basic military training programme. Our secondary aim was to determine the relationship of EA, by both measures, with physical training adaptations (changes in body composition and physical fitness), autonomic adaptation (HRV) and disordered eating. We hypothesised that EEE_{mvpa} would be consistently lower than EEE_{tpa}, the difference representing non-exercise physical activity, thus representing a valid measurement of EA. We hypothesised both EA measures would be associated with concordant changes in physical and autonomic adaptation, and eating behaviour.

3.2.3 Methods

3.2.3.1 Participants and setting

All female Officer Cadets commencing the British Army Officer Commissioning Course at the Royal Military Academy, Sandhurst, (the Course) in April 2017, September 2017 and January 2018 were invited to take part at a routine briefing, 6 to 20 weeks before the Course **(Figure 3-2-1**). The Course consists of three, 14-week terms separated by 1-week adventurous training (e.g. mountaineering, skiing or paddle sports) and 2-3 weeks of leave. During the 44 training weeks, Officer Cadets undergo rigorous infantry-based training with physical, academic and leadership elements. The course is designed to be immersive and intense; working days usually last over 16 hours. Inclusion criteria were: commencing the Course, female sex, aged 18 to 30 years at start of the Course. Before starting training, participants underwent a routine detailed medical screen which included a full history, physical examination and an ECG. Participants must meet exacting predefined standards for participation on the Course (Ministry of Defence, 2018). Conditions screened for, of relevance to this paper, included primary amenorrhoea, thyrotoxicosis, confirmed eating disorder, malabsorption or food intolerance requiring medical intervention.

The study followed a repeated measures design summarised in **Figure 3-2-1**. Height was measured at visit 1 (Seca stadiometer model 217, Birmingham, UK) and weight was measured at each study visit (Seca scales model 874), wearing a T-shirt and combat trousers or shorts (**Figure 3-2-1**). Eating behaviour and body composition were measured at regular study visits at the beginning and end of each term.

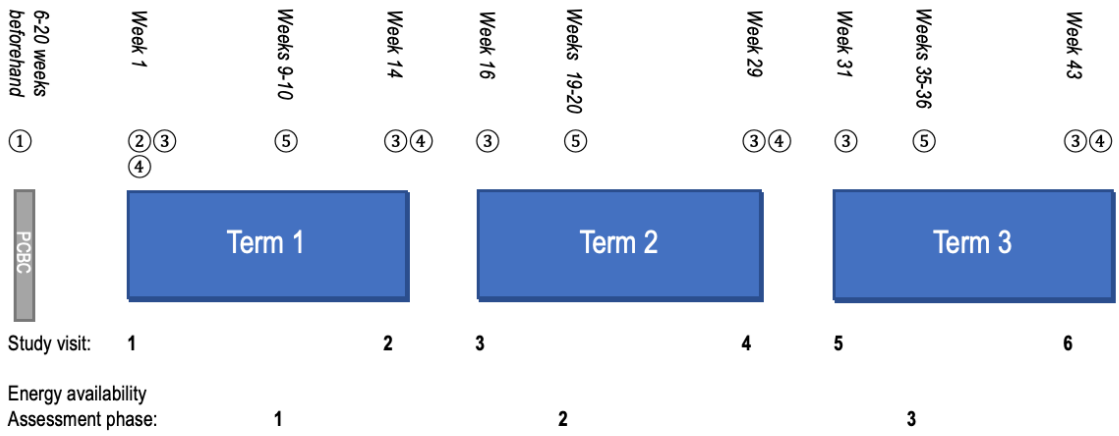


Figure 3-2-1. Scheme of study visits and energy availability assessment phases ① recruitment, ② height, ③ weight, brief eating disorder questionnaire (BEDA-Q), ④ dual-energy x-ray absorptiometry,(DXA) and heart rate variability (HRV), ⑤ 10-day energy availability assessment phase. PCBC, Pre-Course briefing course, 6 to 20 weeks before start of term 1

Body composition was measured by dual-energy X-ray absorptiometry (DXA; GE Lunar iDXA, GE Healthcare Systems, Chalfont St Giles, UK) at study visits 1, 2, 4 and 6, wearing t-shirt and shorts.

A physical activity and diet questionnaire was completed at visit 1 with reference to the preceding 6 months (Appendix 1). Participants were asked to estimate how physically active they were compared with individuals of their own age and gender (rating on a five-point scale of 'much less' to 'much more'), what kind of exercise they undertook (how much hard tiring, moderate and mild exercise lasting more than 15 minutes, per week), how many snacks and meals were eaten or skipped per day, and if they had any dietary preferences (e.g. vegan, vegetarian).

3.2.3.2 Energy availability assessment phases

Total energy expenditure and EA took place during three 10 d Phases during training (once per term, denoted 1 to 3). These Phases were selected in consort with training staff to be representative of the entire Course, capturing different physical aspects. During Phase 1, participants resided in barrack accommodation and underwent programmed drill, physical exercise, field-based study and tactical leadership assessments. Programmed activities (excluding meals) took place on all ten days; median (range) duration was 12.75 (5.50 to 14.25) h/ d. During Phase 2,

participants undertook a 4-d field exercise which involved high physical demands and assessed leadership, strength, stamina and reaction to pressure. Participants rested where possible during the exercise, typically for 4-6 h/ d. The remainder of the Phase comprised weapons training, classroom activities with programmed activities lasting median (range) 12.50 (11.50 to 14.50) h/ d. During Phase 3, participants resided in barracks and underwent predominantly classroom-based lessons. Participants were expected to undertake daily physical exercise outside programmed activities (programmed for nine of the ten days, lasting median (range) 14.25 (11.50 to 16.25) h/ day).

3.2.3.3 Exercise energy expenditure by accelerometry

Exercise energy expenditure from moderate and vigorous physical activity (EEE_{mvpa}) was measured using the wrist-worn GENEActiv Original tri-axial accelerometer (Activinsights, Cambridgeshire, UK). The device was worn on the dominant wrist for 24 h/ d throughout each 10-d Phase and data were recorded at 75 Hz. The accelerometer data were processed with the GENEAread R package from The Comprehensive R Archive Network (2019) using a customised script openly available from the manufacturer (Activinsights, 2019). The data were calibrated (van Hees et al., 2014) and a wear-time assessment excluded days with more than 7 hours non-wear (van Hees et al., 2014; van Hees et al., 2011). The mean absolute gravity-subtracted acceleration was then calculated for each 1-minute epoch within a 24 hour period for each participant per phase.

Acceleration accumulated in sedentary activities (SEE g/min) were separated from acceleration accumulated in moderate and vigorous physical activity (PAEE g/min) using a cut-point of 0.09g (Esliger et al., 2011). Moderate and vigorous activities during each Phase were expressed as metabolic equivalents (METs), based on programmed activities Jette et al. (1990). Exercise energy expenditure was calculated from the accumulated duration spent undertaking moderate and vigorous activity for each Phase as follows:

$$EEE_{mvpa} = t_{mvpa} \times MET \times 3.5 \times 0.0049 \times Weight$$

Where t_{mvpa} is mean daily duration (minutes) of moderate and vigorous physical activity, MET is the mean daily metabolic equivalent of activity, 3.5 is the assumed oxygen cost for one MET (mL/kg/min), 0.0049 is the calorific value (kcal) of 1 mL oxygen and Weight is participant body mass (kg) before the measurement.

3.2.3.4 Exercise energy expenditure by doubly-labelled water

During each EA assessment period, TEE was estimated using doubly labelled water (DLW). In brief, a baseline urine sample was collected, followed by administration of a single DLW dose containing 174 mg/kg BW $H_2{}^{18}O$ and 70 mg/kg BW ^2H_2O. Urine samples were then collected on ten consecutive days (second or subsequent void of the day). Urine was stored at 5ºC for up to 7 d before being returned to MRC Elsie Widdowson Laboratory where they were stored at -20ºC until analysis. Urine samples were analysed for ^{18}O enrichment using the CO_2 equilibration method of Roether (1970). Briefly, 0.5 mL of sample was transferred into 12 mL vials (Labco Ltd., Lampeter, UK), flush-filled with 5% CO_2 in N_2 gas and equilibrated overnight whilst agitated on rotators (Stuart, Bibby Scientific). Headspace of the samples were then analysed using a continuous flow isotope ratio mass spectrometer (IRMS) (AP2003, Analytical Precision Ltd, Northwich, Cheshire, UK). For ^2H enrichment, 0.4 mL of sample was flush-filled with H_2 gas and equilibrated over 6 hours in the presence of a platinum catalyst. Headspace of the samples were then analysed using a dual-inlet IRMS (Isoprime, GV Instruments Ltd, Wythenshawe, Manchester, UK). All samples were measured alongside secondary reference standards previously calibrated against the primary international standards Vienna-Standard Mean Ocean Water (vSMOW) and Vienna-Standard Light Antarctic Precipitate (International Atomic Energy Agency, Vienna, Austria). Sample enrichments were corrected for interference according to Craig (1957) and expressed relative to vSMOW. Analytical precision was 0.3 ppm for ^2H and 0.5 ppm for ^{18}O.

Total production of CO_2 was estimated using the multipoint method of Coward (1988) and converted to TEE using the equations of Elia and Livesey (1988) with an assumed RQ of 0.85. with an assumed RQ of 0.85.

Resting metabolic rate was estimated from FFM according to the equation of Cunningham (1991) and energy expenditure of all physical activity (EEE_{tpa}) was calculated by subtracting RMR from TEE:

$$RMR = \lfloor 370 + (21.6 \times FFM) \rfloor$$

$$EEE_{tpa} = TEE - RMR$$

Where RMR is resting metabolic rate, FFM is fat-free mass, EEE_{tpa} is exercise energy expenditure of total physical activity and TEE is total energy expenditure.

3.2.3.5 Energy Intake and energy availability

Energy intake was measured using a 24-hour food diary. On the first day of each Phase participants were given a full briefing detailing the correct way to populate the food diary and were given an example diary to refer to throughout the Phase. Participants were asked to list all food and drink consumed during the day, along with the brand, method of cooking and estimated portion size. At each evening visit to the lab, individual food diaries were reviewed by the research team and participants were asked to confirm the food diary entries and recall any items that may have been missed. Researchers used questioning to prompt participants to remember any missed items, for example, '*Did you have any dessert after dinner?*' or '*Did you eat any snacks this morning?*' In addition, standardised text messages were sent at 10:00 AM and 15:00 PM each day by the same member of the research team to remind participants to continue filling in food diaries (RD). Every canteen meal for 16 women and 10 men were weighed across all three Phases to create a database of average portion sizes served. Average portions, along with the nutritional content provided by the Royal Military Academy, Sandhurst, were entered into dietary analysis software (Nutritics Ltd., Dublin, Ireland). Using these portion sizes, a large, normal and small portion were entered as 1.5, 1.0 and 0.5 of these weighed average portions. For branded snack food, weight and nutritional content provided by the manufacturer was used. All food diary data were entered into Nutritics by the same member of the research team (RD) to calculate EI.

EA was calculated as follows:

$$EA_{\,mvpa} = \frac{EI - EEE_{mvpa}}{FFM}$$

$$EA_{tpa} = \frac{EI - EEE_{tpa}}{FFM}$$

Where EA_{mvpa} is energy availability after moderate and vigorous physical activity, EI is energy intake, EEE_{mvpa} is exercise energy expenditure of moderate and vigorous physical activity, FFM is fat free mass, EA_{tpa} is energy availability after total physical activity and EEE_{tpa} is exercise energy expenditure of total physical activity.

3.2.3.6 Physical and autonomic training adaptation, eating behaviour

A best-effort 1.5 mile (2.4 km) run test was undertaken during the same weeks as study visits 1, 2, 3, 5 and 6 as part of routine testing during the Course. This field

test is a good indicator of cardiorespiratory fitness; run time correlated strongly with VO_{2max} in a recent meta-analysis (r = 0.79; 95% CI 0.73 to 0.85) (Mayorga-Vega et al., 2016).

Heart rate variability was measured and analysed at study visits 1, 2, 4 and 6, as described previously (Gifford et al., 2018a). In brief, a 5-minute single-lead electrocardiograph (ECG) was measured using portable CheckMyHeart™ devices (DailyCare Biomedical, Taiwan), according to manufacturer's instructions. Participants avoided caffeine for 8 hours beforehand, were sitting upright in a quiet environment and were requested not to talk or move during measurements. Beat-to-beat time series were produced using proprietary software (CheckMyHeart software version 2.2) and inspected manually to ensure appropriate identification of normal-normal intervals. R-R intervals were exported and analysed using Kubios® HRV Premium version 3.2.0 (http://www.kubios.com). We examined mean heart rate, traditional markers of time domain (root mean square of successive differences (RMSSD), percentage of successive normal R-R intervals greater than 50 ms, (pNN50)), and frequency domain (fast-Fourier transformed logarithms of low-frequency (0.04–0.15 Hz) and high frequency (0.15–0.40 Hz) power, LnLF and LnHF, respectively, and their ratio, LF:HF) (Task Force, 1996). Sample entropy, a non-linear measure of chaos within the HRV signal, and indices of parasympathetic and parasympathetic nervous system activity were also assessed. The parasympathetic index represents a synthesis of mean heartrate, RMSSD and the standard deviation of short term HRV (SD1), while the parasympathetic index represents heartrate, stress index (as per Baevsky and Berseneva (2008)) and mean standard deviation of long-term HRV (SD2), both reported to reflect the mean deviation from normal values (Nunan et al., 2010). Parasympathetic and sympathetic index values of zero mean that the parameters are on average equal to their normal values, while positive and negative values reflect a relative increase or decrease, respectively.

Eating attitudes were assessed using the Brief Eating Disorders in Athletes Questionnaire (BEDA-Q) and scored according to the methods of Peric *et al.* (2016) using total scores. The binary items were treated as follows: 'are you dieting' was scored at each study visit, while 'have you ever dieted' was asked at study visit 1 only. Questionnaires were completed using self-rating items on a web-based application (SmartSurvey, Tewkesbury, UK).

3.2.3.7 Statistical analysis

Comparison of EA_{tpa} and EA_{mvpa}. Normality was assessed using the Shapiro-Wilk test and non-normal data were log transformed prior to analyses. Baseline characteristics of participants who completed the study were compared with those who did not complete the study using independent samples t-tests. For each Phase, TEE was compared with EI, EEE_{tpa} with EEE_{mvpa}, and pre-Phase body mass was compared with post-Phase body mass, using paired samples t-tests. Changes in EI, TEE and EEE were evaluated for the 27 participants who completed all phases by repeated measures ANOVAs (Phase 1 × Phase 2 × Phase 3). Since EA has been demonstrated to show a linear dose-response relationship with health and training-related outcomes (Lieberman et al., 2018), it was treated as continuous variable. Since EA_{tpa} was estimated from gold-standard technique (DLW), it was treated as the referent measure against which EA_{mvpa} was assessed. EA_{tpa} and EA_{mvpa} were compared at each Phase using paired t-tests and partial correlations (taking account of repeated measures in the same individuals) comparing averages across all three Phases and individual Phases. Their relationship was described using linear regression. Systematic bias between EA_{mvpa} and EA_{tpa} was assessed by comparing the difference between measurements across the range of values, according the methods of Bland & Altman (1999). Where EA_{mpva} data were missing but EA_{tpa} data were available, EA_{mpva} was imputed from the regression equation for the appropriate Phase (41 Phase exposures, 24%) prior to determining the relationship with adaptation.

Adaptations to training. Study visit body mass, fat mass, fat-free mass, 1.5 mile run time, HRV indices and eating behaviour scores were compared by repeated measures ANOVAs (visit 1 × visit 2 × visit 3 × visit 4). The binary question ‘are you dieting’ was assessed using Spearman’s rank (r_s), investigating the relationship of dieting with training duration.

Relationship of EA_{tpa} and EA_{mvpa} with adaptation. Physical adaptation was calculated as the difference between the closest post- and pre-Phase measurement in FM, FFM and 1.5 mile run time (**Figure 3-2-1**), to give loss of FM, gain of FFM and improvement of 1.5 mile run time, respectively. The associations between EA measures and physical adaptations, post-Phase HRV (measured at the study visit after each Phase, **Figure 3-2-1**) and pre-Phase BEDA-Q continuous score (measured at the study visit before each Phase) was assessed using partial correlation.

Relationships between EA measures and the binary BEDA-Q element (answer to the question “are you trying to lose weight?”) were determined using point-biserial nonparametric correlation (r_{pb}). The EA measures among participants who reported ever dieting was compared with other participants, using an independent samples t test. A p value <0.05 was deemed significant, except for multiple correlations of training adaptations with EA, where Bonferroni adjustment was made for the six comparisons (HRV treated as one comparison) and p<0.0083 was deemed significant. Statistical analyses were carried out in SPSS for Mac version 24 (IBM, New York, NY, USA).

3.2.4 Results

3.2.4.1 Participants

Recruitment and loss to follow up are illustrated in **Figure 3-2-2**. Sixty-eight prospective Officer Cadets enrolled in the study, of whom 59 attended study visit 1 and 47 completed the three Phases. Participants were aged 23.9 years (SD 2.6) and average body mass index at visit 1 was 23.3 kg/m^2 (SD 2.1) with no differences in age, height, BMI, percent body fat or dieting status between participants who completed all study visits and those who withdrew (**Table 3-2-1**).

All participants undertook regular exercise prior to commencing the Course (**Appendix Table 3-2-1**). Fifty-two participants (76%) reported exercising slightly more or much more than other women or men of their own age prior to the Course. All participants reported doing physical activity long enough to work up a sweat (sometimes n=8 (13%), or often n=53 (87%)). The most common exercise types were running (n=49, 80%) and weight training (n=38, 62%). The exercise intensity most frequently undertaken before the Course was ‘mild’, followed by ‘hard, tiring exercise’, then ‘moderate exercise’.

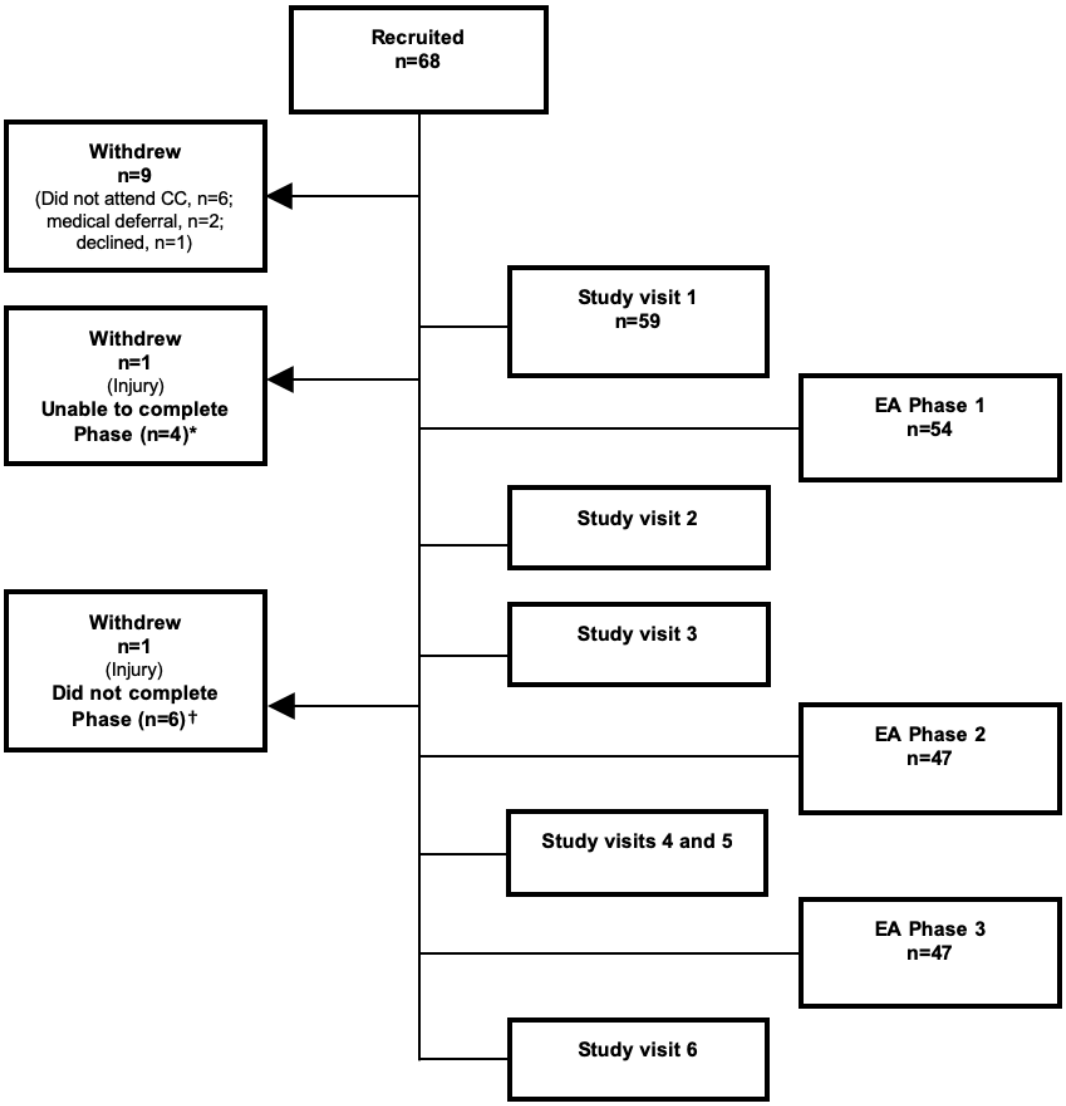


Figure 3-2-2 . **Recruitment and follow up.** EA phase, energy availability assessment phase, where energy requirement was measured using multi-point doubly labelled water and energy intake and exercise energy expenditure were estimated. At 'study visits', weight, HRV, and body composition were measured and questionnaires were completed. * 2 participants declined and 2 provided insufficient urine samples. † 2 declined and 4 provided insufficient urine samples

	Completed three Phases, n=47	Did not complete three Phases, n = 12	*p*
Age, y	23.9 (2.6)	24.9 (2.5)	0.6
Height, cm	169.0 (8.6)	167.8 (3.2)	0.4
BMI, kg/m^2	23.36 (2.12)	22.77 (2.68)	0.9
Fractional fat, %	25.0 (5.1)	25.9 (4.9)	0.6
1.5 mile run time, mm:ss	10:54 (0:54)	10:36 (0:58)	0.9
BEDAQ score, median (IQR)	4 (1,6)	5 (0, 7)	0.6
BEDAQ ever dieted, 'yes', n (%)	27 (57)	6 (50)	0.6

Table 3-2-1. Baseline evaluation of participants at baseline who completed three energy availability assessment phases (Phases) with those who did not. Values are Mean (SD) unless otherwise stated. BMI: body mass index, BEDAQ: brief eating disorders in athletes questionnaire, IQR: interquartile range. P values for independent samples t-test (participants who withdrew versus those who completed the study).

Participants reported eating 3 meals (SD 1) and 3 snacks (SD 1) per day before the Course, and 3 meals (SD 0) and 2 snacks (SD 1) per day at the start of the Course (**Appendix Table 3-2-1**). While 35 participants (51%) reported skipping meals prior to the Course (18 (30%) 'sometimes' and 13 (21%) 'often'), only 2 (3%) reported doing so at the start of the Course (both 'sometimes'). More than half the participants reported their weight had stayed the same or increased prior to the Course (n=33, 56% and n=12, 19%, respectively) and the majority (n=41, 67%) thought the Sandhurst diet was nutritionally adequate at the start of training. Six participants (10%) were vegetarian (4 lacto-ovo and 2 vegan, for definitions see **Appendix Table 3-2-1**).

3.2.4.2 Energy availability assessment phases

Phases 1, 2 and 3 were completed by 54, 47, and 47 women, respectively (**Figure 3-2-2**). Participants gained body mass during Phase 1, lost a similar amount of body mass during Phase 2, and Phase 3 was body mass neutral (**Table 3-2-2**). As expected, EEE_{tpa} was significantly higher than EEE_{mvpa} during Phases 1, 2 and 3 (mean difference 452 kcal/d (SD 358), 504 kcal/d (SD 544) and 506 kcal/d(SD 413),

respectively, all p<0.001). Higher EI was measured during Phase 1 than Phases 2 or 3 (**Table 3-2-2**). TEE was higher than EI during Phases 1, 2 and 3 (mean difference 654 kcal/d (SD 558), 1573 kcal/d (SD 578), and 673 kcal/d (SD 663), respectively, all p<0.001; **Table 3-2-2**). EA_{tpa} was lower than EA_{mvpa}, the mean values for all Phases being 9.8 (13.5) and 19.8 (16.4) kcal/kg/d.

	Phase 1		Phase 2		Phase 3		ηp^2	p
Body mass pre / post, kg	64.4 (7.7)	65.6 (7.7)	65.0 (7.6)	63.7 (7.4)	65.8 (7.7)	65.3 (7.6)		
Difference, kg	+1.2 (1.1), p<0.001		-1.3 (3.0), p=0.003		-0.3 (1.3), p=0.11		0.471	<0.001
EI, kcal/d	2667 (696)		2320 (574)*		2358 (422)*		0.115	0.003
TEE, kcal/d	3332 (424)		3849 (363)*		3041 (286)*		0.321	<0.001
EEE_{tpa}, kcal/d	2228 (355)		2811 (455)*		1963 (325)* †		0.695	<0.001
EEE_{mvpa}, kcal/d	1865 (312)		2253 (536)*		1513 (336)* †		0.752	<0.001
EA_{tpa}, kcal/kg/d	8 (11)		-10 (11)*		9 (12)* †		0.511	<0.001
EA_{mvpa} kcal/kg/d	18 (13)		1 (13)*		23 (15)* †		0.419	0.001

Table 3-2-2 Energy availability measurements. Values are mean (SD) unless otherwise stated. P for repeated measures ANOVA (main effect of time), ηp^2: partial Eta squared, *significant (p<0.05) difference versus Phase 1. † significant vs Phase 2. TEE, total energy expenditure; EI energy intake; EEE exercise energy expenditure – either from total physical activity (tpa; measured by doubly labelled water) or from measured moderate and vigorous physical activity (mpva; measured by accelerometry); EA energy availability for each measure of EEE, expressed as kcal per kg fat-free mass per day.

3.2.4.3 Comparisons of EA from accelerometry with EA from doubly-labelled water

EA_{mvpa} demonstrated robust characteristics compared with EA_{tpa} for all Phases and for each individual Phase, as shown in **Figure 3-2-3 A** and **B**, and described by linear regression equations in **Appendix Table 3-2-2**. The Bland-Altman plot (**Figure 3-2-3 C**) demonstrated a mean negative bias of –10.2 kcal/kg/d (SD 8.3) from EA_{tpa} to EA_{mvpa}, which was consistent across the range of measurement.

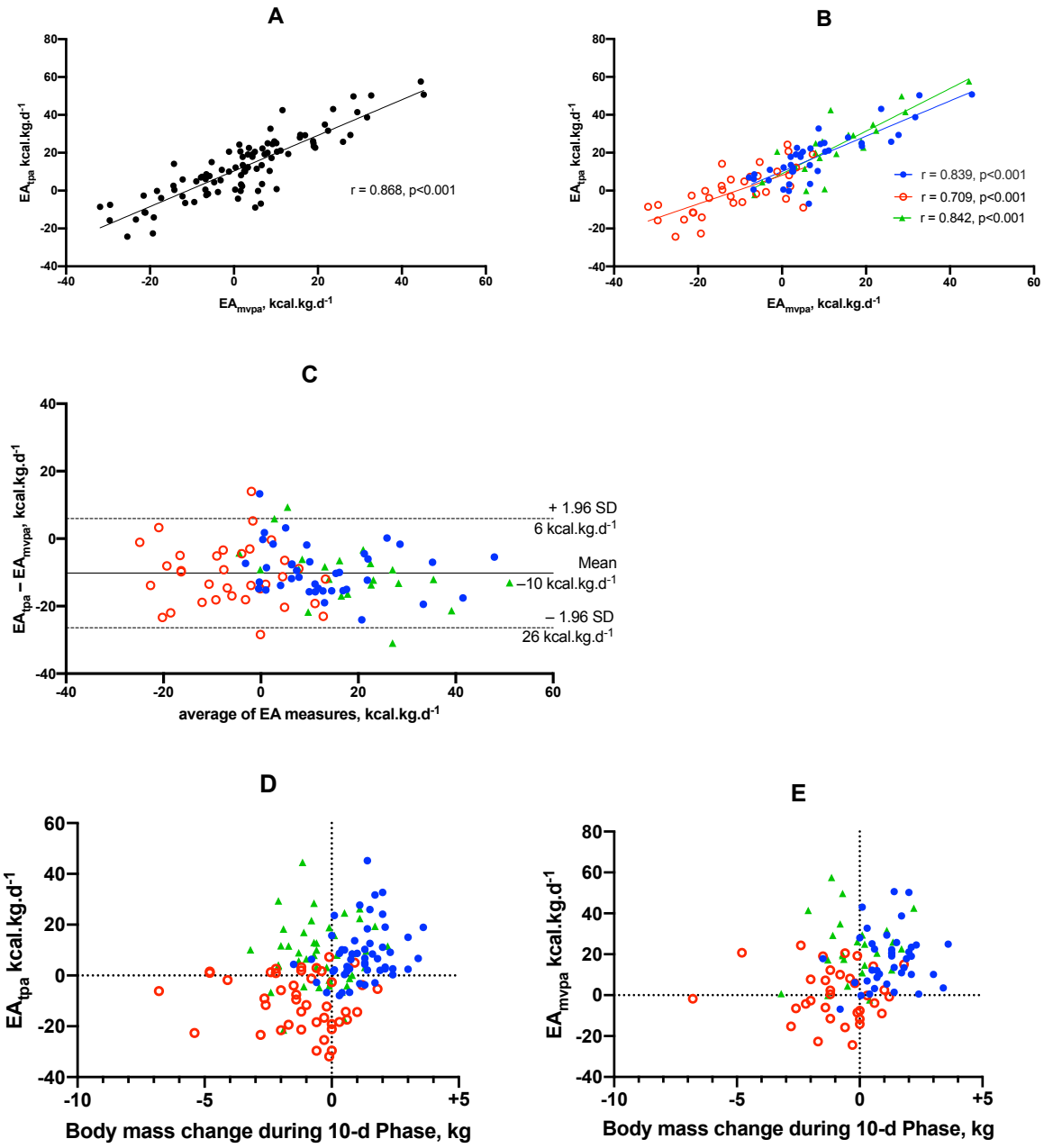


Figure 3-2-3 Comparisons of EA_{tpa} and EA_{mvpa} A: Scatter plot of all paired EA_{tpa} against paired EA_{mvpa} values with overall linear regression equation, B: Phases plotted separately. C: Bland-Altman Plot, demonstrating difference between EA_{mvpa} and EA_{tpa} at the range of values measured. D and E: change in weight during 10-d assessment Phases (weight post – weight pre) plotted against EA_{tpa} and EA_{mvpa}, respectively. In panels B to E, blue circle represent Phase 1, unfilled red circles Phase 2 and green triangles phase 3. EA_{tpa}: energy availability from total physical activity (measurement based on total energy expenditure from doubly-labelled water); EA_{mvpa}: energy availability from moderate and vigorous physical activity (measurement based on accelerometry).

3.2.4.4 Training adaptation and eating behaviour

Physical and autonomic adaptations to training and eating behaviour scores are presented in **Table 3-2-3**. A significant overall effect of time was seen for body mass, fat free mass and fat mass with small effect sizes. Small but significant fluctuations in body mass were demonstrated with pairwise increases from visit 1 to visits 4 and 5 (+0.81 kg (SD 2.65), p = 0.020 and +0.82 kg (SD 2.70), p = 0.031, respectively, compared with visit 1) but no difference between visits 1 and 6 (–0.25 kg (SD 3.03), p=0.60). Fat-free mass increased modestly from visits 1 to 2 (+0.47 kg (SD 1.52), p=0.032) but did not differ from visit 1 at visits 4 or 6 (–0.01 kg (SD 1.26), p=0.90 and +0.23 kg (SD 2.59), p=0.30). Compared with visit 1, FM was lower at visit 2 but higher at visit 4 (–0.89 kg (SD 1.92), p=0.001, and +0.85 kg (SD 2.28), p=0.003, respectively). There was no difference in FM between visits 1 and 6 (+0.04 kg (SD 2.3), p=0.90).

Fitness improved, as evidenced by decreased 1.5 mile run time compared with visit 1 at visits 2, 3 and 6 (–0:30 min (SD 0:30), p<0.001, –0:20 min (SD 0:29), p<0.001 and –0:15 min (SD 0:40), p=0.022, respectively).

Heartrate variability demonstrated beneficial adaptation during training, notably from visits 1 to 2, and more modestly from visits 3 to 4 (**Table 3-2-3**). A significant overall effect of time was seen for time domain and parasympathetic and sympathetic indices with small effect sizes, and for sample entropy with a moderate effect size. Time domain measures (pNN50% and RMSSD) demonstrated a significant rise from visit 1 to 2 followed by a modest decline from visit 2 to 3, but remaining higher than visit 1. Frequency domain measures also suggested an improvement with a decrease in LF:HF power ratio from visit 1 to visits 2 and 3, driven by an increase in HF power. Sample entropy increased at visits 3 and 4 compared with visit 1, indicating increased chaotic variability. The PNS and SNS indices, representing a synthesis of time and domain variables, showed an increased parasympathetic and decreased sympathetic activity, respectively. The BEDA-Q score was low and did not change during the study (**Table 3-2-3**).

	Visit 1	Visit 2	Visit 3	Visit 4	Visit 5	Visit 6	ηp^2	p
Body mass, kg	64.1 (7.9)	63.7 (7.9)	64.5 (7.9)	65.0 (7.7)*	65.2 (7.9)*	63.9 (7.8)	0.073	0.006
Fat free mass, kg	49.6 (5.3)	50.2 (5.4)*		49.4 (5.0)		49.4 (5.0)	0.072	0.041
Fat mass, kg	15.6 (4.0)	14.4 (3.9)**		16. 1 (4.0)*		15.6 (4.2)	0.245	<0.001
1.5 mile run time, mm:ss	10:41 (1:02)	10:08 (0:55)**	10:20 (0:57)**		10:38 (0:58)	10:29 (1:00)	0.399	<0.001
Heart rate	76.4 ±11.9	69.7 ±9.6*		71.2 ±8.9*		68.2 ±10.3**	0.122	<0.001
Time domain HRV								
RMSSD, median (IQR)	35.7 [23.5, 56.4]	51.4 [33.6, 66.81]*		45.9 [37.6 ,55.29]*		47.2 [30.6 ,66.6]*	0.075	0.036
pNN50,%, median (IQR)	14.7 [4.9, 31.6]	28.4 [11.1, 43.8]*		24.4 [16, 35.7]*		27.6 [9.6, 44.9]*	0.079	0.003
Frequency domain HRV (fast-Fourier transformed)								
LF (log)	6.87 ±0.88	7.09 ±0.97		6.84 ±0.77		7.00 ±1.07	0.002	0.70
HF (log)	6.25 ±1.19	6.70 ±1.17*		6.48 ±0.93		6.59 ±1.28	0.028	0.11
LF:HF median (IQR)	1.8 [1.2, 3.3]	1.4 [0.8, 3.1]		1.8 [1.1, 2.5]		1.6 [0.9, 2.5]	0.021	0.17

Non-linear HRV								
Sample entropy	1.61 ±0.26	1.69 ±0.24		1.72 ±0.25*		1.75 ±0.26*	0.149	0.003
Autonomic nervous system indices								
PNS Index median (IQR)	-0.7 [-1.4, 0.2]	0.1 [-0.7, 0.6*]		-0.2 [-0.6, 0.1]*		0.3 [-0.7, 1.2]**	0.090	0.001
SNS Index	0.86 ±1.35	0.14 ±1.09*		0.30 ±0.89*		0.12 ±1.26**	0.083	0.002
BEDA-Q score median [IQR]	3 [1, 4]	3 [1, 5]	3 [2, 5]	4 [2, 6]	3 [1, 5]	3 [2, 5]	0.034	0.70
BEDA-Q dieting ("yes") n (%)	11 (18.6%)	7 (13.0%)	5 (15.6%)	14 (29.2%)	16 (35.6%)	16 (30.8%)	0.139[a]	0.010[a]

Table 3-2-3 Physical and autonomic adaptation and eating behaviour scores during training. Data are mean (SD) unless otherwise stated. BEDA-Q Brief Eating Disorders in Athletes Questionnaire (score was log transformed prior to analysis). IQR, inter-quartile range. For continuous variables, p values refer to repeated measures ANOVA (main effect of time). [a] Spearman's correlation for BEDA-Q dieting (dichotomous) with visit week. RMSSD: root mean square of successive differences, pNN50: percentage of successive normal R-R intervals above 50 ms, IQR: inter quartile range, LF: low frequency power, HF: high frequency power, PNS: parasympathetic nervous system, SNS: sympathetic nervous system.* pairwise difference with visit 1 ($p<0.05$), ** pairwise difference with visit 1 ($p<0.001$)

3.2.4.5 Correlations of EA with training adaptations and eating behaviour

Training adaptations were related to one another: FM loss demonstrated a modest inverse association with FFM loss, a moderate positive correlation with 1.5 mile run improvement, and a modest inverse association with BEDA-Q score. FFM loss was not associated with 1.5 mile run improvement or BEDA-Q, however 1.5 mile run improvement was modestly inversely associated with BEDA-Q score. FM loss, FFM loss, 1.5 mile run improvement and BEDA-Q outcomes demonstrated no association with any HRV parameter. There was a modest positive association between BEDA-Q score and answering yes to “are you trying to lose weight?”.

Associations between EA, adaptations to training, and eating disorder scores are presented in **Table 3-2-4**. Both EA_{tpa} and EA_{mvpa} demonstrated significant, modest associations with pre–post Phase FM loss (i.e. higher EA was associated with greater loss in FM) and modest associations with pre–post Phase 1.5 mile run improvement (i.e. higher EA was associated greater improvement in run time). There were no association between FFM gain and EA, but FFM gain was weakly associated with FM loss. There was no association between EA and any of the HRV parameters. As expected, EA demonstrated a modest but significant negative association with BEDA-Q score (continuous scale or binary question: ‘are you trying to lose weight?’). BEDA-Q score demonstrated a modest negative association with FM loss, and there was a trend towards a negative correlation with 1.5 mile run improvement (after Bonferroni adjustment for multiple comparisons: r –.302, p=0.012, **Table 3-2-4**). EA_{tpa} and EA_{mvpa} were lower among participants who reported dieting at any point during the study or beforehand, compared with those who did not (-0.18 (11.7) kcal/kg/d versus 5.5 (16.1) kcal/kg/d, p=0.016; and 22.0 (13.0) kcal/kg/d versus 29.0 (15.6) kcal/kg/d, p=0.004, respectively).

	EA_{tpa}	EA_{mvpa}	FM loss	FFM gain	1.5 mile run improvement	RMSSD (log)	pNN50 (log)	LF (log)	HF (log)	LF:HF (log)	Sample entropy	PNS Index	SNS Index	BEDA-Q score
FM loss	.376**	.373**												
FFM gain	.161	.213	.222**											
1.5 mile run †	.284**	.252**	.482**	.082										
RMSSD (log)	.203	.208	.007	.154	.087									
pNN50 (log)	.182	.193	–.037	–.23	–.165	*.838***								
LF (log)	–.118	–.097	–.018	–.005	–.227	*.730***	*.484***							
HF (log)	.252	.227	–.006	.201	–.115	*.918***	*.847***	*.678***						
LF:HF (log)	–0.197	–0.182	.044	–.204	–0.067	*–.271*	*–.472***	*.276**	*–.445***					
Sample entropy	.171	.204	.047	.059	–.048	*.2938*	*.054*	*–.302*	*.286*	*–.391***				
PNS Index (log)	.168	.188	.068	0.028	–.042	*.806***	*.823***	*.564***	*.761***	*–.348***	*–.018*			
SNS Index	–.163	–.184	–.035	.05	.074	*–.799***	*–.762***	*–.672***	*–.679*	*.413***	*–.321***	*–.880***		
BEDA-Q score	–.367**	–.321*	–.285*	.036	–.302	–.318	–.19	–.246	–.299	.133	.029	–.157	.038	
BEDA-Q – "are you trying to lose weight?"[pb]	–.211*	–.206*	.037	–.057	.037	.002	.007	.024	–.010	.062	.024	–.088	.120	.343**

Table 3-2-4 Correlations between EA measures and training adaptation. Correlations are between EAtpa, EAmvpa measurements during Energy Assessment Phases, and Pre to post-Phase training adaptation, or between concurrent training adaptation measures. All are partial correlations (taking account of repeated measures in individuals) except marked ^{pb} (point biserial non-parametric correlation). Significant correlations after Bonferroni adjustment: ** p<0.0001; * p<0.008. EA: energy availability measured by $_{tpa}$ (total physical activity, from total energy expenditure) or $_{mvpa}$ (from moderate and vigorous physical activity, from accelerometry), FM and FFM loss and 1.5 mile run improvement: difference in fat mass, fat free mass and 1.5 mile best-effort run time, respectively, from pre to post-EA measurement (pre minus post). Heart-rate variability (HRV) and BEDA-Q measured before each EA measurement. Associations of HRV variables are italicised; these are expected to correlate strongly with one other. RMSSD: root mean square of successive differences, log: transformed by natural logarithm, pNN50: percentage of successive normal R-R intervals above 50 ms, IQR: inter quartile range, LF: low frequency power, HF: high frequency power, PNS: parasympathetic nervous system, SNS: sympathetic nervous system, BEDA-Q: Brief eating disorder in athletes questionnaire.

3.2.5 Discussion

Energy availability based on estimation of moderate and vigorous physical activity using accelerometry demonstrated robust characteristics in comparison with EA based on total physical activity from the gold-standard technique in this setting of arduous military training. EA_{mvpa} measured on average 10 kcal/kg/d higher than EA_{tpa}, which likely represents the energy expenditure difference between 'total physical activity' and 'moderate and vigorous exercise', i.e. non-exercise activity.

Energy availability values according to either EEE_{tpa} or EEE_{mvpa} were far lower than would be anticipated based on other studies that measured EA (reviewed by Burke et al. (2018)). However, TEE and EEE were commensurate with previous reports using DLW in athletes (reviewed by Capling et al. (2017)). On average, EI was 26 (SD 8) % lower than TEE (958 (732) kcal/d) across the three Phases. Based on an average tissue density of 7,000 kcal/kg for adult women (Black et al., 1986) such an energy deficit would be associated with an average body mass loss 1.06 kg per Phase. Instead, we observed no significant body mass change during Phases on average, implying crude energy balance. Therefore, EI was likely to be underestimated by 26%, which is consistent with a systematic review of studies in free-living athletes by Capling et al. (2017). Any underestimation may have been related to inaccurate portion size estimation, omission of food and drinks, altering foods consumed, poor memory or tiredness (Hill and Davies, 2001), while increased frequency of eating, due to time pressure, can be challenging to quantify (Magkos and Yannakoulia, 2003).

Alternatively, homeostatic mechanisms could have contributed to the discrepancy we observed. Acute bouts of exercise are associated with suppression of hunger and negative energy balance, possibly caused by changes in appetite-modulating gut hormones (Martins et al., 2007). Several studies have shown repeated exercise bouts increase anorectic hormones like peptide YY, pancreatic polypeptide and GLP-1 while suppressing orexigenic hormones like ghrelin (reviewed by Braun and Newman, (2019)). Equilibrium between a change in habitual exercise and EI may take weeks or months to achieve, so in RMAS where there are frequent and pronounced changes in physical activity, it is conceivable that appetite and EEE may be chronically mismatched (Blundell et al., 2003).

Non-exercise activity thermogenesis, including activity like writing or typing, or non-purposeful movements like fidgeting, may explain the high TEE we observed. Manual

workers expend more NEAT than office workers, and NEAT can be increased by overfeeding and cold exposure (Levine, 2002). Increase in NEAT could increased non-exercise activity thermogenesis may have been responsible, increasing energy expenditure from non-purposeful movements. Dietary induced thermogenesis

Both EA_{mvpa} and EA_{tpa} were associated with positive training adaptations: loss in FM and faster 1.5 mile run time. These benefits were more marked over the first term, tended to wane over the second term, and were partially recovered by the end of term three. That higher EA was associated with FM loss could be interpreted as a result of energy compensation following higher EEE during Phases, whereby compensatory increases in EI and reduction in non-exercise activity may be a response to periods of lower EA (i.e. the 5 to 9 weeks after the Phase but before the DXA scan). Energy compensation is observed following exercise interventions but is highly variable between individuals (Riou et al., 2015). For example, during controlled, 16-day exercise protocols of varying intensities, Whybrow et al. (2008) found a compensatory EI increase of 30% to 50% among men and women. Schubert et al. (2017) studied 24 athletes (14 female), following high interval exercise training, finding that higher energy compensation was associated with increased FM and reduced FFM. While military training is not an exercise intervention *per se,* the Course includes long durations of moderate physical activity. Sedentary activity (often defined as 1.5 METs or less) is not the focus of military training, and may have inadvertently increased following periods of increased training-associated EEE_{mvpa}. Sedentary activity, as well as EI, is an important determinant of metabolic, cardiovascular and psychological health (de Rezende et al., 2014). Our findings suggest that lower EA may have been associated with altered sedentary activity as well as increased EI, but studies of EA are required which use accelerometry to discriminate EEE_{mvpa} and sedentary activity, as others have done (e.g. Kerrigan et al., 2018), to confirm this.

The overall improvement noted in cardiovascular fitness, indicated by improved 1.5 mile best-effort run time, was positively associated with EA_{mvpa}, but demonstrated a stronger correlation with FM loss ($r = .252$ and $r = .482$, respectively, both $p<0.0001$). Our study design does not allow us to determine causality, although increased energy compensation in following lower EA is a plausible explanation for the relationship observed between favorable adaptations and higher EA. In a synthesis of data from two randomised, controlled trials of exercise for obesity, McNeil et al. (McNeil et al.,

2017) showed a large inter-individual variation in energy compensation, with VO_{2peak} being significantly higher among participants with lower energy compensation.

Military training is associated with rapid changes in the volume and nature of physical activity, and although our participants reported active lifestyles before training, they found the training intensity highly challenging, as evidenced by a parallel study on stress responses (Gifford et al., 2019b). It takes several years to develop feeding habits (Blundell and King, 1999); the immersive nature and abrupt onset of initial military training could be associated with delayed adjustments in EI (Ackerman et al., 2019a). That participants' 1.5 mile run times were longer around the end of term 2 (visit 4) could be explained by increased energy compensation during this term. Several arduous field exercises were conducted during term 2 (including during Phase 2), culminating in an important selection process in week 12 of this term, and were likely to be associated with low EA. In the study of six women crossing Antarctica in **Chapter 4**, despite an energy deficit for 61 days (on average ~10 kg loss), there was no change in FFM (**Section 4.1**), REE or substrate utilisation (Hattersley et al., 2019b), weight-adjusted VO_{2peak} was unchanged and anaerobic capacity improved (Thake et al., 2019). Differences in adaptation from the present study could be explained by the different type of exercise or the nutritional programme which mitigated cyclical energy deficit and surplus with steady through-expedition EI (**Section 4.2**).

Positive autonomic adaptations were demonstrated throughout the Course, but no association was seen with EA. Heartrate variability has been measured in studies of psychological stress as well as exercise: increased HRV is associated with physical fitness (Task Force, 1996), and inversely associated with negative effects of psychological stress (Kim et al., 2018). Increased parasympathetic and decreased sympathetic activity were observed in time domain measures consistently throughout the Course, especially during term 1, although time and frequency domain measures were slightly below means for athletes reported elsewhere (Shaffer and Ginsberg, 2017). We have reported that the Course carried a significant burden of psychological stress throughout (Gifford et al., 2019b), so we conclude that the change in autonomic function (i.e. adaptation) was a benefit of exercise. Our findings imply that autonomic adaptation was not influenced by energy compensation. In the aforementioned study of women before and after an Antarctic traverse, HRV was unchanged immediately after the expedition, with a subsequent increase in sample entropy, time and

frequency domains after 2 weeks of recovery (Gifford et al., 2018a). This finding also implies that beneficial autonomic adaptation of exercise was independent of energy deficit during the expedition.

The BEDA-Q has received increasing recognition (Avliyakulov et al., 2014; Gordon et al., 2017; Knapp et al., 2014; Mountjoy et al., 2018) and offers advantages in its brevity and indirect approach compared with the EDI or derivatives of the Eating Disorders Examination Questionnaire (Cooper et al., 1989; Morgan et al., 1999). In the present study, BEDA-Q elements suggesting increased risk of disordered eating were weakly associated inversely with EA, FM loss and 1.5-mile improvement. EA was also modestly lower among participants who reported trying to lose weight at any point during the study or beforehand. The BEDA-Q therefore demonstrated potential for ongoing use in a military setting.

Assessment of EA places a significant burden on the individual to collect data (Ackerman et al., 2019a). The Course, which involved significant time pressure on participants and prolonged moderate physical activity, was a challenging context for the measurement of EEE. A key strength of our study was therefore to demonstrate the validity of a simple, accelerometry-based measurement of EEE from moderate and vigorous physical activity, in a setting where activity logs were not possible. This measure may have wider applications for assessment of free-living EA, but in other contexts would warrant further validation alongside concurrent DLW or indirect calorimetry.

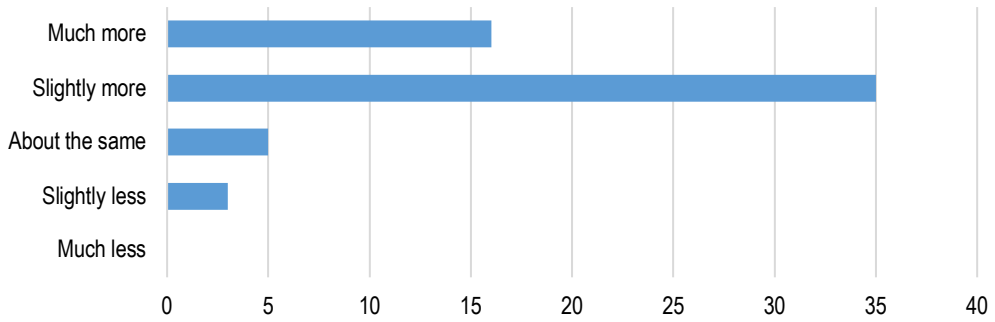
Our study has several limitations. Accelerometry tends to underestimate EEE in free-living environments (Murakami et al., 2016) and may contribute to overestimating EA. However, EA was likely underestimated due underreporting of EI, which could relate to participant motivation and fatigue from reporting over prolonged durations, or the time pressure of the Course. Poor validity of reports of EI is a widespread problem and difficult to address. The gold-standard weighed food analysis is unlikely to capture all food in a free-living environment due to the practicality of weighing irregular meals and snacks (Gemming et al., 2015). Technological innovations have emerged to attempt to address these shortcomings, however these often require in-depth analysis of images and may not capture all EI (Gemming et al., 2015; Martin et al., 2014; Rangan et al., 2016). These may warrant consideration for future studies. We were unable to carry out measurement peak or maximal oxygen uptake, or perform indirect

calorimetry due to time constraints of the Course. It was not feasible to measure HRV in the quiet controlled environment recommended (Task Force, 1996); measurements were necessarily completed at study visits when other investigations were taking place. However, these conditions were broadly consistent between visits, so they are unlikely to have impacted the pattern of our findings. Anecdotally, the 1.5 mile run assessment is affected by volition, learning of technique or course specificsm, and by ulterior motives (e.g. not being streamed into a more difficult exercise group). Thus, VO_{2max} would have been a preferable measure of cardiovascular fitness.

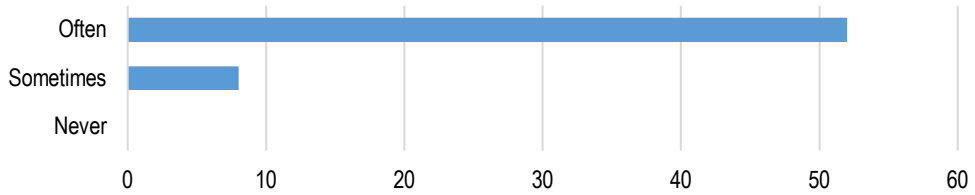
In conclusion, we have demonstrated the validity of a simple measure of EEE based on moderate and vigorous physical activity identified from accelerometry in women undertaking basic military training, against a DLW-based measure. However, EA was underreported due to underreporting of EI of around 26%. Basic military training was associated with beneficial physical and autonomic adaptations over 11 months, of which physical but not autonomic adaptations were associated with EA. EA_{mvpa} warrants further investigation as a contribution to screening for low EA.

3.2.6 Appendix Tables to Section 3-2

During the six months prior to starting the Commissioning Course, how active do you think you have been compared to other men or women your age?



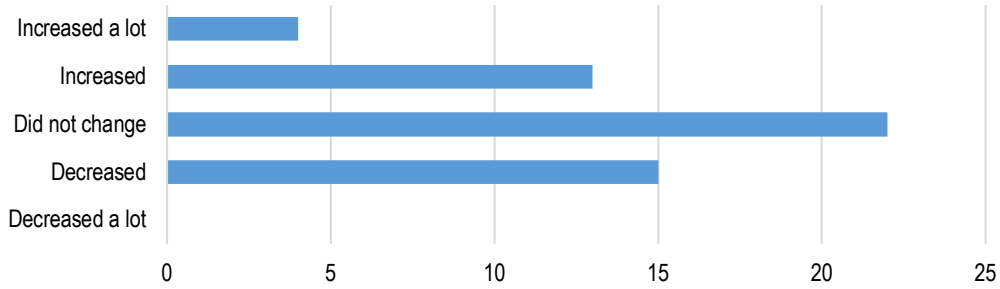
Before you arrived at Sandhurst, how frequently did you do any activity long enough to work up a sweat (heart beats quickly)?



What type of exercise did you do most often before you arrived at Sandhurst?

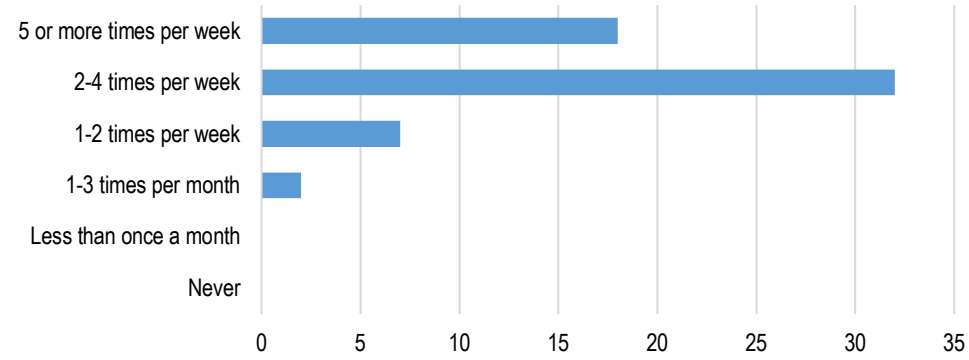
	I didn't exercise	Running	Team sport (football, hockey, rugby, etc.)	Swimming	Cycling	Weight training
	0 (0%)	49 (80%)	16 (26%)	10 (16%)	12 (20%)	38 (62%)

Between PCCBC and starting the Commissioning Course how much did you change the amount of exercise you did compared to the 6 months before that?

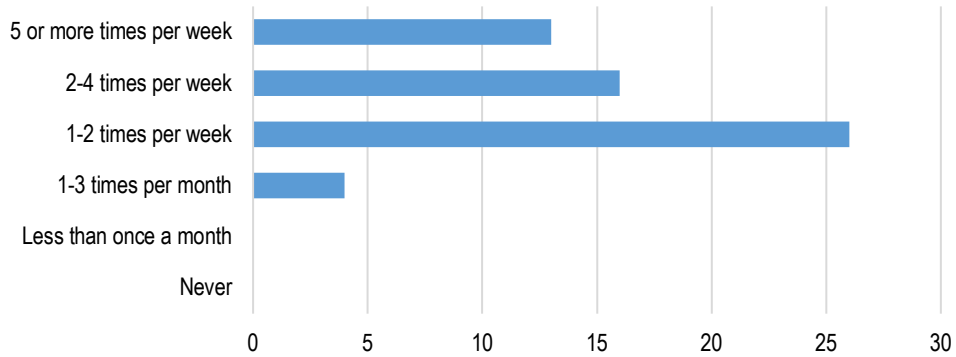


In a typical week before you arrived at Sandhurst, how many times did you do the following kinds of exercise for more than 15 minutes?

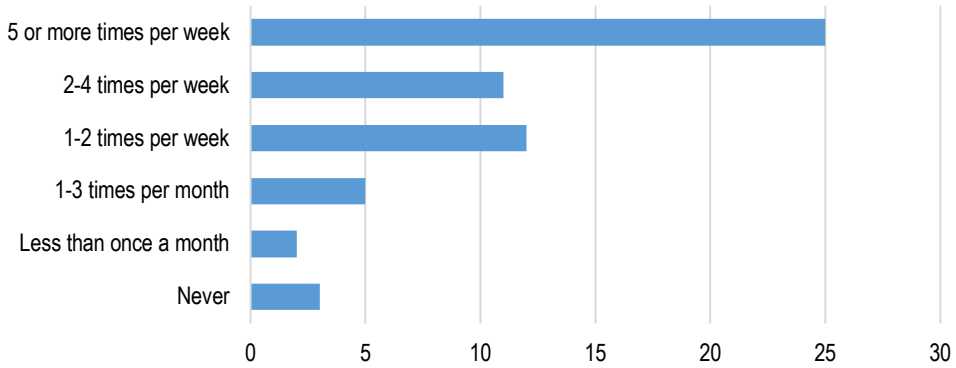
Hard, tiring exercise (heart beats rapidly and get out of breath quickly, e.g. running, jogging, hockey, football, rugby, basketball, hard swimming, long distance cycling, hard weights session, karate)



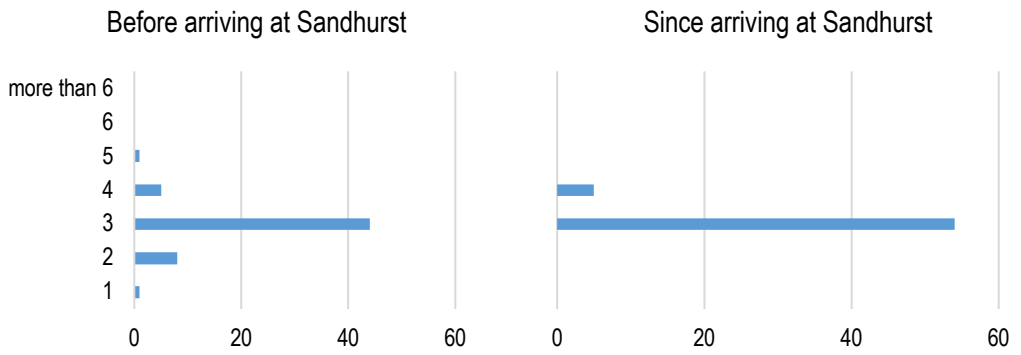
Moderate exercise (fairly tiring but not exhausting, e.g. fast walking, tennis, easy cycling, badminton, easy swimming, volleyball, rock climbing, hill walking, easy weights session, gardening, cricket)



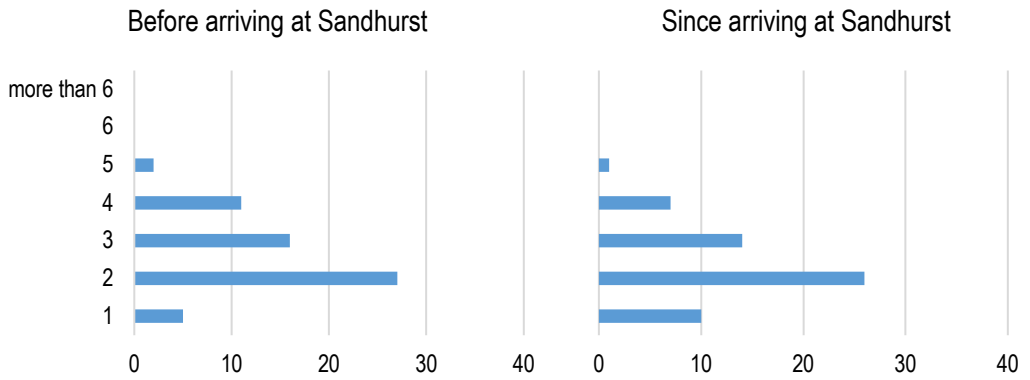
Mild exercise (small amount of effort, e.g. bowling, yoga, easy walking)



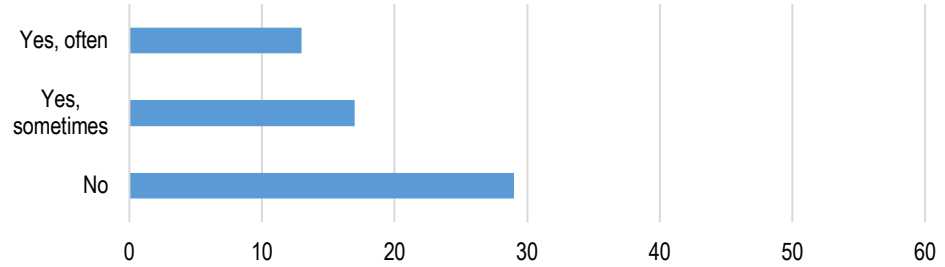
How many meals (i.e. breakfast, lunch, dinner) did you usually eat per day?



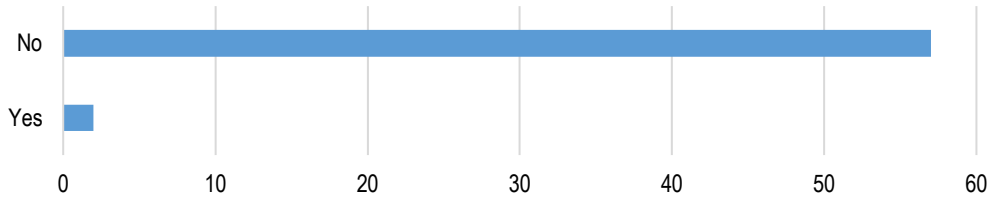
How many snacks (i.e. chocolate bar, sports bar, piece of fruit) did you usually eat per day?



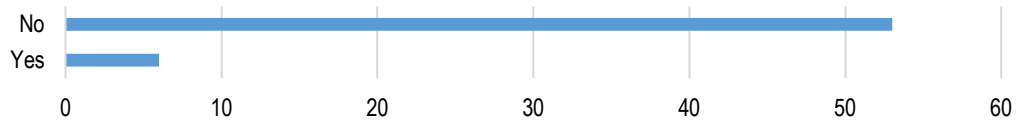
Before arriving at Sandhurst have you skipped meals?



Have you skipped any meals since arriving at Sandhurst?



Are you vegetarian?

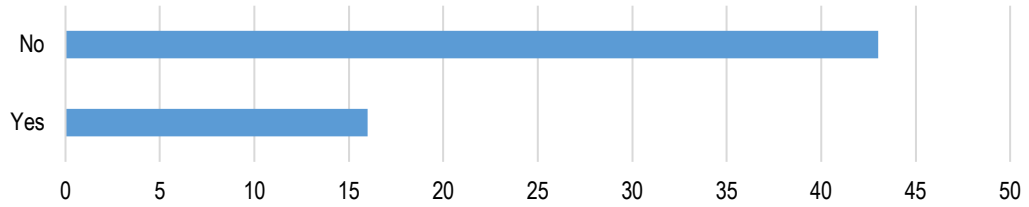


Please indicate which type of vegetarian

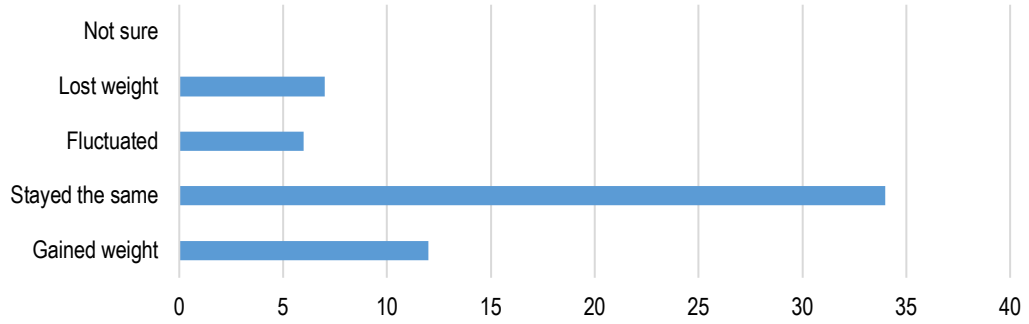
Vegan (does not eat meat, eggs or dairy) – 2

Lacto-ovo (does not eat meat, but does eat eggs and dairy) – 4

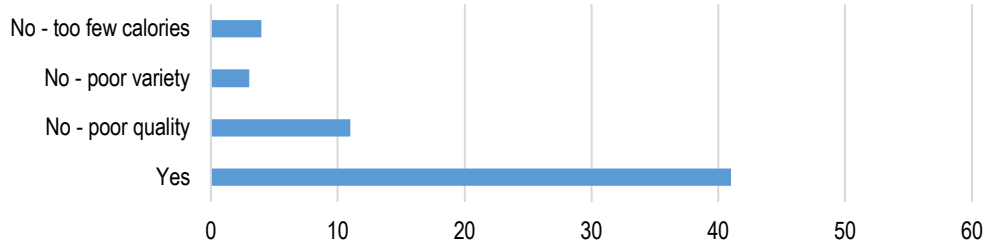
In the past 6 months have you limited the amount or type of food you eat to control your weight?



Since PCCBC what has happened to your weight overall?



Do you think your diet is nutritionally adequate (you have a good diet, eat enough and get what you need from the food you eat)?



Appendix Table 3-2-1. Exercise and diet at the commencement of the study. PCCBC: Pre-Commissioning Course Briefing Course (5 to 13 weeks beforehand).

	Equation	R^2	P
All Phases	Y = 0.94*X + 10.36	0.76	<0.001
Phase 1	Y = 0.93*X + 10.10	0.70	<0.001
Phase 2	Y = 0.76*X + 8.09	0.47	<0.001
Phase 3	Y = 1.12*X + 8.92	0.71	<0.001

Appendix Table 3-2-2. Linear regression equations of energy availability measured by accelerometry (Energy availability from moderate and vigorous physical activity, EA_{mvpa} X) and by doubly labelled water (energy availability from total physical activity, EA_{tpa} , Y)

Where Y = EA_{tpa} and X = $EA_{mvpa.}$ R^2: coefficient of determination.

3.3 Hypothalamic–pituitary–adrenal responsiveness to basic military training in women

The following section is in press with Psychoneuroendocrinology under the title "Positive adaptation of HPA axis function in women during 44 weeks of infantry-based military training", by Dr Robert M Gifford (RG), Dr Thomas J O'Leary (TO), Miss Rebecca L Double (RD), Dr Sophie L Wardle (SW), Dr Luke D Boyle (LB), Dr Natalie ZM Homer (NH), Prof Clemens Kirschbaum (CK), Prof Julie P Greeves (JG), Dr David R Woods (DW) and Dr Rebecca M Reynolds (RR). The study was planned, ethical approval was obtained, and participants were recruited by RG with supervisory input from RR, DW, JG and SW. RG, SW, RD and TO conducted measurements, with assistance from the Wellcome Trust Clinical Research Facility, Edinburgh, led by Jo Singleton. Assays were performed by NH (plasma cortisol), CK (hair cortisol) and LB (cortisol binding globulin). Dr Forbes Howie and Miss Kirsten Wilson (University of Edinburgh Centre for Reproductive Health) performed albumin and saliva cortisol assays. Analyses were conducted by RG with supervisory input from RR and DW. RG wrote the first draft of the manuscript, tables and figures and responses to reviewers' comments; all authors provided editorial oversight.

In summary, this study found women undertaking basic military training demonstrated an initial increase in morning cortisol over the first two training months, likely representing a stress response, followed by habituation to the stress of training over the next 9 months. Fasting plasma cortisol decreased during training, while cortisol response to low dose ACTH increased in non-combined oral contraceptive pill users. Average monthly hair cortisol increased throughout training. Participants reported high levels of stress and sleep deprivation, which manifested in reduced mood and resilience ratings throughout. These findings are consistent with a healthy adrenocortical response to demonstrably stressful training.

3.3.1 Abstract

3.3.1.1 Background

Basic military training (BMT) is a useful model of prolonged exposure to multiple stressors. Eight to twelve-week BMT is associated with perturbations in the hypothalamic-pituitary-adrenal (HPA) axis which could predispose recruits to injury and psychological strain. However, characterisations of HPA axis adaptations during BMT have not been comprehensive and most studies included few if any women.

3.3.1.2 Methods

We studied women undertaking an arduous, 44-week BMT programme in the UK. Anxiety, depression and resilience questionnaires, average hair cortisol concentration (HCC), morning and evening saliva cortisol and morning plasma cortisol were assessed at regular intervals throughout. A 1-h dynamic cortisol response to 1 μg ACTH-1-24 was performed during weeks 1 and 29.

3.3.1.3 Results

Fifty-three women (aged 24 ±2.5 years) completed the study. Questionnaires demonstrated increased depression and reduced resilience during training (F 6.93 and F 7.24, respectively, both $p<0.001$). HCC increased from 3 months before training to the final 3 months of training (median (IQR) 9.63 (5.38, 16.26) versus 11.56 (6.2, 22.45) pg/mg, $p=0.003$). Morning saliva cortisol increased during the first 7 weeks of training (0.44 ±0.23 versus 0.59 ±0.24 μg/dL $p<0.001$) and decreased thereafter, with no difference between the first and final weeks (0.44 ±0.23 versus 0.38 ±0.21 μg/dL, $p=0.2$). Evening saliva cortisol did not change. Fasting cortisol decreased during training (beginning, mid and end-training concentrations: 701 ±134, 671 ±158 and 561 ±177 nmol/L, respectively, $p<0.001$). Afternoon basal cortisol increased during training while there was a trend towards increased peak stimulated cortisol (177 ±92 versus 259 ±13 nmol/L, $p=0.003$, and 589 ±164 versus 656 ±135, $p=0.058$, respectively).

3.3.1.4 Discussion

These results suggest a normal stress response in early training was followed quickly by habituation, despite psychological and physical stress evidenced by questionnaire scores and HCC, respectively. There was no evidence of HPA axis maladaptation. These observations are reassuring for women undertaking arduous employment.

3.3.2 Introduction

Stress can be defined as the response of an individual to a threat or challenge (a stressor) to maintain mental or physical allostasis (Selye, 1946). Basic military training is an ideal setting for the study of stress, since it entails prolonged exposure to multifaceted stressors, such as long days of physical work, restricted food intake and sleep, austere environments, time pressure and increasing responsibility while under continuous assessment by military instructors. Field exercises, a core

component of basic military training, combine strenuous exertion over days or weeks with challenging scenarios of increasing complexity, in an unfamiliar, multi-stressor environment. The overall aims of basic military training are to test leadership and multi-tasking and develop traits like self-awareness and physical and mental robustness.

Cortisol, the main effector hormone of the hypothalamic-pituitary-adrenal (HPA) axis, is an important biological marker of stress. Cortisol is released in a pulsatile manner. Fasted morning plasma cortisol concentrations can be considered a 'stress' response to fasting or venepuncture (Reynolds et al., 2001b), whereas early morning salivary concentrations may provide information about the HPA and neurophysiological response to waking (Chida and Steptoe, 2009). Cortisol concentrations measured in urine or hair give additional information about activation of the HPA axis over longer durations (hours or months, respectively). Morning and evening sampling on the same day allows the diurnal cortisol slope to be calculated, with slope size inversely associated with a wide range of mental and physical health outcomes (Adam et al., 2017). Cortisol can also be measured in response to physiological stimuli (e.g. adrenocorticotrophic hormone, ACTH) to observe isolated HPA axis function, the size of response being associated with traumatic stress exposure (Golier et al., 2014), and increased risk of cardiovascular disease (Reynolds et al., 2001a) and reproductive dysfunction (Ackerman et al., 2013).

Sustained elevations in serum cortisol have been reported following stressful military captivity training (Taylor et al., 2007b). Low concentrations of hair and saliva cortisol in response to social stress predict subsequent development of post-traumatic stress disorder during military deployments (Steudte-Schmiedgen et al., 2015). Variations in cortisol concentrations have complex and multidimensional associations with a variety of biological and psychological disorders. For example, sleep deprivation is associated with relatively low wakening cortisol compared with the following evening (Abell et al., 2016a), while hair cortisol is positively associated with symptoms of depression (Abell et al., 2016b) and stress (Stalder et al., 2017), and negatively with anxiety disorders (Stalder et al., 2017). Higher average overnight serum cortisol is found in anorexia nervosa and functional hypothalamic amenorrhoea (Gordon et al., 2017). Overnight cortisol concentration is associated with lower bone mineral density (Lawson et al., 2009) and reduced gonadotrophin secretion (Ackerman et al., 2013) in women. Increased cardiovascular risk is

associated with lower cortisol response to waking, higher average hair cortisol (Kuehl et al., 2015), and lower morning cortisol concentration compared with evening concentration (Kumari et al., 2011).

Previous studies of cortisol responses to basic military training have only been undertaken during short duration training. There was no effect of 10 weeks basic military training on hair cortisol concentration in male Swiss Army cadets (Boesch et al., 2015), while others have identified increased cortisol in 12-hour urine samples after 4 weeks military training among male Greek recruits (Makras et al., 2005) and in fasting blood samples after 9 weeks of basic military training among Australian Army male and female recruits (Drain et al., 2017). Conversely, Clow et al. (2006) demonstrated a reduction in the cortisol response to waking after 11 weeks of British basic military training in men and women. Some of the discrepancies between studies may be explained by differences in volume and intensity of exercise, a major component of basic military training; both are associated with acutely elevated cortisol concentrations (Skoluda et al., 2015; Zschucke et al., 2015). High intensity interval training during Australian basic military training has been associated with additional plasma cortisol elevations compared with extant, endurance-based training (Drain et al., 2017). Exercise is also associated with elevated hair cortisol concentrations (Skoluda et al., 2012), however overtraining syndromes, which may occur in basic military training (Booth et al., 2006), may be associated with blunted dynamic cortisol responses (Cadegiani and Kater, 2017a). Cortisol response to ACTH and/ or corticotrophin releasing hormone (CRH) may also be reduced by sleep deprivation, a common component of military training (Guyon et al., 2014). Whether long durations of military training are associated with a transient adaptation in the HPA axis, or if stress, and other factors, are associated with reduced dynamic function consistent with overtraining is unknown. Disruption of cortisol secretion may indirectly be related to risk of training-related injury from uncoupling of bone turnover, and in the long-term, reproductive function and mental health problems (Abell et al., 2016a; Ackerman et al., 2013; Gordon et al., 2017). We studied women since women in the military could be at greater risk of reproductive dysfunction (Gifford et al. 2017), are exposed to greater physiological strain (O'Leary et al., 2018), and are at a greater risk of training related injury (Blacker et al. 2008) and stress fracture (Wentz et al. 2011) than men.

This study aimed to comprehensively characterise the HPA response in women to a long and arduous infantry-based basic military training programme in the UK. We hypothesised that compared with the first week of training, ongoing training would be associated with reduced cortisol in the early morning and unchanged or elevated cortisol the preceding evening. Given anticipated effects of sustained psychological stress of adapting to the military environment, intense exercise and restricted sleep, we hypothesised HPA axis responsiveness to ACTH would be reduced, while hair and fasted plasma cortisol would be elevated, and these observations would resolve as training became less arduous.

3.3.3 Methods

3.3.3.1 Setting

This study is part of the Female Endocrinology in Arduous Training research programme, which comprises studies aiming to characterise female endocrine and metabolic responses to military training. This study took place at the Royal Military Academy, Sandhurst UK, where the British Army trains all Officers during the Commissioning Course. The regular Commissioning Course is an immersive, 44-week, infantry-based training programme, taking place in mixed sex platoons. It is designed to be physically and mentally arduous, teaching theoretical and practical leadership and the fundamentals of soldiering. The 44-week course is separated into three terms, each 14 weeks long, with 2 weeks of adventurous training.

3.3.3.2 Participants, inclusion and exclusion criteria

All women commencing the Commissioning Course over three successive intakes (May 2017, September 2017 and January 2018) were invited to participate at a pre-course briefing held 6 to 20 weeks before the start of training. Immediately before starting the Commissioning Course, cadets underwent a medical examination to confirm fitness, including a detailed medical history, review of medical records and physical examination to exclude among other things medically diagnosed psychological disorders in the past year (including anxiety, depression and eating disorders), treated hormone deficiency (except hypothyroidism, which must have been treated and stable for six months beforehand) and arrhythmia. Exclusion criteria were the use of inhaled, oral or topical steroid preparations in the past three months or during the Commissioning Course. Participation in the study was voluntary, and all women provided informed written consent 24 hours after oral and

written briefings. The study was approved by the Ministry of Defence Research Ethics Committee.

3.3.3.3 Procedures

The study used a repeated measures design across the three 14-week terms. Study visits took place at the pre-course briefing (visit Pre), beginning and end of term 1 (visits 1 and 2), end of term 2 (visit 3), and end of term 3 (visit 4). Saliva sampling also took place in weeks 5 or 7 of each term **(Figure 3-3-1).**

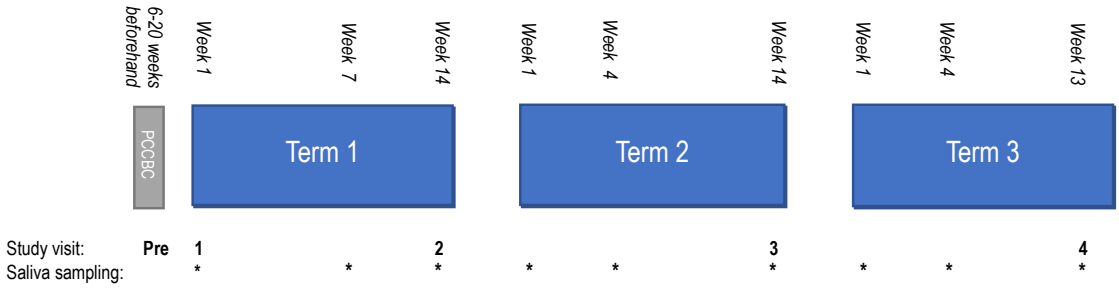


Figure 3-3-1. Schematic of study visits. Study visits (numbered Pre, 1, 2, 3 and 4) and saliva sampling (indicated by *) are indicated below the weeks in which they took place. PCCBC, pre-Commissioning Course Briefing Course

3.3.3.3.1 Questionnaires

A baseline questionnaire was completed at study visit 1 detailing age, ethnicity, education, and reproductive, medical and surgical history. Five questionnaires were undertaken at the pre-course briefing and the beginning and end of each term: the 10-point Connor Davidson Resilience Scale (CD-RISC-10) (Connor and Davidson, 2003), patient health-questionnaire 9 (PHQ-9) (Kroenke et al., 2001), psychosocial stress questionnaire of Rosengren et al. (2004), impact of events scale – revised (IES-R) (Weiss, 1997) and the Beck Anxiety Inventory (BAI) (Beck et al., 1988). Questionnaires were completed on smart phones using SmartSurvey (SmartSurvey, Tewkesbury, UK).

The CD-RISC-10 is a measure of the ability to respond to adversity and comprises 10 items, scored from 0 (“not true at all”) to 4 (“true nearly all the time”), and is abridged from the 25-point CD-RISC on the basis of a thorough factor analysis (Campbell-Sills and Stein, 2007). The scale has demonstrated strong psychometric

properties in young adults (Campbell-Sills et al., 2009) and military populations (Green et al., 2014; Johnson et al., 2011). Permission was granted from the author to use the CD-RISC-10. The PHQ-9 is a measure of low mood, consisting of nine criteria scored from 0 (“not at all”) to 3 (“nearly every day”). The PHQ-9 has demonstrated good validity and reliability as a diagnostic and severity measure in military and general populations (Martin et al., 2006; Wells et al., 2013). We analysed scores on a continuous scale of 0 to 27, to detect subtle differences over time, and used the cut-off ≥10 points, which has 88% sensitivity and specificity for moderate depression (Kroenke et al., 2001). The psychosocial stress questionnaire defined stress as feeling irritable, filled with anxiety, or as having sleeping difficulties because of conditions at work or at home, with the following response options: ‘never’, ‘some periods’, ‘several periods’ or ‘permanent stress’. In asking about the level of financial stress, three options were given: ‘little or none’, ‘moderate’ or ‘high or severe’, while major life events such as major family conflict, divorce or separation were categorised into ‘none’ or ‘one or more’. Participants were then asked to complete the IES-R with reference to any major life event(s) identified. The BAI assesses how much each of 21 anxiety symptoms has bothered participants in the past month on a 4-point Likert scale from 0 (“not at all”) to 3 (“severely – it bothered me a lot”). The BAI has demonstrated high internal consistency in a military population (α coefficient 0.91), adequate test-retest reliability (r = 0.75) and correlates highly with other measures of anxiety (Beck et al., 1988; Nathan et al., 2012).

3.3.3.3.2 Hair sampling

At the pre-course briefing and the end of each term (visits Pre, 2, 3 and 4), a 5 mm diameter section of hair was sampled from the posterior vertex region of the head, as close as possible to the scalp, and stored in aluminium foil at room temperature until transport for analysis by Dresden Lab Service GmbH (Dresden, Germany). Hair samples were divided into 1 cm segments by the laboratory, assuming an average growth rate of 1 cm per month. Up to 7 segments were assayed from visit Pre and four segments from other visits, giving a maximum of 17 1-month hair cortisol concentrations. The number of segments assayed varied according to participant hair length. To account for differing hair lengths, like other studies (e.g. Boesch et al. (2015); McLennan et al. (2016)), we compared average hair cortisol across three-month periods. Participants with hair length ≥ 5 cm and < 5 cm provided five 3-month periods and four 3-month periods, respectively. Subjects using peroxide

treatment were excluded from analysis due to its cortisol-lowering effect (Stalder et al., 2017). Due to the negative association of the combined contraceptive pill (COCP) use with hair cortisol (Stalder et al., 2017), COCP users were considered separately from non-users (**see Section 3.3.3.5**).

3.3.3.3.3 Saliva sampling

Diurnal cortisol slope necessitates morning and evening saliva sampling on the same day, however this was not feasible due to constraints of the training programme. Instead, cortisol was measured from evening and morning saliva samples from saliva samples on consecutive days at the beginning, middle and end of each term. Participants were requested to provide saliva samples using a Sarstedt Salivette ® (Sarstedt, Leicester, UK), by chewing on the synthetic swab for 30 secs, as described elsewhere (Stalder et al., 2016). Saliva samples were collected immediately before bed (before brushing teeth) and immediately after waking the following morning. Sampling instructions were given through live demonstration, videos and written instructions (on paper and by text message). Participants documented the time of sampling on the tube. Reminders were sent to participants' mobile phones by text message at around 10 pm on the evening of sampling and at around 6 am on the morning of sampling (Cadets normally woke shortly before 6 am).

3.3.3.3.4 Basal and dynamic blood sampling

A single sample of blood was collected in EDTA-containing tubes after an overnight fast at visits 1, 3 and 4. Each blood sample was analysed for cortisol binding globulin (CBG), albumin and cortisol. The day after fasted blood sampling on visits 1 and 3, a 1-hour combined dynamic adrenal function test was used to assess adrenal cortex function (Morosini et al., 1989). Due to constraints imposed by training, dynamic testing was completed in the early evening (average time 6.51 pm ±51 min, range 5.20 pm to 8.10 pm) and for each participant was completed at the same time on both occasions. Participants relaxed supine on a bed before a 20 G cannula was inserted into an antecubital fossa vein. A sample of blood was taken from the cannula in EDTA-containing tubes. After 10-15 minutes, 1.0 µL of a 1 µg/mL solution of adrenocorticotrophic hormone ($ACTH_{1-24}$, tetracosactrin acetate as Synacthen®, Mallinckrodt, Dublin, Ireland), freshly diluted on each occasion as described previously (Gifford et al., 2019a), was injected followed by a 10 mL saline flush. Venous blood was sampled from the cannula in EDTA-containing tubes after 20, 30,

40 and 60 min. Basal (afternoon), peak (stimulated), and area under the curve (AUC) dynamic plasma cortisol concentrations were assessed, with AUC calculated using the trapezoidal rule. For plasma cortisol analyses, participants were considered separately if they used a COCP, since synthetic oestrogens would be expected to elevate CBG levels and thus total cortisol. Since sex hormones are expected to alter cortisol responsiveness under psychosocial stress (Stephens et al., 2016) when plasma cortisol was assessed, participants who did not use hormonal contraception were asked the number of days since the first day of their last menstrual period.

3.3.3.4 Laboratory methods

Hair cortisol was assayed from 1 cm samples using methods described elsewhere (Iob et al., 2018). Saliva cortisol was assayed using a commercial ELISA kit (Salimetrics ®, State College, PA). Cortisol quantities in the plasma samples were obtained following extraction and LC-MS/MS analysis. Briefly, a calibration curve of cortisol was prepared alongside plasma samples (200 µL) enriched with isotopically labelled cortisol. Samples were extracted using Supported Liquid Extraction SLE400 cartridges (Biotage, UK) by diluting in 0.5M ammonium hydroxide (200 µL), loading, eluting with dichloromethane/isopropanol (0.45 mL x 3), drying under nitrogen and resuspending in 70:30 water/methanol (100 µL described previously (Spaanderman et al., 2018)). Chromatographic separation was achieved following injection (20 µL) using a gradient on a Shimadzu Nexera UPLC system on a Kinetex C18 (150 x 3 mm; 2 µm) column of mobile phases: 0.1 % FA in water, 0.1 % FA in methanol, 0.5 mL/min, 30 °C, followed by MS/MS analysis on a Sciex Qtrap 6500+ operated in positive ESI, where Mass Spectrometry settings have been described previously (Stirrat et al., 2018). Least squares regression of the peak area ratio, with 1/x weighting, was used to calculate the amount of steroid in each sample within Analyst MultiQuant software (Sciex, UK). Total CBG was assayed from plasma using ELISA as per Lewis and Elder (2011), and albumin was assayed using commercial kits (Alpha Laboratories, Eastleigh, UK) adapted for use on a Cobas Fara centrifugal analyser (Roche, UK).

3.3.3.5 Statistical analyses

Statistical analyses were performed using SPSS 24.0 for Mac (IBM, New York, NY). Data were visually assessed for normality and non-normal data were transformed prior to analysis using parametric tests (CBG and average 3-month hair cortisol

concentrations were transformed by natural logarithms). Baseline demographics of participants who withdrew were compared with those who completed the study using independent samples t-tests and χ^2 for continuous and categorical variables, respectively. Four participants were excluded from analyses of hair and plasma cortisol due to commencing or discontinuing a COCP during the study (change in COCP use precluded repeated measures due to the effect of COCP on CBG); a further two were excluded from analyses of hair cortisol due to peroxide treatment. Fourteen participants used a COCP throughout the study. Missing data (saliva cortisol) were imputed using group means for those time points (159 samples, 17%) before analysis of successive morning and evening concentrations.

Changes in questionnaire scores and days since last menstrual period were assessed using repeated measures ANOVAs (main effect of time [visit 1 vs visit 2 vs visit 3 vs visit 4]), with *post-hoc* uncorrected paired samples t-tests used to assess differences between time-points in the event of a significant main effect. Where statistically significant changes in questionnaire score were identified, scores for individual questions within those questionnaires were compared over time using RM ANOVA with Bonferroni adjustment. Changes in hair and saliva cortisol concentration were assessed using a two-way mixed-design ANOVA (group [COCP user vs non-COCP user] × time). Changes in dynamic cortisol concentration from visit 1 to visit 3 were assessed using paired samples t-tests; comparisons between COCP users and non-COCP users were made using independent samples t-tests. A p-value <0.05 was deemed significant.

3.3.4 Results

3.3.4.1 Participant characteristics

Of 77 women who attended the study briefing, 68 (88%) volunteered to participate (**Figure 3-3-2**). Five participants (8%) completed the baseline visit (visit Pre) but did not commence the Commissioning Course. Ten (15%) withdrew during the Commissioning Course: six during term 1 (two medically discharged on arrival, three due to training-related injury, one chose to withdraw from the study), two during term 2 (training-related injury), and two during term 3 (training-related injury). A total of 53 women completed the study; their baseline characteristics compared with participants who withdrew are presented in **Table 3-3-1**. The age, rate of stressful events and anxiety scores did not differ between those who withdrew and those who

completed the study. There were no correlations between cortisol indices with age, ethnicity or educational qualification.

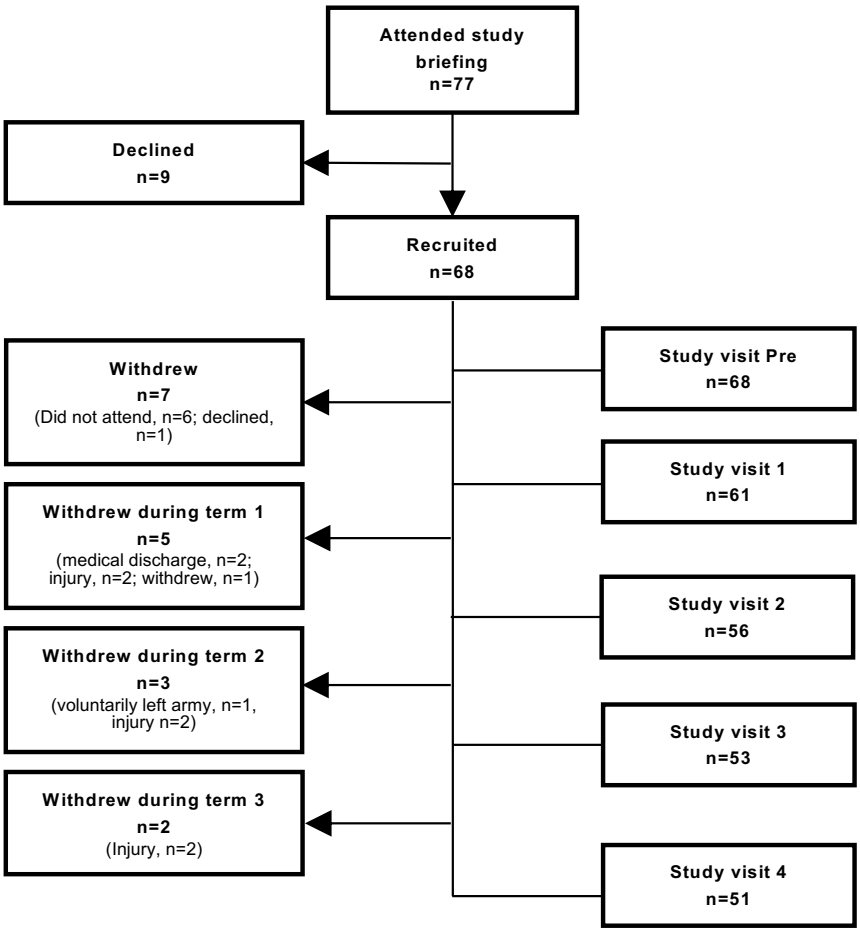


Figure 3-3-2. Recruitment and follow-up

	Completed the study (n=52)	Withdrew (n=18)	p
Age, years	24.0 ±2.5	23.9 ±2.8	ns
Ethnicity, n (%)			
White Scottish. English/ Welsh/ Northern Irish/ British	50 (96)	18 (100)	ns
White Irish	1 (2)	0 (0)	
Other white background	1 (2)	0 (0)	
Highest educational qualification, n (%)			
Master's degree	10 (21)	3 (17)	

Bachelor's degree	36 (70)	9 (50)	
Secondary school	6 (11)	6 (33)	ns
Smoker, n (%)	3 (6)	2 (12)	ns
Drink alcohol (yes) n (%)	48 (91)	12 (80)	ns
Age at menarche, years; median (range)	13 (11-16)	13 (11-16)	ns
Contraception, n (%)			
Combined contraceptive pill	13 (25)	5 (29)	
Progestogen-only	16 (32)	4 (24)	
None or intrauterine device	18 (34)	8 (47)	
Discontinued combined contraceptive pill during study	3 (5)		
Commenced combined contraceptive pill during study	1 (4)		ns
Several periods of psychological stress, n (%)	14 (26)	1 (9)	ns
Permanent, psychosocial stress, n (%)	0 (0)	0 (0)	
Some periods of psychological stress, n (%)	35 (66)	13 (87)	
Never experienced psychological stress, n (%)	4 (8)	1 (9)	
High or severe financial stress, n (%)	1 (2)	0 (0)	-
One or more adverse events, n (%)	17 (32)	3 (27)	ns
IES-R with respect to adverse event	12 ±11	6 ±1	ns
CDRISC 10	30 ±5	30 ±3	ns
PHQ-9	4 ±3	3 ±4	ns
BAI	8 ±7	6 ±3	ns
Peroxide hair treatment	2 (4%)	0 (0)	-
Hair cortisol concentration, 4 to 6 months before Course, median (interquartile range), pg/mg	5.95 (3.53, 13.9)	6.77 (1.91, 15.5)	ns
Hair cortisol concentration, 1 to 3 months before Course, median (interquartile range), pg/mg	8.65 (5.22, 15.4)	5.50 (2.98, 11.9)	ns

Table 3-3-1. Demographics. Data are mean ±Standard deviation, ns: p>0.10, IES-R: impact of events scale – revised, CD-RISC-10: Connor Davidson Resilience Scale-10, PHQ-9: patient health-questionnaire, BAI: Beck Anxiety Inventory.

3.3.4.2 Questionnaires

Questionnaires scores and statistical significance indicators are presented in **Table 3-3-2**. CD-RISC-10 and PHQ9 scores decreased and increased, respectively, across the Commissioning Course with modest to large effect sizes. Post-hoc tests showed significant decreases from visit Pre at all visits (1 to 4) for CD-RISC-10, and increases from visit Pre at visits 2 to 4 for PHQ-9. Question-by-question analysis of CD-RISC-10 (**Section 3-3-6, Appendix Table 3-3-1**) showed modest decreases in measures of traits labelled 'hardiness', specifically the ability to cope with change and illness, injury and hardship, and 'persistence', specifically not giving up and working to attain goals despite roadblocks (Campbell-Sills and Stein, 2007). For the PHQ-9, subsequent analysis showed a significant increase in all domains except concentration and psychomotor function, which were elevated throughout the study (**Appendix Table 3-3-2**) (Kroenke et al., 2001). Forty participants (74%) reported 'feeling tired or having little energy' for 'several days or more' throughout the study, while the number reporting 'feeling tired or having little energy' increased significantly from visit 1 to visits 2 to 4, being highest at visit 2 (49 (92%) reported this "several days" or more). Twelve participants (18%) reached the PHQ-9 cut-off (≥10 points) on one occasion and 4 (6%) on two occasions. Of these participants, ten (83%) also described a stressful life event not related to the training (e.g. death of a loved one or divorce), which may account for higher scores suggesting low mood. More participants described feeling work-related stress (feeling irritable, filled with anxiety, or as having sleeping difficulties) over several periods or permanently during the Commissioning Course compared with before training, and this finding was most marked at visits 2 and 3. Anxiety scores did not change during the study, although the number of participants reporting financial stress and stress due to work increased from visits 1 to 3, and 3 to 4 (**Table 3-3-1**).

	Visit Pre	Visit 1	Visit 2	Visit 3	Visit 4	F	p
CD RISC 10							
	32.6 ±4.1	30.2 ±4.9*	29.3 ±5.5*	30.4 ±5.6*	29.2 ±5.5*	6.93	<0.001
PHQ-9							
	3.1 ±3.0	3.4 ±2.6	5.5 ±3.5*	4.6 ±3.9*	4.6 ±4.4*	7.24	<0.001
Number reaching cut-off score of 10, n (%)	1 (2)	2 (4)	5 (9)	5 (9)	7 (13)		
Adverse events and IES-R							
	6 events in 6 subjects 24.8 ±15.4	8 events in 8 subjects 13 ±10.49	10 events in 10 subjects 25.7 ±15.2	11 events in 9 subjects 24.0 ±16.1	9 events in 9 subjects 25.5 ±15.7		
BAI							
	7.7 ±6.7	6.9 ±5.2	6.3 ±5.8	6.1 ±6.3	5.2 ±6.2	1.10	ns
High or severe financial stress, n (%)							
	1 (1)	2 (3)	2 (3)	5 (9)	4 (8)		
How often have you experienced stress due to work? †, n (%)							
Never	3 (6)	2 (11)	0 (0)	0 (0)	0 (0)		
Some periods	35 (66)	22 (41)	17 (32)	23 (43)	25 (47)		
Several periods	15 (28)	17 (32)	33 (62)	25 (47)	23 (43)		
Permanently	0 (0)	1 (2)	3 (6)	5 (9)	5 (9)		
Not working or at school or university	5 (9)	7 (13)	0 (0)	0 (0)	0 (0)		

Table 3-3-2. Psychological health questionnaires. Data are mean ±Standard deviation, ns: p>0.10. IES-R: impact of events scale – revised, CD-RISC-10: Connor Davidson Resilience Scale-10, PHQ-9: patient health-questionnaire, BAI: Beck Anxiety Inventory. RM ANOVA Repeated measures analysis of variance. †At visits pre- and visit 1 this question was 'how often have you experienced stress due to school, university or work?' * post hoc p<0.05 for paired t-test versus visit Pre.

3.3.4.3 Hair cortisol concentration

Monthly hair cortisol concentrations are shown in **Figure 3-3-3A** and comparisons of 3-month average hair cortisol concentrations between COCP and non-COCP users are displayed in **Table 3-3-3**. There was no COCP use × time interaction for hair cortisol, but the effect of time was significant (p=0.003, **Table 3-3-3**) demonstrating that hair cortisol increased in both non-COCP users and COCP users. Post-hoc t-tests demonstrated hair cortisol was higher at months pre-3 to pre-1 and months 1 to 4 and 9 to 12 of training than months pre-6 to pre-4 (**Table 3-3-3**).

3.3.4.4 Saliva cortisol concentration

Evening and morning saliva sampling recording times were 11.12 pm ±35 min and 6.07 am ±28 min, respectively. Evening saliva concentrations did not change during the Commissioning Course (**Figure 3-3-3B**). There was a main effect of time for morning cortisol (p<0.001), with post-hoc t-tests demonstrating that morning cortisol increased from week 1 to week 7 of term 1 (0.44 ±0.23 versus 0.59 ±0.24 µg/dL, p<0.001, **Figure 3-3-3B**), with no significant differences between any other time-points. Morning salivary cortisol in term 1 week 1 was not different to term 3 week 13 (0.44 ±0.23 versus 0.38 ±0.21 µg/dL, p=0.2). The response of COCP users was not different to non-COCP users (group × time interaction, p=0.4).

3.3.4.5 Basal and dynamic blood tests

Cortisol binding globulin was higher among COCP users than non COCP users (median (interquartile range) at visit 1: 379 (165, 444) ng/mL versus 95 (63, 220) ng/mL, respectively at visit 1, p<0.001) but did not change in either group during the Commissioning Course (p=0.6, **Appendix Table 3-3-3**). Albumin did not differ between COCP users and non-users (34.3 ±2.3 versus 35.0 ±2.6 g/L, respectively at visit 1, p=0.6) and did not change during the Commissioning Course (p=0.7, **Appendix Table 3-3-3**). In non-COCP users, fasting plasma cortisol decreased progressively from visits 1 to visits 3 and 4 (701 ±134, 671 ±158 and 561 ±177 ng/mL, respectively, p<0.001, **Figure 3-3-3C**), with significant post-hoc differences in non-COCP users between visit 1 and visits 3 and 4 (p=0.009 and p<0.001, respectively, **Figure 3-3-3C** and **Appendix Table 3-3-3**). By contrast, in non-COCP users, dynamic function testing demonstrated an increase in afternoon basal cortisol from visits 1 to 3 (177 ±92 and 259 ±103 nmol/L, respectively, p=0.003; **Figure 3-3-3D** and **Appendix Table 3-3-3**) and suggested an increase in peak stimulated

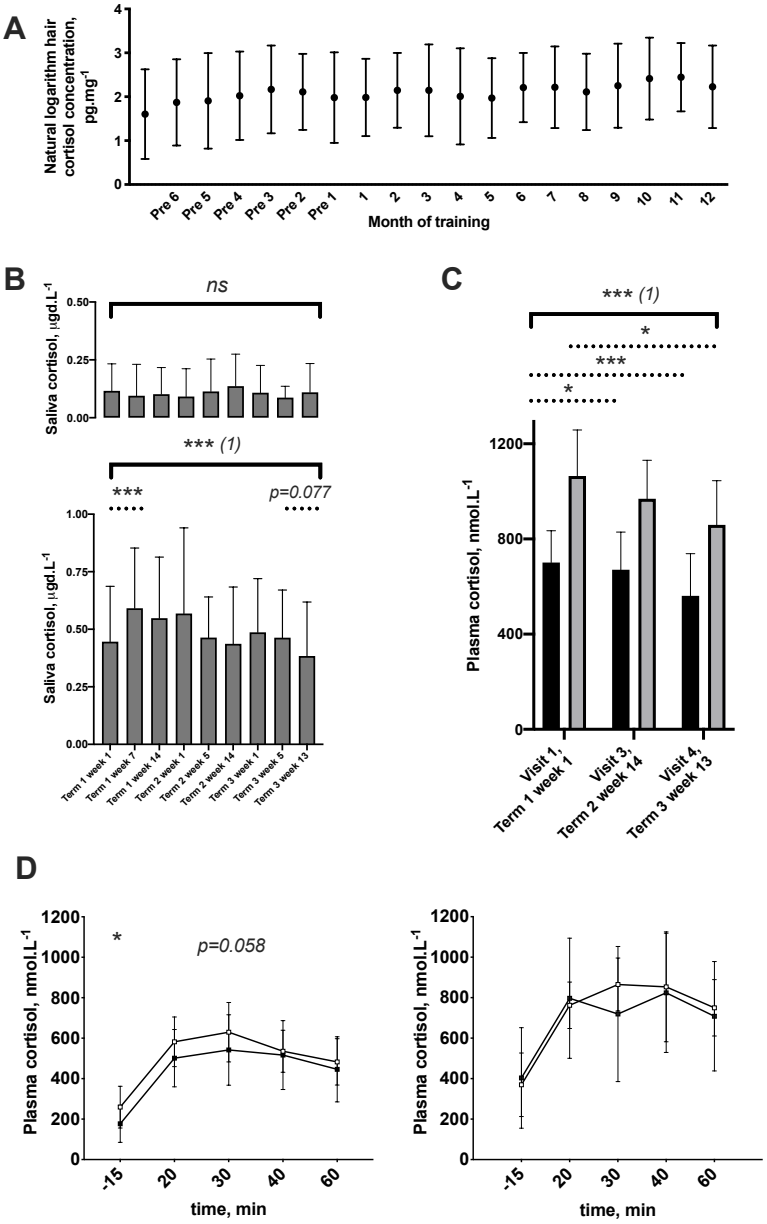


Figure 3-3-3. A: One-month average hair cortisol concentrations prior to and during the Commissioning Course (all participants); month 'pre' was prior to the Course starting. Hair was sampled at study visits 'Pre' (either month Pre 1 or Pre 2), 2 (month 4), 4 (month 8) and 6 (month 12). **B: Evening and morning saliva cortisol concentrations**; top panel: sampled in the evening, bottom panel: sampled the following morning. **C: Fasting plasma cortisol concentration**; non-COCP users (black column) and COCP users (grey column). **D: Mean ±SD total cortisol concentrations during dynamic 1-25 ACTH testing**; non-COCP users (left, n=39) and COCP users (right, n=13) at visit 1 (filled square) and visit 3 (unfilled square). Data are mean ±SD. Solid bracket: mixed two-way ANOVA, Dotted line: significant post-hoc comparisons. *** p<0.001, * p<0.05, ns p>0.10. *** (1) p<0.001 for effect of time; no interaction of group [COCP users vs non-COCP users] × time.

	Months pre-6 to pre-4, pg/mg	Months pre-3 to pre-1, pg/mg	Months 1 to 4, pg/mg	Months 5 to 8, pg/mg	Months 9 to 12, pg/mg	
Non COCP users	7.12 (3.88, 13.90)	9.63 (5.38 ,16.26)*	9.60 (6.80, 15.3)*	10.39 (5.56, 17.28)	11.56 (6.20, 22.45)*	F 4.247
COCP users	4.84 (3.52 ,12.99)	6.23 (4.64 ,12.23)*	7.76 (4.08, 11.26)	10.08 (7.27, 13.36)	13.7 (6.1, 18.63)*	F 3.236
Effect COCP use × time						P=0.70
Effect time						p=0.003

Table 3-3-3 Average hair cortisol concentrations in 3-month periods: Combined contraceptive pill (COCP, n=13) users and non-COCP users (n=33) considered separately due to known association of COCP use with hair cortisol. Data are median (interquartile range). P value for two-way repeated measures ANOVA (main effect of time or group × time). * post hoc p<0.05 for paired t-test versus months pre-6 to pre-4.

cortisol (589 ±164 and 656 ±135 nmol/L, p=0.058, **Figure 3-3-3D** and **Appendix Table 3-3-3**). Fasting plasma cortisol decreased in COCP users from visit 1 to visit 4 (1065 ±193 nmol/L versus 859 ±186 nmol/L, p=0.013, **Figure 3-3-3C** and **Appendix Table 3-3-3**). There was no effect of COCP use for fasting cortisol, (COCP use × time interaction, p=0.9, **Figure 3-3-3C** and **Appendix Table 3-3-3**) and in COCP users afternoon cortisol, peak cortisol response to ACTH and cortisol AUC did not change from visit 1 to visit 3 (**Figure 3-3-3D** and **Appendix Table 3-3-3**). In participants not using hormonal contraceptives, duration of days since last menstrual cycle did not differ between visits 1, 4 and 6 (19 ±19 days,16 ±12 days and 15 ±10 days, respectively, p=0.5).

3.3.5 Discussion

This study comprehensively characterised the HPA axis response to prolonged arduous infantry-based military training in women. We demonstrated a significant rise in morning salivary cortisol concentrations from week 1 to 7 of training, tending to suggest a normal stress response, which is in contrast to the relative decrease in morning cortisol, which we had hypothesised; evening saliva cortisol did not change. Thereafter, saliva cortisol concentrations appeared to demonstrate habituation,

returning to baseline levels by the end of training, corroborated by a decrease in morning fasting plasma cortisol. Peak stimulated cortisol rose modestly (in non-COCP users), suggesting the training was associated with a slight increase in HPA axis responsiveness. Average cortisol concentration in hair demonstrated a modest rise during training.

In our study design we were unable to obtain a true baseline saliva cortisol; participants were already 3 days into training when testing started, so the first sample may have reflected some of the 'shock of capture'. However, our findings of habituation through training are perhaps consistent with those of Clow et al. (2006), who found a latent decrease in cortisol awakening response during 11-week basic military training in male and female British Army recruits. The increase in hair cortisol measured before the course lies within known rates of cortisol washout (29% loss from the most proximal 3 cm to the next most proximal 3 cm segment, from the meta-analysis by Stalder et al. (2017)), so we are unable to determine if there was a true anticipatory rise prior to training. The rise in hair cortisol observed during the Commissioning Course is contrary to Boesch et al. (2015), who found no change in male hair cortisol during training in a single intake of Swiss military cadets. However, Boesch et al. (2015) highlighted shortcomings of their study including the inability to obtain long enough hair samples, which resulted in relatively short hair cortisol exposures, which were interrupted by haircuts, pretraining hair cortisol concentration and affected by seasonal variation. The use of women in our study helped overcome this, while our recruitment over three courses meant pre- and within-training hair cortisol concentrations represented continuous exposures of 27 months. While average hair cortisol concentration did not exhibit the same HPA habituation seen in the morning saliva cortisol (a stress response), the increase throughout training may be explained by regular physical exercise during training (Gerber et al., 2012); chronic stress but not self-report measures of perceived stress could be expected to elicit increased hair cortisol (Stalder et al., 2017). While we found resilience and mood decreased while hair cortisol increased, a recent meta-analysis found no association between various scales of low mood and hair cortisol ($r=–0.059$, $p=0.078$) (Stalder et al., 2017). We conclude the rise in hair cortisol was more likely a reflection of physical activity or energy deficit, than low mood or psychological stress.

Contrary to our hypothesis, we demonstrated a concurrent increase in the plasma cortisol response to ACTH with decreased early morning plasma cortisol. In a similar dynamic function test, veterans with traumatic military experiences demonstrated increased responsiveness to ACTH compared with controls, which was unrelated to anxiety disorders (Golier et al., 2014). Pre-stimulation morning cortisol levels are often elevated while stimulated cortisol may be suppressed in overtraining syndromes (Cadegiani and Kater, 2017a). In our previous study of women undertaking an arduous ski expedition, responsiveness to a similar 1 μg ACTH test was suppressed, with marked sensitivity to central negative feedback, but did not change immediately following or two weeks after a 2 month exercise exposure, compared with 1 month beforehand (Gifford et al., 2019a). In the present study, the increase in cortisol responsiveness was not accounted for by changes in CBG. We postulate our findings represent an increased HPA axis responsiveness during training which could be interpreted as 'healthy'.

Resilience scores were consistent with the upper end of the range reported previously for similar populations throughout the study (Davidson, 2018), despite demonstrating a modest but steady decrease during training. Resilience might appear to be a solely desirable attribute for a Serviceperson. However, resilience could conceivably lead to an individual remaining in a situation of likely defeat, which is more likely to be harmful, instead of escaping. Thus, this decline in resilience may also be interpreted as a beneficial, not detrimental, adaptation. A likely explanation for the slight decrease we observed in resilience, constituting reduced hardiness and persistence ratings, was fatigue. Certainly, the PHQ-9 scores may have been distorted by a lack of sleep. For example, the question 'do you have trouble falling or staying asleep, or sleeping too much' was perhaps confounding, since it was more likely to reflect a tiring training programme than low mood. Where the clinical cut-off of the PHQ-9 was reached (≥10 points), this was generally attributable to a non-course related adverse event, which likely explained the overall increase in PHQ-9 (although the number reporting work-related stress increased from 7 in term 1 to 10 in term 3). Alternatively, it is possible that the changes observed in CD-RISC-10 and PHQ-9 related to the increased ratings of stress from work. It is also possible that the change in CDRISC-10 and PHQ-9 scores represents regression to the true value, whereby the belief that the results are not anonymised subsides over time.

Strengths of our study include the multimodal approach to HPA axis assessment, alongside repeated measures of mood and resilience and the large sample of female military cadets studied during arduous training over a long duration. Participants were well-matched and were undertaking an identical arduous training programme, which will be relevant to women in physically demanding occupations.

Unfortunately, we were limited to diurnal cortisol measurement and were unable to examine cortisol awakening response due to restraints on the participants' time (they were often undertaking programmed activities within 1 hour of waking) and our saliva cortisol findings are, therefore, preliminary. We were also unable to perform dynamic HPA axis testing in the morning, so could not assess central axis sensitivity to dexamethasone to determine whether there were any changes in central negative feedback sensitivity (Reynolds et al., 2001a). The Course was a relatively long military training programme; shorter duration training, which is more common, might provoke pathological activity of the HPA in female military cadets. Therefore, the findings of the present study need to be replicated by further studies providing a different training content to enhance the generalisability of the results.

Our hypothesis that military women would demonstrate maladaptive cortisol responses to basic military training was rejected. Through a comprehensive assessment, the initial rise in morning cortisol and fasting plasma cortisol, appeared to be followed by habituation, and increased HPA axis responsiveness. These responses were observed despite modest reductions in mood and resilience and increased perceived stress during training. The observed increase in hair cortisol during training was possibly related to physical exercise. We interpret these findings as being consistent with a healthy adaptation of the HPA axis during basic military training among women, despite evidence of ongoing perceived stress.

3.3.6 Appendix material to Section 3.3

	Visit Pre	Visit 1	Visit 2	Visit 3	Visit 4	RM ANOVA
I am able to adapt to changes when they occur						
Not true at all	0 (0%)	0 (0%)	1 (2%)	0 (0%)	1 (2%)	p<0.001*
Rarely true	0 (0%)	3 (6%)	5 (9%)	0 (0%)	1 (2%)	ηp^2= 0.099
Sometimes true	2 (4%)	4 (8%)	8 (15%)	7 (13%)	5 (9%)	F= 5.62
Often true	25 (47%)	35 (66%)	21 (40%)	32 (60%)	37 (70%)	
True nearly all the time	26 (49%)	11 (21%)	18 (34%)	14 (26%)	9 (17%)	
I can deal with whatever comes my way						
Not true at all	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	ns
Rarely true	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (2%)	ηp^2= 0.022
Sometimes true	3 (6%)	7 (13%)	12 (23%)	8 (15%)	7 (13%)	F=2.058
Often true	33 (62%)	32 (60%)	24 (45%)	28 (53%)	31 (58%)	
True nearly all the time	17 (32%)	14 (26%)	17 (32%)	17 (32%)	14 (26%)	
I try to see the humorous side of things when I am faced with problems						
Not true at all	0 (0%)	0 (0%)	0 (0%)	1 (2%)	1 (2%)	p=0.022
Rarely true	1 (2%)	3 (6%)	3 (6%)	0 (0%)	1 (2%)	ηp^2= 0.098
Sometimes true	10 (19%)	12 (23%)	12 (23%)	16 (30%)	14 (26%)	F= 5.56
Often true	26 (49%)	21 (40%)	27 (51%)	23 (43%)	27 (51%)	
True nearly all the time	16 (30%)	17 (32%)	11 (21%)	13 (25%)	10 (19%)	
Having to cope with stress can make me stronger						
Not true at all	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	p=0.045
Rarely true	0 (0%)	0 (0%)	2 (4%)	1 (2%)	1 (2%)	ηp^2= 0.046
Sometimes true	8 (15%)	8 (15%)	6 (11%)	11 (21%)	7 (13%)	F=2.49
Often true	24 (45%)	24 (45%)	30 (57%)	22 (42%)	38 (72%)	
True nearly all the time	21 (40%)	21 (40%)	15 (28%)	19 (36%)	7 (13%)	
I tend to bounce back after illness, injury and other hardships						
Not true at all	0 (0%)	1 (2%)	0 (0%)	1 (2%)	1 (2%)	p=0.001*
Rarely true	0 (0%)	2 (4%)	0 (0%)	2 (4%)	1 (2%)	ηp^2= 0.088
Sometimes true	3 (6%)	8 (15%)	6 (11%)	2 (4%)	9 (17%)	F=4.95
Often true	23 (43%)	21 (40%)	26 (49%)	27 (51%)	30 (57%)	
True nearly all the time	27 (51%)	21 (40%)	21 (40%)	21 (40%)	12 (23%)	
I believe I can achieve my goals, even if there are obstacles						
Not true at all	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	p=0.009

Rarely true	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	ηp^2= 0.064
Sometimes true	3 (6%)	7 (13%)	8 (15%)	5 (9%)	11 (21%)	F= 3.480
Often true	26 (49%)	25 (47%)	28 (53%)	30 (57%)	28 (53%)	
True nearly all the time	24 (45%)	21 (40%)	17 (32%)	18 (34%)	14 (26%)	
Under pressure, I stay focused and think clearly						
Not true at all	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	p=0.040
Rarely true	0 (0%)	2 (4%)	2 (4%)	0 (0%)	2 (4%)	ηp^2= 0.048
Sometimes true	10 (19%)	16 (30%)	19 (36%)	17 (32%)	14 (26%)	F=2.56
Often true	30 (57%)	26 (49%)	23 (43%)	25 (47%)	28 (53%)	
True nearly all the time	13 (25%)	9 (17%)	9 (17%)	11 (21%)	9 (17%)	
I am not easily discouraged by failure						
Not true at all	0 (0%)	0 (0%)	0 (0%)	1 (2%)	1 (2%)	p<0.001*
Rarely true	0 (0%)	3 (6%)	6 (11%)	3 (6%)	3 (6%)	ηp^2= 0.103
Sometimes true	16 (30%)	24 (45%)	25 (47%)	17 (32%)	17 (32%)	F=5.72
Often true	23 (43%)	20 (38%)	17 (32%)	24 (45%)	25 (47%)	
True nearly all the time	14 (26%)	6 (11%)	5 (9%)	8 (15%)	7 (13%)	
I think of myself as a strong person when dealing with life's challenges and difficulties						
Not true at all	0 (0%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	p<0.001*
Rarely true	0 (0%)	0 (0%)	0 (0%)	1 (2%)	1 (2%)	ηp^2= 0.110
Sometimes true	2 (4%)	7 (13%)	7 (13%)	10 (19%)	9 (17%)	F=6.30
Often true	21 (40%)	32 (60%)	28 (53%)	21 (40%)	28 (53%)	
True nearly all the time	30 (57%)	14 (26%)	17 (32%)	21 (40%)	15 (28%)	
I am able to handle unpleasant or painful feelings like sadness, fear, and anger						
Not true at all	0 (0%)	0 (0%)	1 (2%)	0 (0%)	1 (2%)	ns
Rarely true	1 (2%)	3 (6%)	4 (8%)	1 (2%)	2 (4%)	ηp^2= 0.032
Sometimes true	11 (21%)	12 (23%)	15 (28%)	13 (25%)	14 (26%)	F=1.71
Often true	27 (51%)	26 (49%)	22 (42%)	27 (51%)	29 (55%)	
True nearly all the time	14 (26%)	12 (23%)	11 (21%)	12 (23%)	7 (13%)	

Appendix Table 3-3-1. Question-by-question analysis of the Connor Davidson Resilience Scale-10. RM ANOVA, repeated measures analysis of variance for score per question. Percentages rounded to the nearest whole percentage point, so some do not add to 100. * p<.0050 is deemed significant after Bonferroni adjustment. ns, not significant. Hp^2 partial Eta square

	Visit Pre	Visit 1	Visit 2	Visit 3	Visit 4	χ^2
Little interest or pleasure in doing things						
Not at all	40 (75%)	39 (74%)	26 (49%)	29 (55%)	33 (62%)	p=0.001*
Several days	12 (23%)	13 (25%)	24 (45%)	20 (38%)	13 (25%)	ηp^2= 0.089
More than half the days	1 (2%)	1 (2%)	2 (4%)	2 (4%)	6 (11%)	F=4.89
Nearly every day	0 (0%)	0 (0%)	1 (2%)	2 (4%)	1 (2%)	
Feeling down, depressed or hopeless						
Not at all	44 (83%)	41 (77%)	27 (51%)	39 (74%)	34 (64%)	p<.001*
Several days	9 (100%)	12 (100%)	24 (92%)	11 (79%)	12 (63%)	ηp^2= 0.098
More than half the days	0 (0%)	0 (0%)	2 (6%)	1 (4%)	6 (20%)	F= 5.451
Nearly every day	0 (0%)	0 (0%)	0 (0%)	2 (4%)	1 (2%)	
Trouble falling or staying asleep, or sleeping too much						
Not at all	21 (40%)	22 (42%)	32 (60%)	22 (42%)	23 (43%)	ns
Several days	27 (84%)	26 (84%)	17 (81%)	24 (77%)	23 (77%)	ηp^2= 0.017
More than half the days	4 (17%)	3 (17%)	1 (13%)	3 (17%)	3 (19%)	F= 0.850
Nearly every day	1 (2%)	2 (4%)	3 (10%)	4 (9%)	4 (8%)	
Feeling tired or having little energy						
Not at all	19 (36%)	13 (25%)	4 (8%)	11 (21%)	9 (17%)	p<0.001*
Several days	28 (53%)	31 (58%)	22 (42%)	30 (57%)	35 (66%)	ηp^2= 0.185
More than half the days	6 (11%)	5 (9%)	17 (32%)	9 (17%)	6 (11%)	F= 11.35
Nearly every day	0 (0%)	4 (8%)	10 (19%)	3 (6%)	3 (6%)	
Poor appetite or overeating						
Not at all	41 (77%)	34 (64%)	20 (38%)	25 (47%)	23 (43%)	p<0.001*
Several days	10 (19%)	15 (28%)	22 (42%)	20 (38%)	22 (42%)	ηp^2= .131
More than half the days	1 (2%)	4 (8%)	9 (17%)	7 (13%)	6 (11%)	F= 7.55
Nearly every day	1 (2%)	0 (0%)	2 (4%)	1 (2%)	2 (4%)	

Feeling bad about yourself, or that you are a failure or have let yourself, your colleagues or your family down						
Not at all	43 (81%)	42 (79%)	30 (57%)	36 (68%)	34 (64%)	p<0.001*
Several days	9 (17%)	11 (21%)	16 (30%)	14 (26%)	15 (28%)	ηp^2= 0.104
More than half the days	1 (2%)	0 (0%)	7 (13%)	1 (2%)	2 (4%)	F=5.78
Nearly every day	0 (0%)	0 (0%)	0 (0%)	2 (4%)	2 (4%)	
Trouble concentrating on things, such as reading the newspaper or watching television						
Not at all	29 (55%)	31 (58%)	29 (55%)	29 (55%)	33 (62%)	ns
Several days	22 (42%)	20 (38%)	20 (38%)	20 (38%)	15 (28%)	ηp^2= 0.010
More than half the days	1 (2%)	2 (4%)	4 (8%)	3 (6%)	5 (9%)	F= .492
Nearly every day	1 (2%)	0 (0%)	0 (0%)	1 (2%)	0 (0%)	
Moving or speaking so slowly that other people have noticed, or the opposite- being so fidgety or restless that you have been moving around a lot more than usual						
Not at all	44 (83%)	47 (89%)	42 (79%)	45 (85%)	45 (85%)	ns
Several days	7 (13%)	5 (9%)	9 (17%)	7 (13%)	7 (13%)	ηp^2= 0.026
More than half the days	2 (4%)	1 (2%)	2 (4%)	1 (2%)	1 (2%)	F=1.34
Nearly every day	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	
Thoughts that you would be better off dead or hurting yourself in some way						
Not at all	53 (100%)	53 (100%)	53 (100%)	51 (96%)	52 (98%)	p<.001*
Several days	0 (0%)	0 (0%)	0 (0%)	2 (4%)	1 (2%)	ηp^2= 0.099
More than half the days	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	F=5.63
Nearly every day	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	

Appendix Table 3-3-2. Question-by-question analysis of Patient Health Questionnaire-9. RM ANOVA, repeated measures analysis of variance for score per question. Percentages rounded to the nearest whole percentage point, so some do not add to 100. * p<.0056 is deemed significant after Bonferroni adjustment. ns, not significant. ηp^2 partial Eta squared. χ^2 Chi squared

	Visit 1	Visit 3	Visit 4	p
Cortisol binding globulin, ng/mL				
non-COCP users	94.9 (62.9 ,220.3)	104.9 (73.4 ,162.7)	90.1 (71.1 ,192.6)	
COCP users	378.5 (165.4 ,443.6)	350.0 (228.8 ,449.9)	401.0 (183.7 ,535.0)	
Effect of COCP use × time				ns
Effect of time				ns
Albumin, g/L				
non-COCP users	35.0 ±2.6	34.5 ±2.6	34.5 ±2.5	
COCP users	34.3 ±2.3	33.9 ±2.5	33.6 ±2.3	
Effect of COCP use × time				ns
Effect of time				ns
Total fasting cortisol, nmol/L				
non-COCP users	701 ±134	671 ±158*	561 ±177**	
COCP users	1065 ±193	969 ±162	859 ±186*	
Effect of COCP use × time				ns
Effect of time				<0.001
Afternoon basal cortisol, nmol/L				
non-COCP users	177 ±92	259 ±103		0.003
COCP users	403 ±248	369 ±157		ns
Peak stimulated cortisol, nmol/L				
non-COCP users	589 ±164	656 ±135		0.058
COCP users	921 ±320	954 ±241		ns
Cortisol AUC				
non-COCP users	27385 ±7986	29851 ±5825		ns
COCP users	42072 ±13981	43417 ±8103		ns

Appendix Table 3-3-3. Plasma protein and cortisol concentrations. Fasting tests (visits 1, 3 and 4), p for two-way mixed ANOVA. Dynamic tests (visits 1 and 3), p for paired samples t-test. COCP, combined contraceptive pill users (n=13), non-COCP users (n=33). Ns: p>0.10, AUC: area under the curve, *: p<0.05 post-hoc test versus visit 1, **: p<0.001 post-hoc test versus visit 1.

3.4 Hypothalamic–pituitary–gonadal axis and metabolic adaptation to basic military training

Section 3.4 is in submission to the Journal of Clinical Endocrinology and Metabolism under the title “Maladaptive metabolic, pituitary and gonadal responses to military training in women” but Dr Robert M Gifford (RG), Dr Thomas J O’Leary (TO), Dr Sophie L Wardle (SW), Miss Rebecca Louise Double (RD), Prof Julie P Greeves (JG), Prof Richard A Anderson (RA), Prof David R Woods (DW) and Prof Rebecca M Reynolds (RR). RG designed the study and recruited the participants with supervisory support from RR, DW, JG and RA. Study measures were conducted by RG (blood and urine sampling), RD, SW and TO (dual x-ray absorptiometry imaging and urine sampling). Blood and urine sampling was assisted by nurses from the Wellcome Trust Clinical Research Facility, Edinburgh. Assays were completed by Dr Forbes Howie (MRC Centre for Reproductive Health), Dr Takeshi Fujisawa (BHF/ University of Edinburgh Centre for Cardiovascular Science) and Dr Susan Johnston (NHS Greater Glasgow and Clyde). RG conducted the analysis and wrote the manuscript with supervisory support from RR, DW and RA. All authors provided editorial input into the final version.

In summary, in women undertaking 11-month basic Officer training at the Royal Military Academy, Sandhurst, this study found marked anovulation, suppression of pituitary gonadotroph function and ovarian follicular dysregulation. This was not explained by energy deficit: fat mass fluctuated but did not decrease overall, leptin increased and insulin resistance occurred relative to baseline. Reproductive function and metabolic maladaptation in military training are explained by factors other than low energy availability

3.4.1 Abstract

Background: Low energy availability (LEA) in female athletes can result in HPG axis suppression. Basic military training (BMT) is physically arduous and associated with amenorrhoea and low-trauma fractures. We hypothesised that women undergoing BMT would demonstrate evidence of LEA and suppressed HPG function.

Design: Prospective study of 61 women undertaking 11-month BMT. Subjects acted as their own controls at baseline (all measures). Body composition measurement (DXA) was repeated after 3, 7 and 11 months; fasting blood (glucose, insulin (for

homeostatic model of insulin resistance, HOMA2), leptin, inhibin B, oestradiol, anti-Müllerian hormone (AMH), and FSH) after 7 and 11 months, and dynamic 1-hour 10 μg GnRH test measuring LH and FSH after 7 months. Menstruation and ovulation were assessed in non-contraceptive pill-users (n=22) using diaries and weekly urinary progesterone: creatinine ratio, respectively.

Results: 52 women, aged 24.0±0.3 years, completed the study. Fat mass decreased 0.8kg from baseline to month 3, increased 1.8kg to month 7 and reverted to baseline by month 11 ($p<0.001$). Fat-free mass did not change ($p=0.13$). HOMA2 and leptin increased (both $p<0.001$), as did oestradiol and inhibin B ($p<0.05$) while AMH was unchanged ($p=0.6$).

	Baseline	7 months	11 months
HOMA2	1.77±0.52	1.85±0.30*	2.06±0.61*
Leptin, ng/mL	8.09±3.11	11.37±4.10*	12.52±4.12*
Inhibin B, pg/mL	26.8±10.9	54.0±30.8*	42.3±27.6*
Estradiol, pmol/L	83±46	145±73*	95±64
AMH, pmol/L	24.1±18.6	22.5±14.3	22.4±14.8

Maximum and area-under-the-curve fold-responses of LH and FSH to GnRH were suppressed after 7 months (both $p<0.001$). Findings were unaffected by contraceptive use (effect × time $p=0.8$). Seven participants (32%) became oligo/amenorrhoeic. 87% of regular (23–35d) cycles were anovulatory.

Conclusion: Evidence of adiposity-related adaptation suggests non-LEA stressors contributed to HPG axis suppression and follicular dysgenesis. Further studies are required to delineate causes of reproductive dysfunction and associated pathology in military women.

3.4.2 Background

Women undergoing military training appear to be at higher risk of menstrual disturbance than civilian women or military men, respectively. Many studies have demonstrated the propensity for women to develop reproductive dysfunction during training involving exercise, nutritional deficit and/ or psychosocial challenges

(Lauder et al., 1999; Loucks, 2013; Schneider et al., 1999; Warren, 1980). It has been suggested that in the context of sports and exercise, low energy availability (EA) is a key mechanism underpinning a range of endocrine deficiencies in athletes including anovulation and bone turnover uncoupling (Mountjoy et al., 2014) which may underlie the observed increased risk in stress fractures (Ackerman et al., 2015). Military training, while being physically arduous, is more complex than exercise or sports and the cause of reproductive dysfunction and associated pathology in military women is not fully understood (Gifford et al., 2017).

Direct assessment of EA is hampered by imprecision in energy intake measurement, as was identified in the study in **Section 3-2**. Studies of arduous training have therefore tended to measure biomarkers of low EA alongside changes in weight or body composition, rather than directly measuring energy intake and exercise energy expenditure (Ackerman et al., 2013; Friedl et al., 2000; Henning et al., 2014; Mancini et al., 2019; Nindl et al., 2011; Opstad, 2001; Tharion et al., 2005). Hormonal and metabolic biomarkers of low EA may be reproductive (e.g. low oestradiol), bone-related (e.g. high C-terminal telopeptide of type I collagen, CTX, low N-terminal propeptide of type I procollagen, P1NP, or low osteocalcin) and metabolic (e.g. low triiodothyronine (T3), leptin, insulin, insulin-like growth factor 1 IGF-1, and high cortisol, ghrelin, growth hormone and peptide YY) (De Souza et al., 2019a) although many of these clearly have low specificity for low EA.

Studies of endocrine and metabolic adaptations to multi-stressor environments have not examined sufficiency of micronutrient intake or metabolic consequences of high fat diets. For example, vitamin D and trace elements like zinc, since these appear to be important for healthy bone turnover (Berger et al., 2015; Gaffney-Stomberg et al., 2019; Välimäki et al., 2005), and may be insufficient in military training (Andersen et al., 2010; Lutz et al., 2019). Iron sufficiency is important for physical and cognitive performance; requirements are higher for premenopausal women than men and may not be met during military training (Martin et al., 2019). Macronutrient dietary components in military settings may be fat and calorie-dense (Beals et al., 2015), so a comprehensive assessment of metabolic adaptation to training should also include measures of glucose regulation and lipid metabolism (e.g. glucose, C-peptide, insulin and nonesterified fatty acids, NEFA).

Hypothalamic amenorrhoea (HA) manifests in infrequent or absent menses with a nonorganic, reversible cause (hence sometimes labelled as 'functional') (Gordon et al., 2017). Accordingly, studies in athletes and the military have focused on identifying menstrual disruption (reviewed by Gifford et al. (2017)). However, HA is a disorder of anovulation, resulting from reduced GnRH and kisspeptin drive, and does not always manifest as oligo/ amenorrhoea. Hormonal markers have therefore allowed studies to identify subtle changes in ovulation, from urinary progesterone and oestrogens (Beitins et al., 1991) or their metabolites (Adler et al., 2015) or repeated sampling of LH (Loucks and Thuma, 2003) or a combination (Andrews et al., 2015). Longitudinal biochemical assessments allow subtle but important adaptations like luteal phase defects to be identified in the absence of changes to menstrual cycles.

Field studies of endocrine adaptations to exercise generally rely on early morning blood or urine sampling to detect day-to-day changes, notably in progesterone (De Souza et al., 2010; Williams et al., 2015) and LH (Ackerman et al., 2013; Cano Sokoloff et al., 2016; Hoyt et al., 2006; Opstad, 2001; Taylor et al., 2016). It is important to assess basal androgens. Hyperandrogenism is associated with elevated LH:FSH ratio, abnormal follicular maturation and anovulation in a subset of women many of whom will have PCOS (Javed et al., 2015).

Sole reliance on basal hormone measurements limit the findings of field studies: concentrations are highly variable within and between days, it is difficult to fit field study visits around reproductive cycles, which may become irregular during exercise, and subtle but important functional changes may be missed. Furthermore, participants using hormonal contraceptives are generally excluded from such studies since they suppress gonadotrophins and alter sex steroids. This undermines real-world relevance. In the UK and USA, many if not most women of reproductive age use hormonal contraception, altering the prevailing 'normal' hormonal milieu (Batig, 2017; Firman et al., 2018).

New approaches to examine functional changes in the HPG axis are therefore required. Dynamic endocrine testing, although widely used in clinical endocrine practice, has not been used in field studies until recently (Cadegiani and Kater, 2017b, 2018b). HA results from reduced GnRH drive via downregulation of kisspeptin/ neurokinin B/ dynorphin neurons, which when prolonged may impair

gonadotroph responsiveness to GnRH (Veldhuis et al., 2008). We developed a novel application for pituitary gonadotroph function testing using a low-dose GnRH (Gonadorelin ®) test to identify subtle changes in pituitary function following to arduous exercise in military women (Gifford et al., 2019a).

Metabolic and reproductive adaptations should also be measured alongside changes in bone mineral density and turnover. While bone turnover markers highlight relative coupling of formation and resorption at the time of measurement, regional and total changes in BMD are important end-organ effects, reflecting bone turnover over the medium to long term (O'Leary et al., 2018). Impaired bone health resulting from deficiencies in nutrition and/ or ovarian steroidogenesis is of key relevance both to athletes and the military (Reed et al., 2015).

The Commissioning Course at the Royal Military Academy, Sandhurst, UK, trains all Officers entering the British Army *ab initio* each year. The considerable physical demands it places on Officer Cadets are consistent across course intakes, making an ideal setting for a longitudinal cohort study (Richmond et al., 2012). Training takes place in mixed sex platoons, such that physical activity and psychosocial stressors are similar among male and female Officer Cadets, although the proportion of female Officer Cadets is typically low (around 8-12%).

We examined the effect of 42–week military training at the Royal Military Academy on female body composition and biomarkers of low EA, reproductive function (comprising menstrual function, ovulation, androgenisation and pituitary gonadotroph responsiveness) and bone health (bone turnover and BMD). We hypothesised that evidence of low EA would be associated with reproductive dysfunction and impaired bone health.

3.4.3 Methods

Officer Cadets enrolling in the Commissioning Course between February 2017 and January 2018 were invited to participate in the study across nine pre-Commissioning Course briefings (5 to 15 attended each briefing). Inclusion criteria were: commencing Commissioning Course, aged 18-30 years, female sex. Exclusion criteria included pregnancy, known history of adrenal, ovarian or gonadotrophin releasing hormone (GnRH) insufficiency, pituitary disease, thyroid disease in the past year, diabetes, hyperparathyroidism, osteopenia, oral, inhaled or topical glucocorticoid use or ongoing musculoskeletal injury. Participant health status was

confirmed by entry medical examination prior to enrolment, including history and examination, completed according to entry requirements for UK Defence (Ministry of Defence, 2018). Women using hormonal contraceptive were eligible for inclusion in the study. All participants provided informed consent. Ethical approval was obtained from the Ministry of Defence Research Committee (790/MoDREC/16) and the study was conducted in accordance with the Declaration of Helsinki.

3.4.3.1 Study setting

The study design consisted of two dynamic endocrine function tests three, serial basal (fasting) tests and four dual-energy x ray absorptiometry (DXA) body composition scans, and continuous, longitudinal tracking of menstrual function and ovulation during the Commissioning Course (**Figure 3-4-1**). The Commissioning Course is a residential programme during which men and women are trained together to lead as Army Officers over three 14-week terms in mixed sex groups. The focus is to develop knowledge, skills and attitudes for leadership and the basics of soldiering, using immersive infantry-based training. Components include field exercises, academic study and rigorous physical activity. Programmed activities often last from early morning until late at night and include many weekends. Most entrants have little or no previous military experience. Officer Cadets are initiated into the army in the first term and undergo competitive selection for regiments in the latter stages of the second term, based on their performance. The first two terms are particularly onerous. In the final term, military knowledge is built through classroom based activities and skills are applied in field exercises.

Throughout the Commissioning Course, Officer Cadets consumed a diet provided by the Army, including field rations when on exercise. Officer Cadets may choose to supplement this diet as they are free to purchase additional food locally.

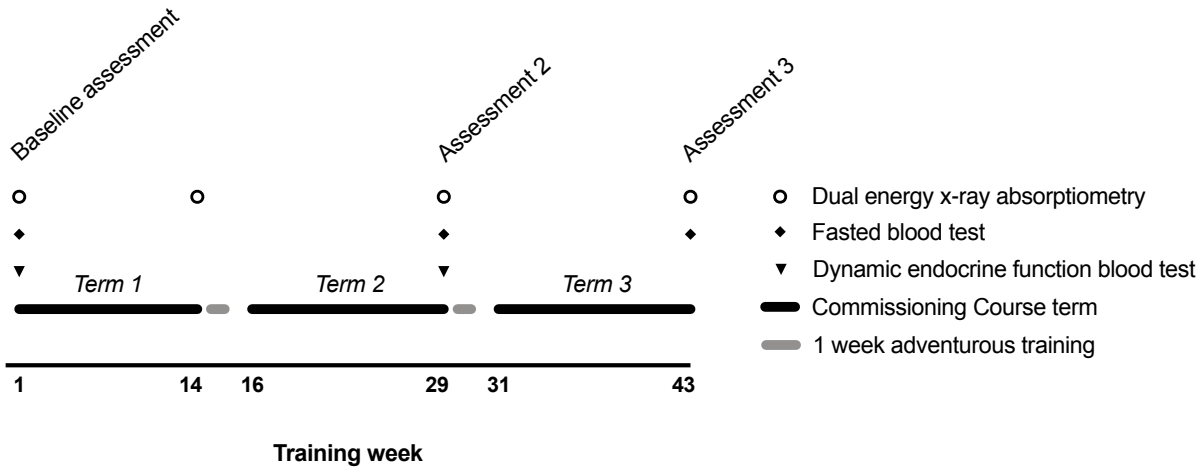


Figure 3-4-1 Study design

3.4.3.2 Baseline assessment and questionnaire

At study baseline, information including alcohol consumption, smoking and a comprehensive reproductive and medical history were recorded, including indication and type of any hormonal contraception used, or menstrual history if not using hormonal contraception, and year of menarche. Height and weight were assessed using stadiometer and scales (Seca models 213 and 874, respectively, Seca, Birmingham, UK).

3.4.3.3 Body composition

At the baseline assessment and at the ends of terms 1, 2 and 3, body composition was assessed using dual energy x-ray absorptiometry (DXA) (GE Lunar iDXA, GE Healthcare, Chalfont St Giles, UK) with participants wearing shorts and t-shirts. Whole body regional lean mass, fat mass and bone mineral content were computed using enCORE Body Composition software (GE Healthcare). Predefined regions were: arms, legs, trunk, android (area between the ribs and pelvis) and gynoid (pelvis and upper thighs). An additional module (CoreScan ®, GE Healthcare) assessed mass and volume of visceral adipose tissue (VAT) within the android region.

3.4.3.4 Menstrual function and ovulation

Participants not using hormonal contraceptives completed a pocket diary with six months of blank dates, printed on waterproof and rip-proof paper, indicating dates of menstrual bleeding throughout the study. Incomplete diary data (e.g. due to diaries being lost), were addressed using retrospective recall at the end of each term: date

of last menstrual period, menstrual period number and regularity, as well as current contraception use. Definitions used for categorisation as eumenorrhoeic, oligomenorrhoeic and amenorrhoeic were cycle lengths 26–35 days, 36–89 days and ≥90 days, respectively.

Ovulation was detected from changes in progesterone: creatinine ratio from serial urine aliquots. This method demonstrated 90% sensitivity and specificity compared with transvaginal ultrasound or gold-standard pregnanediol glucuronide (**Section 3.1**). Participants provided daily urine samples for 30 days at the start of the study, to attempt to identify a typical cycle (index cycle). For the remainder of the study, a weekly sample was provided on the same day each week for the remainder of the study. Ovulation was identified from our previously defined threshold values of 5.31 pg/mol (weekly samples from ovulatory and anovulatory women) and 3.62 pg/mol (daily samples from ovulatory women). Luteal phase defect (LPD) was more defined as 2-fold rise from the follicular phase plus a urine progesterone between 2.00 and 3.61 pg/mol. Luteal phase duration was not assessed after the index cycle. Non-cyclical progesterone concentrations or concentrations ≤2 nmol/mmol were defined as anovulatory.

3.4.3.5 Bone mineral density

Bone mineral density (BMD) was assessed from DXA scans at baseline and the ends of each term.

3.4.3.6 Blood sampling

In the first week of term 1 and final weeks of terms 2 and 3, at around 8 a.m., after fasting from 10 p.m., a single venous blood sample was drawn into EDTA, serum-separating gel and fluoride oxalate tubes (Monovette®, Sarstedt, Nümbrecht, Germany).

Since the exercise and stress involved in military training would be expected to have a suppressive effect on endocrine axes (for example, in studies of US Army Rangers (Friedl et al., 2000; Nindl et al., 2007)) we chose a stimulatory test. At week 1 of term 1 and the final week of term 2, a low-dose 1-hour gonadotrophin releasing hormone (GnRH) test was used to investigate subtle differences in anterior pituitary (gonadotroph) function (Morosini et al., 1989). Due to constraints imposed by the training schedule, dynamic testing was completed in the late afternoon. Participants were allowed to relax before a 20G cannula (B Braun, Dublin, Ireland) was inserted

into an antecubital fossa vein. A sample of blood was taken from the cannula in EDTA-containing tubes. After 10-15 minutes, 10μL gonadorelin hydrochloride (Intrapharm, Maidenhead, UK) was injected followed by a 10 mL saline flush. Venous blood was sampled from the cannula in EDTA-containing tubes after 20, 30, 40 and 60 minutes.

3.4.3.7 Laboratory methods

Blood samples were centrifuged at 3,550 g for 5 minutes and immediately stored at -80°C until the day of measurement.

Plasma LH, FSH, oestradiol, and progesterone, and urine creatinine and progesterone were assayed by Abbott Architect ® (Abbot, Longford, Ireland) according to manufacturer's instructions. The progesterone assay marketed for plasma/serum was also used to assay concentrations in urine, in tandem with urine creatinine, as described in **Section 3.1**. Thyroid stimulating hormone (TSH), free thyroxine (fT4) and total T3, IGF-1, insulin, C-peptide, anti-Müllerian hormone and sex hormone binding globulin (SHBG) were measured from gel-separated serum using Roche ® Cobas e411 analyser (Roche Diagnostics, Welwyn Garden City, UK). Creatinine, albumin, calcium, zinc and magnesium were determined from gel-separated serum and glucose from plasma containing fluoride oxalate using commercial kits (Alpha Laboratories, Eastleigh, UK) adapted for use on a Cobas Fara centrifugal analyser (Roche, UK). Leptin and inhibin B were measured from plasma by ELISA (Quantikine, USA and Beckman Coulter, High Wycombe, UK, respectively).

Intra- and inter-batch coefficient variations were <2.5% for Architect assays. Dihydroepiandrostenedione (DHEA), androstenedione 17-OH progesterone and testosterone were quantities in plasma were obtained following extraction and LC-MS/MS analysis. Briefly, a calibration curve for these steroids was prepared alongside plasma samples (200 μL) enriched with isotopically labelled analytes. Samples were extracted using Supported Liquid Extraction SLE400 cartridges (Biotage, UK) by diluting in 0.5M ammonium hydroxide (200 μL), loading, eluting with dichloromethane/isopropanol (0.45 mL x 3), drying under nitrogen and resuspending in 70:30 water/methanol (100 μL described previously (Spaanderman et al., 2018)). Chromatographic separation was achieved following injection (20 μL) using a gradient on a Shimadzu Nexera UPLC system on a Kinetex C18 (150 x 3

mm; 2 µm) column of mobile phases: 0.1 % FA in water, 0.1 % FA in methanol, 0.5 mL/min, 30 °C, followed by MS/MS analysis on a Sciex QTrap 6500+ operated in positive ESI, where Mass Spectrometry settings have been described previously (Stirrat et al., 2018). Least squares regression of the peak area ratio, with 1/x weighting, was used to calculate the amount of steroid in each sample within Analyst MultiQuant software (Sciex, UK). Inter-assay %CV was <4% for Architect ®, e411, Fara assays and intra-assay %CV <10% for all ELISAs.

3.4.3.8 Statistical analyses

Data are expressed as mean (standard deviation, SD) for normally distributed variables and median (IQR, interquartile range) for other data. The Homeostatic Model Assessment of insulin resistance 2 (HOMA2) was calculated using software freely available from the Oxford Centre for Diabetes Endocrinology and Metabolism Diabetes Trials Unit (https://www.dtu.ox.ac.uk/homacalculator/). Non-normal data were log transformed prior to analysis (leptin, inhibin B, LH, FSH and their ratio, oestradiol, serum progesterone, SHBG, insulin, AMH). Participants were excluded from follow-up analysis of reproductive function testing if they changed hormonal contraceptive between tests. High concentrations of synthetic sex steroids would be expected to exert negative feedback on the hypothalamus; thus, participants were categorised into groups depending on hormonal contraception: combined oral contraceptive pill (COCP, n=13), progestogen-containing contraception (Prog, e.g. progestogen-eluting contraceptive implant, progestogen-only pill or intramuscular medroxyprogesterone, n=16) and no hormonal contraceptive (Nil, n=14) or intrauterine system (IUS, n=4, e.g. Mirena ® or Jaydess ®). The Nil/ IUS group (n=18) were pooled since IUS delivers progestogen locally and does not impact the HPG axis centrally or interrupt ovulation (Apter et al., 2014).

Group characteristics were compared at baseline using Chi-square (binomial variables) or one-way ANOVA (continuous variables). Within-group changes in DXA variables and fasting blood markers were assessed using repeated measures ANOVA (main effect of time) and mixed repeated measures ANOVA (effect of group $\times$ time), respectively. Post-hoc paired t-tests comparing baseline values to each subsequent measurement were performed, where a significant main effect of time was identified. Effect sizes were reported using partial Eta squared (η_p^2, the proportion of true variance (as sum of squares) attributable to the effect, i.e. variance attributable divided by the total variance which could have been attributable

to the effect). η_p^2 values above 0.06 and 0.14 are medium and large effect sizes, respectively. The bone turnover ratio was calculated as ratio between P1NP and CTX, by as per Papageorgiou et al. (2017): a higher ratio would suggest bone turnover favouring formation over resorption.

Urine progesterone concentrations and menstrual periods were plotted against time for Nil/ IUS. Participants were categorised as eumenorrhoeic, oligomenorrhoeic or amenorrhoeic each term based on the menstrual pattern observed. Participants were categorised as ovulatory if 2 or more ovulatory cycles were detected, LPD if 2 or more LPD cycles met LPD criteria (**Section 3.4.3.4**) and anovulatory if 0 or 1 cycles were ovulatory, each term. Menstrual status and ovulation status were compared across terms using Chi squared tests. Since it was not possible to control the time of blood testing for reproductive cycle or contraceptive pack phase, the dynamic responses of FSH and LH to GnRH were analysed as the fold difference from immediately before GnRH administration: peak fold rise and area under the curve of fold response to GnRH (AUC), calculated by the trapezoidal rule. Peak and AUC of fold rise changes in LH and FSH were log transformed before the effects of group, time and group × time were measured by two-way ANOVA with Bonferroni post-hoc tests.

Statistical analysis was performed using SPSS version 24 for Macintosh (IBM, New York, USA). Significance was set at $p<0.05$.

3.4.4 Results

3.4.4.1 Participants, baseline assessment

Of 77 women who attended the study briefing and were eligible, 68 (88%) volunteered to participate (**Figure 3-4-2**). Of these, seven (10%) did not participate in the study *ab initio*, so did not undergo baseline testing. Nine participants (15%) withdrew subsequently due to injury, declining to participate or voluntary resignation (**Figure 3-4-2**). A further five were excluded from follow-up analysis of LH, FSH, oestradiol and androgens because they changed contraceptive during the study (commenced (n=4) or discontinued (n=1) a COCP).

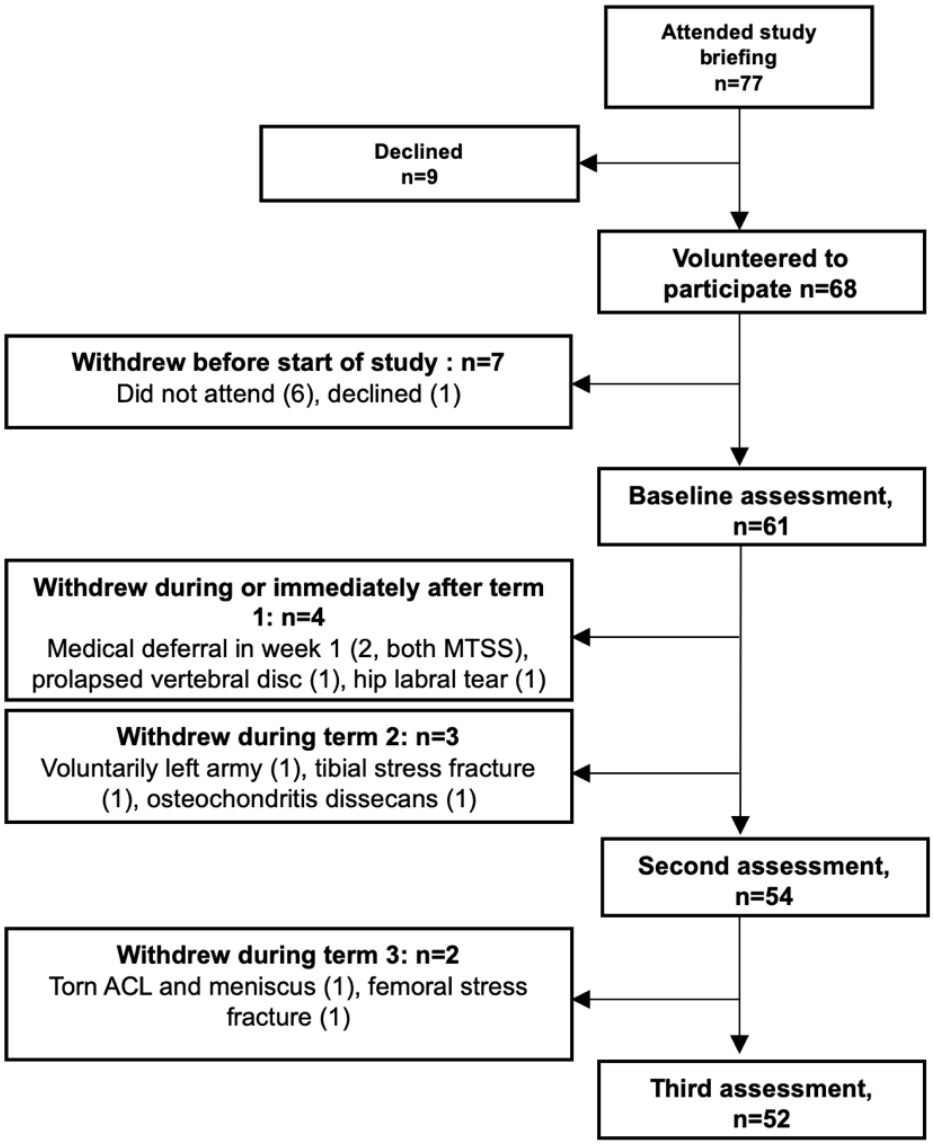


Figure 3-4-2. Recruitment, follow-up and withdrawal. MTSS: medial tibial stress syndrome ("shin splints"), ACL: anterior cruciate ligament.

The characteristics at baseline of participants who withdrew are compared with those who completed the study in **Table 3-4-1.** There were no differences in demographic, anthropometric, reproductive or lifestyle factors. Plasma or serum nutritional, reproductive and bone turnover markers did not differ except progesterone which was slightly higher among participants who withdrew. Fasted TSH, free T4, prolactin, LH, FSH, LH:FSH ratio, androstenedione, free and total testosterone, sex hormone binding globulin, dihydroepiandrostenedione, 21-OH progesterone and creatinine were within normal limits prior to participation in all participants at baseline (**Table 3-4-1**).

	Withdrew (n=9)	Completed study (n=52)	t or χ^2	p
Age, years	24.2 ±2.9	24.0 ±2.5	0.08	0.97
BMI, kg/m^2	23.5	23.2	-.36	0.73
Contraceptive group, n (%)				
COCP	2 (22)	13 (25)	0.68	0.72
Nil/ IUS	4 (44)	18 (35)		
Prog	3 (33)	16 (31)		
Discontinued COCP		4 (8)		
Commenced COCP		1 (2)		
Gynaecologic age, y	11.2 ±2.9	13.0 ±2.5	-0.40	0.61
Cycle length 6 months prior to participation (women not using hormonal contraception, median (IQR), d	48 (34, 68)	30 (30, 45)	-0.83	0.40
Ethnicity, n (%)				
White British	9 (100)	50 (96)	n/a	n/a
White Irish		1 (2)		
Other white background		1 (2)		
Teetotal (n, %)	3 (33)	5 (10)	χ^2 1.52	0.12
Smoker (n, %)	2 (22)	8 (15)	0.320	0.60
Fat mass, kg	15.7 ±3.95	15.2 ±3.3	0.340	0.70
Lean mass, kg	48.4 ±3.25	46.0 ±5.25	-1.390	0.21
Leptin	6.1 (4.18, 11.06)	7.87 (5.32, 11.2)	-0.013	0.98
IGF-1, ng/mL	233.0	237.1	0.210	0.64
Free T4, nmol/L	15.9 ±1.5	16.0 ±3.2	0.170	0.92
Total T3 nmol/L	2.03 ±0.35	2.07 ±0.49	0.917	0.92
TSH mU/L	2.81 ±1.54	2.60 ±1.00	1.790	0.19
PO_4, mmol/L	0.77 ±0.08	0.78 ±0.09	-0.240	0.84
Mg, mmol/L	118.7 ±6.24	119.69 ±8.38	-0.350	0.71
Zn, µg/dL	34.63 ±2.48	34.51 ±2.36	0.150	0.95
Albumin	67.5 (51, 93)	76 (57, 107)	0.590	0.61
Ferritin, g/L	20.86 ±6.18	20.83 ±4.89	0.180	0.98
Iron, µg/dL	2.46 ±0.42	2.73 ±0.62	-1.300	0.23
Transferrin, g/L	0.72 (0.6, 0.87)	0.79 (0.69, 0.9)	0.390	0.70

Lactate, mmol/L	5.62 (4.55, 9.45)	7.72 (5.77, 10.1)	-1.52	0.80
Insulin, uU/mL	4.52 ±0.46	4.65 ±0.37	-0.970	0.34
Glucose, mmol/L	1.65 ±0.65	1.79 ±0.45	-0.920	0.41
C peptide, ng/mL	80 (76, 91)	77 (77, 86)	1.180	0.29
Ca (adjusted) mmol/L	2.56 ±0.12	2.55 ±0.11	0.130	0.86
OHD3, nmol/L	72.1 ±25.7	81.6 ±35.9	2.162	0.30
Prolactin, mU/L	553 ±343	646 ±368	0.310	0.38
NEFA, mmol/L	0.74 ±0.4	0.94 ±0.45	-1.320	0.21
SHBG, nmol/L	74 (49.43, 110)	67.5 (51.42, 71)	-0.290	0.68
AMH, pmol/L	21.6 (12.2, 34.1)	19.5 (11.2, 27.3)	0.630	0.50
Inhibin, pg/mL	39.6 (22.1, 47.0)	28.5 (8.95, 61.0)	-0.730	0.51
DHEA, nmol/L	30.5 ±14.8	29.5 ±9.9	1.160	0.28
FAI	1.68 ±1.17	1.25 ±0.76	2.530	0.12
Progesterone, pmol/L	1.27 (0.95, 3.50)	1.11 (0.63, 1.59)	-2.340	0.02
Estradiol, pmol/L	139 (50, 312)	83 (36, 180)	0.140	0.9
FSH, IU/L	3.48 ±2.11	2.90 ±1.82	-0.867	0.87
LH, IU/L	4.26 (1.48, 4.75)	2.63 (0.85, 5.82)	0.910	0.42
CTX, ng/mL	0.56 (0.43, 0.92)	0.55 (0.41, 0.65)	1.508	0.14
Sclerostin, pmol/L	32.7 (29.9, 35.0)	37.4 (32.3, 43.1)	-1.593	0.12
BAP, U/L	21.84 ±5.13	19.32 ±4.85	1.490	0.14
P1NP, ng/mL	83.9 ±29.8	75.3 ±24.9	0.640	0.22

Table 3-4-1 Comparison of participants who completed the study with those who withdrew. p value for independent samples t- test (with t statistic) or Chi squared (χ^2, with χ^2 statistic), for continuous and categorical variables, respectively. BMI: body mass index, COCP: combined contraceptive pill, Prog: progesterone only contraception, Nil/ IUS: no hormonal contraception or intrauterine system user, IQR: interquartile range, IGF-1: insulin-like growth factor 1,T4: thyroxine, T3: triiodothyronine, TSH thyroid stimulating hormone, OHD3: 25-hydroxy vitamin D3, NEFA: nonesterified fatty acids, SHBG: sex hormone binding globulin, AMH: anti-Müllerian hormone, DHEA: dihydroepiandrostenedione, FAI: free androgen index, FSH: follicle stimulating hormone, LH: luteinising hormone, CTX: C-terminal telopeptide of type I collagen, P1NP: N-terminal propeptide of type I procollagen. BSAP: bone specific alkaline phosphatase

3.4.4.2 Nutrition and metabolic function changes

3.4.4.2.1 Body composition

Detailed regional body compositional changes are shown in **Table 3-4-2**. In the first term participants lost 1.1 ±3.6 kg in fat mass and gained 0.6 kg in fat-free mass, but by the end of term 2, fat mass increased 1.7 ±3.8kg and fat-free mass was no different from baseline. Fat-free and fat mass were no different from week 43. Regional changes demonstrated loss followed by gain of fat in the android region, and to a lesser extent in the trunk. Visceral adipose tissue (VAT) volume and mass increased consistently at each visit. Regional fat-free mass changes were modest but gains reached statistical significance in the arms, trunk and gynoid regions during term 1.

	Baseline	Week 14	Week 29	Week 43	ηp^2	p
Total Mass	64.2 ±7.8	63.8 ±7.78	65.0 ±7.72*	64.6 ±7.74	0.093	0.002
Fat Mass						
Arms	1.67 ±0.45	1.73 ±0.43	1.91 ±0.42*	1.84 ±0.44*	0.230	<0.001
Legs	6.46 ±1.61	6.27 ±1.6	6.73 ±1.68*	6.56 ±1.7	0.145	<0.001
Trunk	6.76 ±2.12	5.80 ±1.86*	6.85 ±1.9	6.52 ±2	0.299	<0.001
Gynoid	3.23 ±0.72	3.04 ±0.76*	3.39 ±0.78*	3.29 ±0.82	0.22	<0.001
Android	0.87 ±0.39	0.69 ±0.31*	0.87 ±0.34	0.80 ±0.33*	0.636	<0.001
Total	15.7 ±3.95	14.6 ±3.73*	16.3 ±3.78*	15.7 ±3.94	0.242	<0.001
VAT mass, g	92.3 ±84.9	96.9 ±71.6	122 ±88.3*	128 ±71.4*	0.112	0.008
VAT volume, cm^3	95.6 ±90.1	101.1 ±75.0	129 ±92.8*	134.7 ±75*	0.112	0.008
Fat-Free Mass						
Arms	5.06 ±0.8	5.19 ±0.74*	5.27 ±0.77*	5.13 ±0.72	0.101	0.001
Legs	16.7 ±2.2	16.8 ±2.09	16.6 ±2.07	16.6 ±1.87	0.043	0.086
Trunk	23.3 ±2.64	23.7 ±2.56*	23.2 ±2.42	23.6 ±2.54	0.113	<0.001
Gynoid	7.47 ±0.93	7.67 ±0.94*	7.53 ±0.91	7.63 ±0.95*	0.149	<0.001
Android	3.21 ±0.39	3.25 ±0.36	3.18 ±0.38	3.26 ±0.39	0.012	0.4
Total	48.6 ±5.48	49.3 ±5.23*	48.7 ±5.16	48.8 ±4.94	0.083	0.004

Table 3-4-2 Body composition. p for repeat measures ANOVA main effect of time; ηp^2: partial Eta squared, VAT: visceral adipose tissue, android: area ribs and pelvis; gynoid: pelvis and upper thighs

3.4.4.2.2 Markers of nutritional status and metabolic function

There was a significant increase in leptin, glucose, C-peptide and, accordingly, HOMA2 and decrease in NEFA, from baseline to the ends of terms 2 and 3 (**Figure 3-4-3** and **Table 3-4-3**). IGF-1 did not change during the study, free T4 decreased with significant decrements at each successive visit, while TSH increased significantly among non-COCP users (**Figure 3-4-3** and **Table 3-4-3**). Other markers of macro- and micronutrient status did not demonstrate any change changes with the exception of zinc, which decreased during the study overall. Iron and ferritin did not change although transferrin demonstrated a significant increase and decrease from baseline by the ends end of terms 2 and 3, respectively. Lactate was modestly higher elevated at the end of term 3 than previous visits. The effect interaction of contraceptive group with time was not significant for metabolic markers with the exception of zinc, which demonstrated decreases among COCP and Prog groups but not Nil/IUS, and total vitamin D, which decreased throughout the study but did not reach insufficiency (Thacher and Clarke, 2011) (**Table 3-4-3** and **Figure 3-4-3 D**).

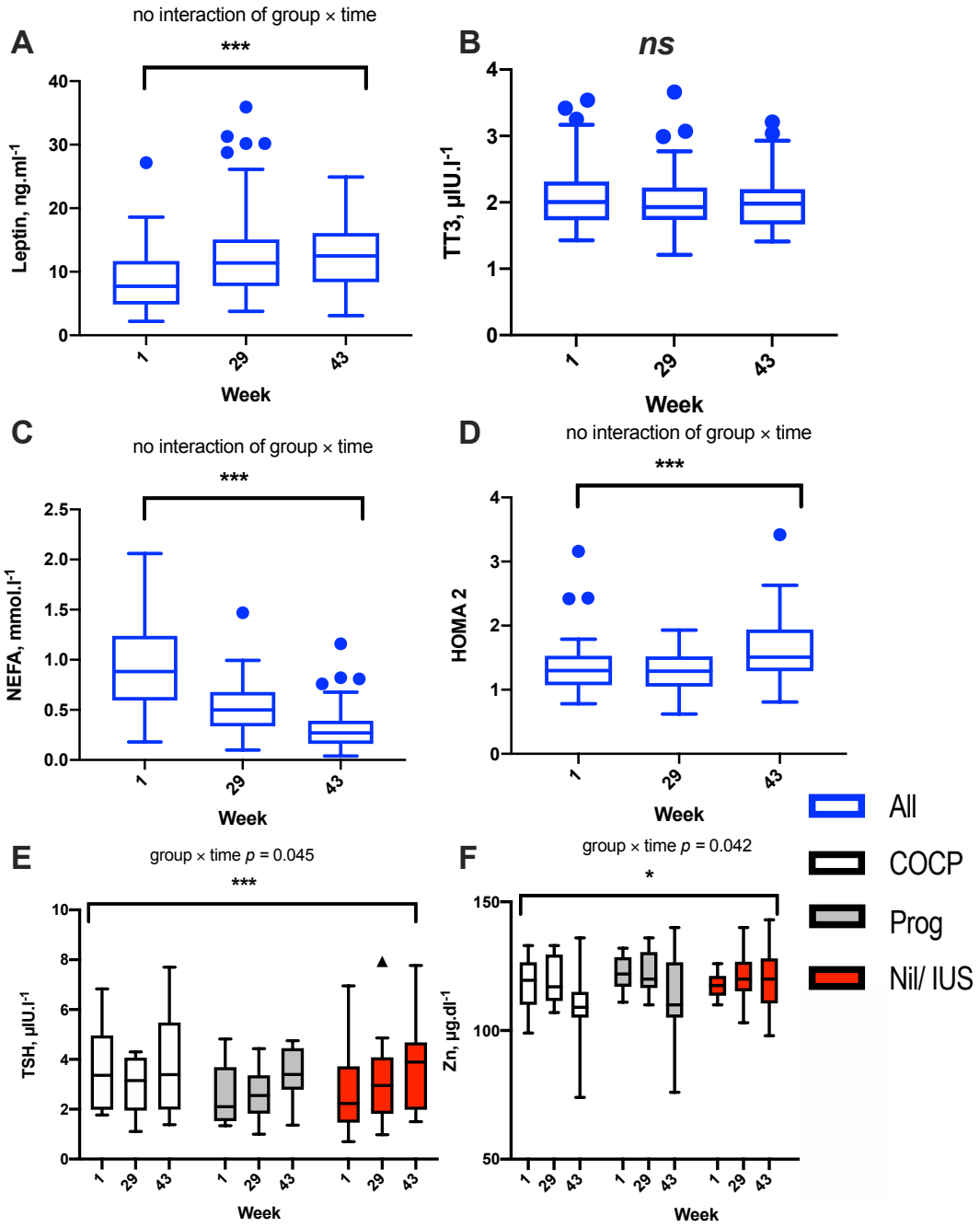


Figure 3-4-3. Nutritional changes during training. Changes from baseline (week 1) to ends of terms 2 and 3 (weeks 29 and 43, respectively) in fasting plasma or serum leptin (A), total T3 (B), non-esterified fatty acids (NEFA, C), homeostatic modelling assessment of insulin resistance 2 (HOMA2, D) thyroid stimulating hormone (TSH, E) and zinc (Zn, F). All: all participants (n=52); COCP combined contraceptive pill users(n=13) Prog progesterone only contraception users (n=16) Nil/ IUS intrauterine system or non-contraception users (n=18). * p<0.05, *** p<0.001

	week 1	week 28	week 42	ηp^2	p
Insulin, uU/mL median (IQR)	5.73 (4.31, 7.78)	6.23 (5.64, 8.28)	8.08 (7.12, 10.7)	0.060	0.17
Glucose, mmol/L	4.64 ±0.37	4.74 ±0.43**	4.75 ±0.6**	0.087	0.002
C peptide, ng/mL	1.86 ±0.49	1.85 ±0.43*	2.23 ±0.68**	0.281	<0.001
IGF-1, ng/mL	230 ±54	234 ±56	241 ±53	0.024	0.37
Free T4, nmol/L	15.93 ±1.55	14.37 ±1.77**	14.08 ±1.59**	0.440	<0.001
Total T3, nmol/L median (IQR)	1.96 (1.72, 2.27)	1.93 (1.76, 2.21)	1.94 (1.67, 2.17)	0.044	0.12
PO_4, mmol/L	1.61 ±0.15	1.58 ±0.17	1.63 ±0.14	0.057	0.18
Mg, mmol/L	0.78 ±0.1	0.75 ±0.09	0.77 ±0.07	0.034	0.44
Albumin, g/L	34.7 ±2.46	34.0 ±2.56	34.1 ±2.51	0.059	0.17
Ferritin, g/L, median (IQR)	78.5 (59.0, 117.5)	89.5 (62.0, 129)*	67.0 (49.5, 114)*	0.240	<0.001
Iron, µg/dL	20.6 ±4.99	20.7 ±4.4	19.4 ±4.82	0.040	0.31
Transferrin, g/L	2.75 ±0.61	2.68 ±0.49	2.73 ±0.53	0.028	0.43
Lactate, mmol/L	0.80 ±0.32	0.82 ±0.26	0.95 ±0.28*	0.112	0.002
Creatinine, mmol/L median (IQR)	79.8 (73.15, 86.45)	88.3 (78.9, 92.1)*	85.5 (78.9, 92.1)**	0.155	0.008
Ca (adjusted), mmol/L	2.49 ±0.11	2.54 ±0.11	2.47 ±0.45	0.024	0.5
OHD3, nmol/L	71.9 (3.36)	66.2 (2.90)	59.4 (3.01)**	0.138	0.036
Prolactin, mIU/L median (IQR)	637 (412, 875)	646 (434, 893)	599 (483, 798)	0.008	0.84

Table 3-4-3 Markers of metabolic and nutrition status. p value for repeated measures ANOVA, main effect of time. Group x time was insignificant for all variables in table. *p<0.05 vs baseline. **p<0.01 vs baseline. ηp^2: partial Eta squared. IQR: interquartile range, IGF-1: insulin-like growth factor 1,T4: thyroxine, T3: triiodothyronine, TSH thyroid stimulating hormone, OHD3: 25-hydroxy vitamin D3

3.4.4.3 Reproductive function

The contraceptive groups comprised 13 participants who used a COCP, 16 who used Prog (9 progestogen-eluting contraceptive implant, 6 progestogen-only pill and 1 intramuscular medroxyprogesterone), 14 non users and 4 who used an IUS.

3.4.4.3.1 Menstrual function

Of 24 women for whom menstrual function was recorded, two commenced a COCP and three withdrew during the study. For these five women, menstrual function was only recorded for median (range) 4 (2 to 8) months only. Menstrual cycle data were not analysed for the four women who used an IUS because of the local effects of levonorgestrel, resulting in erratic or infrequent (two women) or absent menses (two women). Menstrual function is shown **Figure 3-4-4** panels A and B. Eleven participants (65%) and 6 (43%) were oligo/amenorrhoeic in terms 2 and 3, respectively, compared with 5 (25%) in term 1. There was no significant effect of study visit on menstrual function (χ^2 23.4, p=0.32).

3.4.4.3.2 Ovulation

The higher progesterone: creatine threshold to confirm ovulation of 5.31 pg/mol (1.67 µmol/mol; **Section 3.1**) was not reached, except in single samples by three separate women (one each in terms 1, 2 and 3). The lower threshold of 3.62 pg/mol (1.14 µmol/mol) was achieved in 25 cycles (16%) in 12 women. Ovulatory function is shown in **Figure 3-4-4** panels C and D. Three (13%), one (5%) and 2 (17%) participants were ovulatory in terms 1, 2 and 3, respectively. Ovulatory function decreased significantly during the study (χ^2 18.5, p=0.005; **Figure 3-4-4** C and D). Representative examples of menstrual and urine progesterone profiles are given in **Appendix 3-4-1**. These illustrate a range of menstrual regularity which was not accompanied by ovulation and/ or a luteal phase.

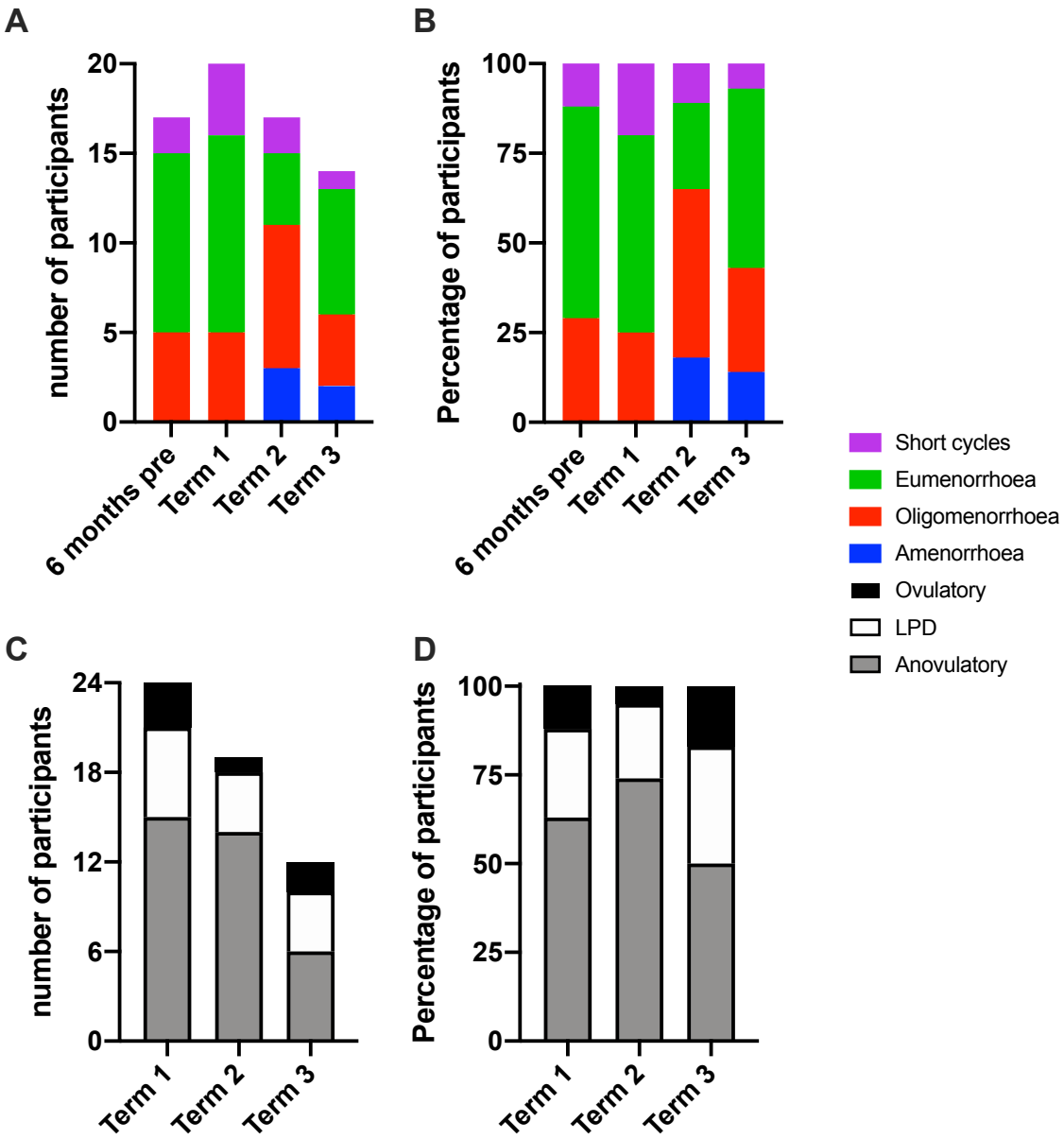


Figure 3-4-4. Menstrual and ovulatory function during the study. Panels A and B: menstrual function; B and C: ovulatory function. ■short cycles ■eumenorrhoea ■oligomenorrhoea ■amenorrhoea ■ovulatory □luteal phase defect (LPD) ■ anovulatory

3.4.4.3.3 Basal reproductive function markers

Baseline assessment, assessment 2 and assessment 3 were completed on cycle day 19 (±19 days),16 (±12 days) and 15 (±10 days), respectively, for non–hormonal contraception users; p=0.5. Inhibin B and FSH were lower among COCP users than other groups and increased during the study in all groups (significant main effect; no

interaction of group × time). Oestradiol demonstrated a trend towards increase across study visits, but no change was observed in AMH (**Figure 3-4-5**).

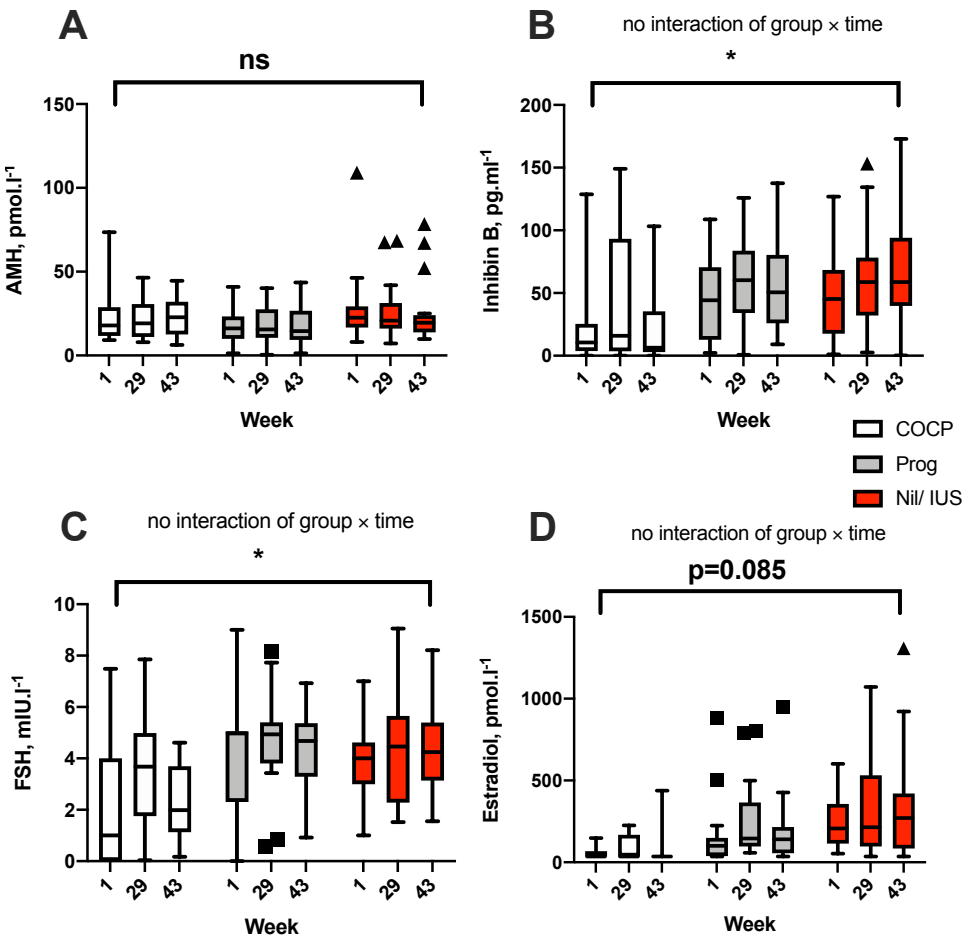


Figure 3-4-5. Markers of reproductive function. Panel A: Fasting anti-Müllerian hormone (AMH), panel B: inhibin B, Panel C: follicle stimulating hormone (FSH) and Panel D:oestradiol at weeks 1, 29 and 43. COCP combined contraceptive pill users (n=13) Prog: progesterone only contraception users (n=16) Nil/ IUS intrauterine system or non-contraception users (n=18) * p<0.05

Other basal reproductive function markers are detailed in **Table 3-4-4**. SHBG was higher among COCP users than other participants. While the main effect of time was significant, demonstrating increased SHBG overall by week 29, and the effect of group × time showed SHBG decreased in COCP and Prog, but not Nil/IUS, by week 43, post-hoc paired t-tests did not show differences in groups compared with baseline. FAI was lower among COCP, however FAI, androstenedione and DHEA did not change during the study. Progesterone and LH:FSH ratio did not change during the study.

	Week 1	Week 29	Week 43	Time (main effect)		Group × time interaction	
				η_p^2	p	η_p^2	p
SHBG, nmol/L							
COCP	193 (90, 230)	201 (156, 251)	127 (99, 229)	0.204	0.008	0.103	0.053
Prog	61 (39, 80)	64 (41, 88)	56 (42, 66)				
Nil/ IUS	52 (48, 74)	62 (49, 87)	61 (51, 74)				
Free androgen index							
COCP	0.58 (0.47, 0.78)	0.56 (0.42, 0.65)	0.72 (0.45, 0.92)	0.014	0.53	0.012	0.91
Prog	1.5 (1.15, 1.8)	1.57 (1.15, 1.8)	1.66 (1.05, 1.97)				
Nil/ IUS	1.4 (1.19, 1.62)	1.70 (1.32, 2.1)	1.57 (1.35, 1.91)				
Dihydroepiandrostenedione, nmol/L							
COCP	27.2 (18.9, 30.7)	20.8 (19.6, 30.7)	25.5 (18.2, 32.6)	0.034	0.23	0.029	0.64
Prog	35.6 (22, 38.9)	22.6 (17.8, 35.7)	22.9 (19.3, 34)				
Nil/ IUS	31.6 (23.6, 36.2)	29.4 (23.9, 35.3)	28.1 (21.6, 31.9)				
Androstenedione, nmol/L							
COCP	3.95 (3.86, 4.25)	3.51 (3.09, 4.4)	3.84 (3.21, 4.26)	0.054	0.095	0.042	0.46
Prog	4.11 (3.85, 5.66)	4.54 (2.74, 5.2)	4.12 (2.95, 5.16)				
Nil/ IUS	5.03 (3.84, 6.04)	5.36 (4.27, 7.08)	4.66 (3.56, 5.98)				
Progesterone, pmol/L							
COCP	1.27 (0.95, 1.59)	0.95 (0.64, 1.27)	0.79 (0.64, 0.95)	0.064	0.091	0.010	0.87
Prog	1.27 (0.95, 12.1)	0.95 (0.64, 1.27)	1.11 (0.64, 1.59)				
Nil/ IUS	2.23 (1.27, 15.9)	0.95 (0.79, 3.5)	5.41 (1.11, 14.9)				
LH:FSH							
COCP	0.35 (0.29, 0.58)	0.67 (0.34, 0.75)	0.65 (0.21, 0.84)	0.028	0.32	0.070	0.21
Prog	0.8 (0.47, 1.27)	0.81 (0.46, 1.22)	0.96 (0.48, 1.32)				
Nil/ IUS	1.36 (0.8, 1.9)	2.07 (1.29, 2.82)	1.28 (0.57, 2.06)				

Table 3-4-4. Fasting reproductive markers. All are measured from plasma. COCP: combined contraceptive pill, Prog: progesterone only contraception, Nil/ IUS: no hormonal contraception or intrauterine system user, IQR: interquartile range, SHBG: sex hormone binding globulin, FSH: follicle stimulating hormone, LH: luteinising hormone. η_p^2 partial Eta squared

3.4.4.3.4 Dynamic HPG axis function

The responses of LH and FSH to GnRH are shown in **Table 3-4-5** and **Figure 3-4-6** shows the maximum and AUC fold-wise response of LH and FSH to GnRH. All four measures of gonadotroph response became suppressed by the end of term 2 (effect of time $p<0.05$). The direction of response did not differ between groups (no effect of group × time).

	Week 1	Week 29	Time (main effect)		Group × time interaction	
			F	p	F	p
LH, IU/L						
COCP	0.44 (0.03, 2.13)	1.3 (0.6, 2.73)	5.13	0.028	0.85	0.43
Prog	3.02 (1.06, 4.04)	3.36 (1.84, 4.21)				
Nil/ IUS	3.61 (2.06, 6.6)	4.20 (3.73, 6.37)				
FSH, IU/L						
COCP	0.81 (0.14, 4.81)	3.57 (2.64, 4.13)	8.33	0.006	2.92	0.11
Prog	3.73 (2.92, 5.18)	4.63 (3.44, 5.32)				
Nil/ IUS	3.25 (1.92, 4.36)	4.19 (2.78, 5)				

Table 3-4-5. Afternoon, pre-test LH and FSH levels in week 1 and week 29. COCP: combined contraceptive pill, Prog: progesterone only contraception, Nil/ IUS: no hormonal contraception or intrauterine system user, IQR: interquartile range, SHBG: sex hormone binding globulin, FSH: follicle stimulating hormone, LH: luteinising hormone, CTX: C-terminal telopeptide of type I collagen, P1NP: N-terminal propeptide of type I procollagen. BSAP: bone specific alkaline phosphatase, η_p^2 partial Eta squared.

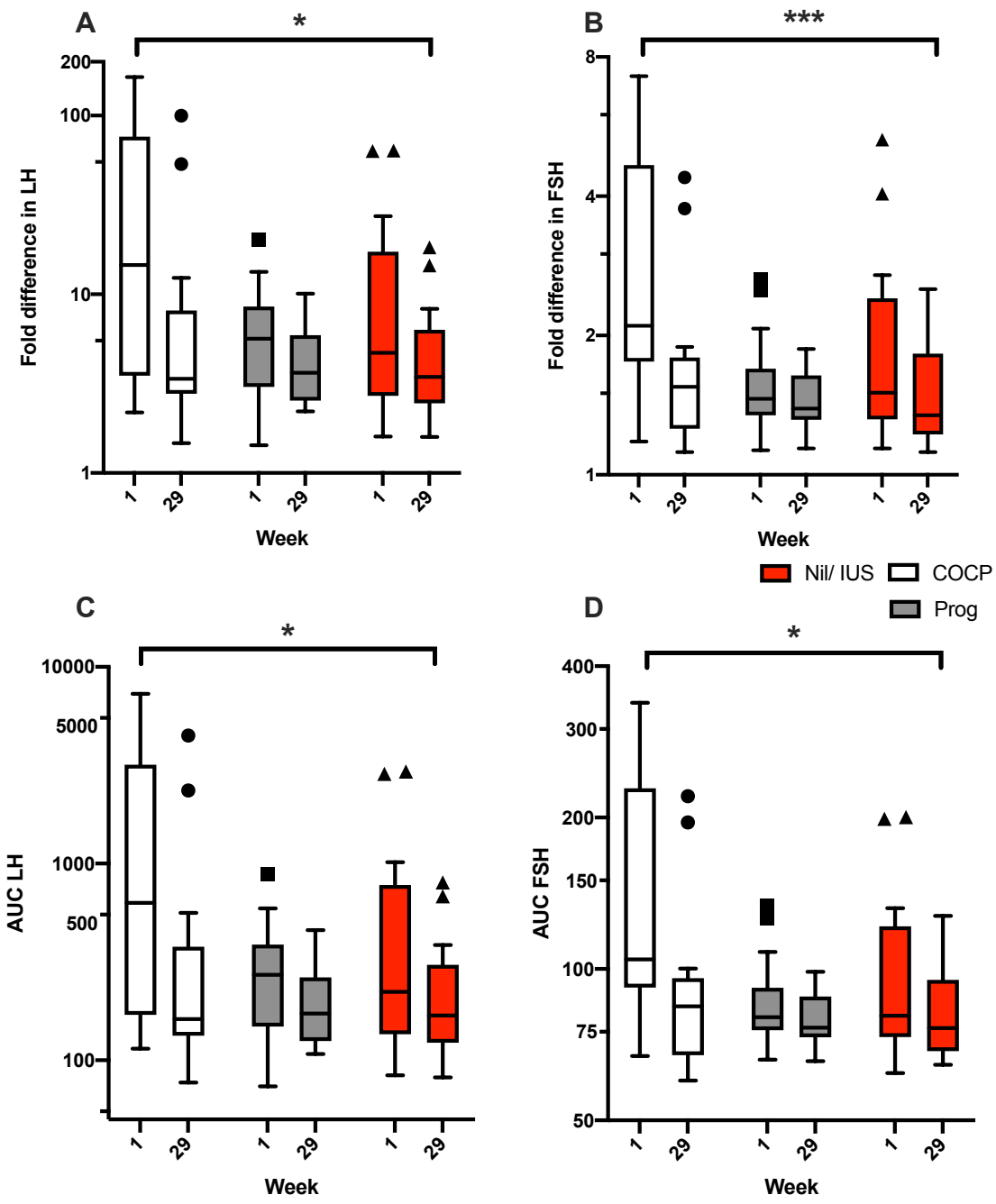


Figure 3-4-6. Response to GnRH test at week 1 and 29 of training. Panels A and B show maximal fold differences in luteinising hormone (LH) and follicle stimulating hormone (FSH), respectively. Panels C and D show area under the curve (AUC) response over 1 hour in LH and FSH, respectively. Due to the wide range of responses between groups, y axes are shown as a log scale. COCP: combined contraceptive pill users(n=13) Prog: progesterone only contraception users (n=16) Nil/ IUS: intrauterine system or non-contraception users (n=18) * $p<0.05$, *** $p<0.001$.

3.4.4.4 Bone adaptation

3.4.4.4.1 Bone mineral density

There was a modest overall decrease in BMD from baseline to week 43 (**Table 3-4-6**). Significant increases in arms and pelvis BMD were accompanied by a trend towards decreased ribs BMD, however effect sizes were all small.

	Week 1	Week 14	Week 29	Week 43	ηp^2	p
Total BMD, $g.cm^3$	1.23 ±0.08	1.23 ±0.08	1.22 ±0.08	1.22 ±0.09*	0.078	0.013
Arms BMD, $g.cm^3$	0.87 ±0.1	0.89 ±0.1*	0.88 ±0.12	0.83 ±0.13	0.084	0.009
Head BMD, $g.cm^3$	2.34 ±0.22	2.35 ±0.22	2.33 ±0.21	2.36 ±0.21	0.046	0.10
Legs BMD, $g.cm^3$	1.26 ±0.09	1.26 ±0.1	1.25 ±0.08	1.25 ±0.08	0.045	0.11
Pelvis BMD, $g.cm^3$	1.12 ±0.12	1.13 ±0.11	1.13 ±0.11	1.13 ±0.11	0.043	0.12
Ribs BMD, $g.cm^3$	0.89 ±0.06	0.87 ±0.07	0.88 ±0.07	0.88 ±0.07	0.045	0.059
Spine BMD, $g.cm^3$	1.12 ±0.11	1.12 ±0.1	1.12 ±0.1	1.11 ±0.1	0.011	0.73
Trunk BMD, $g.cm^3$	1.04 ±0.09	1.04 ±0.09	1.04 ±0.09	1.04 ±0.09	0.024	0.44

Table 3-4-6. Total and regional bone mineral density (BMD). p for repeated measures analysis of variance (RM ANOVA). * significant ($p<0.05$) pairwise comparison versus baseline, where RM ANOVA main effect of time was significant.

3.4.4.4.2 Bone turnover markers

Bone turnover markers are illustrated in **Table 3-4-7.** Resorption markers CTX decreased significantly, while sclerostin tended to increase after 29 weeks. The effect size was greater for CTX. Bone specific alkaline phosphatase and P1NP did not change significantly during training, although bone turnover ratio increased from baseline to week 29, reverting to baseline by week 43. There was no effect of contraceptive group × time for any bone turnover marker.

		Week 1	Week 29	Week 43	ηp^2	p
Resorption	CTX, ng/mL	0.58 (0.42, 0.71)	0.53 (0.39, 0.63)**	0.53 (0.43, 0.67)	0.138	0.015
	Sclerostin, pmol/L	38.4 (32.9, 46.8)	40.6 (34.9, 48.2)*	41.9 (31.8, 47.7)	0.085	0.077
Formation	BSAP, U/L	19.3 ±4.89	19.3 ±4.48	19.6 ±5.46	0.021	0.84
	P1NP, ng/mL	73.2 (57.4, 88.1)	77 (62.1, 91.9)	73.7 (61, 87.3)	0.050	0.12
	BTR	134.3 ±32.9	168.4 ±72.4*	136.5 ±36.5	.0134	0.003

Table 3-4-7. Bone turnover markers. P for repeated measures ANOVA (main effect of time). Effect of contraceptive group × time was not significant for any variable in this table. * p<0.05 ** p<0.01 for paired t-test versus baseline. CTX: C-terminal telopeptide of type I collagen, P1NP: N-terminal propeptide of type I procollagen. BSAP: bone specific alkaline phosphatase BTR: bone turnover ratio.

3.4.5 Discussion

We undertook a detailed assessment of the HPG axis in women during arduous training. Our hypothesis was rejected; contrary to our expectations, evidence suggested nutritional sufficiency, yet anovulation occurred almost universally in non-contraceptive users. Gonadotroph response to GnRH was suppressed in both contraceptive users and non-users. Progressive elevations in inhibin, FSH and an upward trend in oestradiol, alongside unchanged AMH, suggested ovarian follicular dysregulation. Bone turnover ratio showed increased formation after 29 weeks but was no different to baseline by 43 weeks of training.

Fasting leptin, insulin resistance, oestradiol and bone turnover markers demonstrated changes in the opposite direction expected during low EA (reviewed by Elliott-Sale et al. (2018)). Other markers such as IGF-1, albumin, iron, calcium and other micronutrients (except zinc) did not change during training. There was no change in TT3. In low EA, T3 is often decreased with unaffected TSH (Elliott-Sale et al., 2018). In the present study, TSH levels were higher in contraceptive users, possibly due to TSH-stimulating effect of oestrogens, while the progressive TSH increase in non-COCP users might have been related to the observed rise in oestradiol (Weeke and Hansen, 1975).

Paradoxically, decrease in fat mass and increases in lean mass were greatest during term 1, when fewer arduous field exercises took place and the training focus was on building soldiering skills, fitness and strength. In the second term, when the

most arduous field exercise assessments took place, body fat increase was greatest. It is likely that 'energy compensation' occurred between field exercises, whereby energy intake increases and expenditure decreases following a period of energy deficit, which was likely to have been exacerbated by restricted sleep (Riou et al., 2015). Body fat reverted towards baseline by the end of term 3, which was measurably less physically and psychologically taxing than term 2 **(Section 3.3)**; however, HOMA2, reflecting pancreatic insulin resistance, continued an upward trend, possibly caused by a modest increase in VAT. Visceral adipose tissue is highly metabolically active and associated more closely with pancreatic insulin resistance than total body fat (Goodpaster et al., 1999). In obese individuals, NEFA concentrations reflect insulin resistance (Johnston et al., 2018), but the progressive decrease in NEFA in this study may have been an acute response of (highly insulin-sensitive) adipose tissue to increasing insulin levels in women with normal BMI.

We found zinc concentrations decreased overall with training in users of hormonal contraception. Zinc is an important cofactor for many metalloprotein enzymes including bone specific alkaline phosphatase (BSAP), and zinc insufficiency has been shown to uncouple bone turnover (King et al., 2016). Mean serum zinc did not fall reach insufficiency (the lower cutoff for women being 70 ng/mL, (King et al., 2016)) and BSAP concentrations were preserved, suggesting zinc insufficiency did not contribute to bone outcomes. A similar relationship between hormonal contraceptive use and zinc was made in one cross-sectional study: Simoes et al. (2015) found zinc and BSAP concentrations were 11% and 28% lower among COCP users than non-users, respectively, despite identical dietary intake (progesterone only contraceptives were not studied). Zinc is important for DNA synthesis; its role in spermatogenesis is well-understood (Zhao et al., 2016) and zinc sufficiency also appears to be important for oocyte development and ovulation (Ebisch et al., 2006). Military women appear to be at risk of micronutrient deficiencies including zinc, vitamin D and iron (Andersen et al., 2010; Lutz et al., 2019), like athletes (McClung et al., 2014), hence the interaction between zinc, sex hormones and bone turnover would warrant further investigation.

Whether recruits undergoing basic military training may be considered athletes is moot. As was seen in Section 3.2 and has been shown elsewhere (Richmond et al., 2010; Siddall et al., 2019), high physical demands are placed on Officer Cadets. During three representative 10-day assessments using doubly-labelled water and

accelerometry, we found higher levels of moderate physical activity than vigorous physical activity (mean duration per day 5.0–7.5 hours versus 20–45 minutes).

Studies demonstrating HPG axis suppression in athletes have established a widely-accepted paradigm of HA induced by low EA, accompanied by suppressed frequency and amplitude of LH and FSH secretion, low oestradiol and reduced BMD (Ihle and Loucks, 2004; Papageorgiou et al., 2017; Southmayd et al., 2019). Subjects with low EA have reduced GnRH pulsatility with preserved or increased pituitary gonadotroph responsiveness compared with controls (Loucks et al., 1989; Loucks and Thuma, 2003). Conversely, our study identified decreased pituitary responsiveness after 28 weeks of training, impacting LH proportionately more than FSH, with oestradiol concentrations tending to increase. There was little or no biochemical evidence of low EA, so low EA does not explain our observations.

The rates of menstrual disturbance we observed were commensurate with previous studies of military training in USA (reviewed in **Section 1.2**). The incidence of anovulation in women not taking contraceptives was concerningly high, using the more conservative cut-off for ovulation identified in **Section 3.1**. In the few instances of recorded ovulation, the luteal phase did not appear to be sustained and such that progesterone insufficiency was likely.

The increase we observed in inhibin B but not AMH suggests follicular dysregulation, which occurred irrespective of contraceptive use. Inhibin B, a marker of larger follicle dominance, was likely to be driven by increased basal FSH (Telfer and McLaughlin, 2007). The duration of the FEAT study was longer than previous studies in athletes and it may be that the increased FSH was due to prolonged durations spent in the follicular phase over 11 months. FSH also drives increased ovarian oestradiol production. The increase in oestradiol was a novel finding in exercising women, and ran contrary to data from athletic women with low EA, in whom oestradiol, LH and FSH are commonly suppressed. For example, a study detailing gonadotrophin and sex steroid profiles in women with low EA showed LPD and oligomenorrhoea were associated with suppression of normal LH (and FSH) surges at the end of the luteal phase and lower oestradiol (De Souza et al., 1998). The authors explained these as reduced follicular recruitment and maturation the athlete with low EA, and many studies have since made similar findings in exercising women with low EA (reviewed by De Souza et al., (2014)). The

development of smaller follicles, for which AMH is more specific, appeared to be unchanged (Li et al., 2011).

While we did observe increased insulin resistance, there was no evidence of androgenisation during training, Androgenisation is an important contributor to reproductive dysfunction in a subset of athletes, often compounded by low EA, and is closely related to insulin resistance in polycystic ovary syndrome (PCOS) (Javed et al., 2015). Studies of exercise in non-athletes demonstrate benefits to the HPG axis in patients with PCOS and obesity (Mena et al., 2019). Exercise appears to improve ovulation by reducing insulin resistance, although when exercise is arduous (60 minutes or more per day) it is associated with upregulation of the HPA axis and suppression of the HPG axis, exacerbated by low EA (Hakimi and Cameron, 2017).

The concept of within-day energy deficit has emerged, whereby hour-to-hour energy balance can swing from deficit to surplus while across 24 hours, total energy balance is maintained. Within-day energy deficit has been shown to be associated with menstrual disturbance in female athletes in whom 24-hour EA and body composition do not differ from normally menstruating controls (Fahrenholtz et al., 2018). The same authors made similar observations in men, albeit less conclusively (Torstveit et al., 2018). However, the number of hours of within-day energy deficit was associated with elevated cortisol and lower oestradiol and total T3 (Fahrenholtz et al., 2018), whereas we found cortisol decreased (**section 3.3**), oestradiol tended to increase, and total T3 did not change. Moreover, while it is plausible that within-day energy deficit may account for reproductive dysfunction where there is no difference in net (across day or days) energy balance, it is difficult to envisage within-day energy deficit made a significant contribution to reproductive dysfunction when weight gain occurred.

One interpretation of our findings is that they reflect a latent effect of several periods of low EA lasting days or weeks, interspersed by periods of excess energy compensation, leading to reversal of biomarker expression of low EA and overall increases in insulin resistance and leptin. In **Section 3.2**, 47 participants from this study demonstrated lower EA during a period encompassing a military field exercise than periods which did not, and higher EA was associated with loss in fat mass. However, the clinical effects of low EA in athletes are temporary energy-conserving

adaptations; upon restoration of adequate EA it is anticipated that reproductive function is restored (De Souza et al., 2019a).

However, it seems more likely that our findings represent a maladaptive response of reproductive function to military training, which was hypothesised in **Section 3-1**. As demonstrated in **Section 3.**3, the HPA axis was upregulated, with increased monthly average cortisol concentrations during training and increased plasma cortisol response to ACTH from baseline to the end of term 2. Fasting plasma and morning saliva cortisol reduced from term 1 throughout training, which likely represented a habituation to the stress of venepuncture and training, respectively. We observed increased perceived stressors and adverse psychological effects (increased anxiety and decreased mood), which reflected tiredness likely caused by sleep disturbance or deprivation.

The stress induced by exercise itself has been shown not to suppress the HPG axis when EA is sufficient (Loucks, 2013). Low EA decreases leptin and ghrelin, which mediate HA both by direct effects on GnRH secretion in the median eminence of the hypothalamus (Hofmann et al., 2017), and by upregulating the HPA axis (Misra and Klibanski, 2014). Activation of the HPA axis has been shown to modulate the HPG axis per se (Vulliemoz et al., 2008). We hypothesise that HPA axis activation caused by other pressures of training disrupted gonadotrophin associated with metabolic maladaptation.

Strengths of our study included combining dynamic and basal testing of reproductive function over a meaningful duration (seven to 11 months). The majority of our study population used hormonal contraception, making our findings generalisable to many women undertaking arduous training, not just non-users (Batig, 2017). Measuring EA in practice is notoriously challenging; our combined physical and biochemical measures provided meaningful corroborative evidence of metabolic status. In future these tests may constitute a ‘panel’ of EA surrogates for clinicians and researchers. Likewise, weekly urinary progesterone is a promising technique for larger field studies aiming to detect ovulation, and highlighted the importance of biochemical assessment over reliance on menstrual cycle function. Near-patient testing of pregnanediol (Leiva et al., 2019) and even LH (Dhillo et al., 2019) may soon be possible. The GnRH test we used was an effective field study measure of HPG axis

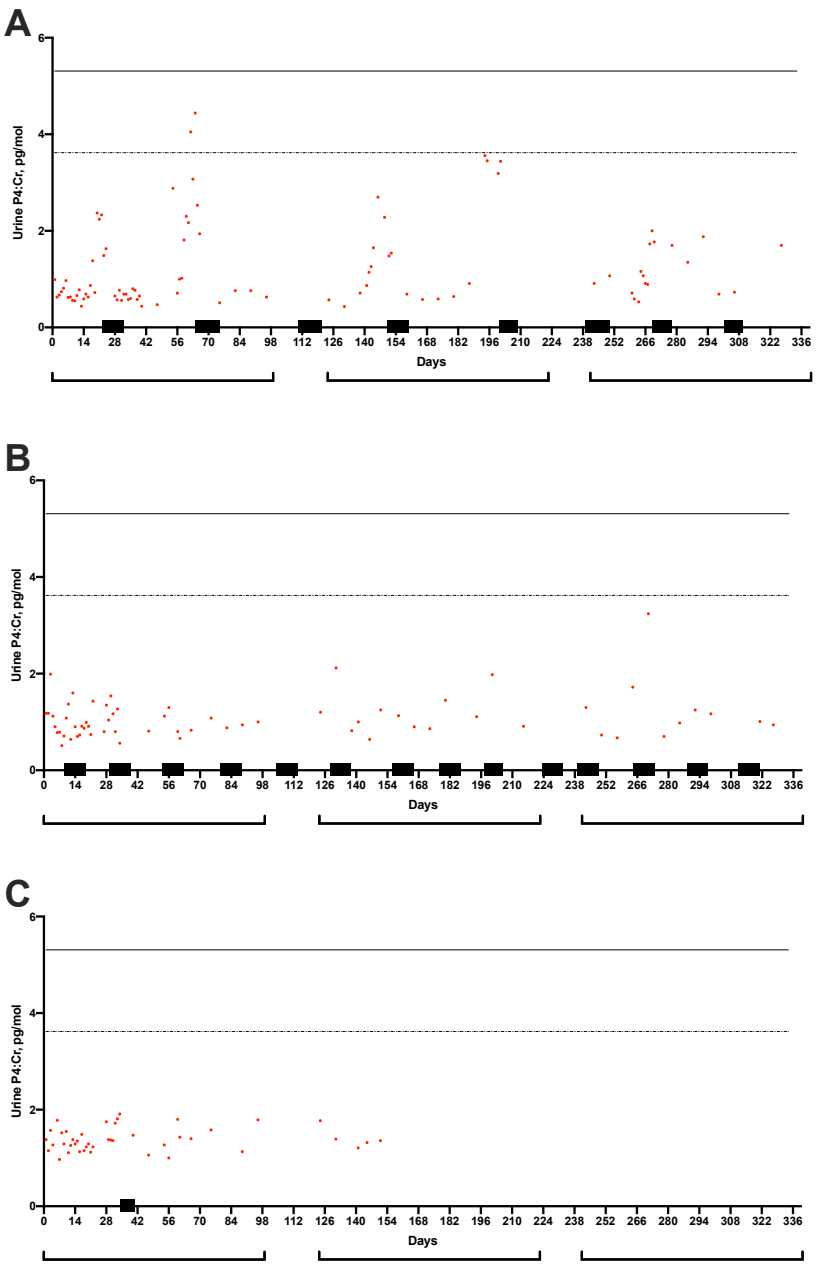
function and allowed study visits to be organised around training schedule, not the reproductive cycle.

Diagnosing LPD is difficult (Mesen and Young, 2015); weekly sampling, while sensitive and specific for identifying ovulation, is insufficient to detect LPD using the diagnostic criterion of a luteal phase duration of 9 days or less (Smith et al., 1984). A four day sampling interval would be required, or more reasonably sampling twice per week. Unfortunately, it was not possible to determine EA in **Section 3.3** due to inaccuracy in EI measurement, hence we have relied on fasting biomarkers. Future studies might aim to assess the hypothalamic tier of the HPG axis using kisspeptin testing (Stieg et al., 2015).

Tracking urinary progesterone from its ratio to creatinine assumes a constant creatinine level, however during heavy exercise creatinine may vary (Beitins et al., 2001). Heavy exercise therefore could have led to progesterone: creatinine ratios below thresholds for ovulation or LPD in women with regular cycles, falsely. Future studies could compare progesterone with contemporaneous specific gravity, which may be less prone to fluctuations during exercise (Miller et al., 2004). In the present study, we found storage of urine samples at –80 ℃ precipitated calcium salt crystals, which could have altered specific gravity.

Arduous training in women was associated with metabolic maladaptation, pituitary gonadotroph suppression and ovarian dysregulation. Suppression of the HPG axis may have been compounded by other elements of training via upregulation of the HPA axis, including sleep deprivation and psychological stress. Focused studies are required to explore mechanisms underlying our observations. Understand sex-specific impacts of activities formerly only undertaken by one sex or gender, like soldiering, is essential. It is not likely that one factor (e.g. low EA) was responsible for the HPG axis suppression we observed and a combination of stressors contributed, e.g. exercise, psychological stress, intermittent energy insufficiency and excess, and sleep deprivation. As wider studies of women in military employment are planned (e.g. the ARMI Study (Hughes et al., 2019)), our results make a valuable contribution to interpreting the effect of arduous training on HPG axis function.

3.4.6 Appendix material to Section 3-4



Appendix 3-4-1. Representative menstruation and ovulation patterns. • Urinary P4:Cr (progesterone: creatinine concentration), ■ recorded menses. Top line ··· ovulatory P4:Cr threshold identified from daily samples in healthy women, lower dashed line - - - ovulatory P4:Cr threshold identified from weekly samples in anovulatory and ovulatory women (see Section 3.1). ⌙ three 14-week terms. Panel A: two ovulatory cycles and lengthening of cycles; evidence of luteal insufficiency. B: no ovulatory cycles identified despite regular menstruation; C: oligo/ amenorrhoeic from the start of term 1 until week 5 of term 2 (at this point the participant left the study due to a tibial stress fracture).

Chapter 4 Responses to arduous exercise in extreme cold

The following chapter refers to Ex ICE MAIDEN, the first all-female unassisted team crossing of Antarctica, completed by a team of six from November 2017 – January 2018. The team endured temperatures below –50º C and wind speeds up to 70 mph, finishing the 1000-mile ski expedition 15 days ahead of schedule. The 'Ice Maidens' were the largest team of any gender to complete the crossing, the first polar exploration novices to do so, and they increased the total number of women who had crossed Antarctica from four to 10.

A study of hormonal function before and after the expedition, using many of the same measures as **Chapter 3,** found preserved function of the hypothalamic-pituitary-gonad (HPG) and hypothalamic-pituitary-adrenal (HPA) axes, while fasting markers suggested evidence of metabolic benefit from the expedition, despite around 10 kg weight loss (**Section 4.1**). Weight loss comprised fat mass not lean mass, which was putatively important for the observed preservation of hormonal function. Since mitigation factors for the expedition were of interest to the training of WGCC, these were explored in a retrospective study of physical changes and dietary preparation during before the expedition (**Section 4.2**).

4.1 Female endocrine adaptation to an Antarctic traverse expedition

Section 4.1 was published in Medicine and Science in Sports and Exercise under the title "Female Reproductive, Adrenal, and Metabolic Changes during an Antarctic Traverse" by Dr Robert M Gifford (RG), Dr Thomas J O'Leary (TO), Ms Rinn Cobb (RC), Dr Jodie Blackadder-Weinstein (JB), Miss Rebecca Double (RD), Dr Sophie Wardle (SW), Prof Richard A Anderson (RA), Dr Doug Thake (DT), Dr John Hattersley (JH), Prof Christopher Imray (CI), Prof Adrian Wilson (AW), Prof Julie P Greeves (JG), Prof Rebecca M Reynolds (RR) and Prof David R Woods (DW) (Gifford et al, 2019a). RG sought ethical approval, designed the study and recruited the participants with supervision from DW, RR, JG and RA. RG conducted measurements with the assistance of Ms Jo Singleton (research nurse at the Wellcome Trust Clinical Research Facility, Edinburgh), except for imaging studies which were conducted by TO, SW and RD, and through-expedition questionnaires which were collected by JB. RG conducted statistical analysis, wrote the first draft of

the manuscript and responded to reviewers' comments. All authors provided editorial oversight of the final manuscript.

4.1.1 Abstract

Purpose: To explore the effects of the first all-female transantarctic expedition on hormonal axes pertinent to reproductive and metabolic function.

Methods: Six females (age 28–36 y; body mass index, 24.2 ±0.97 kg/m^2 hauled 80 kg sledges 1700 km in 61 d. Estimated average energy intake was 20.8 ±0.1 MJ/d (4970 ±25 kcal/d). Whole and regional body composition was measured by dual-energy x-ray absorptiometry 1 and 2 months before and 15 d after the expedition. Body fat was also estimated by skinfold and bioimpedance immediately before and after the expedition. Basal metabolic and endocrine blood markers and, after 0.25 mg dexamethasone suppression, 1-h 10.0 µg gonadorelin and 1.0 µg adrenocorticotropin-(1–24) tests were completed, 39–38 d pre-expedition and 4 to 5 d and 15 to 16 d post expedition. Cortisol was assessed in hair (monthly average concentrations) and saliva (five-point day curves and two-point diurnal sampling).

Results: Average body mass loss was 9.37 ±2.31 kg ($p<0.0001$), comprising fat mass only; total lean mass was maintained. Basal sex steroids, corticosteroids, and metabolic markers were largely unaffected by the expedition except leptin, which decreased during the expedition and recovered after 15 d, a proportionately greater change than body fat. Luteinizing hormone reactivity was suppressed before and during the expedition, but recovered after 15 d, whereas follicle-stimulating hormone did not change during or after the expedition. Cortisol reactivity did not change during or after the expedition. Basal (suppressed) cortisol was 73.25 ±45.23 mmol/L before, 61.66 ±33.11 mmol/L 5 d post expedition and 54.43 ±28.60 mmol/L 16 d post expedition ($P = 0.7$). Hair cortisol was elevated during the expedition.

Conclusions: Maintenance of reproductive and hypothalamic-pituitary-adrenal axis function in women after an extreme physical endeavour, despite energy deficiency, suggests high female biological capacity for extreme endurance exercise.

4.1.2 Introduction

Women undertake increasingly physically demanding sports and employment but sex-related biological consequences of arduous exercise are poorly understood.

Over the past 20 years, emphasis on energy availability (EA, defined as energy intake minus exercise energy expenditure) has established low EA as a putative cause of the 'female athlete triad': hypothalamic pituitary gonad (HPG) axis suppression in athletes, leading to functional hypothalamic amenorrhoea (FHA) and/ or impaired bone health (**Appendix A**). The term 'female athlete triad' has been questioned, since these phenomena can also affect men (Mountjoy et al., 2014), however women may have greater sensitivity to the effects of low EA than men, and there is a higher prevalence of disordered eating among women than men (Gibbs et al., 2013).

In the setting of military employment, it has been suggested that women may be at higher risk of psychological problems than men, such as post-traumatic stress disorder (**Appendix A**; Yehuda, 2002). There appears to be evidence suggesting a greater incidence of primary infertility in military women than age-matched civilians (Department of Manning (Army), 2016). While these observations remain unexplored in terms of aetiology, we recently proposed FHA in military women might contribute to menstrual dysfunction, hypothesizing this could be mediated by a complex alteration in hormonal milieu, including reduced EA (**Appendix A**). Aspects of military training and employment other than exercise and reduced EA may also be likely to contribute to HPG axis suppression, for example, sleep deprivation and psychological stress (**Appendix A**; Hoyt and Friedl, 2006; Nindl et al., 2007).

Field studies of military training generally measure the effects of multiple concurrent stressors, making it difficult to delineate the effects of individual components like low EA, sleep deprivation or psychological stress (Hoyt and Friedl, 2006). One highly researched model of the endocrine effects of a multi-stressor environment is US Army Ranger Training. Predominantly undertaken by men, Ranger training involves 61 days of strenuous exercise, sleep deprivation, total energy expenditure of around 4000-5000kCal/ day, routine energy deficit and widespread metabolic and hormonal deficiencies, e.g. elevated fasting cortisol, reduced total testosterone and IGF-1 (Henning et al., 2014; Nindl et al., 2007). Such changes have been demonstrated to be reversible upon re-feeding, cessation of stress and sleep derestriction (Henning et al., 2014). However, extremes of arduous exertion lasting this duration have not been widely researched in women.

We undertook an exploratory, observational study of the concurrent acute response and short-term recovery of female HPG and HPA axes (using basal and dynamic testing) in women undertaking an unprecedented, extremely arduous expedition to cross the Antarctic continental landmass, of similar duration to US Army Ranger training. The purpose of the crossing was to attempt to become the first all-female team to complete an unassisted Antarctic traverse using muscle power alone, and was not competitive, primarily research-focused or done to achieve a political or military training objective. The *a priori* hypothesis was that this expedition would induce an energy deficit, despite a comprehensive programme of physical and nutritional preparation, with concurrent disturbances in HPG and HPA axes.

4.1.3 Methods

4.1.3.1 Participants

Six women participating in an unassisted Antarctic ski traverse expedition were invited to participate in the study three months beforehand. This was the first all-female team to attempt an unassisted Antarctic traverse. Individuals planned to haul sledges weighing 80kg for 1700 km, expecting the crossing to take around 75 days. Selection and training for the expedition lasted 2 years, the final team being selected from a pool of 250 women. While none of the participants had been to Antarctica before, all had partaken in three preparatory expeditions in Norway, which aimed to simulate the crossing's intensity and conditions (details can be found at http://exicemaiden.com/). Participation in the study was voluntary and independent of the expedition. All six women volunteered and provided written informed consent. Ethical approval was received form the Ministry of Defence Research Ethics Committee (827MoDREC/17). The study was conducted in accordance with the Declaration of Helsinki.

4.1.4 Experimental design

The study design consisted of two pre-expedition measurement sessions, 64 and 39 days prior to the expedition (visit pre-1 and pre-2, respectively) (**Figure 4-1-1A**). Additional body composition measurements were undertaken separately from formal study visits, 16 days before and 5 days after the expedition. Follow-up visits were conducted 4 days after the expedition (immediately after arrival in Punta Arenas, Chile from Antarctica), and 15 days after the expedition, 36 hours after return to the UK (visits post-1 and post-2 respectively). As part of a broader preparation schedule, participants were advised to gain 0.5 kg of body mass per week between

visit pre-1 and the expedition (64 days or 9 weeks; 4.5kg). The expedition altitude profile and distance are indicated in **Figure 4-1-1B**. The maximum elevation above sea level (ASL) was 2950 m.

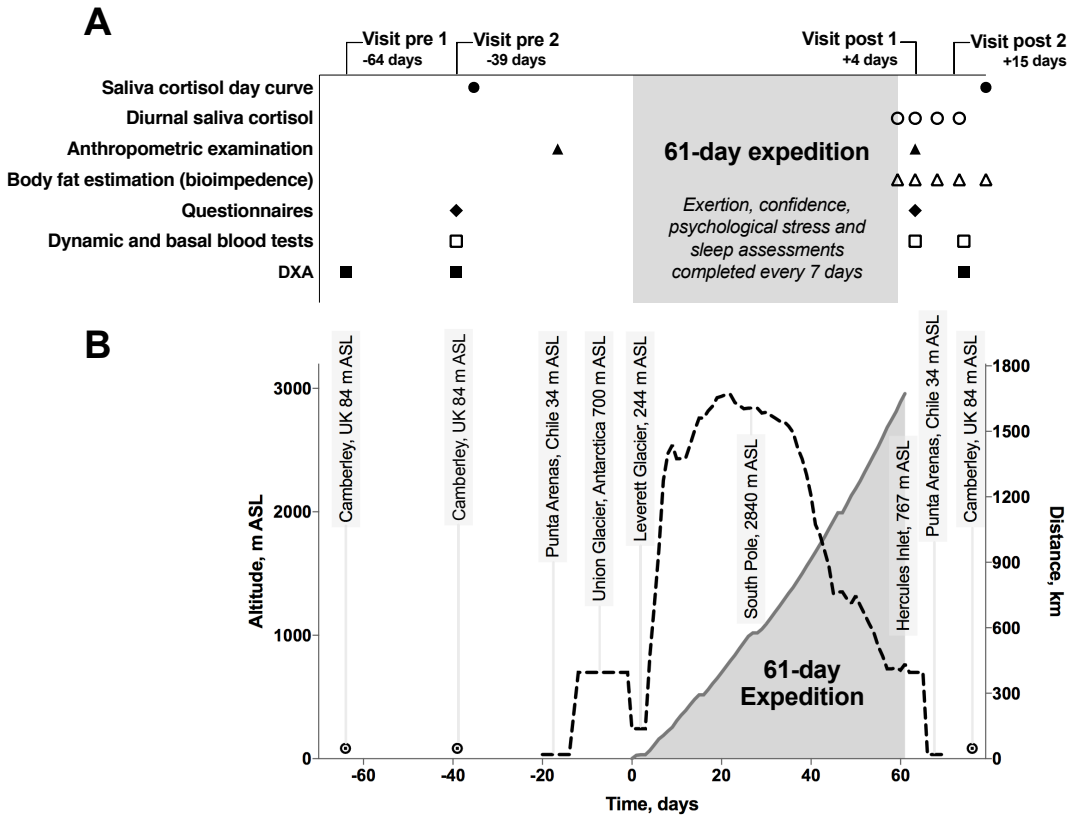


Figure 4-1-1. Overview of experimental design and expedition. A. Timeline summary of major investigations. Saliva cortisol: 5-point day curve was measured 40-34 days pre- and 18-24 days post-expedition (filled circle); morning and evening sampling undertaken 1, 5 and 10 days post-expedition (unfilled circle). Anthropometric examination: weight and skinfolds were undertaken 16 days pre- and 5 days post-expedition (filled triangle). Body fat was estimated by bioimpedance 1,5,10 and 20 days post-expedition (unfilled triangle). Questionnaires were undertaken 39 days pre- and 5 days post-expedition (diamond). Dynamic and basal blood tests: fasted blood sampling and dexamethasone-suppressed combined GnRH and ACTH-(1-24) test, 39 days pre, and 4-5 and 15-16 days post-expedition (unfilled square). Body composition measured by DXA scan 64 and 39 days pre and 15 days post-expedition (filled square). B: Altitude profile of study. Dashed line: altitude. Solid line: elapsed ski distance. Target icons indicate altitude of study visits in Camberley, UK. GnRH, gonadotrophin releasing hormone; ACTH adrenocorticotrophic hormone, DXA dual-energy x-ray absorptiometry, ASL above sea level

4.1.4.1 Dietary provision

Dietary provision for the expedition was estimated from changes in body mass during three training expeditions. During the expedition participants were provided with a complete diet providing average 20.9 ±0.1 MJ per day (4970 ±25 kcal per day, or 70.8 ±0.35 kcal/kg/day), comprising ~45 % carbohydrate (7.7 ±0.32 g/kg/day), ~45 % fat (3.6 ±0.07 g/kg/day) and ~10 % protein (1.7 ±0.35 g/kg/day). It is estimated (verbal communication) that participants consumed median 85 % (range 70 % - 99 %) of the diet provided over the course of the expedition and did not share rations.

4.1.4.2 Procedures

The schedule of measurements is illustrated in **Figure 4-1-1**. At visit pre-1, information including ethnicity, education, smoking habits, alcohol consumption, and a comprehensive medical reproductive and medication history taken including use and type of, and indication for, hormonal contraceptives was recorded. Reproductive and medication history and use of contraceptive questions were repeated after the expedition (visit post-1).

4.1.4.2.1 Psychological assessment

Questionnaires comprising six validated self-rating items on a web-based application (SmartSurvey, Tewkesbury, UK) were completed at visits pre-2 and post-1 (**Figure 4-1-1**). The psychosocial stress questionnaire was completed in an identical manner to Rosengren *et al*, assessing the six-month period prior to visit pre-2, and the four-month period prior to visit post-1 (Rosengren et al., 2004). Participants were asked to complete the Impact of Events Scale – Revised (IES-R) with reference to any major life event(s) identified (Weiss, 1997). The Patient Health Questionnaire 9 (PHQ-9) (Kroenke et al., 2001) was chosen as a robust measure of depressive symptoms in military and civilian populations (Wells et al., 2013). We analysed results on a continuous scale, to identify subtle differences in a low number of participants. The Beck Anxiety Inventory (BAI) and Connor Davidson Resilience Scale 10 (CDRISC 10) demonstrate similar consistency measuring anxiety and resilience, respectively, and were analysed in the same manner (Campbell-Sills and Stein, 2007; Johnson et al., 2011). The BEDA-Q assesses risk of disordered eating concisely and consistently (Martinsen et al., 2014), and was scored according to the methods of Peric *et al.* (2016). Total scores from each questionnaire were used for further analysis.

4.1.4.2.2 Weekly intra-expedition assessments

During the expedition, a weekly questionnaire was completed in the same manner as previous studies of female transantarctic expedition (**Figure 4-1-1**) (Atlis et al., 2004; Kahn and Leon, 1994). This documented average perceived exertion, psychological stress, restfulness of sleep and confidence the team would complete the expedition (all on a Likert type-scale ranging from 1 [not at all] to 10 [the most possible]), and the average number of hours slept per night.

4.1.4.2.3 Body composition

Stature was measured at visit pre-1 (SECA Stadiometer 213, Birmingham, UK) and body mass was measured at every study visit (SECA Scales 874). Whole body and regional lean mass, fat mass and bone mineral content were measured using dual energy x-ray absorptiometry (DXA) was measured with participants wearing shorts and t-shirts at visits pre-1, pre-2 and post-2 (GE Lunar iDXA, GE Healthcare, Chalfont St Giles, UK) (**Figure 4-1-1**).

Sixteen days prior to the expedition (separately from main study visits), and at visit post-1, skinfolds were measured at four sites (bicep, triceps, sub-scapular, supraspinatus) to the nearest mm by the same examiner using Harpenden callipers (BodyCare, UK) according to the method of International Society for the Advancement of Kinanthropometry (Stewart et al., 2011). The average of three measurements taken from each site was used to calculate percentage body fat (Stewart et al., 2011).

Body fat was measured by four-point bioimpedance (Omron BF511, Milton Keynes, UK) upon waking in the morning, 1, 5, 10, 15 and 18-24 days after the expedition.

4.1.4.2.4 Basal blood samples

After an overnight fast, a venous blood sample was collected at visits pre-2, post-1 and post-2 for measurements of metabolic, nutritional, reproductive and adrenal function.

4.1.4.2.5 Dynamic reproductive and adrenal cortex function

Dynamic reproductive and adrenal cortex function was measured at visits pre-2, post-1 and post-2. Participants first ingested 0.25 mg dexamethasone at 2200h before a second overnight fast. This dose has been used to assess the sensitivity of the HPA axis to a near-physiological level of central negative feedback and to

attempt to reduce the baseline variation in morning fasting cortisol prior to the pre-stimulation test cortisol (Kajantie et al., 2003; Reynolds et al., 2001a). At 0800 the following morning, a 21-gauge cannula was inserted into an antecubital or dorsal hand vein and a baseline blood sample was obtained before 10 µg Gonadorelin hydrochloride (Intrapharm, Maidenhead UK), followed by 1.0 µg ACTH-(1-24) (tetracosactrin acetate as ®, Mallinckrodt, Dublin, Ireland), were injected followed by a 10 mL saline flush. ACTH-(1-24) was freshly diluted using 249 mL 0.9% NaCl (Baxter, UK), to which ® 250 µg in 1mL had been added, shaken thoroughly and 1 mL of this mixture was injected using a 5 mL syringe to minimise contact with plastic. Venous blood was sampled through the cannula in EDTA-containing tubes 20, 30, 40 and 60 min after drug administration. The doses of Gonadorelin, dexamethasone and ACTH-(1-24) were selected to mimic physiological levels of stimulation, as opposed to stimulation tests used clinically (and recommended in various clinical practice guidelines) which are intended to induce maximal axis stimulation and exclude endocrine insufficiency (e.g. 100 µg, 1 mg and 250 µg, respectively) (Reynolds et al., 2001a).

4.1.4.2.6 Hair and saliva cortisol

A 0.5cm diameter hair sample was taken close to the scalp for measurement of cortisol at visit pre-2 (6 x 1cm segments) and visit post-1 (4 x 1cm segments). Hair grows at 1cm per month, thus 1cm represents 1 month of cortisol exposure (Stalder and Kirschbaum, 2012).

Saliva was sampled by chewing on a synthetic swab for 1 minute, which was placed in a plastic collection tube (Salivette®; Sarstedt, Nümbrecht, Germany). A detailed saliva day curve was measured at visits pre-2 and post-2 as follows: participants were woken at 07:00 and saliva sampled at 07:10, 08:20, 09:00, 09:30, 12:15, 13:30, 17:20 and 21:50. Evening and morning saliva sampling (last thing at night before going to sleep and immediately after waking the following morning) were also measured 1, 5 and 10 days after the expedition.

4.1.4.3 Laboratory methods

Blood was collected in EDTA, serum-separating gel and fluoride oxalate tubes (Monovette®, Sarstedt, Nümbrecht, Germany) and centrifuged at 5,000 rpm for 5 minutes. Plasma and serum were stored at −80°C (after dry ice shipment to the UK of samples taken in Chile) until measurement.

4.1.4.3.1 Metabolic and nutritional markers

Thyroid stimulating hormone (TSH), unbound thyroxine (fT4) and total T3 (tT3) were measured from gel-separated serum using Abbott ® Architect analyser (Abbott, Maidenhead, UK) according to manufacturer's instructions. Insulin-like growth factor 1 (IGF-1), ferritin, insulin and C-peptide were determined from gel-separated serum using Roche ® Cobas e411 analyser (Roche Diagnostics, Welwyn Garden City, UK) according to manufacturer's instructions. Creatinine, albumin, transferrin, calcium, zinc, iron and magnesium were determined from gel-separated serum and glucose and lactate from plasma containing fluoride oxalate using commercial kits (Alpha Laboratories, Eastleigh, UK) adapted for use on a Cobas Fara centrifugal analyser (Roche, UK). Leptin was measured by ELISA (Quantikine, USA). Urea was determined from gel-separated serum using a commercial kit (Randox laboratories, UK) adapted for use on a Cobas Fara centrifugal analyser. Homeostatic modelling assessment 2 (HOMA2) for beta cell function (HOMA-B) insulin sensitivity (HOMA-S) and insulin resistance (HOMA-IR) were calculated according to the methods of Levy *et al*. (1998).

Additional data including resting energy expenditure and substrate utilisation from direct calorimetry pre- and post-expedition are being published elsewhere.

4.1.4.3.2 Reproductive markers

Luteinizing hormone (LH), follicle stimulating hormone (FSH), progesterone and oestradiol were determined from plasma containing EDTA using Abbot Architect ® analyser according to the manufacturer's instructions. Inhibin B was measured by ELISA (Beckman Coulter, High Wycombe, UK). Sex hormone binding globulin (SHBG) and anti-müllerian hormone (AMH) were determined from gel-separated serum using Roche ® Cobas e411 analyser according to manufacturer's instructions. The rationale for these methods are summarised in Appendix Box 1.

4.1.4.3.3 Adrenal markers

Cortisol, 17OH progesterone, testosterone, dihydroepiandrostenedione (DHEA) and androstenedione were measured using liquid chromatography mass spectrometry (LC/ MS), by modifying internal standards from a protocol described previously (Stirrat et al., 2018). Hair was divided into 1cm segments and powdered prior to cortisol extraction in each segment, representing 1 month averages, for a total of 10 months. Extraction and analysis by LC/ MS was completed as described by

Kirschbaum *et al*. (2009). Saliva was stored at -80 C within 7 days of collection and was extracted and analysed by LC/ MS as described by Miller *et al. (2013).*

Inter-assay %CV was <4% for Architect ®, e411, Fara assays and blood gas analyser, and intra-assay %CV <10% for all ELISAs.

4.1.4.4 Statistical Analysis

Data are presented as individual data, or mean ±SD or median (IQR) for group comparison. Normality was assessed using Shapiro-Wilk test and non-normally distributed data were log transformed prior to statistical analysis. Due to the small sample size, variables are presented as mean (95% confidence interval [CI]). Repeated measures ANOVA was used to compare change in variables over time and pairwise comparisons were used where appropriate for statistically significant results. Paired t tests were used to compare the two pre-expedition DXA scans, and single post-expedition variables with baseline. Pre- and post-expedition dichotomous questionnaire data were compared using Chi squared test. One individual was excluded from analyses of basal reproductive hormones as she had commenced a combined contraceptive pill immediately prior to the expedition. Serum LH and FSH concentrations following injection of GnRH and ACTH were described as absolute values, and as percentage change, by dividing concentrations after injection by the baseline concentration. This was done to allow comparison of within-subject change, since hormone-containing contraceptive use influenced baseline values. Area under the curve (AUC) was calculated using the trapezoidal rule. Within-subject changes in peak and AUC of cortisol and fold-rise in LH and FSH from baseline were compared from before to after the expedition.

Statistical analysis was performed using SPSS version 23.0 for Mac (IBM, USA). Significance was set at $p<0.05$. For multiple variables assessed in the same domain, Bonferroni adjustment was made as follows: body composition, $p<0.01$; basal reproductive markers, $p<0.005$, adrenal markers $p<0.05$, metabolic markers, $p<0.002$.

4.1.5 Results

4.1.5.1 Description of participants

Baseline characteristics of the cohort are shown in **Table 4-1-1**. The median (range) age was 32.8 (28.6 to 36.1) years. Baseline questionnaires demonstrated high

resilience, low depression and anxiety scores and normal patterns of eating behaviour. Fasting TSH, free T4, total T3, prolactin, LH:FSH ratio, androstenedione, total testosterone, DHEA, 17-OH progesterone, urea, sodium, potassium and creatinine were within normal limits prior to the expedition.

Age, years; median (range)	32.7 (28.6 to 36.1)
Reproductive characteristics	
Age at menarche, years; median (range)	13 (11-16)
Medical suppression of menstruation Levonorgestel 20mcg per 24h intrauterine device (Mirena ®) only – 4 (67%) Mirena ® plus ethinylestradiol 30 mcg/ levonorgestrel 50 mcg – 1* (17%) 68 mg subcutaneous implant (Nexplanon ®) – 1 (17%)	
Body composition	
Body mass, kg, mean (SD)	72.8 (4.00)
BMI, kg/m^2, mean, (SD)	24.2 (0.97)
% fat by DXA, kg, mean (SD)	20.92 (2.12)
Lean mass by DXA, kg, mean (SD)	53.5 (3.06)
Psychological assessments	
Several periods of psychological stress, mean (SD)	5 (83)
Permanent, psychosocial stress, n (%)	0
Some periods of psychological stress, n (%)	1 (17)
Never experienced psychological stress, n (%)	0
One or more adverse events, n (%)	4 (67)
High or severe financial stress, n (%)	0
IES-R, median (range)	36 (9 – 52)**
PHQ-9, median (range)	3 (0 – 11)
BAI, median (range)	11 (2 – 15)
CDRISC 10, median (range)	34 (31 – 36)
BEDA-Q score, Median (range)	4 (0-6)
BEDA-Q “Are you trying to lose weight now?” =yes	0
BEDA-Q “Have you ever tried to lose weight?”=yes If so, number of times (n [%])	3 (50) 3-5 (2 [33]) >5 (1 [17])

Table 4-1-1. Participant characteristics at baseline. Data are mean (SD) unless otherwise stated. BEDA-Q brief eating disorders in athletes questionnaire. IES-R Impact of events scale (revised), PHQ-9 adjusted patient health questionnaire 9, BAI Beck Anxiety Inventory, CDRISC10 Connor Davidson Resilience Scale 10, N/A not applicable * One participant using Mirena ® also commenced ethinylestradiol 30 mcg/ levonorgestrel 50 mcg once daily immediately prior to expedition until after testing was completed. ** Applies to four subjects who experienced a significant event

All participants used hormonal contraceptives during the expedition, intending to induce amenorrhoea. One individual commenced levonorgestrel 150 mcg/ ethinylestradiol 30 mcg immediately prior to the expedition. One individual used Nexplanon ® contraceptive implant while all others used a Mirena ® intrauterine device. Five participants were amenorrhoeic during the expedition and one menstruated twice, stating this was less frequent than normal, within 4-10 days of due date.

4.1.5.2 Intra-expedition rating scales

Average scores for physical exertion scale were 5.5 ±2.3 /10 and stress level 3.7 ±1.94 /10, and level of confidence the team would complete the expedition 6.73 ±1.81)/10; (6.35 ±1.93 in weeks 1-3 and 7.11 ±1.32 in weeks 5-8, p=0.09). Average duration of sleep was 6.73 ±1.75 hours and rating of restfulness of sleep was 5.53 ±2.05 /10. Questionnaires following the expedition suggested moderately lower levels of psychosocial stress and financial stress, and fewer significant adverse events than prior to the expedition (p=0.079, **Appendix Table 4-1-2**).

4.1.5.3 Body composition, metabolic and nutritional changes

Physical changes during the study are presented in **Figure 4-1-2** and Appendix **Tables 4-1-1** and **4-1-2**. All participants gained body mass during the two months prior to the expedition, (average increase 2.56 ±0.79 kg, or 3.69 ±1.12 % of body weight, p=0.006), consisting of body fat (average increase 4.05 ±0.96 %, p<0.0001), and lost body mass during the expedition (average loss 9.37 ±2.31 kg, or 12.9 ±3.17 % of body weight, p<0.0001). Body composition measured by DXA demonstrated a significant increase in total fat mass before (13.2 ±2.11 vs 17.5 ±2.52 kg, p<0.001) and loss during the expedition (fat mass at visit post-2 was 12.1 ±1.37 kg, p<0.001), with these changes reflected in most regions (**Appendix Table 4-1-2**). However, there was no difference in total lean mass or bone mineral content between visit pre-2 and visit post-2 (52.3 ±2.10 vs 51.5 ±3.04, p=0.27), despite a 6.10% loss in lean

mass from the legs. In the 15 days between the expedition and the follow-up DXA scan, fat mass estimated by bioimpedance tended to increase (**Appendix Table 4-1-1**). Regional DXA analysis showed statistically significant but modest decreases in android (area between the ribs and pelvis), gynoid (pelvis and upper thighs) and leg lean mass between visits pre-1 and pre-2, and loss of leg lean mass during the expedition (average 6.05 ±1.11 % decrease), but these did not impact the change in total lean mass (**Appendix Table 4-1-2**). There was a small but statistically significant increase in total bone mineral content prior to the expedition (2.75 ±0.13 kg vs 2.80 ±0.13) kg, p=0.005, but no change between visits pre- 2 and post- 2 (2.77 ±0.12, p=0.19 (**Appendix Table 4-1-2**).

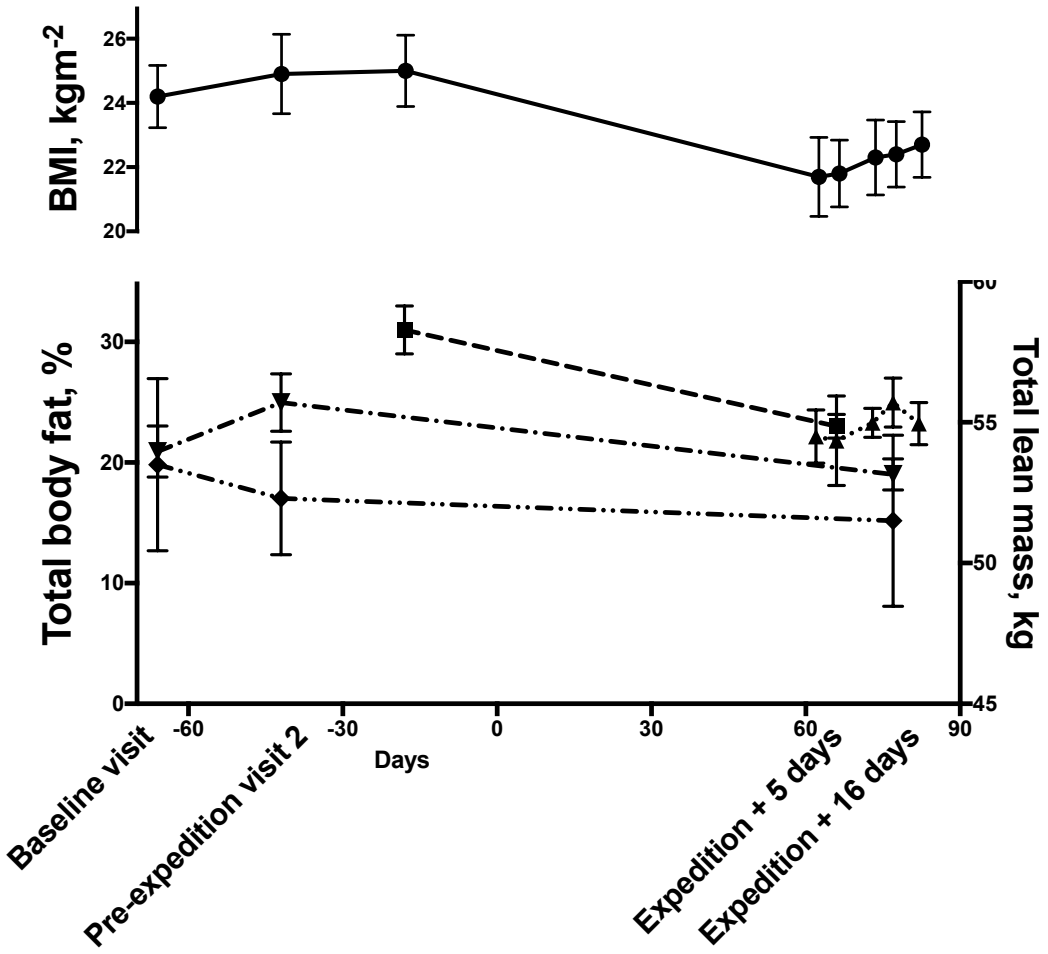


Figure 4-1-2. Anthropometric changes during the expedition. Data are mean ±SD. Shaded rectangle: Duration of expedition. Circle with solid line: BMI. Square with dashed: total body fat (%) by skinfold. Triangle pointing upwards with dash-dot-dot line: total body fat (%) by bio-electrical impedance. Triangle pointing downwards with dotted line: Total body fat (%) by dual energy x-ray

absorptiometry. Diamond with dash-dot line: total lean mass (kg) by dual energy x-ray absorptiometry. BMI, body mass index

Leptin decreased significantly following the expedition, thereafter, increasing two-fold from visits post-1 to post-2 (**Table 4-1-2**). Post-hoc tests showed the change between visit pre-2 and post-1 was significant (p=0.005), while there was no difference between pre-2 and post-2 (p=0.39). Thyroid stimulating hormone, free T4 and total T3 were normal pre-expedition and remained unchanged after the expedition (**Table 4-1-2**). Fasted glucose, HOMA-B, HOMA-S and HOMA-IR, adjusted calcium, magnesium and phosphate did not change during or after the expedition (**Table 4-1-2**).

	Visit pre 2	Visit post 1	Visit post 2	Mean (95% CI) difference visit pre 2 vs post 1	Mean (95% CI) difference visit pre 2 vs post 2	p
Oestradiol, pmol/L	227 (176)	163 (144)	394 (183)	64.3 (-156, 284)	-168 (-427, 92)	0.043
LH, IU/L	5.36 (2.03)	5.13 (3.70)	3.42 (1.43)	0.23 (-4.15, 4.61)	1.94 (-1.02, 4.91)	0.3
FSH, IU/L	5.83 (1.09)	5.30 (1.36)	3.50 (0.48)	0.53 (-1.47, 2.53)	2.33 (-0.82, 5.48)	0.16
Androstenedione, nmol/L	9.56 (2.98)	7.33 (2.61)	8.91 (2.78)	2.23 (0.45, 4.01)	0.65 (-2.92, 4.22)	0.148
Testosterone, nmol/L	1.46 (1.56)	0.59 (0.81)	0.54 (0.4)	0.87 (-0.8, 2.55)	0.92 (-0.39, 2.22)	0.18
DHT, nmol/L	2.93 (2.12)	2.18 (1.24)	1.33 (0.78)	0.75 (-0.5, 1.99)	1.6 (-1.09, 4.28)	0.18
DHEA, nmol/L	360.24 (85.34)	370.37 (46.32)	412 (42.3)	-10.1 (-81.9, 61.6)	-52.4 (-114, 9.62)	0.13
17-OH P, nmol/L	3.81 (5.04)	1.54 (1.51)	7.97 (4.53)	2.27 (-2.18, 6.72)	-4.15 (-13.1, 4.74)	0.071
SHBG, nmol/L	59.5 (25.4)	100 (53.0)	69.0 (25.0)	-40.7 (24.6, -104)	13.8 (-44.9, 25.9)	0.13
Prolactin, mU/L	338 (45.0)					N/A
LH:FSH ratio	1.05 (0.22)	0.93 (0.32)	1.32 (0.57)	0.12 (-1.10, 1.34)	-0.26 (-2.12, 1.60)	0.8
AMH, pmol/L	12.2 (3.85)	9.44 (2.61)		2.79 (-1.38, 6.97)		0.15
Unsuppressed cortisol, mmol/L	552 (67.3)	434 (74.2)	519 (19.4)	117 (-123, 358)	32.2 (-122, 186)	0.3
Suppressed cortisol, mmol/L	73.3 (45.2)	61.7 (33.1)	54.4 (11.7)	11.6 (-28.1, 51.3)	18.8 (-5.21, 42.9)	0.3

Albumin, g/L	35.4 (2.06)	33.1 (0.84)	34.9 (1.58)	0.94 (-0.14, 4.71)	0.92 (-1.88, 2.84)	0.075
Glucose, mmol/L	4.93 (0.62	4.57 (0.25)	4.42 (0.75)	0.37 (-0.2, 0.97)	0.52 (-0.2, 1.19)	0.13
HOMA %B	96.4 (27.8) 113 (31.3)	114 (39.6)	118 (63.0)	15.8 (-59.1, 22.5)	21.6 (-77.1, 34.0)	0.6
HOMA %S	0.94 (0.24)	130 (82.2)	140 (39.5)	42.8 (-127, 92.8)	18.5 (-75.3. 19.8)	0.7
HOMA IR		1.00 (0.46)	0.76 (0.22)	0.26 (-0.74, 0.62)	0.10 (-0.08, 0.43)	0.5
Leptin, ng/mL	10.8 (4.84)	2.71 (1.57)	4.93 (3.58)	8.09 (3.64, 12.5)	5.87 (1.04, 10.7)	0.002*
IGF-1, ng/mL	46.0 (18.8)	29.1 (11.9)	46.0 (18.8)	33.1 (-44.4, 110)	-19.5 (-79.3, 40.3)	0.12
Iron, µmol/L	23.3 (6.05)	28.3 (8.87)	19.0 (4.29)	2.50 (-11.4, 1.43)	1.76 (-0.20, 8.87)	0.091
Ferritin, ng/mL	59.7 (22.8)	55.3 (35.3)	55.2 (26.7)	4.33 (-11.1, 19.8)	4.50 (-14.6, 23.6)	0.8
TSH, mU/L	2.51 (0.57)	3.53 (1.84)		-1.02 (-2.69, 0.66		0.18
Total T3, nmol/L	1.43 (0.08)	1.42 (0.09)	1.38 (0.07)	-0.17 (-0.40, 0.37)	0.05 (-0.20, 0.30)	0.9
Free T4, pmol/L	12.5 (0.34)	11.8 (0.54)		0.67 (-0.42, 1.75)		0.18
Zinc, µg/dL	134 (6.15)	122 (12.3)	137 (15.7)	5.60 (-1.91, 26.9)	4.31 (-13.7, 8.41)	0.041
Sodium, mmol/L	141 (10.2)	143 (8.0)	139 (5.09)	-2.67 (-15.1, 9.73)	1.17 (-9.5, 11.84)	0.8
Potassium, mmol/L	4.27 (0.34)	4.45 (0.24)	4.08 (0.330	-0.18 (-0.4, 0.05)	0.18 (-0.4, 0.74)	0.14
Magnesium, mmol/L	0.79 (0.06)	0.81 (0.04)	0.80 (0.73)	0.02 (-0.06, 0.17)	0.01 (-0.04, 0.25)	0.4
Creatinine, µmol/L	63.7 (5.89(	63.3 (3.98)	69.0 (5.66)	0.33 (-4.8, 5.42)	-5.33 (-9.9, -0.8)	0.023
Creatine kinase, U/L	130 (16.9)	137 (33.2)	153 (71.8)	77.3 (-145.6, 300)	-118 (-371, 134)	0.4
Lactate, mmol/L	0.75 (0.17)	0.85 (0.25)	0.57 (0.12)	-0.1 (-0.4, 0.17)	0.04 (-0.4, 0.44)	0.5
Calcium (adjusted), mmol/L	2.56 (0.08)	2.55 (0.06)	2.56 (0.04)	0.01 (-0.06, 0.09)	-0.01 (-0.11, -0.01)	0.8
25 OH D, nmol/L	112 (25.3)	75.8 (21.3)		35.8 (14.3, 57.4)		0.008*
Phosphate, mmol/L	1.19 (0.05)	1.08 (0.04)	1.34 (0.09)	0.07 (-0.06, 0.29)	0.01 (-0.35, 0.03)	0.017

Table 4-1-2. Biochemical and hormonal parameters at baseline, 4 and 14 days after the expedition. Values are mean (SD) unless otherwise stated. 17 OH-P: 17 hydroxyprogesterone, AMH, antimüllerian hormone; DHEA, Dehydroepiandrosterone; DHT dihydrotestosterone, FSH, follicle stimulating hormone; HOMA, homeostatic modelling assessment ; IGF-1 insulin-like growth factor 1; %S insulin sensitivity; %B –beta cell function; -IR- insulin resistance; LH, Luteinizing hormone; T3, triiodothyronine; T4, thyroxine; TSH, thyroid stimulating hormone; 25 OH D 25 hydroxyl vitamin D. P value for repeated measures ANOVA. * denotes statistical significance after Bonferroni adjustment: p<0.005 for reproductive markers, p<0.05 for cortisol, p<0.002 for metabolic and nutritional markers.

Questionnaire data demonstrated a marginal increase in BEDA-Q scores after the expedition, consistent with higher markers of disordered eating risk (**Appendix Table 4-1-3**). Markers of nutritional status (albumin, magnesium, phosphate, iron, zinc), urea (Ln transformed) and electrolytes did not change during or after the expedition (**Table 4-1-2**).

4.1.5.4 Reproductive function

Basal markers of reproductive function are displayed in **Table 4-1-2**.Oestradiol tended to be lower at visit post-1, with a recovery noted by visit post-2. No differences between other sex steroids, LH or FSH were shown. Inhibin B and AMH did not differ between baseline and immediately after the expedition (p=0.71 and p=0.15, respectively, **Table 4-1-2**).

Dynamic LH and FSH responses before and after the expedition are shown in **Figure 4-1-3**. Fold rise in FSH and FSH AUC were log transformed prior to statistical analysis. LH and FSH fold rise and AUC during the test did not differ between visit pre-2 and visit post-1. At visit post-2, FSH had not changed from visit pre-1 (**Figure 4-1-3C**, Appendix **Table 4-1-4**), while there was a marked upward trend in LH, measured by AUC fold rise and peak fold rise (p=0.055 and p=0.071, respectively; **Figure 4-1-3D**, **Appendix Table 4-1-4**).

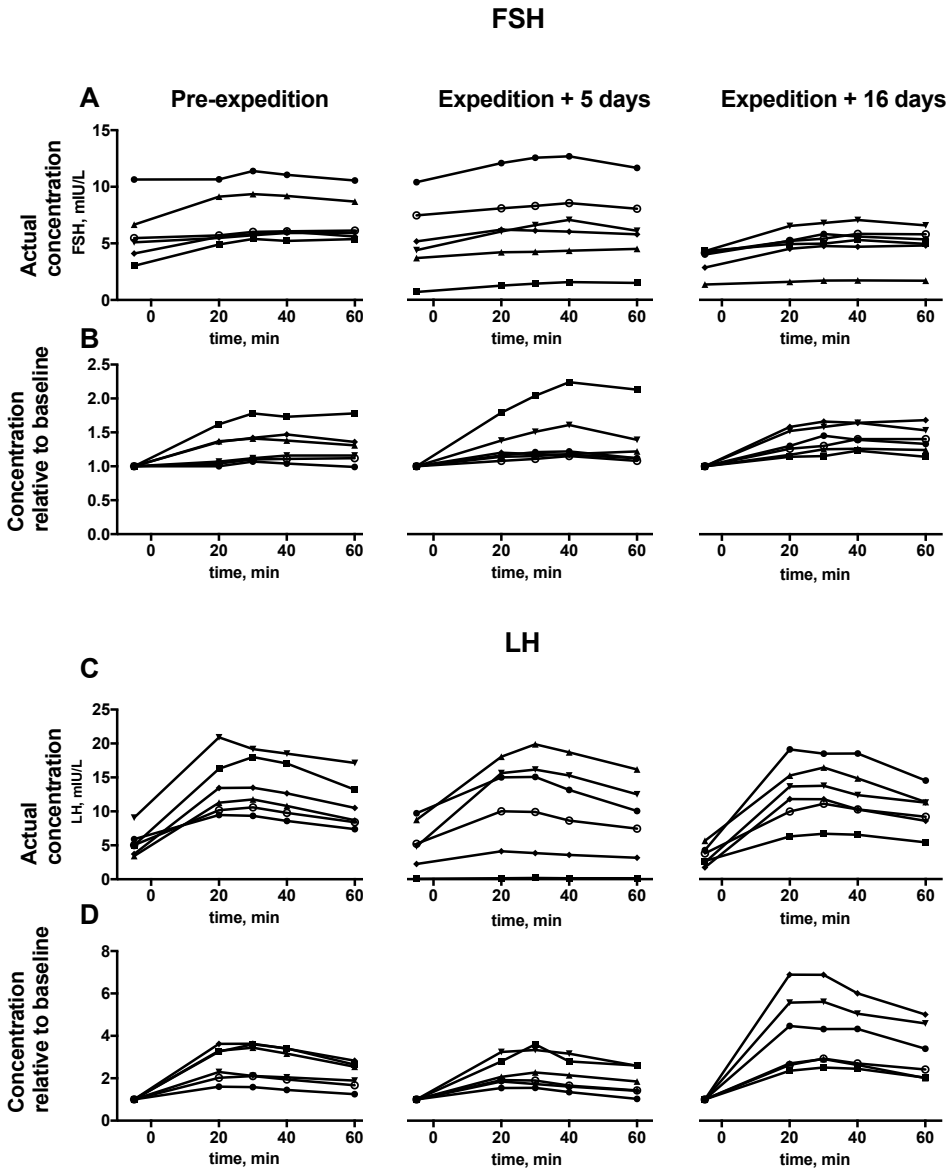


Figure 4-1-3. Dynamic gonadotrophin function before and after the expedition. Individuals represented by symbols. Actual concentrations (top row) and fold difference from baseline concentrations (middle row) after 10µg GnRH administration before, 5 and 16 days after the expedition for FSH (A) and LH (B). The bottom row shows change in AUC and peak concentrations following 10µg GnRH administration at the same 3 time points for FSH (C) and LH (D). FSH AUC fold rise and peak fold rise did not change across visits (p=0.71 and p=0.55, respectively). There was an upward trend in LH AUC fold rise and peak fold rise (p=0.055 and p=0.071, respectively). FSH, follicle stimulating hormone; LH, luteinising hormone; GnRH, gonadotrophin releasing hormone; AUC, area under the curve. One individual (filled square) commenced levonorgestrel 150 mcg/ ethinylestradiol 30 mcg immediately prior to the expedition. One individual (unfilled circle) used Nexplanon ® contraceptive implant while all others used a Mirena ® intrauterine device

4.1.5.5 Adrenal cortex function

Basal plasma cortisol did not change significantly during or after the expedition (**Table 4-1-2**).

Average hair cortisol before and during the expedition is shown in **Figure 4-1-4B**. Mean values are shown in **Appendix Table 4-1-4**. Most participants demonstrated a significant increase in average cortisol levels *during* the expedition.

Individuals' dynamic plasma cortisol responses before and after the expedition are shown in **Figure 4-1-4A**. Both AUC and peak cortisol did not change between the three time points (p=0.12 and p=0.45, respectively, **Figure 4-1-4B**). Subjects demonstrated marked suppression of early morning cortisol following low-dose dexamethasone administration

One participant demonstrated a more suppressed baseline in plasma cortisol than others (filled square symbol, **Figure 4-1-4**). This individual also demonstrated markedly higher hair cortisol concentration through the expedition and two months beforehand.

Salivary cortisol in the days immediately following the exercise was blunted but by day 10 had recovered (**Figure 4-1-4C**), reflected in a normal day curve which was unchanged from baseline (**Figure 4-1-4D**).

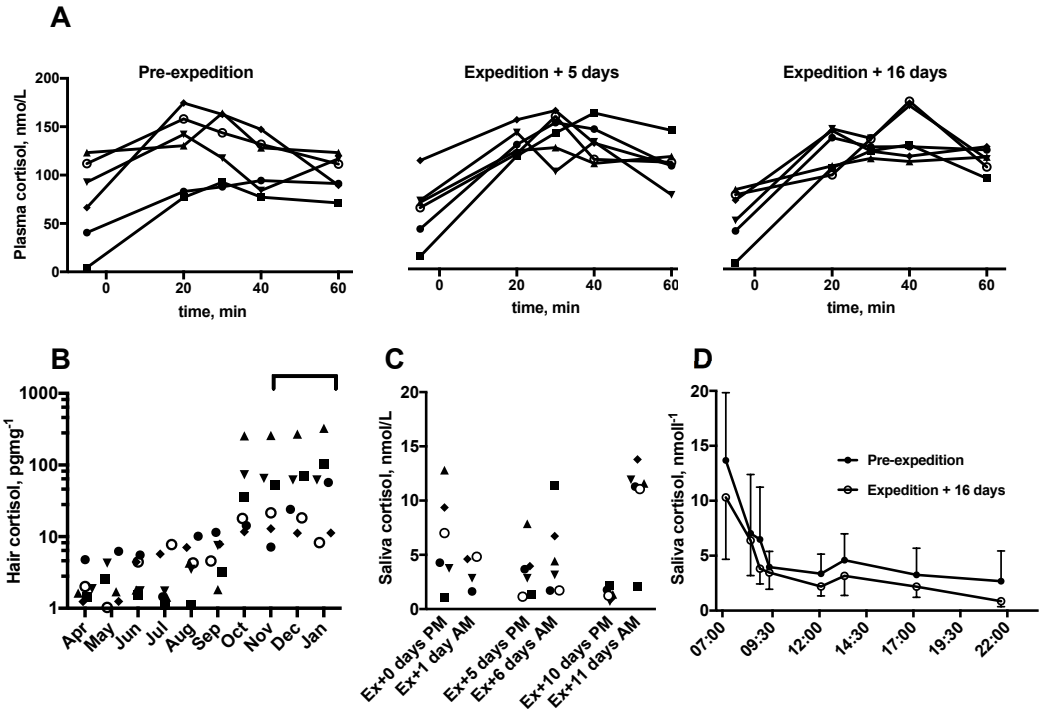


Figure 4-1-4. Dynamic, monthly average hair and diurnal saliva cortisol concentrations A: adrenal response to (1-24) adrenocorticotropin, 10 hours after central suppression with 0.25 mg dexamethasone, before, and 5 days and 16 days after the expedition. Top row: cortisol concentrations. Bottom row: fold difference in cortisol from baseline. Area under the curve and peak cortisol did not change between the three time points (p=0.12 and p=0.45, respectively, panel B). B: average monthly cortisol from 1cm hair segments prior to and during the expedition (expedition represented by bracket). C: change in AUC and peak concentrations during the dynamic test before, and 5 and 16 days after the expedition. D: Saliva cortisol 36-40 days pre-expedition and 18-24 days post expedition (left panel) and diurnal cortisol 1,4 and 10 days post-expedition (right panel).Individuals represented by symbols. Time: after ACTH-(1-24) administered. ACTH, adrenocorticotropin; F, cortisol.

4.1.6 Discussion

With on-going debate as to whether women can endure extreme physical activity without detrimental effects on hormonal axes, given the finding of HPA and HPG axis suppression in extremely arduous exercise in men (e.g. in US Army ranger training (Henning et al., 2014; Nindl et al., 2007)), we exploited the opportunity to examine the HPA and HPG axes among six women who completed a 1700km ski expedition hauling 80 kg sledges up to 2950 m elevation. In doing so, the team broke several records including being the first all-female team to cross the Antarctic unsupported. Our data demonstrate HPG and HPA axis resilience during extreme

exertion despite significant fat loss. HPA axis basal function, sensitivity to central suppression and adrenal reactivity to ACTH did not change during or after the expedition, but demonstrated greater sensitivity to suppression from dexamethasone than anticipated from other studies using a similar protocol in older participants (Kajantie et al., 2003; Reynolds et al., 2001a). Hair cortisol rose during the expedition as would be expected with sustained arduous exercise (Gerber et al., 2012).

Coincidentally, the expedition duration (61 days) was identical to US Army Ranger training. Trainee Rangers are expected to cover around 322 km, carrying 30-41 kg. While the expedition comprised a different form of exercise (skiing rather than walking or running), it was arguably non-inferior in terms of effort or endeavour. One crucial difference is the 0-5 hours of sleep per day expected during Ranger training (Nindl et al., 2007), and deliberate psychological stress (Friedl et al., 2000). Which contrasts with the average 6.73 ±1.75 hours of sleep per night, albeit with poor perception of restfulness (in 24-hour daylight), and modest weekly and whole-expedition stress ratings.

The primary drivers of adverse endocrine and metabolic changes in Ranger training appear to be nutritional deprivation (with loss of lean mass), psychological stress, sleep deprivation and exercise intensity. Nindl et al. (2007) showed a 12.6% loss of body mass, 6% lean mass and 50% fat mass. The endocrine effects of negative energy balance are well-documented adaptations for survival and include suppression of the HPG axis and hypercortisolaemia (**Appendix A**). In their meta-regression of field studies of arduous training, Murphy *et al.* showed that the combination of training duration and low EA were inversely associated with physical performance (Murphy et al., 2018), although it is difficult to delineate EA as a cause from the other factors described here.

A carefully calculated provision of approximately 21 MJ/ day (5000 kcal/ day; ~45% carbohydrate, ~45% fat and ~10% protein), with significant fat gain prior to the expedition, plus a relatively low altitude and preservation of sleep, meant participants lost only fat mass, not lean mass. Sustained, submaximal exertion appears to have had the effect of preserving total lean mass, although leg lean mass reduced by 6.10%. This may relate to muscle fibber pennation rather than reduced mass *per se;* we were unable to confirm this by biopsy. Thus, weight loss

was healthy, reinforcing the importance of appropriate nutrition preventing loss of lean mass and/ or hormonal disturbances, as has been shown in overtraining syndrome (Cadegiani and Kater, 2018a). As insufficient nutrition has been shown to cause multiple endocrine deficiencies in sports and exercise (Mountjoy et al., 2014), we hypothesize that sufficient and appropriate nutrition had an important role in preventing changes to the HPA and HPG axes.

Calbet et al. (2017) demonstrated that exercise maintains lean mass, during a 4-day extreme energy deficit in overweight men. Protein supplementation alone (1.5g/kg body mass/day) did not preserve lean mass, compared with carbohydrate. However, as demonstrated by Smith et al. (2016) in obese, sedentary women, a protein intake of 1.2 g/kg/day mitigated loss of lean mass, compared with low protein intake (0.8 g protein/kg/day) during 10% weight loss over 27 weeks. In men undertaking arduous military training, a mixed dietary supplement (5.1MJ/day (1220 kcal/day); ~45% carbohydrate, ~40% fat, ~15% protein) prevented 2 kg loss in lean mass, over 8 weeks, compared with non-supplemented controls (Fortes et al., 2011). Despite a caloric deficit (indicated by weight loss), our participants maintained total lean mass, with an average protein intake of around 1.6 g/kg/day.

Low ambient temperatures induce brown adipose tissue (BAT) thermogenesis, mediated by catecholamine upregulation, acting as a sink for glucose and fatty acid uptake (Sidossis and Kajimura, 2015). Adaptive thermogenesis is upregulated by β-3 adrenergic receptors, which are expressed in fat but not in muscle (Krief et al., 1993). Thus, the cold Antarctic environment could partially explain the high selectivity of substrate.

In mixed sex Norwegian Ranger training involving seven-day food and sleep deprivation, women demonstrated greater fat utilisation and glycogen preservation than men, implying greater capacity for endurance exercise (Hoyt et al., 2006). Oestrogens appear to be responsible for this substrate dimorphism (Tarnopolsky, 2008), while women subjectively claim better patrolling performance than men perhaps because of this metabolic advantage.

In addition to exercise and nutrition, the modest altitude of the expedition environment could have mitigated the loss in lean mass, compared with arduous expeditions at extreme altitudes, where hypobaric hypoxia contributes to loss of lean mass (Wandrag et al., 2017). Likewise, insufficient sleep, whether at altitude or as a

programmed part of arduous training, could impede absorption of macronutrients and reduces gut readiness for daytime absorption (Potter et al., 2016), and it could be postulated that preservation of sleep contributed to the maintained total lean mass we observed.

No suppression of metabolic parameters such as thyroid hormones or elevated cortisol were seen during or after the expedition. Lean mass exerts a greater effect on resting metabolic rate and appetite than fat mass (Stubbs et al., 2018), and demonstrates a greater bidirectional relationship with androgens, and to a lesser extent oestrogens, than fat mass (Carson and Manolagas, 2015). Thus, preservation of total lean mass might mitigate against some of the endocrine sequelae of negative energy balance. The decrease in leptin, followed by recovery post-expedition, was more pronounced than the changes we observed in body fat. Cold exposure itself may reduce leptin in women (Ricci et al., 2000), but this appears to become effective only when cold exposure is sustained (Iwen et al., 2011). The change in HPG axis function we observed did not correlate with leptin, as has been reported previously (Corr et al., 2011).

Dynamic attenuations in LH and sex steroids following an energy deficit may confer immediate survival benefits but may be associated with maladaptive suppression of hormonal axes and reproductive, bone or psychological sequelae if sustained (Mountjoy et al., 2018). Luteinizing hormone was relatively suppressed prior to and during the expedition (reflecting hormonal contraception usage), but recovered by post-exercise visit pre-2. There was no change in FSH before, during or after the expedition; this is consistent with studies of overtraining syndrome which generally demonstrate relatively normal FSH levels when LH is suppressed (reviewed by Cadegiani and Cater (2017a)), and laboratory studies of reduced EA, which show normal levels relative to suppression of LH in response (Loucks and Thuma, 2003).

Cortisol reactivity and diurnal salivary cortisol were blunted relative to other studies, and may be an appropriate response to a high intensity of training (Kajantie et al., 2003; Reynolds et al., 2001a). Alternatively, similar responses have been noted in dynamic testing of athletes during dysfunctional overtraining, also associated with elevated basal cortisol (reviewed in Cadegiani and Cater (2017a). Elevated hair cortisol concentrations are associated with exercise *per se;* whether the marked elevation during the expedition may represent an overtraining syndrome would be a

pertinent question for future studies (Gerber et al., 2012). The response of the HPA axis to central negative feedback is greater than has been described elsewhere (Reynolds et al., 2001a; Yehuda, 2002). Yehuda reviewed the use of low-dose dexamethasone suppression in post-traumatic stress disorder (PTSD), showing PTSD was associated with increased central axis sensitivity (Yehuda, 2002). No suggestion of PTSD was noted from the psychological stress or IES-R assessments before or after the expedition, thus this may relate simply to age, fitness and lower volume of distribution of these participants compared with previous studies.

Similar exercise-associated patterns in the HPA and HPG axis were seen following restricted carbohydrate intake with aerobic and resistance activity (average 46 ±9.1 MET and 4.7 ±0.7 sessions per week, respectively), in normal BMI women over 20 weeks (Hulmi et al., 2017). This regimen achieved a 11.9% weight loss with unchanged lean mass, and was associated with increased menstrual dysfunction, reduced testosterone, oestradiol, free T3 and TSH and unchanged cortisol compared with weight-stable, exercising controls. While the degree of weight loss was similar to the present study, this intervention was achieved primarily through dietary restriction, since the exercise was less intense. The investigators also assessed recovery, demonstrating partial normalisation of sex and thyroid hormones and leptin after 18 weeks. As in the current study, mood profile was unaffected by the intervention, which might possibly account for the apparently stable cortisol responsiveness we observed.

Other correlates of overtraining syndrome include sleep deprivation and psychological stress (Meeusen et al., 2013). Psychological stress is a prominent feature of extreme physical endeavour. Therefore, while both stress and reduced EA may be shown to cause reproductive endocrine dysfunction independently, their impact in this context may be synergistic and it may be impossible to draw a distinction between them (**Appendix A**). The expedition required both significant mental and physical exertion, although perceived stress levels were modest through the expedition and anxiety, depression and psychosocial risk factor assessments did not change after the expedition.

It has been suggested the psychological stress of Ranger training results from nutrient and sleep deprivation, which serve to increase the arduousness of many military training formats (Hoyt and Friedl, 2006). Sleep deprivation in isolation is

associated with elevated evening cortisol, flattened cortisol day curve, reduced androgen secretion and higher sympathetic nervous system activity (Potter et al., 2016). Female sex hormones appear to be protective of the effect of sleep deprivation on cortisol blunting after psychosocial stress (Bassett et al., 2015). The sustained moderate to high exercise intensity needed for a polar traverse represents a different form of exertion compared to US Ranger training, including its sustained, repetitive nature, austere environment, safety concerns and isolation. The stress and physical exertion scores reported during the expedition were consistent with previous arduous expeditions (Atlis et al., 2004; Kahn and Leon, 1994), while sleep diaries showed significantly longer sleep duration than would be expected in Ranger training (Friedl et al., 2000; Nindl et al., 2007), albeit of low perceived restfulness. Both increased sleep and the sustained, submaximal intensity of exercise could also account for the biological resilience we observed. Together with the nutritional strategy taken, and relatively reduced energy expenditure in women compared with men, these factors might have contributed to mitigating some of the negative psychological effects.

The major strength of our study is the unique nature of the expedition; this likely represents the first opportunity to study a cohort of female participants complete an endeavour of such a prolonged, arduous nature. Mitigating against low EA in women is important, since women appear to be at greater risk of low EA and its consequences than men (**Appendix A**; Mountjoy et al., 2014). Previous studies of prolonged, arduous training have focused on male cohorts and recovery rates in women have not been studied. Furthermore, the effects of exercise or low EA on the dynamic of the HPA and HPG axes have not previously been studied in either sex.

Limitations to our study include the small number of participants. This is unavoidable on such extreme expeditions; we have attempted to mitigate this by a comprehensive characterisation of the participants. The team is larger than any previous female-only transantarctic attempts, increasing the number of women who have skied across the continent from four to 10 (Atlis et al., 2004; Kahn and Leon, 1994). Other limitations include the natural limitations of a field study, such as four day delay in testing after the expedition. Every effort was made to overcome these using study visits shortly after the expedition arrived in Chile with imaging undertaken as soon as reasonably possible following the participants return to the UK. It was not logistically possible to repeat imaging immediately before and after

the expedition, or use the same examiner to perform skinfolds in the UK and Chile, so we used the best feasible measures of body composition. The use of hormonal contraceptives, while representative of real-world hormonal milieu, do limit the interpretation of LH responses. For logistical and ethical reasons, dynamic tests of the HPA axis at a higher level (e.g. insulin tolerance test, corticotrophin releasing hormone test, desmopressin test) were not possible, however in future studies a maximal or two-bout exercise test could be considered. Calculation of cortisol awakening response would add merit to our study, but was not possible since participants were woken 10 minutes before the first saliva sample taken in the pre- and post-expedition day curves.

In conclusion, no short term adverse effects were demonstrated from an unprecedented, successful transantarctic expedition in women. Cortisol reactivity and pituitary gonadotrophin reactivity were not impaired. We hypothesize these findings related to and pre- and intra-expedition nutrition, sleep provision, on the background of desirable selection characteristics, so that participants did not rate the expedition as subjectively stressful and lean mass was maintained.

4.1.7 Appendices to Section 4.1

17 OH Progesterone is an important steroid precursor hormone and is elevated in common forms of congenital adrenal hyperplasia (CAH). It is commonly checked to exclude CAH.

Androstenedione is a weak adrenal androgen and precursor of testosterone and oestradiol. It is also produced in the ovaries under influence of gonadotrophins and higher levels may predict recovery from FHA (Falsetti et al., 2002).

Anti-müllerian hormone is a biomarker of ovarian reserve. It peaks during puberty, then correlates inversely with age from around age 25 years (Lie Fong et al., 2012).

Cortisol is a glucocorticoid produced by the hypothalamic-pituitary-adrenal axis. It has important roles in mobilizing energy stores and may be released in response to external stimuli, such as physical or psychosocial threats or challenges.

Oestradiol is the major feminizing sex hormone, responsible for the development of secondary sexual characteristics. It is produced from oestrone or testosterone, predominantly but not exclusively in the ovaries.

Inhibin B is produced in the ovaries in response to FSH and is reflects early-follicular phase follicle activity (McNeilly, 2012).

LH, FSH are secreted in a pulsatile manner by the anterior pituitary in response to GnRH, and serve to control gonad function. The **LH:FSH ratio** is elevated in conditions with elevated androgen levels, such as polycystic ovary syndrome.

Prolactin is secreted by the anterior pituitary and is included as part of a complete anterior pituitary function test.

Sex hormone binding globulin is produced by the liver and binds androgens and oestrogens, limiting the amount of biologically available hormone. It is produced in response to oestrogens while its production is reduced by androgens and IGF-1.

Testosterone is the main androgen, produced in in men and to a lesser extent women. It is activated to **dihydrotestosterone**, which has higher androgenic effect, by 5αreductase

Appendix 4-1-1. Description of hormonal markers tested. LH: luteinising hormone, FSH: follicle-stimulating hormone

	Pre-expedition			Post-expedition				
	Pre-1 (-64 d)	Pre-2 (-39 d)	- 16 d	+1 d	Post-1 (+5 d)	+ 10 d	Post-2 (+15 d)	+ 18 to 24 d
BMI, kg/m^{2,}	24.2 (0.97) [22.8 – 25.45]	24.9 (1.24) [23.11 – 26.72]	25.0 (1.11) [23.48 – 26.73]	21.7 (1.23) [19.68 – 22.91]	22.3 (1.03) [20.59 – 23.27]	22.3 (1.17) [20.37 – 23.53]	22.4 (1.02) [20.56 – 23.42]	22.7 (1.02) [20.81 – 23.72]
Body mass, kg	70.47 (4.51) [62.7 – 74.2]	72.6 (1.75) [65.85 – 76.83]	72.8 (3.99) [65.8 – 77.2]	63.2 (4.74) [55.5 – 68.4]	64.9 (4.20) [58.0 – 69.2]	65.0 (4.03) [58.4 – 68.9]	65.3 (4.21) [58.3 – 69.1]	66.2 (4.11) [59.5 – 69.7]
% fat, DXA	20.92 (2.12) [19.2 to 25.4]	24.97 (2.39) [23.3 to 30.3]					19.02 (1.28) [17.2 to 20.4]	
% fat, skinfold			31.0 (2.0) [0.28 – 0.33]		23.0 (1.0) [0.21 – 0.24]			
%fat, BIA				22.15 (2.22) [19.4 – 26.0]	21.8 (3.71) [18.1 – 28.8]	23.3 (1.20) [22.5 – 25.7]	25.0 (2.04)	23.2 (1.74) [22.2 – 28.1]
Total body mass, DXA, kg	70.6 (4.5) [62.7 to 74.3]	72.6 (3.9) [65.9 to 76.8]					66.4 (4.2) [59.7 to 70.7]	
Total lean mass, DXA, kg	53.5 (3.06) [48.8 to 57.3]	52.3 (2.00) [48.9 to 54.7]					51.5 (3.04) [47.3 to 54.2]	
Total bone mineral content, DXA, kg	2.75 (0.13) [2.62 to 2.97]	2.80 (0.13) [2.67 to 3.02]					2.77 (0.12) [2.64 to 2.97]	

Appendix Table 4-1-1 Anthropometric changes during the expedition. Data are Mean (SD) [range]. BMI, body mass index. DXA, dual x-ray absorptiometry; BIA, bio-electrical impedance.

	Visit pre-1	**Visit pre-2**	**Visit post-1**
Lean mass (kg)			
Arms	5.44 (0.57)	5.15 (0.45)*	4.99 (0.36)
Legs	18.79 (1.05)	18.08 (0.98)*	16.98 (1.22)*
Trunk	26.15 (1.96)	26.06 (1.5)	26.52 (1.84)
Android	3.52 (0.24)	3.73 (0.25)*	3.67 (0.31)
Gynoid	8.33 (0.61)	8.40 (0.59)	8.22 (0.54)
Total	51.57 (3.06)	52.34 (1.99)	51.55 (3.04)
Fat mass (kg)			
Arms	1.44 (0.12)	1.94 (0.16)*	1.49 (0.19)*
Legs	6.4 (1.13)	7.26 (1.47)*	5.05 (0.79)*
Trunk	5.59 (1.09)	7.47 (1.16)*	4.79 (0.66)*
Android	0.61 (0.17)	0.9 (0.15)*	0.49 (0.08)*
Gynoid	2.92 (0.52)	3.71 (0.59)*	2.49 (0.43)*
Total	14.23 (2.11)	17.5 (2.52)*	12.13 (1.37)*
Total (%)	1.44 (0.12)	1.94 (0.16)*	1.49 (0.19)*
Bone mineral content (kg)			
Arms	0.34 (0.02)	0.35 (0.03)	0.35 (0.03)
Legs	1.05 (0.06)	1.05 (0.06)	1.05 (0.05)
Trunk	0.83 (0.07)	0.86 (0.07)*	0.83 (0.07)
Android	0.05 (0.00)	0.06 (0.01)*	0.05 (0.01)
Gynoid	0.30 (0.02)	0.30 (0.02)*	0.30 (0.02)
Total	2.75 (0.13)	2.80 (0.13)*	2.77 (0.12)

Appendix Table 4-1-2. Regional lean, fat and bone mass changes during the expedition. Data are mean (SD). Android: the area between the ribs and pelvis, gynoid: pelvis and upper thighs. * $p<0.05$ vs previous visit (paired t test)

	Score Pre	Score Post ('during expedition')	*p*
Several periods of psychological stress	5	2	0.079
Permanent psychosocial stress	0	0	1.0
Some periods of psychological stress	1	3	0.2
Never experienced psychological stress	0	1	1
One or more adverse events	4	1	0.079
High or severe financial stress	0	0	1.0
IES-R, median (range)	36 (9 – 52)*	41**	N/A
PHQ-9, median (range)	3 (0 – 11)	4 (1 – 6)	1.0
BAI, median (range)	11 (2 – 15)	6 (2 – 11)	0.386
CDRISC 10, median (range)	34 (31 – 36)	31 (29 – 35)	0.076
BEDA-Q			
Score, median (range)	4 (0-6)	7 (2-8)	0.009
BEDA-Q part B			
"Are you trying to lose weight now?"			
Yes, n (%)	0 (0.0)	1 (16.7)	1.00
"Have you ever tried to lose weight?"			
Yes, n (%)	3 (50)	4 (66.7)	0.558
"If so, how many times?"			
	3-5 (2) >5 (1)	>1 (1) 3-5 (2) >5 (1)	

Appendix Table 4-1-3. Comparison of Pre and Post- Expedition psychological testing. BEDA-Q: brief eating disorders in athletes questionnaire, IES-R: Impact of events scale (revised), PHQ-9 adjusted patient health questionnaire 9, BAI Beck Anxiety Inventory, CDRISC10 Connor Davidson Resilience Scale 10, N/A not applicable * Applies to four subjects who experienced a significant event **Applies to one subject who experienced a significant event

FSH concentration (Figure 4-1-3A)					
	Baseline	20 min	30 min	40 min	60 min
Pre-Ex	5.83 (2.66)	6.91 (2.38)	7.28 (2.49)	7.25 (2.33)	7.05 (2.1)
Ex + 5 d	5.3 (3.33)	6.31 (3.65)	6.55 (3.77)	6.71 (3.79)	6.27 (3.42)
Ex + 16 d	3.5 (1.18)	4.67 (1.65)	4.91 (1.73)	5.03 (1.8)	4.86 (1.68)
FSH concentration relative to baseline (Figure 4-1-3B)					
Pre-Ex		1.24 (0.25)	1.32 (0.28)	1.31 (0.26)	1.29 (0.28)
Ex + 5 d		1.29 (0.26)	1.37 (0.36)	1.43 (0.44)	1.34 (0.4)
Ex + 16 d		1.33 (0.18)	1.40 (0.20)	1.43 (0.18)	1.39 (0.19)
LH concentration (Figure 4-1-3C)					
Pre-Ex	5.36 (2.04)	13.59 (4.35)	13.73 (4.03)	12.90 (4.03)	10.89 (3.68)
Ex + 5 d	5.13 (3.70)	10.49 (7.10)	10.84 (7.63)	9.92 (7.12)	8.25 (5.95)
Ex + 5 d	3.42 (1.43)	12.67 (4.42)	13.03 (4.18)	12.12 (4.15)	10.04 (3.08)
LH concentration relative to baseline (Figure 4-1-3D)					
Pre-Ex		2.68 (0.82)	2.74 (0.91)	2.57 (0.86)	2.14 (0.63)
Ex + 5 d		2.24 (0.64)	2.40 (0.87)	2.12 (0.72)	1.82 (0.65)
Ex + 5 d		4.10 (1.86)	4.19 (1.75)	3.86 (1.49)	3.24 (1.31)
Cortisol concentration (Figure 4-1-4A)					
Pre-Ex	73.2 (45.23)	127.5 (39.77)	128.1 (33.8)	110.5 (28.9)	100.6 (19.9)
Ex + 5 d	61.7 (33.11)	130.5 (14.62)	140.1 (23.5)	131.8 (19.5)	110.2 (21.0)
Ex + 5 d	54.4 (28.60)	122.2 (21.50)	125.8 (8.02)	137.6 (26.9)	113.2 (12.0)

Hair cortisol by month (Figure 4-1-4B)

Month*	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan
	8.36 (2.89)	11.40 (1.78)	10.83 (3.43)	11.96 (4.89)	11.58 (4.68)	9.50 (4.94)	35.22 (29.31)	38.86 (36.2)	41.48 (48.56)	54.79 (75.41)

Diurnal saliva cortisol post-Ex (Figure 4-1-4C)					
Ex+1 d, pm	Ex+2 d, am	Ex+4 d, pm	Ex+5 d, am	Ex+10 d, pm	Ex+11 d, am
6.38 (4.24)	6.89 (7.74)	3.47 (2.45)	4.86 (3.73)	1.37 (0.55)	10.29 (4.15)

Hair cortisol day curve (Figure 4-1-4D)								
Time	07:10	08:30	09:00	09:30	12:15	13:30	17:20	21:50
Pre-Ex	13.68 (6.16)	7.03 (5.37)	6.48 (4.76)	3.97 (1.43)	3.36 (1.79)	4.6 (2.39)	3.26 (2.42)	2.68 (2.75)
Post-Ex	10.31 (5.63)	6.39 (3.2)	3.82 (1.39)	3.47 (1.52)	2.2 (0.85)	3.17 (1.79)	2.18 (0.97)	0.86 (0.50)

Appendix Table 4-1-4. Mean (SD) values for dynamic endocrine function tests. Baseline: immediately prior to dynamic function test, LH: luteinising hormone, FSH: follicle stimulating hormone, Ex: expedition

4.2 Experience from the selection and nutritional preparation for Expedition ICE MAIDEN

Section 4.2 was published under the title “Experience from the selection and nutritional preparation for Exp ICE MAIDEN: the first successful all-female unassisted Antarctic traverse” in the Journal of the Royal Army Medical Corps by Dr Natalie Taylor (NT), Dr Robert Gifford (RG) [joint first authors], Ms Rinn Cobb (RC), Dr Sophie Wardle (SW), Mr Steve Jones (SJ), Dr Jodie Blackadder-Weinstein (JB), Dr John Hattersley (JH), Prof Adrian Wilson (AW), Prof Chris Imray (CI), Prof Julie Greeves (JG), Prof Rebecca Reynolds (RR) and Prof David Woods (DW) (Taylor et al, 2019). RG, NT and RC designed the experiment. NT drew up the shortlist for Ex ICE MAIDEN participants and RC designed the pre-expeditionary nutritional programme (“Operation fois gras”) and expeditionary diet. NT and RG both wrote the first draft, which RG finalised with supervisory support from DW and RR. Data collection was performed by NT, RC and RG. RG performed the analysis with assistance from RC and SW, under the supervision of DW and RR. RG responded to reviewer comments. All authors provided editorial oversight of the manuscript.

This study evaluated physical changes during training among applicants for Ex ICE MAIDEN, and found more beneficial adaptations among eventual team-members among women who were not selected. The nutritional planning are described and compared with data from previous expeditions to aid planning for future expeditions.

4.2.1 Abstract

Introduction. Ex ICE MAIDEN was the first all-female unsupported crossing of Antarctica. We describe the prerequisite selection and training, comparing those who formed the final team with other participants, and discuss how the expedition diet was established.

Methods. All women serving in the British Army were invited to participate. Following initial assessments, successful women completed three training/selection ski expeditions. Between expeditions 1 and 2, participants completed 6 months rigorous UK-based training. Weight was measured before and after the 6 months UK-based training, expeditions 2 and 3, and body composition by skinfold before and after expedition 2. Participant feedback, body composition and weight changes were applied to modify the expedition diet and provide weight gain targets prior to Ex ICE MAIDEN.

Results. Following 250 applications, 50 women were assessed and 22, 12 and 7 women attended training expeditions, 1, 2 and 3, respectively. The final team of 6 women lost more weight than other participants during UK-based training (mean (SD) change –1.3 (1.5) kg vs –0.5 (1.6) kg, respectively, p=0.046) and during expedition 2 (–2.8 (0.8) kg vs –1.7 (0.4) kg, respectively, p=0.048), when they also gained more lean mass (+2.1 (0.8) kg vs +0.4 (0.7) kg, respectively, p=0.004). The Ex ICE MAIDEN diet provided 5000 kCal/day, comprising approximately 45% carbohydrate, 45% fat and 10% protein. Median (range) weight change between expedition 3 and Ex ICE MAIDEN was +8.7 (-1.9 to +14.3) kg.
Conclusions. The selected Ex ICE MAIDEN team demonstrated favourable training-associated body composition changes. Training-associated weight loss informed the expeditionary diet design.

The first all-female team crossing of Antarctica using muscle power alone, Ex ICE MAIDEN, was completed by six women from the British Army in January 2018
The team's success was widely publicised, while a series of studies has been published demonstrating biological resilience to the arduous expedition and weight loss.
This manuscript details the how the selection and training of these women were carried out.
We noted more favourable physical changes during training among women who undertook Ex ICE MAIDEN compared to other volunteers.
The diet developed for Ex ICE MAIDEN was key to their success and is compared with diets from previous expeditions.

Box 4-2. Key points to Section 4-2.

4.2.2 Introduction

Ex ICE MAIDEN was the first successful all-female unassisted Antarctic traverse, completed by six women from the British Army in January 2018. The team demonstrated preserved pituitary, adrenal, gonadal and autonomic function five and 15 days post expedition (**Section 4.1**). Only small changes in energy expenditure and substrate utilisation were identified between pre- and post-expedition measurements (Hattersley et al., 2019b), while bone turnover was only modestly impaired with no impact on estimated lower limb fracture threshold (O'Leary et al., 2018). These were achieved despite an energy deficit of at least 685 kCal/d

(Hattersley et al., 2019b). The team also displayed psychological resilience and robust team dynamics (Blackadder-Weinstein et al., 2019). We attributed the success of Ex ICE MAIDEN from a hormonal and metabolic perspective to the team's approach to planning, selection and training, and appropriate provision of nutrition and sleep (**Section 4.1**; Hattersley et al. (2019b)).

Antarctic crossings from Shackleton to the present day have been highly arduous. Extremes of isolation, physical activity and climactic conditions can lead to profound energy deficits and weight loss (Halsey and Stroud, 2012). While body mass loss during these expeditions have been described before (Castrission, 2013; Saunders and Maddison, 2015; Skog, 2011; Stroud et al., 1996), little appears to be known about team mates or training partners who did not take part, through choice, circumstance or deselection. For Ex ICE MAIDEN, the final team was chosen from large number of volunteers, through several training expeditions in the UK and abroad. This was perhaps unusual. Given the expedition's success, there may be value to future expeditions in describing this selection process and comparing physical characteristics of the final team with other volunteers during training and selection.

Polar explorers have tended to consume significantly more fat than would be recommended for the general population, and have exceeded 50% of recommended dietary intake (Loewen and Bancroft, 2001) (UK guidelines are maximum 78g or 35% of energy intake per day for women (Public Health England, 2016)). The aim of such diets is to increase the ratio of energy content: food mass and counteract body fat losses. This strategy has not always been successful. The fat rich, 5,000 kCal diet consumed by the American Women's Expedition to Antarctica was insufficient to prevent weight loss, fatigue and energy depletion, which became significant problems (Loewen and Bancroft, 2001). Caroline Hamilton's account of a five-woman expedition identified problems from food not tasting the same on Antarctica as it had at home and being difficult to eat (Hamilton, 2000, p. 116, 129, 247 and 253). Guidelines for endurance sports recommend high carbohydrate intakes, and are based on studies of events like ultra-triathlons and multi-day running races (Bushman, 2017; Kerksick et al., 2018; Maughan and Burke, 2011). In such settings, athletes depend on carbohydrate to replenish lost muscle and liver glycogen. It is recommended that endurance athletes consume similar or slightly higher amounts of fat than is recommended for non-athletes, although this can change in certain

situations such as high-volume training (Kerksick et al., 2018). However, these guidelines predominantly rely on data from men and do not take into account the importance of food mass or the harsh environment encountered on polar expeditions.

Striking a balance between energy content of food and the food mass has been a key challenge to polar explorers (Saunders and Maddison, 2015). Nutritional strategies aim to prevent the lean mass loss which could hamper physical performance and undermine outcomes of the expedition. Some have opted to maximise pre-expedition weight gain (Skog, 2011), while others carry more additional rations (Halsey and Stroud, 2011; Saunders and Maddison, 2015; Stroud et al., 1996).

This manuscript describes the selection and training process for Ex ICE MAIDEN, including three training expeditions. We aimed to provide exploratory comparisons of physical characteristics between the six participants who proceeded to the undertake Ex ICE MAIDEN with those who were selected to undertake training but were not selected for the expedition. Since the Ex ICE MAIDEN team demonstrated favourable physiological changes after the expedition, we also aimed to examine the nutritional strategy, and demonstrate how data collected during the planning process were used to create individual targets for pre-expedition weight gain and how the rations for Ex ICE MAIDEN were designed.

4.2.3 Methods

4.2.3.1 Participants

All women serving in the British Army (Regular and Reserve) were invited to apply to participate in Ex ICE MAIDEN. Electronic applications were judged independently by two expedition leaders based on motivation and aptitude. Ethical approval was received for the study immediately before and after Ex ICE MAIDEN (827/MoDREC/17). All participants provided informed consent prior to undergoing any measurements.

4.2.3.2 Selection and training

The selection and training process was designed by two team leaders experienced in endurance activities and by consensus with experts at Antarctic Logistics Expeditions (Salt Lake City, UT, USA). After each expedition, participants were

selected to proceed by the expedition leaders based on opinion-based judgments of their technical skills, physical ability and team work.

The timeline for the selection process is shown in **Figure 4-2-1**. Applicants selected following from the initial application process were invited to attend a selection and training weekend in Capel Curig, UK in September 2015, where they underwent physical exercise, public speaking, team work and interview assessments. Onward selection from this training weekend was focussed around the demonstration of physical and mental resilience; those who proceeded to the next phase were not necessarily fastest or physically strongest but demonstrated adaptability in a small team. Expedition leaders were also looking for the willingness and aptitude to learn skills necessary for survival in a polar environment. These factors were assessed by both expedition leaders independently before a consensus was reached.

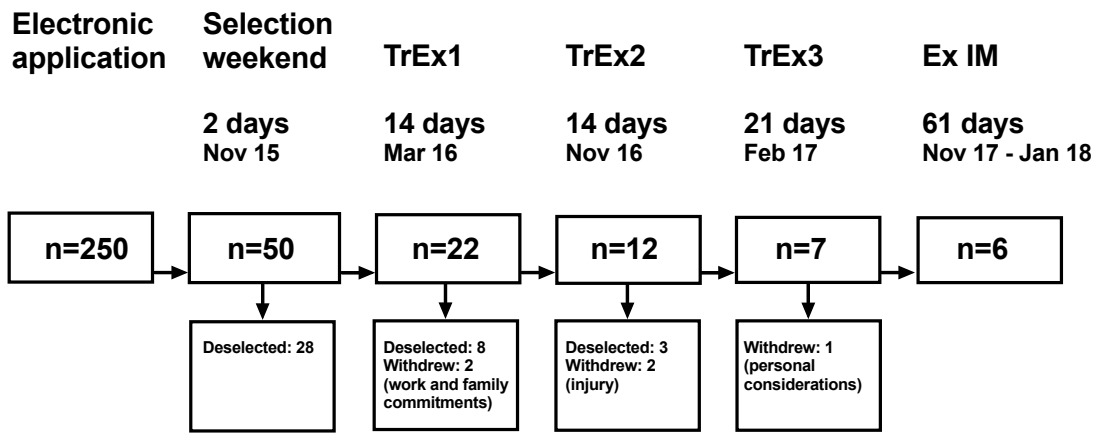


Figure 4-2-1. Overview of training and selection programme. TrEx: Training Expedition, TrEx1 and TrEx2: Norwegian Arctic Circle.TrEx3: Hardangervidda, Norway. Ex IM: Ex ICE MAIDEN.

After each subsequent training expedition (TrEx, numbered 1, 2 and 3), participants were selected to proceed based on resilience, endeavour and cohesion, assessed by two expedition leaders by consensus. Twenty-two women attended TrEx1, which took place in the Norwegian Arctic Circle in March 2016, comprising 7 days of ski training, cold weather survival and sleeping in snow holes followed by a 7-day expedition where these skills were applied.

Participants selected to continue beyond TrEx1 undertook a 6 month training programme in the UK from April – October 2016 (**Figure 4-2-1**), developing strength, aerobic fitness and technical skills. TrEx2 took place in Lakselev, North

Norway in November 2016, lasting 14 days; 4 days Nordic ski training with pulks (sledges) was followed by a 10 day self-sufficient ski tour. Participants each pulled pulks whose weight was gradually increased from 25 kg to 50 kg as they felt able.

TrEx3 took place on the Hardangervidda, Norway in February 2017 with seven participants and lasted 21 days. Training content built on previous expeditions, with participants pulling 50 kg pulks over increasing distances and attempting to become accustomed to the daily routine of an Antarctic traverse and task-specific roles. By the middle of TrEx3, participants were skiing for 9 to 10 h/day in a pattern of 60 min skiing followed by a 5 min break and rested 7 to 10 h/ night. Winter expeditions in northern Norway were chosen since darkness lasting ≥18 h/day would increase arduousness and, it was hoped, prepare psychologically for sleep disturbance (during Ex ICE MAIDEN, 24 h daylight was anticipated; however, this could not be replicated during training for logistical reasons). The final team of six started Ex ICE MAIDEN in November 2017.

4.2.3.3 Assessment of dietary intake and provision

UK military rations were consumed during TrEx1 to familiarise participants with food commonly consumed on polar expeditions and the volume required to be ingested daily (**Appendix Table 4-2-1**). Personal feedback from participants on TrEx1 who were going on to TrEx2 was used to design the diet for TrEx2. Dietary intake was assessed during TrEx2 using a standardised food diary, completed once daily using a buddy-buddy system for prompting. Dietary intake was compared with food provision. The diet for TrEx3 was modified according to measured body mass and composition changes during TrEx2 and further participant feedback. The Ex ICE MAIDEN diet was modified to increase protein intake, based on studies from endurance athletes (Kerksick et al., 2018), and the ratio of energy content: weight carried.

Energy expenditure during TrEx2 was estimated as energy intake plus 9 kCal/g body fat mass loss (see Anthropometric measurements). The study dietitian discussed estimated energy expenditure with each participant, explored possible reasons for high weight loss, and directly monitored energy intake during TrEx3. Ex ICE MAIDEN rations were based on a modification of TrEx3 rations, to increase calorie: food mass ratio, while meeting minimum carbohydrate and protein requirements for endurance exercise: 8–12g/kg/day and 1.2 – 1.7g/kg/day,

respectively, based on existing guidelines (Volpe and Stanzione, 2017; Maughan and Burke, 2011).

A daily multivitamin (Healthspan Gold A-Z, Guernsey, UK) was provided to supplement micronutrients potentially missing from freeze-dried meals. This supplement, listed on the Informed Sport website,(Informed-Sport, 2018) was selected to ensure it contained no illegal and prohibited substances (e.g. amphetamine). It was taken daily for 3 months before Ex ICE MAIDEN, and once weekly during the expedition due to stringent rationing of space inside sledges .

4.2.3.4 Anthropometric measurements, pre-expedition weight gain

Weight and height were measured 10 days after TrEx1 (Seca scales model 874 and stadiometer model 213, Birmingham, UK). Participants were weighed in April 2016 (prior to the 6 month exercise programme), before and after TrEx2, TrEx3 and Ex ICE MAIDEN. Body composition was measured by a single International Society for the Advancement of Kinanthropometry-trained examiner at six sites, before and after TrEx2 and Ex ICE MAIDEN, as described in Section 4-1. Percent body fat was estimated using the equation of Durnin & Womersley (1974). Which was derived from measuring four peripheral skinfold sites (triceps, biceps, subscapular and suprailiac crest) , substituting the log of their sum as follows (for women aged 30-39):

$$D = 1.1423 - (0.0632 \times L)$$

Where D = predicted density of the body (g/mL), and L = log total of the 4 skinfolds (mm). The density value was converted to percentage body fat using the Siri Equation (1974).

TrEx3 (duration 21 days) attempted to simulate the arduous exercise of Ex ICE MAIDEN (anticipated duration 70 days). Assuming linearity, projected weight loss for Ex ICE MAIDEN was estimated as:

$$\frac{70}{21} \times [\text{TrEx3 weight loss}]$$

Individualised pre-Ex ICE MAIDEN weight gain targets were set, based on projected weight loss. Weight gain was evaluated by the dietician regularly for the 4 months preceding Ex ICE MAIDEN. Several participants found this difficult and the dietician

provided remedial advice or support. Psychological difficulties with weight gain and body image were managed by peer support via group messaging on smartphones.

4.2.3.5 Statistical analysis

Statistical analysis was performed using SPSS for Mac version 24 (IBM, New York, USA). A visual inspection of data was made for outliers. Data are presented as mean (SD) unless otherwise stated. Weights (kg) and skinfold measurements (mm) before and after each expedition were compared using paired samples t test. Group comparison (final Ex ICE MAIDEN team versus other participants) was made by independent samples t-tests. All t-tests were 2-tailed. Statistical significance was set at $p<0.05$

4.2.4 Results

4.2.4.1 Participants, selection and training

The number of participants at each stage of selection and training is illustrated in **Figure 4-2-1**. A total of 250 women submitted electronic applications. Fifty (20%) attended the training weekend. Twenty-two (44%) attended TrEx1. Fourteen (64%) were invited to attend TrEx2. Of these, two withdrew during Summer 2016 (**Figure 4-2-1**) due to work and domestic commitments. Three withdrew following TrEx2 due to sports injuries sustained following TrEx2 and two withdrew due to work and family commitments. Seven proceeded to TrEx3. One participant withdrew voluntarily after TrEx3 and six completed Ex ICE MAIDEN.

4.2.4.2 Dietary provision and intake

In TrEx1, all participants consumed UK military rations for Arctic environments, providing high calorie to weight ratio (high fat) comprising dehydrated meals, chocolate bars, oat bars, nuts, dried fruit, and hot chocolate and energy drink powders (detailed in **Appendix Table 4-2-1**). Participant feedback identified these rations were identified as being too bland. During TrEx2, 12 different 24 h rations were provided, comprising mean (SD) 3,871 (365) kCal/day (52 (2) % carbohydrate, 39 (3) % fat and 9.4 (1) % protein by calorie content). Participant feedback was positive towards freeze dried meals including Real Turmat (Drytech, Tromsø, Norway), MX3 Adventure (Olonne sur Mer, France) and Extreme Food meals (Lancaster, UK) due to their taste and high energy: weight ratio. Meals were supplemented with ProCal Powder (Vitaflo, Liverpool, UK) and recovery milkshake powder (For Goodness Shakes, UK) to meet protein requirements. During TrEx3,

six daily rations were designed following the feedback from TrEx2, comprising mean (SD) 4,563 (300) kCal/day (53 (2)% carbohydrate, 35 (2)% fat 10 (1)% protein by calorie content). Rations for Ex ICE MAIDEN provided mean (SD) 4,987 (25) kCal/day (71 (0) kCal/kg/day), containing 44 (2) % carbohydrate (7.7 (0) g/kg/day), 45 (1) % fat (3.6 (0) g/kg/day), and 9.4 (2) % protein (1.7 (0) g/kg/day).

Food diaries during TrEx2 and TrEx3 were used to formally advise the participants on their dietary intake during training which helped improve compliance with consuming the total daily calorie requirement, which some participants found difficult. A typical menu is displayed in **Table 4-2-1**; the Ex ICE MAIDEN team chose six menus which rotated during the 61 day crossing. Main meals and snacks made up a roughly even share of dietary energy content by weight.

	Item size, g	Energy content, kCal	Carbohydrate content, g	Fat content, g	Protein content, g
Extreme foods porridge	100	500	50.0	27.3	12.0
ProCal powder	30	200	8.4	16.6	4.0
Beyond the beaten track hot chocolate	60	244	48.6	3.8	2.7
Beyond the beaten track hot chocolate	60	244	48.6	3.8	2.7
Sports drink	47	178	44.0	0.0	0.0
For Goodness Shakes	72	261	49.6	0.3	16.5
Lindt Chocolate Bar	100	560	48.4	36.2	7.8
Fuzion Flapjack	170	764	121	23.3	11.0
Macaroon	70	329	42.7	15.9	2.3
Sesame seed bar	50	269	20.0	16.5	7.0
Salt and Vinegar peanuts	100	614	5.6	51.0	30.0
Vitaflo chicken Soup	50	309	11.3	26	6.0
Extreme Moroccan Chicken	100	500	41.2	29	17.4
TOTAL	1,009	4,972	539.4	249.7	119.4
	(4.5 kCal/g)		51%	37%	10%

Table 4-2-1. Representative 24-h ration menu on Ex ICE MAIDEN.

4.2.4.3 Body composition

At the start of the 6-month exercise regimen prior to TrEx2, mean (SD) body mass index (BMI) was 23.1 (1.7) kg/m^2, ranging from 19.4 to 25.5 kg/m^2, and did not differ significantly between the final Ex ICE MAIDEN team and other participants (23.4 (1.4) kg/m^2 versus 22.7 (2.1) kg/m^2, p=0.53). During the six-month training phase before TrEx2, mean (SD) weight decreased from 67.1 (4.7) kg to 65.8 (5.0) kg, p=0.01. Weight loss during these six months was lower among participants who formed the final Ex ICE MAIDEN team than other participants (mean (SD) 2.3 (1.5) kg vs 0.5 (1.6) kg, p=0.046).

At the start of TrEx2, participants weighed mean (SD) 67.3 (6.2) kg (mean (SD) BMI 23.0 (1.7) kg.m^{-2}) with no difference between participants who ultimately formed the Ex ICE MAIDEN team (n=6) and others (n=6) (68.3 (3.4) kg and 66.2 (8.2) kg, respectively, p=0.6; **Figure 4-2-2A**). Similarly, there were no differences between the Ex ICE MAIDEN team and others in body composition (percent fat: (28.0 (2.9) kg and 29.3 (7.2) kg, respectively, p=0.7; **Figure 4-2-2B**, lean mass: (38.2 (1.5) kg and 37.5 (3.8) kg, respectively, p=0.7; **Figure 4-2-2C**). Both Ex ICE MAIDEN team and other participants lost weight during TrEx2, but the weight loss in the Ex ICE MAIDEN team was greater than other participants (mean (SD) loss 2.8 (1.0) kg versus 1.7 kg (0.9) kg, respectively p=0.006). During TrEx2, participants gained lost fat mass and gained lean mass. Again, the changes were greater among the Ex ICE MAIDEN team than other participants (percent fat loss: 3.9 (1.6) % versus 0.74 (0.80) %, respectively, p=0.006, lean mass gain: 2.0 (0.88) kg versus 0.40 (0.43) kg, respectively, p=0.004).

During TrEx 3, participants lost mean (SD) 3.0 (1.4) kg (**Figure 4-2-2D**). Weight loss did not discriminate the final Ex ICE MAIDEN team from the participant who withdrew, who lost 3.1 kg during TrEx3.

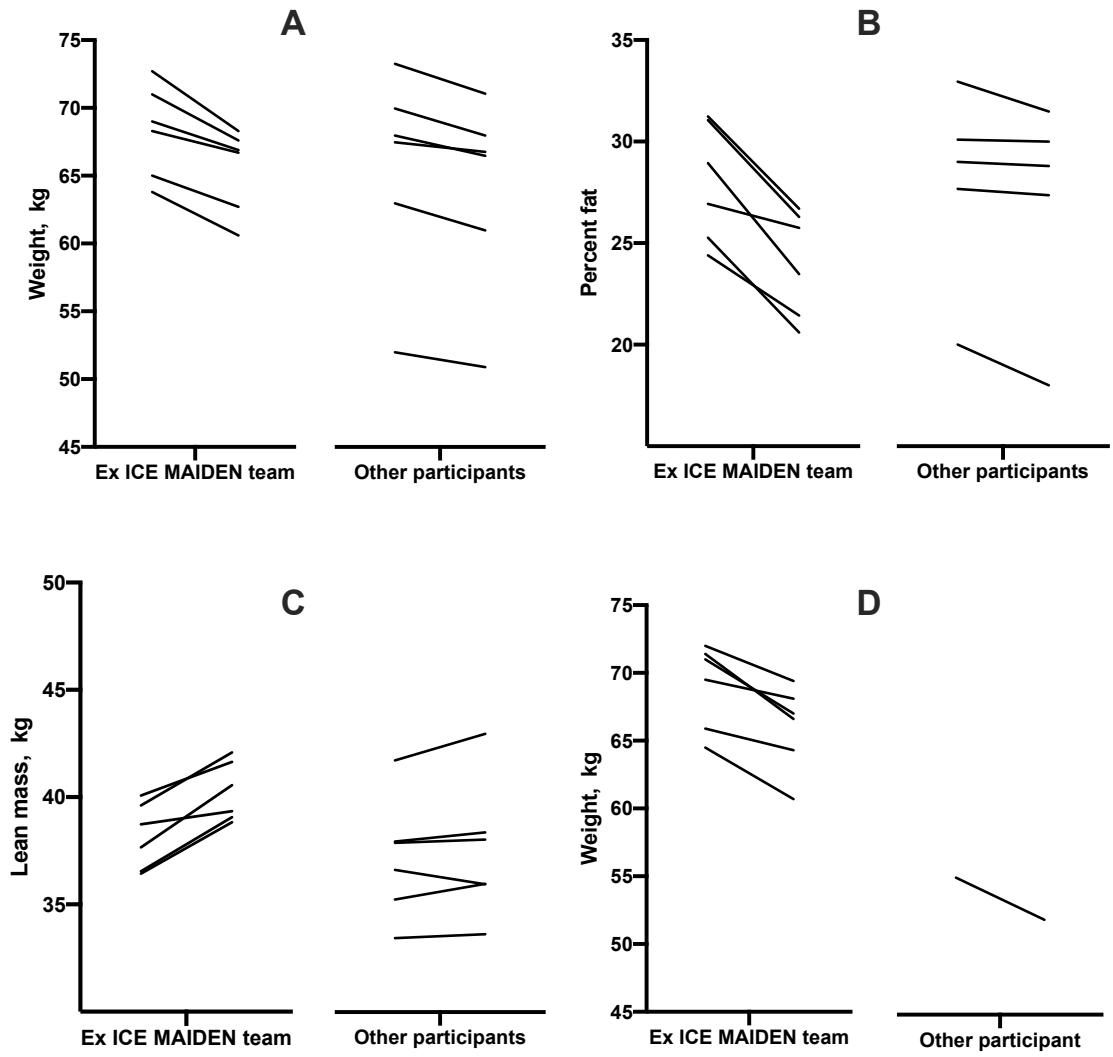


Figure 4-2-2. Weight and body composition before and after training exercises (TrEx) in Ex ICE MAIDEN team (left panels) and those participants who did not undertake Ex ICE MAIDEN (right panels). TrEx2 changes in weight (A), lean mass (B) and percent fat (C). TrEx3 changes in weight (D).

The median (range) pre-expedition weight gain target set 4 months before Ex ICE MAIDEN was 8 (6 to 10) kg. The actual weight change achieved was median (range) +7.0 (-3.6 to +13.4) kg. For two participants, active weight gain was discontinued due to build-up of subcutaneous fat around the thighs, which caused irritation by rubbing trousers when skiing and was thought to increase the likelihood of developing polar thigh syndrome (Rhodes and Sorenson, 2017).

After travelling to Antarctica, weather conditions necessitated a 14 day postponement, where significant exercise was not possible but plenty of food was available. Further weight gain occurred; additional median (range) change in body

mass was 1.0 (-0.2 to +2.7) kg, making total median pre-Ex weight change +8.7 (-1.9 to +14.3) kg. As discussed in Section 4-1, participants lost mean (SD) 9.6 (1.2) kg, p<0.0001, ranging from 5.70 kg to 14.5 kg, and constituting fat mass (mean (SD) loss -8.2 (1.7) %, p=0.001) but not lean mass (mean (SD) pre-Ex lean mass was 38.0 (2.7) kg vs post-Ex 37.9 (2.6), p=0.9).

The macronutrient intake during Ex ICE MAIDEN presented alongside extant data from other expeditions are shown in **Table 4-2-2**.

Expedition	Pre-expedition body mass, kg	Dietary content						
		Energy	Carbohydrate		Fat		Protein	
		kCal	%	g/kg	%	g/kg	%	g/kg
Scott 1912 (Male)	NK	4,200	47	NK	24	NK	29	NK
Stroud & Fiennes 1993 (Male)	95/ 97	5,000	35	4.6/ 5.8	57	3.3/ 3.3	8	1.0/ 1.3
Saunders 2015 (Male)	93	5,875	50	7.9	37	2.6	11	1.7
Castrission & Jones 2012 (Male)	101/93	5,700	50	7.0/ 7.7	30	1.9/ 2.0	12.3	1.7/ 1.9
ICE MAIDEN 2018 (Female) median [range]	72.8 [65.8 to 75.3]	5,000	45	7.7 [7.5 to 8.5]	45	3.5 [3.3 to 3.8]	10	1.7 [1.6 to 1.9]

Table 4-2-2. Polar expeditions daily nutritional intake and pre-expedition body mass. Data from Maddison and Saunders (2015), Stroud et al., (1996), and Castrission (2013), and kindly supplied by Antarctic Logistics Expeditions Ltd. (Salt Lake City, UT, USA). NK, not known

4.2.5 Discussion

We report the experience of preparing an all-female team for a polar expedition de novo. The final team of six women completed the 1700km ski traverse in 61 days, becoming the first all-female team to do so unassisted (without using kites), taking the total number of women who have skied across the continent from four to ten. The women were also the first to complete such an expedition when none had previously either been to Antarctica or experienced this type of expedition prior to commencing training. As section 4-1 explored, the expedition was a success

biologically, with preserved metabolic and dynamic pituitary and adrenal function. Autonomic function was similarly preserved (**Section 4.1**). The expedition diet evidently provided sufficient macronutrients and micronutrients; as we have previously reported, lean mass and nutritional markers including calcium, magnesium, phosphate, zinc and iron were unchanged following the expedition.

Comprehensive preparation was also evidenced by low scores of stress and perceived physical exertion and sufficient sleep during the expedition, while the high self-confidence scores demonstrated perhaps underline the importance of the rigorous selection for non-technical traits (Blackadder-Weinstein et al., 2019). The stepwise, concurrent approach to selection, training and nutrition allowed the participants to optimise team dynamics, improve team members' skillset and resilience, while creating and trialling a bespoke expedition nutrition plan. Selection performed by the two team leaders primarily focussed on subjectively demonstrating traits such as determination to succeed, endeavour and resilience. These are reminiscent to the qualities sought in Shackleton's original advertisement to men:

> "…wanted for hazardous journey… constant danger, safe return doubtful, honour and recognition in case of success" (Horn, 2013).

As others have suggested, the principal non-technical factor responsible for this success may have been individual motivation (Blackadder-Weinstein et al., 2019). Selection for non-technical characteristics may incidentally have identified the participants who were gaining the most physical benefit from training. While body mass changes during training did not contribute to selection, it is interesting to note that those participants who proceeded to undertake Ex ICE MAIDEN lost more weight during the six-month, self-directed training prior to TrEx2. Similarly, during TrEx2, the final Ex ICE MAIDEN team lost more fat mass and gained more lean mass than other participants. It is not possible for us to distinguish cause and effect, i.e. whether more rigorous training among these women led to greater physiological adaption, or if a greater propensity for physiological adaptation may have led to 'successful' training. Increasing lean mass whilst reducing fat mass increases the body's efficiency at undertaking physical tasks (Bosy-Westphal et al., 2015). The thermogenic properties of lean mass are also important for polar expeditions; skeletal muscle contributes to temperature maintenance during cold exposure (Payne et al., 2018).

A personalised diet can improve nutrition effectiveness during prolonged expeditions (Halsey and Stroud, 2012), and may help counteract health-consciousness among female athletes (Garthe et al., 2011). Both dietary volume and composition can be tailored to ensure high energy requirements are met. However, consuming more high-fat foods can be challenging, since this may represent a change from a normally healthy diet (Varnes et al., 2013), while increased adiposity may represent a significant psychological hurdle **(Appendix A).**

The diets for TrEx2 and TrEx3 contained a higher relative carbohydrate content than previous (male) expeditions. While the Ex ICE MAIDEN fat content was increased and carbohydrate reduced from TrEx3, it remained relatively carbohydrate-rich, second only to Ben Saunders's solo crossing in 2015. The Ex ICE MAIDEN team consumed CHO commensurate with existing guidelines for endurance exercise, which recommend 8-12 g/kg/day for ≥4-5 h moderate to high intensity exercise (Maughan and Burke, 2011).

The nutritional strategy for Ex ICE MAIDEN may inform 'extreme exercise' generally for women and polar expeditions specifically for both sexes. American College of Sports Medicine guidelines recommend protein requirements for endurance as 1.2-1.7 g/kg/day and that fat should not make up less than 20% total energy intake (Volpe and Stanzione, 2017). Due to the nature of polar expeditions, namely their very high energy expenditure versus the weight of rations, it is generally anticipated that an energy deficit will occur. This has led polar explorers to increase their body fat stores prior to undertaking these types of enduring expeditions, which can then be used as a fuel source during the expedition.

The pioneers of polar exploration did not have remotely the same polar-suited rations available now, with increased caloric value: weight penalty ratio, allowing greater travelling efficiency. Scott's explorers consumed about 4,200 kCal/day compared to Saunders's 5,875 kCal/day (Saunders and Maddison, 2015). Fiennes's and Stroud's 1993 expedition diet contained 8% protein while Scott had 29% (Halsey and Stroud, 2011; Halsey and Stroud, 2012). In their study of 13 Antarctic ski racers (12 male), Paulin et al. (2015) demonstrated that protein intake and energy intake are independently associated with lean mass gains in this relevant setting (both $p=0.03$), while lean mass gains were associated with faster finishing. Similarly, during TrEx2, we found that lean mass gains were greater among the Ex

ICE MAIDEN team than others, although both TrEx2 and Ex ICE MAIDEN were at a slower pace than the race studied by Paulin et al. (and were twice the distance). Unsupported expeditions, such as Fiennes's and Stroud's 1993 Antarctic traverse, or Cecile Skog and Ryan Waters's 1,746 km crossing in 2009, may result in loss of lean mass as well as fat (Stroud et al., 1996). Without resupply points, such expeditions necessarily haul far greater loads (initially up to 220kg for Stroud and Fiennes, compared with 80 kg for Ex ICE MAIDEN) (Halsey and Stroud, 2011; Halsey and Stroud, 2012).

Estimates of energy expenditure during polar expeditions like Ex ICE MAIDEN vary dramatically. Stroud et al. (2007) reported 11,000 kCal/day, based on doubly labelled water, although isotope enrichment and other experimental concerns mean the measurement could be an overestimate and should be treated with caution. Other measurements reported by Stroud et al. in the same paper are much closer to the figure of 6,100 kCal/day estimated by Acheson (1975), although the experimental evidence to support this figure is unclear. Even if it were possible to carry enough food to meet requirements significantly over 6,000 kCal/day, it would be impossible to consume such a quantity so most expeditions budget for 4,000-6,000 kCal/day. Since it was not practical to create individualised rations for Ex ICE MAIDEN participants, the team dietician provided individualised pre-expedition weight increase targets, based on anthropometric measurements before and after TrEx2 and TrEx3. A total target energy intake of around 5,000 kCal/day was chosen for Ex ICE MAIDEN based on estimated energy expenditures from TrEx2, weight loss during TrEx3 and observation of how much the participants could reliably consume on each day. This strategy was commensurate with recent expeditions (Castrission, 2013; Skog, 2011). Participants reported they had consumed around 85% (range 70 to 99%) of the rations carried (Section 4-1).

It is notable that only fat mass changed during Ex ICE MAIDEN, not lean mass, which is likely due to sufficient protein during the expedition and increases in body fat beforehand. Previous polar expeditions support the value of anticipatory weight gain (Briffa, 2005; Jardine, 2009). Most weight gain occurred in the final weeks before Ex ICE MAIDEN, *en* route to Antarctica. Fortuitously, severe weather necessitated a 2 week wait in Antarctica where training was restricted, and a median additional 1.0 kg body mass could be gained. We envisage that lean mass might have been lost as well as fat mass if this weight gain had not occurred.

Strengths of our study include its novelty in providing the first review of how a polar expedition developed the participants' nutrition strategy. During TrEx2, anthropometric measures were taken in the extreme cold, so were limited to peripheral measurements only. This was mitigated by using the formula of Durnin & Womersley (1974), which permits the use of skinfolds from multiple sites to estimate body fat. Our projection of body fat loss from TrEx2 to Ex ICE MAIDEN which informed pre-expedition weight gain necessarily assumed linearity of weight change, since only pre- and post-exercise measurements could be taken. In reality weight loss following arduous exercise is unlikely to be linear (Marek et al., 2017). Our observations from these six highly selected women may not be widely generalisable, given normal variation in motivation, resilience and food preference. For a variety of reasons, no two teams will expend the same energy during polar expeditions (Halsey and Stroud, 2011; Halsey and Stroud, 2012; Paulin et al., 2015), or other forms of extreme exercise (Reid et al., 2017). Nevertheless, our data provide an important 'proof of concept' for women and men seeking to undertake extremely arduous training.

4.2.6 Summary

The military adage 'train hard, fight easy' might go some way to explain the success of Ex ICE MAIDEN. This success was multifactorial, and the factors outlined here may be relevant to women's training and employment in Defence as well as for those planning future polar expeditions. The pre-Ex ICE MAIDEN weight gain and intra-expedition diet successfully mitigated against lean mass loss; the expertise afforded by an embedded dietitian made an important contribution to this success. None of the participants in Ex ICE MAIDEN were elite or professional athletes, nor were they selected solely on physical fitness or strength. Participants came from within the British Army and were exceptionally well-motivated, thus our findings may be generalisable to women seeking to train in arduous military roles.

Chapter 5 Does female gender confer greater risk of heat illness? A systematic review and meta-analysis

The following chapter was published in abridged form in Environmental Research under the title “Risk of heat illness in men and women: A systematic review and meta-analysis” by Robert M Gifford (RG), Tommaso Todisco (TT), Mike Stacey (MS), Michael Allerhand (MA), David R Woods (DW) and Rebecca M Reynolds (RR) (Gifford et al., 2019c). RG conceived the study question in response to a request from WGCC for a review on the relative rates of HI among men and women. RG and RR designed the study; RG, TT and RR were first, second and third reviewers, respectively; RG collated and RG and MA analysed the data. RG wrote the first draft of the manuscript under the supervision of RR. All authors critically appraised the manuscript prior to submission. Two reviewers and a third statistical reviewer requested clarification of inclusion and exclusion criteria, study weighting, rationale to support the hypothesis and implications for public health.

In summary, this work demonstrated that published rates of heat illness were consistently lower among women than men, across a range of age, occupation and severity. This was contrary to our *a priori* hypothesis, that women would be at higher risk due to differences in hormonal milieu which contribute to higher core temperature both at rest and during exercise. We discussed possible reasons for the observed gender gap, which include normative behaviours and risk awareness. The implications for public health are that interventions to mitigate heat illness should continue to focus on men. For military policymakers, there is no reason to suspect that women would be at greater risk of heat illness than men in ground close combat roles.

5.1 Abstract

Background: Heat illness (HI) is a growing global concern; its incidence has risen dramatically across the world in recent years. The individual factors whereby elevated core temperature produces HI are not well-understood. Given known physiological differences between men and women pertaining to temperature regulation, we hypothesised that women would be at increased risk of HI than men.

Objectives: We aimed to determine the relative risk of HI in women compared with men through an exhaustive literature review and meta-analysis.

Methods: We searched PubMed and Ovid Medline databases from inception to April 2017. Search terms included all permutations of sex and heat illness (including heatstroke and exertional heat illness) with no language restrictions. We included adult or adolescent human data reporting comparable male and female HI rates. One reviewer identified and screened titles and abstracts. Two independent reviewers applied eligibility criteria. Disagreements were resolved with a third reviewer.

Results: Of 5888 articles identified by searches, 36 were included in the systematic review and 22 in the meta-analysis. The mean (standard deviation) quality score was 3.31 (1.25) / 5. Overall the rate among women was consistently lower than men across the lifespan. The male: female pooled IRR was 2.28 ($p<0.001$, 95% CI: 1.66–3.16). There was modest heterogeneity (between-studies variance (τ^2) = 0.02). The rates did not differ significantly when corrected for severity or occupation.

Discussion: The rate of HI was significantly increased in men compared with women. Risk for HI might be conferred by psychological and behavioural factors rather than physiological ones. Further research is required to delineate which groups are at greatest risk, leading to the development of mitigation strategies against HI.

5.2 Introduction: what is heat illness and how is it caused?

Heat stress describes conditions that favour an increase in body temperature and encompasses factors leading to increased metabolic heat production (e.g. strenuous physical activity) as well as those reducing heat dissipation from the body surface, as with hot and/or humid external environments (**Table 5-1**). When the rate of heat production rises above that at which it can be dissipated, heat storage occurs and body temperature rises. Heat-related illness (HI) may be defined occupationally as incapacitation caused by this rise in core body temperature (Headquarters of the Surgeon General, 2017) and is most severely manifested in heatstroke, which may be fatal.

Exertional heat illness (EHI) Incapacitation resulting from a rise in body temperature (**Headquarters of the** Surgeon General, 2017), during or soon after strenuous physical exertion.
Heat acclimation Experimentally-induced heat adaptation, generally limited to a single set of conditions (Taylor, 2014).
Heat acclimatisation Heat adaptation brought about by residing within a hot climatic region (Taylor, 2014).
Heat adaptation Morphological, chemical, functional, and genetic changes that reduce physiological strain when exposed to heat stress (Periard et al., 2015; Taylor, 2014).
Heat stress Conditions that favour an increase in body temperature (Leon and Bouchama, 2015). Categorised according to whether steady-state conditions prevail (compensable heat stress) or Tc rises without reaching equilibrium (uncompensable heat stress).
Heat strain The physiological responses that ensue as a consequence of heat exposure (Leon and Bouchama, 2015). Encompasses changes in body temperature, as well as sustained changes in regulatory systems e.g. the cardiovascular system.
Heat stroke Defined clinically as a life-threatening illness, characterised by elevated Tc >40 °C and central nervous system (CNS) dysfunction causing delirium, convulsions, or coma (Anderson et al., 1983) Defined pathophysiologically as a form of hyperthermia associated with a systemic inflammatory response, leading to a syndrome of multiorgan dysfunction in which encephalopathy predominates (Bouchama and Knochel, 2002).

Heat stroke, Classic (or non-exertional) Heat stroke resulting from exposure to a high environmental temperature (Bouchama and Knochel, 2002), unrelated to strenuous exercise.
Heat stroke, Exertional Heat stroke resulting from strenuous exercise (Bouchama and Knochel, 2002). May result from inadequate heat loss, excessive endogenous heat production, or both factors in combination.

Table 5-1 Definitions of heat-related illness

Morbidity that may arise with heat stress is often considered as a spectrum, from the more modest heat cramps or rash, to heat syncope (brief loss of consciousness), heat exhaustion (more sustained incapacitation), and heat stroke (HS; a catastrophic rise in core temperature accompanied by severe neurological manifestations) (Raukar et al., 2015). This is not a clinical continuum and illness directly attributable to core temperature rise (i.e. true heat illness) is limited to heat stroke and some cases of heat exhaustion. It is common, however, for other disorders that may arise under conditions of heat stress to be including with reporting of heat illness in the literature, in which case the term heat-related illnesses may be applied. Heat-related illnesses including heatstroke are rising in incidence globally with the warming climate, in developed as well as developing countries (Gu et al., 2016; Nelson et al., 2011; Tran et al., 2013). In Japan, the annual heatstroke number of cases has risen consistently for at least four decades and the incidence of HI in 2010 was 20-fold that in 1970 (Nakai, 2012), while in the United States (US), numerous studies have demonstrated a close association between extreme heat and HI-related mortality, making heat the leading weather-related cause of death (Gubernot et al., 2014, 2015). It is important to note that heat-related illnesses are only a small part of potential heat-related health impacts and epidemiological studies have linked heat exposure with numerous other causes of excess mortality (Bai et al., 2014; Davis et al., 2003), and this is projected to increase with the warming climate (Gasparrini et al., 2017).

Our understanding of the factors conferring greater risk of HI and heat-related illness in military settings derives substantially from studies of male recruits during initial training. Identified risk factors may be usefully divided into extrinsic and intrinsic categories (Stacey et al., 2014). Extrinsic risk factors include increased ambient temperature and humidity, increased intensity of activity and reduced food, water and sleep. Intrinsic risk factors pertain to an individual's predisposition to HI at

any given time and include intercurrent illness (Epstein, 1990), previous HI (Armed Forces Health Surveillance Center, 2017), higher BMI (Wallace et al., 2006), lower fitness test score (Druyan et al., 2012; Kazman et al., 2015), poor acclimatisation (Armstrong et al., 2007), sweat gland function (Gagnon and Kenny, 2012b), alcohol (Ellis, 1972) and consumption of some drugs and nutritional supplements (Lopez and Casa, 2009). With the opening of ground close combat roles to women, it is necessary to consider the relative risk of HI of gender – an intrinsic factor – and its interaction with extrinsic factors from the wider literature encompassing civilians, since so few women have hitherto undertaken the most arduous military roles. No systematic review of the effect of female sex on risk of HI has been published. Given the recent surge in publications of health outcomes related to heatwaves, we undertook such a review with meta-analysis.

Evidence from prospective studies suggest women would be at increased risk of HI. Physical factors such as body mass index, body surface area and percent fat strongly influence body temperature responses to the heat and affect both sexes equally (Notley et al., 2017). However, cyclical changes in oestrogen and progesterone specific to females are associated with thermoregulatory changes: core temperature is higher and there is a rightward shift in the temperature threshold for the onset of vasodilation in the luteal phase or in women using many hormonal contraceptives (Wong and Hollowed, 2017). Evaporative cooling at higher levels of heat stress is reduced in women than men (Gagnon and Kenny, 2012a) and cardiovascular fitness (which may be associated with protection against HI) is on average lower in women than men (Bedno et al., 2014; Wallace et al., 2006). In pregnancy, heat dissipation and thermal stability appear enhanced, conferring fetal protection in the heat (Lindqvist et al., 2003; Vaha-Eskeli et al., 1991). Given the overall thrust of temperature physiology studies, which emphasize differences in thermoregulation caused by biological sex, our *a priori* hypothesis was that women would be at greater risk of HI than men. Behavioural gender differences in physical activity in the heat would be expected to influence HI rates, for example where more physical activity is undertaken in the heat, or more preventive measures (such as active cooling) tend to be made by a particular gender and age. However, this did not impact our hypothesis, since an exhaustive review would encompass a wide range of ages and normative sociocultural gender roles. Understanding the true sex difference might improve understanding of pathophysiological mechanisms or provide a basis for effective, targeted prevention strategies.

5.2.1 Aim

The aim was to detect any signal in the extant literature as to the relative risk of female sex on HI. We hypothesised that women would be at increased risk of HI compared to men. We chose three outcomes; EHI, heatstroke and mortality from heatstroke since these endpoints are the most pertinent (in terms of targeting prevention), most widely studied and well-defined, respectively. To improve generalisability of our findings for targeting prevention, we aimed to perform a sub analysis of age as a covariate.

We then aimed to describe the factors potentially contributing to any observed differences between the sexes through a critical literature review. Such factors are discussed in three categories, namely the physical, physiological and psychological characteristics in which women differ from men. Recommendations for further research are made.

5.3 Methods

The systematic review was registered with PROSPERO international prospective register of systematic reviews (CRD42017064739) and reported according to MOOSE criteria for meta-analysis of observational studies (Stroup et al., 2000).

All studies reporting unselected population-level cases or rates of HI in both men and women were included, where these were comparable: men and women were from the same population, exposed to the same meteorological conditions, and not explicitly or implicitly undertaking different levels of physical activity. We included cross-sectional, prospective and observational study designs, since these fulfilled these criteria. We included adult and adolescent data, using the Medline (OvidSP) and PubMed criteria of adult and adolescent (14 years or older), since children exhibit a different ability to thermoregulatory ability from adolescents or adults and sex differences pertaining to hormonal milieu would not be expected in pre-pubertal subjects (Sinclair et al., 2007; Wong and Hollowed, 2017). Studies not reporting a source population, either stating the population size, or stating case rates (n per *N* population per year; n/N/year), were excluded. Studies where cases were reported without a source population size or case rate, but where the male: female ratio of the source population could be deduced or reasonably assumed, were included in the systematic review but excluded from the meta-analysis. Studies not reporting HI by sex, studies aiming to report HI in the context of another medical condition or

substance use, studies reporting all-cause mortality (not explicitly HI), case reports, case series and case-control studies were excluded.

We searched PubMed and Medline (OvidSP) on 3 April 2017 with no date limitation or language restriction. Our search was limited to humans aged over 14 years (the databases' cut-off for adolescent and adult). Our search strategy and inclusion criteria (**Appendix Tables 5-1** and **5-2**, respectively) aimed to identify articles reporting HI by sex based on an analysis of medical subject headings, terms, and key text words for all permutations of 'sex', 'gender' AND 'heat illness'. Bibliographies of pertinent articles were hand-searched. One investigator (RMG) reviewed all titles to exclude duplicates and no relevant literature. Two reviewers (RMG and TT) then independently judged eligibility of all abstracts. The selection and inclusion process is outlined in **Figure 5-1**. Multiple reports on the same dataset were collated as one. Disagreements were resolved by referral to a third reviewer (RMR). Full-text copies of potentially relevant studies were sourced and independently assessed for compliance with the selection criteria. Articles in languages not spoken by the authorship were translated by colleagues at our institution (see acknowledgements).

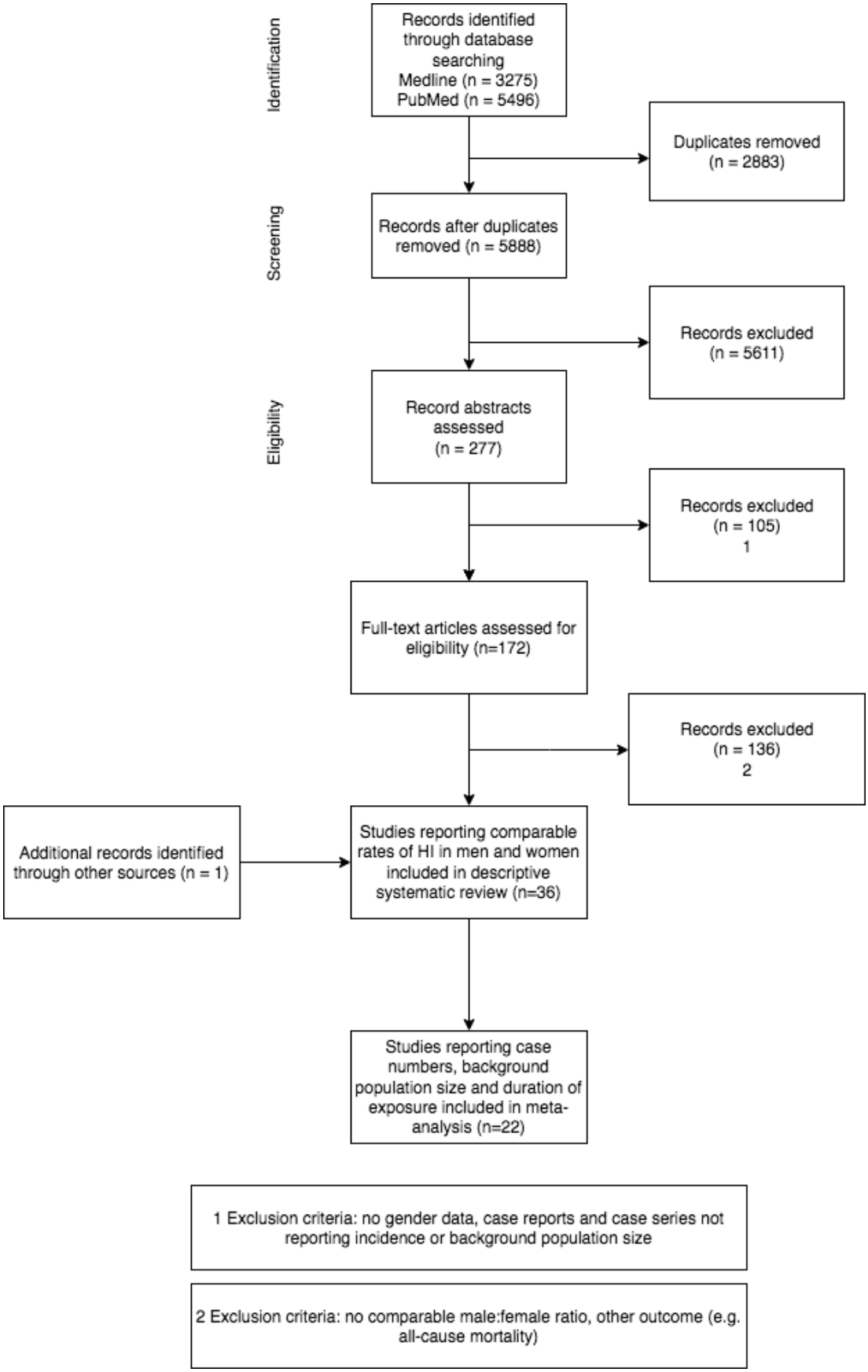


Figure 5-1. Search, selection and exclusion process. HI: heat illness

5.3.1 Data collection, quality and bias assessment

Two investigators (RMG and TT) extracted data independently using a standardised form in Google Forms (Mountain View, California, USA). Extracted data included date and location of study, nature of exposure (e.g. heatwave event duration), male and female cases and population size and characteristics. Where data were alluded to in figures or narrative but were unquantified or incompletely reported, we contacted the authors of those studies up to three times. Fourteen authors were contacted, of whom eight responded with data. Studies reporting numbers of male and female cases, an estimate of background population size and a duration of exposure for HI incidents were included in the final meta-analysis. Where a study reported data from separate datasets and/or more than one outcome type, these were treated as distinct outcomes. A five-point quality and bias assessment scale was completed by the same investigators, awarding one point each for clarity of study objective, validity of diagnostic coding, study population characteristics (and any exclusion criteria likely to introduce bias), generalizability and clarity of statistical methods (**Appendix Table 5-3**). Studies were scored as generalisable if they assessed HI in unselected populations and were not deemed generalisable if they assessed niche populations, such as workers in some specific industry or sportspeople.

5.3.2 Data synthesis

A meta-analysis for incidence-rate ratio was completed using R package (R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.) “Metaphor” version 1.9-9, (Viechtbauer, 2019) A log incidence rate ratio (IRR) was calculated for each study, an effect size representing the ratio of male: female (M:F) incidence rates. Smaller effects indicate less difference between male and female incident rates. Confidence intervals were based on the sample estimate of the population variation; smaller samples were less precise and their confidence intervals were consequently wider. Weighting of each study’s contribution was by its inverse variance, which depended on sample size. A random-effects model was then fitted to calculate the average effect size across studies, assuming heterogeneity. Clinical heterogeneity was tested using τ^2, *I2* Cochrane’s Q statistic. Two sub-analyses were performed: 1) according to four severity outcome types (1 – HI causing death, 2 – heat stroke or hospitalisation, 3 – other HI or emergency department (ED) visit without

hospitalisation, 4 – mild HI), and 2) according to whether HI was occupational or non-occupational, where this information was available. Heterogeneity due to severity of outcome was assessed using an omnibus test comparing pairwise regression for severity types 2, 3, and 4 against type 1, (the referent) (Viechtbauer, 2007). Age-adjusted data were requested from authors, collated and presented in a single format but were not aggregated (due to variation in outcomes and source populations).

5.4 Results

5.4.1 Systematic review.

The searches identified 5,888 results (**Figure 5-1**). Of these, 277 were considered potentially relevant based on title screening and proceeded to abstract review. A total of 172 full text articles were assessed for eligibility and 36 were included, including one article identified from bibliographies (Harduar Morano et al., 2015). Individual outcome data are presented for HI and mortality (**Tables 5-2** and **5-3,** respectively). Definitions used are listed in **Appendix Table 5-4**. Three articles were collated since these reported the same dataset (Nakai, 2012, 2015; Nakai et al., 1999). Thirty-one non-fatal HI outcomes reported in 22 studies demonstrated higher rates of HI in men than women, (**Table 5-2**), of which eight applied to specialist populations (seven from the military and two from sports). Nine non-fatal HI outcomes reported in nine studies demonstrated higher rates in women than men, eight of which applied to specialist military populations (**Table 5-2**). Higher mortality rates in men than women were demonstrated universally (16 outcomes from 15 studies, all reporting general population data; **Table 5-3**).

Reference	Source population and location	Time of study	Description	Outcome		Results	
						Male	Female
[a] Armed Forces Health Surveillance Center (2011, 2012, 2013, 2014, 2015, 2016, 2017)	US Armed Forces (population size NR; cases expressed in person-years)	2010	Medical database interrogation active component US Army, Navy, Air Force, Marine corps or Coast Guard personnel requiring medical intervention (excluding cases treated at field medical facilities)	Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0	285 (0.23)	26 (0.12)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5	2,091 (1.67)	485 (2.32)
		2011		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0	341 (0.27)	21 (0.10)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5	2,099 (1.68)	553 (2.63)
		2012		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0	334 (0.27)	31 (0.15)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5	1,765 (1.44)	492 (2.35)
		2013		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0	293 (0.24)	31 (0.15)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5	1,429 (1.19)	272 (1.30)
		2014		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0	314 (0.27)	30 (0.14)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5	1,410 (0.21)	273 (1.31)
		2015		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0; ICD 10 T67.9	384 (0.35)	33 (0.16)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5; ICD-10: T67.3–T67.5	1 ,625 (1.48)	308 (1.54)

		2016		Cases (rate per 1,000 person-years) of HS	ICD-9-CM: 992.0; ICD 10 T67.9	363 (0.33)	28 (0.19)
				Cases (rate per 1,000 person-years) of other HI	ICD-9-CM: 992.9, 992.3-992.5; ICD-10: T67.3–T67.5	1,749 (1.61)	386 (1.90)
[a] Bai et al., (2014)	Ningbo province, China, (population 7,605,689 in 2010)	3 heatwaves 2011-2012 (30 d total)	Daily collection of data from hospitals via electronic surveillance system	All HI cases	LC	2242	1620
[a] Carter et al., (2005)	US Army (population size NK)	1980-2002	US Army-wide database survey	Incidence density ratio of HI and dehydration (95% CI)	ICD9 992.0-992.9, 276.0, 994.4, 994.5	referent	1.21 (1.09–1.40)
[a] Ellis et al., (1980)	Birmingham, UK (1976 population est. 1,061,800)	24 Jun – 8 July 1976	Survey of local hospital records	Number of recorded admissions for HI	NK	2	3
[a] Fortune et al., (2013)	Workforce of Ontario, Canada (population est. 6,227,900)	2004-2010	Searched national databases of work-related visits to ED for HI	HI presentations (Incidence rate per 1,000,000 FTE months [95% CI])	ICD-10-CA T67, X30	612 (2.2 [2.0-2.4])	173 (0.8 [0.7-0.9])
			Searched state-wide database of work-related health insurance claims for HI	Incidence rate of insurance claim for HI, per 1,000,000 FTE months (95% CI)		419 (1.9 [1.7-2.1])	193 (1.4 [1.2-1.6])
[a] Gu et al., (2016)	China (population NR; cases given as rate/100,000)	1 Jun – 31 Aug 2011–2013	National database survey fed by regional disease information centres	HI cases (rates per 100,000 per y)	LC	3779 (0.55)	1979 (0.42)
[a] Hess et al., (2014)	USA (source population ranges from 298,379,912 for 2006 to 309,349,689 for 2010)	2006-2010	Nationwide retrospective ED survey: stratified sample of 29% of US Eds	Cases of non-hospitalised HS and EHI (frequency per 100,000 per year)	ICD-9-CM 992.0–992.9	197,043 (25.5)	90,676 (12.1)
				Frequency of hospitalised HS and EHI per 100,000 pop/ year		28,591 (3.7)	9801 (1.31)
Huffman et al., (2008)	USA. Nationally representative sample of high schools;	2005-2007	Prospective cohort study in junior athletes – sample of 1,000 US high schools,	Rates of dehydration and heat illness per 100,000 athlete exposures	NK	2.34	0.42

	3,550,141 'athlete exposures' (participation in a practice or competition)		variety of sports. Data from national injury surveillance system.				
Hughson et al., (1980)	Canada (source population size NK)	1978	Track side assessment in 10,000m race.	EHI cases per number of male/ female runners		1.33%	0.88%
Miyake et al., (2012)	Japan (source population size NK)	2010-2011	Japanese database of 94 Eds	Non-exertional heat related illness cases from Japanese national database	NK	1234	503

Table 5-2. Comparison of heat illness rates between men and women.[a] Denotes inclusion in meta-analysis. HS: heatstroke, (E)HI: (exertional) heat illness, ED: emergenc department, LC: local classification; NR not reported, NK: unknown, FTE: full time equivalent

Reference	Source population and location	Time of study	Description	Outcome		Results	
						Male	Female
Centers for Disease Control and Prevention(2006)	USA (population size NK)	1999-2003	Public health records database	deaths resulting from extreme exposure to heat	ICD-10 X30, ICD-10 T67	2,271	1,135
Centers for Disease Control and Prevention(2000)	USA (population size NR; rate expressed per 1,000,000)	1979-1997	Public health records database	Death rate per million due to "extreme weather conditions" in excessive heat	ICD9 E900.0, E900.1, E900.9	0.8	0.4
Ellis et al., (1972)	USA (population size NK)	1952-1967	Surveyed registry records of death	Deaths from excessive heat and insolation (rate per 100,000 population)	ICD E931, 992x	3,381	1,505
[c] Fowler et al., (2013)	Maryland, USA, Ohio, USA, Virginia, USA (population size NR; rate expressed per 100,000)	June 30-Jul 13 2012	surveyed death certificates and medical records	Deaths (rate per 100,000 population) due to HI. HI was underlying cause in 72% and contributing factor in 28%.	ICD9 992.0–992.9 or E E900.0	23 (0.79)	9 (0.31)
	USA (Population NK)	1999-2009				4,955	2278
Gu et al., (2016)	China (population size NK)	(June 1 to August 31) of 2011–2013	National database survey fed by regional disease information centres	Deaths (percentage of all deaths)	LC	124 (3.28)	62 (3.13)

[a] Herbst et al., (2014)	Adelaide, S Australia (population est. 1,600,000)	26 Jan – 7 Feb 2009	Manual search of state coroner data in S Australia	Deaths with heat as cause or contributing factor (annual rate per 1,000 population)	NK	36 (1.26)	18 (0.63)
[a] Mirabelli et al., (2005)	N Carolina, USA	1977-2001	Medical examiners records studied	Work related deaths due to HI	ICD9 692.71, 992.0–992.9, E900.0, E900.1, and E900.9	40	0
				Non-work related deaths due to HI		74	47
[a] Morano et al., (2016)	Florida, USA (all residents from 2005-2012: 148,760,809)	May-Oct 2005-2012	Records from hospitals	Non-work-related heat-related deaths (rate per 100,000 person-years) [rate ratio (95% CI)]	ICD9-CM, ICD 10	94 (0.26) [2.18 (1.53, 3.11)]	45 (0.12) [referent]
	Florida, USA workforce (67,770,884 from 2005-2012)			Work related heat-related deaths (rate per 100,000 person-years) [rate ratio (95% CI)]		22 (0.12) [19.77 (2.67, 146.71)	1 (0.01) [referent]

[c] Nakai et al., (2012, 2015; 1999)	Japan (population NK)	1968-2015	Ministry of Health and Welfare of Japan published and unpublished mortality data	Deaths classified as HI (number)	ICD-8 and ICD-9 992,0 to 992.9; ICD10 T67	7936	5794
[a] Nashold et al., (1996)	Wisconsin, USA (population size NR, reported as rate per 100,000)	Jun 20 – Aug 19, 1995	Coroner findings, National health statistics and , department of health and social services	Heat related deaths (Rate per 100,000 population)	NK	90 (3.6)	64 (2.5
Taylor et al.. (2000)	Alabama, USA (population size NK)	1987-1998	Survey of Public Health records	Cases of heat related death	ICD9 900	71	37
[a] Wheeler et al., (2013)	New York City, USA (source population NR; rate given per 1,000,000	2000-2011	Retrospective survey of death certificates	All death attributed to heat	ICD 9-CM 992.0-992. E900.1, ICD10 X30, T67	85	66

Table 5-3. Comparison of HI mortality reported in men and women. [a] Denotes inclusion in meta-analysis [b] Data from Maryland, Ohio and Virginia included in meta-analysis; USA-wide data not included in meta-analysis. [c] Includes previously unpublished data from continued collection from the Japanese Government Statistics and Information Department database as per Nakai et al. (1999), kindly supplied by S. Nakai. HI, (exertional) heat illness; LC, local classification; NK unknown

5.4.1.1 Sub analysis: age

Age-adjusted data (eight articles, three collated) demonstrate a consistent pattern of higher HI rates among men than women across the lifespan, except for 4 studies (50%) where rates were higher among women than men aged over 70 years (Nakai, 2012, 2015; Nakai et al., 1999; Centers for disease Control and Prevention, 2006) (**Figure 5-2**).

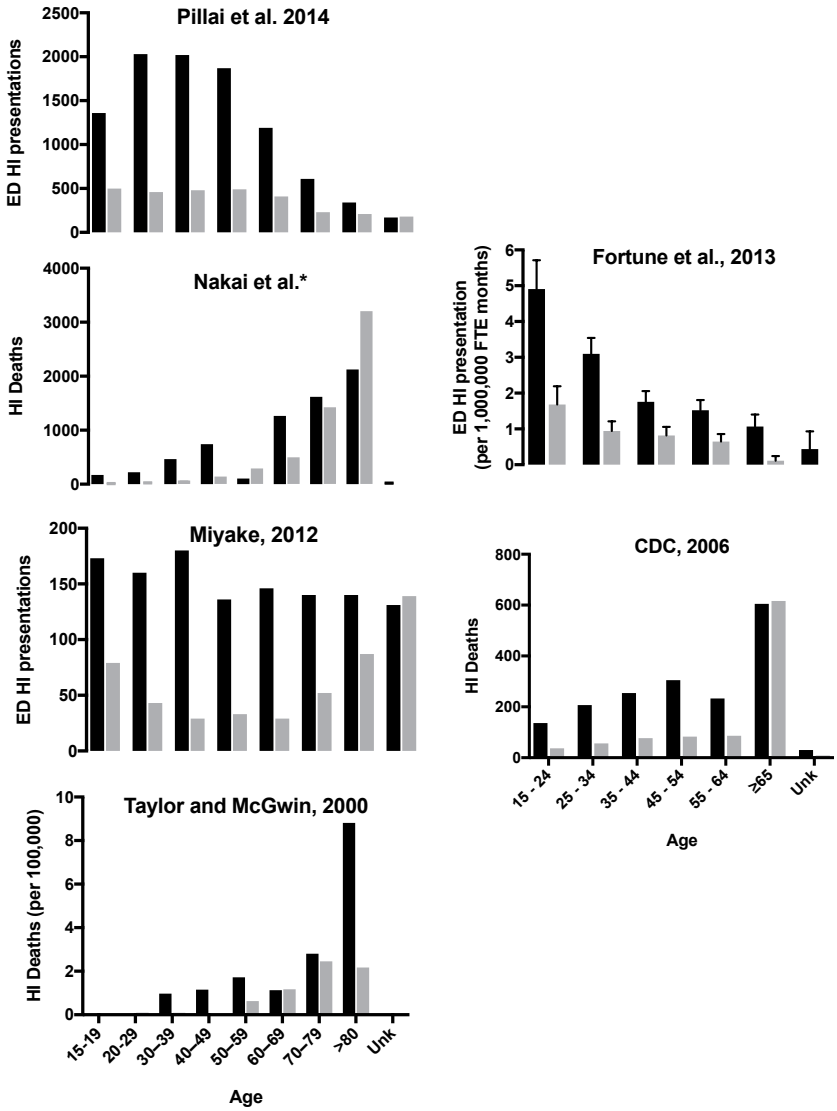
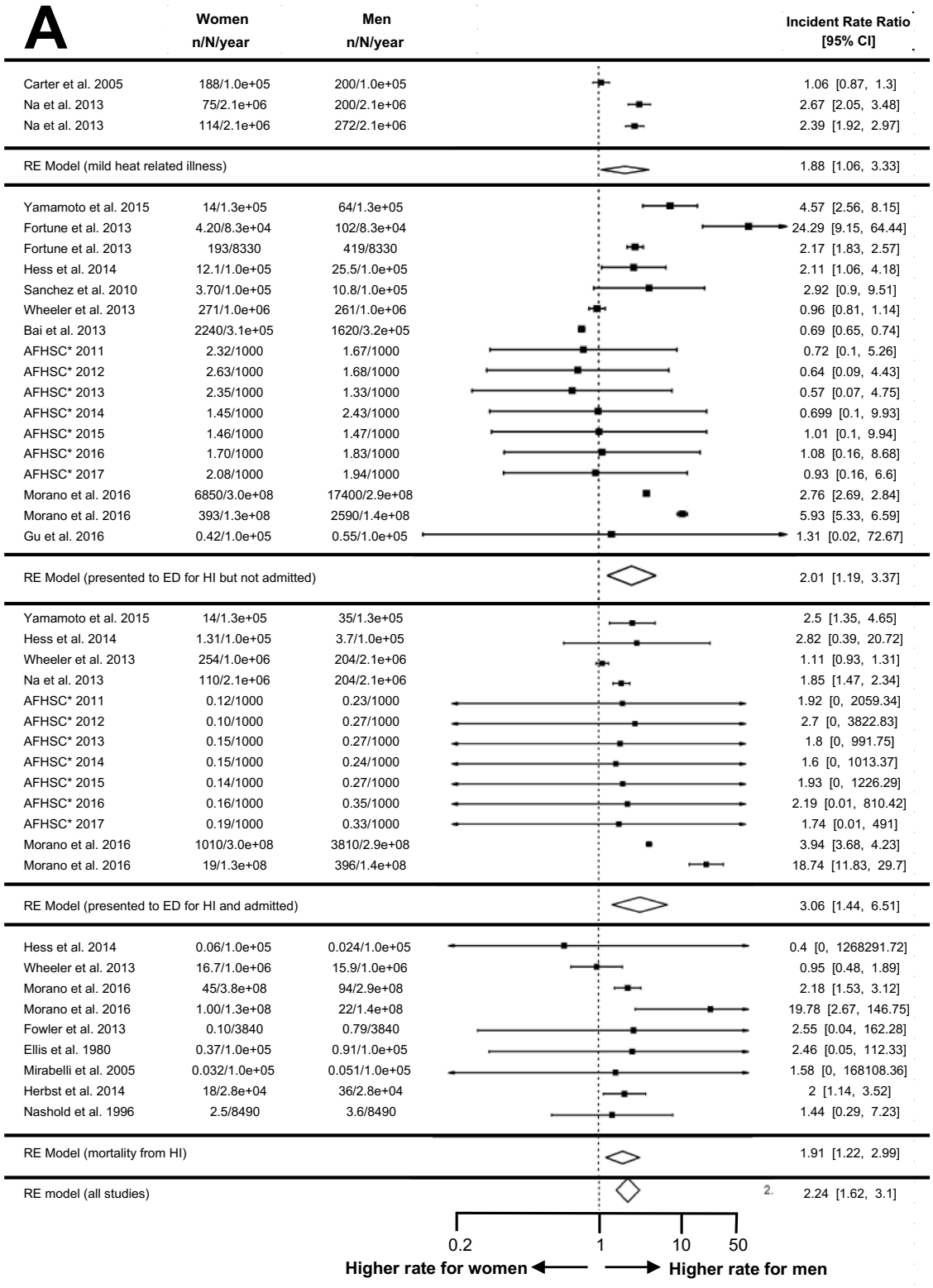


Figure 5-2. Comparison of HI by sex and age where reported. Black bars represent male HI cases, grey bars female. * Synthesis of three publications plus additional data kindly provided by lead author Prof S. Nakai; all other data presented by kind permission of the authors. Left panel: studies grouped where age was reported in decades from age 20. Pillai et al. (2014): comparison of HI-related ED presentations, Georgia, USA (2002-2008), Nakai et al: deaths caused by HS in Japan, 1968-2015 (Nakai, 2012, 2015; Nakai et al., 1999). Miyake (2012): HI presentations to ED, Japan, 2011. Taylor

and McGwin (2000): deaths due to HI in Alabama, USA, 1987-1998 (rate per 100,000). Right panel: HI comparison by sex grouped where age reported in decades from 15. Fortune et al. (2013): work-related presentations to ED for HI in Canada, 2004-2010 (rate [95% CI] per 1,000,000 full-time employment months). Centers for Disease Control and Prevention (2006): deaths caused by HI, USA (1999-2003) [redrawn with kind permission and assistance from Dr Cathy Lansdowne and Dr George Luber, Centers for Disease Control and Prevention]. HI, heat illness; ED emergency department, CI confidence interval; FTE, full-time employment.

5.4.2 Meta-analysis.

Twenty-two articles were included in the meta-analysis. The overall M:F IRR was 2.24 (p<.001, 95% CI: 1.62, 3.10, **Figure 5-3A**). The IRR according to severity of HI was: mild HRI 1.88 (95% CI 1.06, 2.22), moderate HI presented to ED but discharged 1.88 (95% CI 1.24, 2.85), HI admitted to hospital 3.08 (95% CI 1.44, 6.51), and HI-related mortality 1.89 (95% CI 1.22-2.92). The M:F IRR of studies reporting occupational HI (n=3) was 5.66 (95% CI 2.53, 12.64). In studies reporting non-occupational HI (n=2) the M:F IRR was 2.96 (95% CI 2.14, 4.10) (**Figure 5-3B**). There was modest heterogeneity (between-studies variance (τ^2) = 0.02). The percentage of between-studies variance due to heterogeneity (I^2) was 98%; Cochrane's "Q" statistic Q(df=37) = 2404, (p<.001). There was no significant difference in HI severity; QM(df=3) = 1.1180, p= 0.77.


A
Women n/N/year
Men n/N/year
Incident Rate Ratio [95% CI]
Carter et al. 2005 188/1.0e+05 200/1.0e+05 1.06 [0.87, 1.3]
Na et al. 2013 75/2.1e+06 200/2.1e+06 2.67 [2.05, 3.48]
Na et al. 2013 114/2.1e+06 272/2.1e+06 2.39 [1.92, 2.97]
RE Model (mild heat related illness) 1.88 [1.06, 3.33]
Yamamoto et al. 2015 14/1.3e+05 64/1.3e+05 4.57 [2.56, 8.15]
Fortune et al. 2013 4.20/8.3e+04 102/8.3e+04 24.29 [9.15, 64.44]
Fortune et al. 2013 193/8330 419/8330 2.17 [1.83, 2.57]
Hess et al. 2014 12.1/1.0e+05 25.5/1.0e+05 2.11 [1.06, 4.18]
Sanchez et al. 2010 3.70/1.0e+05 10.8/1.0e+05 2.92 [0.9, 9.51]
Wheeler et al. 2013 271/1.0e+06 261/1.0e+06 0.96 [0.81, 1.14]
Bai et al. 2013 2240/3.1e+05 1620/3.2e+05 0.69 [0.65, 0.74]
AFHSC* 2011 2.32/1000 1.67/1000 0.72 [0.1, 5.26]
AFHSC* 2012 2.63/1000 1.68/1000 0.64 [0.09, 4.43]
AFHSC* 2013 2.35/1000 1.33/1000 0.57 [0.07, 4.75]
AFHSC* 2014 1.45/1000 2.43/1000 0.699 [0.1, 9.93]
AFHSC* 2015 1.46/1000 1.47/1000 1.01 [0.1, 9.94]
AFHSC* 2016 1.70/1000 1.83/1000 1.08 [0.16, 8.68]
AFHSC* 2017 2.08/1000 1.94/1000 0.93 [0.16, 6.6]
Morano et al. 2016 6850/3.0e+08 17400/2.9e+08 2.76 [2.69, 2.84]
Morano et al. 2016 393/1.3e+08 2590/1.4e+08 5.93 [5.33, 6.59]
Gu et al. 2016 0.42/1.0e+05 0.55/1.0e+05 1.31 [0.02, 72.67]
RE Model (presented to ED for HI but not admitted) 2.01 [1.19, 3.37]
Yamamoto et al. 2015 14/1.3e+05 35/1.3e+05 2.5 [1.35, 4.65]
Hess et al. 2014 1.31/1.0e+05 3.7/1.0e+05 2.82 [0.39, 20.72]
Wheeler et al. 2013 254/1.0e+06 204/2.1e+06 1.11 [0.93, 1.31]
Na et al. 2013 110/2.1e+06 204/2.1e+06 1.85 [1.47, 2.34]
AFHSC* 2011 0.12/1000 0.23/1000 1.92 [0, 2059.34]
AFHSC* 2012 0.10/1000 0.27/1000 2.7 [0, 3822.83]
AFHSC* 2013 0.15/1000 0.27/1000 1.8 [0, 991.75]
AFHSC* 2014 0.15/1000 0.24/1000 1.6 [0, 1013.37]
AFHSC* 2015 0.14/1000 0.27/1000 1.93 [0, 1226.29]
AFHSC* 2016 0.16/1000 0.35/1000 2.19 [0.01, 810.42]
AFHSC* 2017 0.19/1000 0.33/1000 1.74 [0.01, 491]
Morano et al. 2016 1010/3.0e+08 3810/2.9e+08 3.94 [3.68, 4.23]
Morano et al. 2016 19/1.3e+08 396/1.4e+08 18.74 [11.83, 29.7]
RE Model (presented to ED for HI and admitted) 3.06 [1.44, 6.51]
Hess et al. 2014 0.06/1.0e+05 0.024/1.0e+05 0.4 [0, 1268291.72]
Wheeler et al. 2013 16.7/1.0e+06 15.9/1.0e+06 0.95 [0.48, 1.89]
Morano et al. 2016 45/3.8e+08 94/2.9e+08 2.18 [1.53, 3.12]
Morano et al. 2016 1.00/1.3e+08 22/1.4e+08 19.78 [2.67, 146.75]
Fowler et al. 2013 0.10/3840 0.79/3840 2.55 [0.04, 162.28]
Ellis et al. 1980 0.37/1.0e+05 0.91/1.0e+05 2.46 [0.05, 112.33]
Mirabelli et al. 2005 0.032/1.0e+05 0.051/1.0e+05 1.58 [0, 168108.36]
Herbst et al. 2014 18/2.8e+04 36/2.8e+04 2 [1.14, 3.52]
Nashold et al. 1996 2.5/8490 3.6/8490 1.44 [0.29, 7.23]
RE Model (mortality from HI) 1.91 [1.22, 2.99]
RE model (all studies) 2.24 [1.62, 3.1]
0.2
1
10
50
Higher rate for women
Higher rate for men

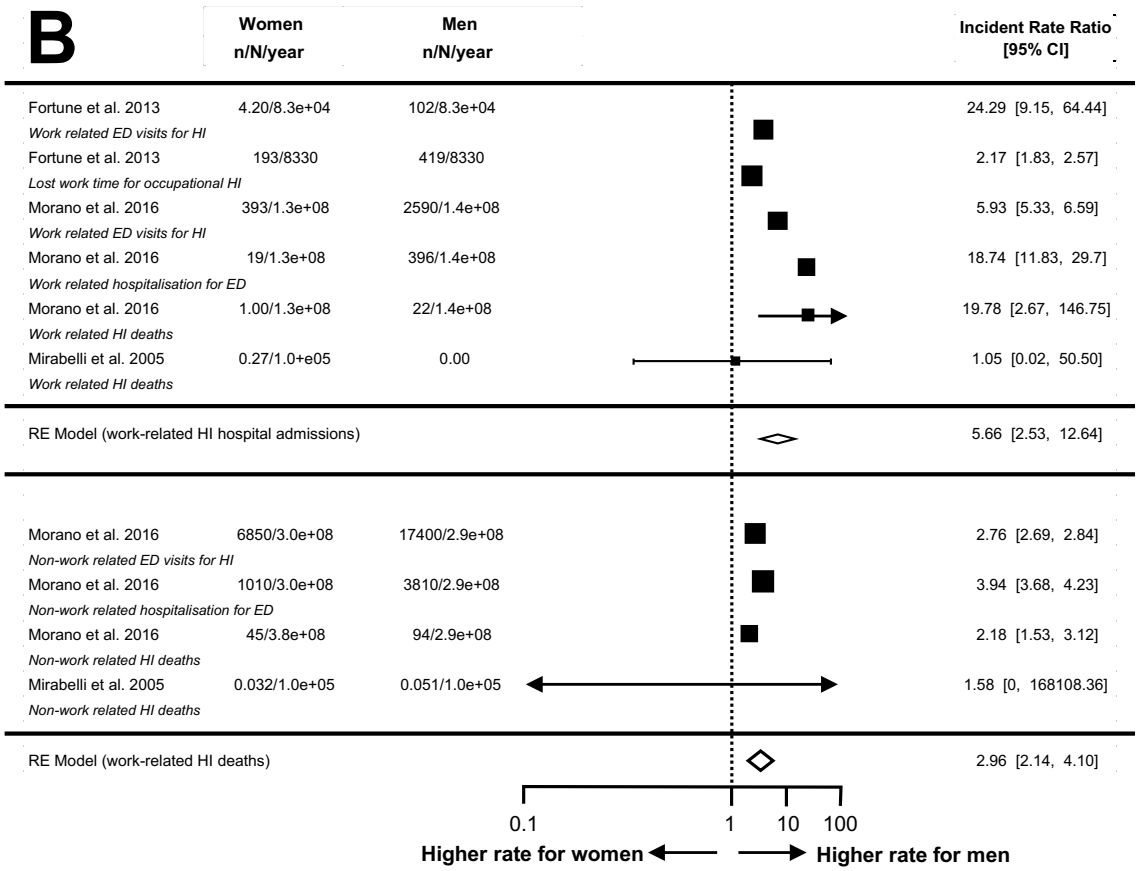


Figure 5-3: Meta-analysis: incident rate ratio for HI outcomes in men compared with women. In some cases, e.g. Gu et al., AFHSC, Hess et al. and Mirabelli et al., CIs are wide due to very small numbers of cases reported relative to population size. 3A: Meta-analysis: incident rate ratio for four heat-related illness outcomes in men compared with women. 3B: Meta-analysis of studies reporting occupational and non-occupational HI. Top panel: occupational HI only. Bottom panel: non-occupational HI only. ED: emergency department, HRI: heat-related illness, CI: confidence interval, n/N/year: incident cases per number population per year, RE: relative effect, AFHSC: Armed Forces Health Surveillance Center.

5.4.3 Quality and bias assessment.

The mean (SD) quality and bias score was 3.31(1.25) of a possible 5 (see *Data collection, quality and bias assessment*) (**Appendix Table 5-3**). Twenty-nine (80.5%) stated a study objective, 23 (63.9%) used validated clinical diagnostic codes, 22 (61.1%) had appropriate population characteristics (no exclusions were made which would introduce bias), 21 (58.3%) were widely generalisable and 21 (58.3%) stated statistical methods clearly.

5.5 Discussion

In our systematic review, higher rates of HI in men than women were evidenced by 27 of 29 outcomes reported from 36 studies. All 13 studies reporting HI-related mortality found higher rates in men than women. The meta-analysis of 22 studies found a M:F IRR of 2.24 (95% CI 1.62 – 3.10). The age-stratified difference in HI events and mortality, where reported, was greatest at a younger age (**Figure 5- 2**) but a greater reported number of events in men largely persisted across the lifespan. Based on a consensus assessment of bias and generalizability, there was moderate quality evidence to a meaningfully increased M:F IRR of HI.

Our findings are consistent with the earliest population studies of HI. Shattuck and Hilferty (1933) observed the HI mortality rate in the USA among men aged 30-70 was threefold that in women, 85 years ago. Since then, with increasing reporting of heat events, studies have reported greater all-cause mortality rates in women than men or no significant difference between sexes (Fortune et al., 2013). Such differences have appropriately been attributed to longevity in women and hence a greater representation of older women with chronic diseases than men contributing to these rates. Our approach excluded studies reporting excess all-cause mortality and consequently data from some heat events, e.g. the 2003 heatwave affecting Europe and elsewhere, is not reported (Huang et al., 2010; Michelozzi et al., 2005; Shen et al., 1998; Toulemon and Barbieri, 2008). We focused on HI specifically due to known physiological sex differences related to temperature. In formulating our hypothesis (that the IRR would favour more HI in women than men), we did not assume that gender-associated differences in occupational heat stress would be consistent across countries (e.g. greater heat stress among men) or would apply to extremes of age. Those typically most at risk of HI are the young and very old, hence we hypothesised a predominance of female cases in these age categories would reflect an increased IRR among women overall. Our hypothesis was rejected, even after accounting for occupational HI and age, where possible.

This is the first published systematic review and meta-analysis examining the impact of sex on rates of HI. This study is relevant and timely; global temperatures and rates of HI are rising (Gasparrini et al., 2017), yet risk factors for the occurrence of HI are poorly understood. Both our systematic review and meta-analysis encompassed a broad range of datasets, including from specific populations in the US military and sports events, occupational registries, hospitals, and local, regional

and national population data. There were insufficient data to undertake meta-analyses for each type of dataset. Military studies demonstrated more mild HI or dehydration among women but more severe HI among men. Sports studies found HI and dehydration were both commoner among men. Similarly, regional hospital databases from Florida and medical examiner records from N Carolina, USA, both found work-related *and* non-work related ED visits and hospitalisations for HI were also commoner among men. While these observations might reflect greater physical exertion among men, more research is indicated to understand the reasons for this, since in the case of sports and military training, men and women could be expected to undergo similar levels of exertion.

Strengths of our study include hard endpoints (such as death and diagnostic coding, as opposed to self-reported symptoms) and quantitative analysis of data, mostly from large health databases, across several countries. Our wide inclusion criteria allowed assessment of the maximum number of HI cases to obtain a true IRR and was intended to reduce selection and confirmation bias. While different populations could theoretically be exposed to different sex-associated HI rates (e.g. occupational registries, military databases and hospital records), in order to be exhaustive, we did not exclude studies unless the heat stress exposure was stated to differ according to sex. We deliberately grouped HI outcomes crudely by severity (HI-related death, heatstroke/ hospital admission, presented to hospital but not admitted, mild HI). We could not categorize illness further by aetiology (e.g. EHI or classical HS), because it these data were frequently unavailable in large datasets. Instead, we documented international classification of diseases (ICD) codes in the systematic review. We were also unable to account for repeat episodes in the same individual. Thus, while we could not reliably state whether studies were measuring precisely the same outcomes, within individual studies, identical outcomes were measured in men and women and both sexes would have been exposed to the same heat event. We excluded studies where the heat exposure to both men and women was unequal. Therefore, we are confident in the result of our meta-analysis, expressed as a ratio. We did not assess the absolute incidence of HI since this would vary greatly (e.g. by varying heat stress).

An unavoidable weakness of our study was that all datasets were formulated from diagnosis by clinicians and subsequent coding (ICD codes are listed in **Appendix Table 5-4**). Identifying HI may be problematic since heat increases risk of

deterioration in chronic diseases and associated mortality (Gubernot et al., 2014). We excluded studies explicitly investigating the relationship between decompensated chronic illnesses (e.g. heart failure) and heat-related hospital admissions, as often it was not clear whether HI featured as a discrete diagnosis in the clinical episodes reported. The investigators of some of the studies included in our meta-analysis used broad outcomes which grouped HI with related diagnoses (e.g. heat-stress associated dehydration (Carter et al., 2005) or sunburn (Mirabelli and Richardson, 2005)) using death, admission or discharge as surrogate measures of severity. The ICD versions 9 and 10 delineates HI some of these related conditions, but nevertheless at the point of diagnosis or classification conflation of HI and other illness may have occurred. The ICD definition of HS changed in 1999 with the introduction of ICD10 clinical modification (CM), although in a longitudinal study examining deaths due to HI diagnoses from 1974 until 2015, ICD10-CM did not affect rate of HS diagnosis (Nakai, 2012).

A further limitation of this review is that we could not consider HS and EHI separately, as most studies did not quantify physical activity prior to the HI event, and none compared activity in men and women. By nature, EHI is a difficult outcome to analyse, since rates of exertion vary greatly and diagnosis tends to be made by bystanders without the benefit of a secondary care medical facility. Greater activity in men could confound the observed M:F IRR. This is particularly pertinent to occupational HI. For example, Gubernot et al. (2015) described a higher occupational heat-related mortality rate in men than women (RR 32.0 (95% CI 17.0-60.0)) but did not describe the gender distribution of workers in different industries, for which RR of heat related deaths varied dramatically (the authors kindly provided us with further data but were unable to clarify HI rates by both employment type and sex). Were men to undergo greater average heat production through their occupation, this would be associated with a higher rate of HI. We have attempted to address this confounder by conducting a sub-analysis of 3 studies where incident-rate data are classified as non-occupational (where a gender imbalance might be less likely) and/or occupational (Fortune et al., 2013; Harduar Morano et al., 2016; Mirabelli and Richardson, 2005). The male to female IRR was higher for occupational HI, of which a significant proportion would be expected to derive from physical exertion, but the same ratio did persist in non-occupational data. The investigators of other studies likely to represent a significant degree of EHI (especially those in athletes and the military) similarly were contacted and could not

provide male and female data adjusted for activity (Armed Forces Health Surveillance Center, 2011, 2012, 2013, 2014, 2015, 2016, 2017; Carter et al., 2005; Huffman et al., 2008; Hughson et al., 1980). However, with increasing age the observed increased rates among men did persist. As age of presentation with HI increases, EHI might be less likely (either from occupation or recreation) and classical heatstroke more so (Miyake, 2013). When data were subdivided by age, we found the greater rate in men persisted across all ages. While differences in occupation would not be expected to affect gender differences in children or the elderly, men are more physically active than women in most countries across the lifespan (Hallal et al., 2012).

5.5.1 Unanswered questions and future research

More prospective research is required to address specific causal factors for the observed sex differences in the development of HI. Only one prospective study was eligible for inclusion in the systematic review and none in the meta-analysis (Hughson et al., 1980). While controlled prospective studies assessing development of HI are generally not ethically viable, there is a plethora of laboratory studies assessing thermal, sudomotor and cardiovascular response to heat in men and women, particularly during exercise. Several excellent reviews have been published summarizing sex-dependent aspects of such studies (Charkoudian and Stachenfeld, 2016; Gagnon and Kenny, 2012a; Kenny and Jay, 2013; Kenny et al., 2016). Sex differences in thermal physiology do not explain our findings, since women have lower body surface area and cardiovascular fitness (Vogel and Friedl, 1992) and a reduced sweating rate compared with men (Gagnon et al., 2013; Nadel et al., 1974), which would increase propensity for HI (Cramer and Jay, 2016; Wallace et al., 2006). However such differences may be of diminished real-world relevance to HI risk (Che Muhamed et al., 2016). Heat stroke is an inflammatory disorder occurring in the context of substantial elevation in core temperature and future research might prospectively examine sex dimorphism in HI pathophysiology rather than temperature *per se*, examining relevant novel markers such as copeptin (Leon and Bouchama, 2015; Stacey et al., 2018b).

It may be that the overall effect favouring more HI in men is blunted during military activities. This could relate to externalised locus of control in the military, where equal work undertaken by men and women could place women at a physiological

disadvantage in the heat, attenuating any behavioural advantage. It will be important to closely observe HI rates as women enter arduous GCC roles.

5.5.2 Physical Differences between the sexes

Physical factors are generally modifiable and include body mass index (BMI), body surface area, fat mass and fat-free mass and VO_{2max} or fitness. Body surface area lends the capacity to dissipate heat while body mass represents the capacity to store heat and varies with body composition. Energy expenditure and metabolic heat production are proportional to body mass. On average men are heavier, have higher body surface area and higher VO_{2max} than women, which would lend a reduced propensity to developing HI (Gagnon and Kenny, 2012a; Vogel and Friedl, 1992).

In a study of 676 US Marine recruits (49 female) who experienced HI and 1802 controls, Wallace et al. (2006) found physical fitness was inversely correlated with risk of HI in both men and women. In men, increasing BMI was similarly associated with increased risk of HI, while in women, compared with a low BMI (<19.8), medium (19.8-23.4) and high (≥23.4) BMI were associated with reduced risk of HI (0.48 (0.21-1.12) and 0.91 (0.30-2.72), respectively). The authors observe this could be due to tighter entry BMI requirements for women than men and postulate that were BMIs higher in the female population, the risk of HI might trend upwards as in men. In similar studies of US Army recruits, increased body fat and reduced fitness were more strongly associated with HI, however too few women were studied for comparison by sex (Bedno et al., 2010; Bedno et al., 2014).

A number of laboratory studies have attempted to correct for physical factors, e.g. by asking participants to exercise at a proportion of VO_{2max} (Frye and Kamon, 1981; Ichinose-Kuwahara et al., 2010; Moran et al., 1999). It is important to note that observations from all such studies, which describe interval changes in haemodynamic parameters such as temperature or heart rate, do not equate to the proinflammatory state of HI, which would of course be unethical to induce in an experimental setting.

Such studies give an indication as to sex differences in response to heat. Interpreting many studies as a true representation of sex difference apart from physical factors is problematic due to confounding by other physical factors

(Gagnon and Kenny, 2012a). Kazman et al. (2015) studied heat tolerance in women and men, defined by reaching predefined cut-offs in heart rate, temperature, symptoms such as nausea, or volitional exhaustion. Women experienced greater heat intolerance than men, however the authors' logistic regression models showed these results could be explained by correction for physical factors: VO_{2max}, which was significantly greater in male participants (52.3 ±7.4 mL/kg/min vs 45.3 mL/kg/min), and greater skinfold body fat (28.4 ±5.1 % vs 18.7 ±5.1 %) and bioimpedance (30.1%+/-21.6% vs 21.6+/-5.1%) in women. In serial experiments performed by 28 males of wide-ranging physical traits, Cramer and Jay (2015) demonstrated 54–71% of individual variability in temperature change and sweating was related to fitness or fatness using stepwise regression. Similar observations were made by Notley et al. (2017) in 36 men and 24 women, with heterogeneity related to specific surface area rather than sex. While it is difficult to distinguish physical factors experimentally, some evidence suggests 'fitness' may be of greater importance than 'fatness'. In an experiment comparing 20 obese women with non-obese female controls at fixed heat production and hydration, Adams et al. (2015) found no significant increase in core temperature among obese women.

5.5.3 Physiological factors

5.5.3.1 Cardiovascular and sudomotor response to heat

While physical factors may be partly responsible for differences observed between men and women in haemodynamic responses to heat stress, they are for the large part modifiable and therefore might be mitigated. The question of a true physiological difference between men and women necessitates controlling for physical factors. Gagnon and Kenny (2011, 2012b) were apparently the first to achieve this by using fixed rates of metabolic heat production, corrected for body surface area and body mass, in conditions permitting evaporation. Their male and female subjects were well-matched for body mass and body surface area with modest differences in VO_{2max}. The authors found greater sudomotor activity and thermosensitivity (i.e. responsiveness of sweating to temperature increase) in men than women, particularly at high at higher requirements for heat loss, suggesting a true physiological difference. Women recruited marginally more sweat glands than men, indicating a significantly greater individual gland output in men. The temperature threshold of sweating onset did not differ between men and women. Vasodilation did not differ between sexes (again when corrected for physical

factors). The cause for this increased sweat output may be central, peripheral or both (Gagnon et al., 2013; Nadel et al., 1974). That sweating and not vasodilation differs between sexes is relevant to heat illness, since sweat evaporation is the more important thermoeffector response in a hot environment (Cramer and Jay, 2016).

Small changes in basal metabolic rate or thermoeffector (2%) responses account for significant change in temp (0.5ºC) (Charkoudian and Joyner, 2004). However it is possible that the sex differences in sweating may be less pertinent to HI risk than other factors discussed below in real terms, since the efficacy of evaporative heat loss is largely determined by the environment. With increasing heat stress and relative humidity (RH), as can occur in a heatwave, evaporation of sweat during exercise and associated heat loss becomes progressively more impaired (Che Muhamed et al., 2016). Sweating alterations undoubtedly contribute to temperature increase which may contribute to heat illness (Yaqub and Al Deeb, 1998), however it may be that in common scenarios of heat stress (in particular elevated RH), the meaningful net effect of reduced effective evaporative heat loss in women is questionable (Gagnon and Kenny, 2012a).

Above a critical core temperature threshold, the area under the temperature-time curve seems to drive the occurrence and severity of heatstroke (Bouchama and Knochel, 2002). Cyclic changes in steroid sex hormones influence both core temperature (Israel and Schneller, 1950) and the threshold for onset of thermoeffector responses (Wong and Hollowed, 2017). Oestradiol is highest in the late follicular phase. This is associated with a slight reduction both in core temperature and the threshold for sweating (Wong and Hollowed, 2017). Oestradiol rises again to a lesser degree in the luteal phase, however a rise in progesterone concentration in the luteal phase (typically 5 to 10-fold, compared with a 2-4 fold rise in oestradiol) is associated with a rise in core temperature of 0.3-0.7C, possibly serving to facilitate implantation of the zygote (Stephenson and Kolka, 1993). Correspondingly, core temperature thresholds for the onset of vasodilation and sweating increase during the luteal phase, although the sensitivity to change in temperature is unaffected versus the follicular phase (Gagnon and Kenny, 2012b; Stephenson and Kolka, 1985).

5.5.3.2 Hormonal adaptation

In order to study thermoeffector responses in isolation, most of the aforementioned laboratory studies were undertaken in the early follicular phase, when concentrations of both oestradiol and progesterone are low (Gagnon and Kenny, 2012a). Thus, the applicability of such studies to HI rates may be reduced, since they do not take into account the broader premenopausal hormonal milieu (Stephenson and Kolka, 1999). Some investigators, such as Charkoudian et al. (1999) have investigated the effect of endogenous cyclical changes in progesterone and oestradiol concentration on thermoeffector response directly. They noted enhanced vasodilation associated with elevated oestradiol and progesterone compared with the early follicular phase, albeit to a modest degree.

Sex hormone status is influenced by body composition and size. Jain et al. (2007) noted that morbid obesity (mean (SD) BMI 48.6 (1.4) kg/m^2) was associated with markedly lower progesterone but no significant difference in oestradiol versus controls with relatively athletic BMI (2 control groups, 21.3 (0.4) and 21.8 (0.5) kg/m^2), over 1 menstrual cycle. While this observation was made at an extreme of BMI, it may be that the higher oestradiol: progesterone ratio protects women against temperature change and/or HI at higher levels of body fat (Friedl, 2005).

Exogenous progestin and oestrogen are taken commonly as hormonal contraceptives; progestins (analogous to progesterone) are the major component in the majority. Progestin-only preparations (such as depot injection and subcutaneous or intrauterine uterine implants) are often preferred by women. The net effect of such contraceptives is to increase core temperature and onset thresholds for thermoeffector responses (Israel and Schneller, 1950; Stachenfeld et al., 2000). Houghton et al. (2005) showed that the bioactivity of progestins is likely to be associated with the temperature threshold for thermoeffector responses.

Overall, the increase in core temperature caused by progesterone does appear to impact the temperature at which women experience HI, but it does not increase the overall risk compared with men. For example, in their study of continuous hot weather training on EHI, Wallace et al, (2005; 2006) found a slightly higher wet bulb globe temperature (a measure of heat stress from temperature and humidity) at the time of HI incident in male than female cases (79.5 ±6.9°F males versus 78.8 ±9.0°F females, no p value given) but overall rates were related to fitness and BMI rather

than sex. Combined contraceptive pills contain both progestins and oestradiol. Stachenfeld et al. (2000) demonstrated that the addition of the oestradiol component reversed the effects of progestin-only contraceptives. Their crossover design in contraceptive-naïve subjects showed no significant thermal difference exercising in heat between the follicular phase and a combined contraceptive. Charkoudian et al. (1997) noted that the core temperature and threshold for vasodilation was increased in users of combined contraceptive, compared with the 5 day placebo phase. These findings in long-term combined contraceptive users indicate that taking the active pill is analogous in thermoeffector terms to the luteal phase. A subsequent study of heat acclimatisation by Armstrong et al. (2005) compared eumenorrhoeic ovulating women with those taking oestradiol plus one of several synthetic progestins or parenteral depot medroxyprogesterone acetate (MPA) alone. They found that after acclimatisation, the onset temperature of sweating was in fact reduced by 0.6 (SD 0.2)ºC in the combined contraceptive group but not the MPA or ovulating groups. However, the overall effectiveness of the training programme and ability to exercise in the heat did not differ between groups, suggesting that the change in thermoeffector response did not infer heat intolerance.

Oestrogen has a protective role, favouring lower body temperature in normal ovulation, when co-administered with progestins and even in women taking lower bioactivity progestin-only preparations (Houghton et al., 2005). Investigators have sought to demonstrate the means by which oestrogen reduces temperature and thermoeffector thresholds (Hayashi et al., 1995; Houghton et al., 2005; Stachenfeld et al., 2000). Changes in core temperature and sweating are likely to be mediated by effects of sex hormones on central autonomic nuclei, while vasodilatory effects of oestrogen appear to be mediated locally (Charkoudian and Stachenfeld, 2016). The effects of oestrogen are reversed by inhibitors of nitric oxide synthase (NOS), indicating that the protective effect of oestradiol is via upregulation of NOS (Johnson and Kellogg, 2010; Wong and Hollowed, 2017).

Overall a potential protective effect of oestrogens against HI, which appear to be only partially modified by progesterone in the midluteal phase, may go some way to explaining our findings. The clinical relevance of the effects of female sex hormones is not well understood. Further research is needed to detect HI rates across a range of variables including BMI, body composition, menstrual cycle timings and hormonal contraceptive use. This would need to be prospective and of sufficient size to allow

linear regression analysis adjusted for multiple variables. It would also be important to study newer and commonly prescribed hormonal contraceptives. Unfortunately, such study would need to be very large and would be unlikely to be feasible.

5.5.4 Behavioural differences

It has been argued that the nuances of female physiology discussed here might not be of meaningful relevance to EHI rates when corrected for modifiable, physical factors (Marsh and Jenkins, 2002). Behavioural differences between men and women have been less widely studied than physiological ones. They pertain to what have previously been considered 'extrinsic' risk factors, i.e. the effect of a task given to the individual in the heat, and the nature in which this is given, which can predispose individuals to HI (Moore et al., 2016). It is known that in the setting of deaths from EHI, a key aetiological factor can be intentional exercise; this may be volitional or imperative (Moyce et al., 2016). However, the approach an individual takes to task completion in the heat is intrinsic to that individual. For example, personality is an important modifier of sympathetic drive and therefore the physiological response to the heat (LeBlanc et al., 2003). It therefore seems appropriate to discuss broad differences in heat-related attitude and behaviour between the sexes.

Attitudes to heat have been studied across the world. In a study of perception of heat waves in the Licheng district of China, Li et al. (2016) demonstrated greater knowledge about heatwaves and their risk among men than women, yet a significantly more circumspect attitude and behaviour with respect to HS prevention in women than men. Similar gender-related differences in risk perception (i.e. an understanding of risk that affects behaviour) were inversely correlated with incidence of HI elsewhere in China (Liu et al., 2013). Heat-related mortality occurring during or after use of a Finnish sauna was far higher among men than women, which was likely due to a greater tendency for reckless behaviour (e.g. excessive alcohol consumption) among men (Kenttamies and Karkola, 2008). A survey of online respondents in the UK showed women were clearly more likely to undertake a range of personal or home protection measures (for example, closing curtains during the day, opening windows at night, avoiding physical exertion or using fans or air conditioners) than men in the heat (Khare et al., 2015).

The gender gap in risk taking is widely accepted. In a systematic review and meta-analysis of gender and risk perception, Byrnes et al. (1999) show a consistently greater propensity for risk taking in men than women, especially in physical and intellectual tasks (despite a significant heterogeneity in both study design and definition of risk). While the overall impact of risk perception on heat-related mortality is not understood (Boeckmann and Rohn, 2014), different behavioural responses to heat between men and women might nevertheless contribute to the observed rates of incidence and mortality.

Differences in tasks undertaken vary significantly by gender. For example, women in SE Asia are less likely to be employed outdoors in strenuous physical activity than men (Burkart et al., 2014; Tawatsupa et al., 2013). In the USA, a nationwide study of Emergency Departments reported that men and boys were far more likely to be undertaking strenuous activity in the heat leading to HI than women or girls (Hess et al., 2014). Such observations are probably largely due to sociocultural normative behaviours, but it is possible that there is also a natural volitional difference between genders during heat. With the gender gap of employment opportunities closing and normative gender roles blurring in many societies today (including the opening of ground close combat roles to women in the military) it may be that the differences observed in this review reduce in future. However, sex differences in task completion in the heat have not been studied widely. The choice of task in the heat may relate to sex-associated intrinsic factors, such as increased risk perception and/or increased sensitivity to the symptoms of early HI among women.

Qualitative studies reveal important gender differences in perception of heat symptoms. For example, older women were more likely to have experienced symptoms during a heat wave than men in Adelaide (Nitschke et al., 2013). In a cross sectional study of 9 cities with a wide socioeconomic and age demographic in Canada, Belanger et al. (2014) found women were more likely to report adverse symptoms pertaining to heat illness than men (OR 1.7 (1.5-2.0)), possibly reflecting increased sensitivity or awareness of HI. Data from the (Armed Forces Health Surveillance Center, 2011, 2012, 2013, 2014, 2015, 2016, 2017) indicate men in the military appear more likely to sustain HS while women are more likely to sustain other HI. While the major factor contributing to this observation is likely to be the difference in activity undertaken by sex, one could postulate it may reflect men presenting later due to different perceptions of symptoms (and/or perception of risk).

More subjective areas such as heat symptom perception and heat-associated risk perception may be important contributors to the risk of HI and warrant further investigation.

5.5.5 Implications for policymakers

Given our findings, behavioural gender differences could be of at least as much importance to real world HI prevention than physical or physiological sex differences (Marsh and Jenkins, 2002). Our findings have implications for public health policy, especially in hot climates. For example, in settings of high occupational heat stress, HI might be prevented in male-dominant workforces through policies like regular break enforcement. Broader legislative changes protecting men in the heat may be considered by public health advisers, and fiscal considerations, such as the cost to certain sectors of the economy of greater HI among men, may be weighed against cost barriers of air conditioning or other domestic or workplace cooling options (Lundgren-Kownacki et al., 2018). While the data we have reported come from a wide range of countries, regulatory considerations ought to be proportionate to need (absolute HI rates) and the prevailing gender normative. It could be beneficial to measure the effectiveness of such interventions using database studies, comparing gender differences in HI before and after.

In the studies incorporated in our meta-analysis, it may be that a greater proportion of men than women were subject to heat stress arising from volitional activities (recreational, or self-paced working) or non-volitional tasks (e.g. military exposures such as marching at group-pace in formation), in which gender-dependent factors such as risk perception and sociocultural normative behaviours operated to predispose greater male risk of HI (Burkart et al., 2014; Stacey et al., 2015; Tawatsupa et al., 2013). Qualitative studies reveal important gender differences in perception of heat symptoms, with women more likely to report symptoms in a heat event than men (Belanger et al., 2014; Nitschke et al., 2013). This may lead to later presentation among men and increase the likelihood of men experiencing more severe HI than women (Armed Forces Health Surveillance Center, 2011, 2012, 2013, 2014, 2015, 2016, 2017). As the gender employment gap closes and normative gender roles blur in many societies today, it may be that the differences observed in this review reduce in future. Further qualitative research is required to explore gender-associated differences in HI symptom perception and risk perception

and their associations with hard endpoints such as heat-related mortality (Boeckmann and Rohn, 2014).

The studies included are population average outcomes, hence we have assessed association not causation. Our data cannot delineate if the factors outlined here or others, such as differences in behaviour, body mass index, hydration, nature intensity and duration of exercise, medication use, ambient temperature, humidity or clothing are responsible. However, our consistent observation from diverse populations reduces the likelihood of any one factor being culprit.

We hypothesize our findings relate to a pan-cultural, perhaps even innate, gender normative. Factors that relate to differences in heat illness, including both physiological and behavioural factors (such as occupation), are likely to contribute to our finding. The implications are two-fold. For researchers, prospective studies of well-matched men and women are required to understand causal risk factors for HI; ours is not the appropriate study design to assess the aetiology of heat illness. For policymakers, the immediate implication of our findings is that a public health message could be indicated, targeting men. It is not clear if it is the tasks being undertaken in the heat, or the manner in which tasks are undertaken, which are responsible for the increased rate of HI. Trials of increased awareness and intervention strategies focused on men should be considered.

5.6 Conclusion

We found a reduced overall observed rate of HI in women compared to men at the population level. This applied to deaths, hospital admissions, hospital presentations and mild HI across the lifespan. This observation was consistent across four decades of data. The increased rate of HI and mortality in men may relate to a range of factors described here and further research is required to define these more clearly. A greater understanding of the factors increasing morbidity and mortality from HI is needed as global temperatures continue to climb. It may be that men might benefit from developing behavioural strategies which protect women from HI, such as earlier reporting of symptoms, protective behaviours and perception of the HI threat.

5.7 Appendices

Heat stress disorders (MeSH)	AND:
OR heat illness	female (MeSH)
OR environmental heat illness	OR women (MeSH)
OR heat injury	OR wom$
OR heat stress	OR Sex
OR heat stroke	OR Gender
OR thermal stress	
OR exertion-associated collapse	
OR exertion-associated hyponatremia	
Limited to: age greater than 14/ adolescent and adult, human studies	
No location filters/ exclusions	

Appendix Table 5-1. Search Strategy. OR: odds ratio, MeSH medical subject heading

Inclusion criteria	**Exclusion criteria**
Adult or adolescent	Studies not reporting HI by sex
Cross sectional, prospective or observational (unselected population-level data)	Different heat exposure or definitions between men and women
Comparable male and female HI rates reported	All case reports, case series, case control studies (relative male/female HI rates not reported)
	Studies reporting all-cause or categorised all-cause mortality (e.g. categorised by 'cardiovascular, respiratory or external causes') not explicitly heat illness.
	Defined sex distribution of exposed population (assumed to be approximately 50:50 in general population studies)

Appendix Table 5-2 Inclusion and Exclusion Criteria. HI, heat illness

		Is study objective stated clearly?	Valid clinical diagnosis code (e.g. ICD10)?	Appropriate study population characteristics (representative - e.g. any excluded) /size?	Are the findings widely generalisable?	Data analysis: and statistical method described clearly?	Score
AFHSC (2011)*	SR MA	1	1	0	0	1	3
AFHSC (2012)*	SR MA	1	1	0	0	1	3
AFHSC (2013)*	SR MA	1	1	0	0	1	3
AFHSC (2014)*	SR MA	1	1	0	0	1	3
AFHSC (2015)*	SR MA	1	1	0	0	1	3
AFHSC (2016)*	SR MA	1	1	0	0	1	3
AFHSC (2017)*	SR MA	1	1	0	0	1	3
Bai et al. (2013)	SR MA	1	0	1	1	0	3
Carter et al. (2005)	SR MA	1	1	0	0	1	3
Centres for Disease Control and Prevention (2000)	SR	0	0	1	1	0	2
Centres for Disease Control and Prevention (2006)	SR	0	0	1	1	0	2
Ellis et al. (1980)	SR MA	1	0	0	0	0	1
Ellis et al. (1972)	SR	1	0	0	0	0	1
Fortune et al. (2013)	SR MA	1	1	0	0	1	3
Fowler et al. (2013)	SR MA	1	0	1	0	1	3
Gu et al. (2016)	SR MA		0	1	1		2
Herbst et al. (2014)	SR MA	1	0	1	1	1	4
Hess et al. (2014)	SR MA	1	1	1	1	1	5

Huffman (2008)	SR	1	1	1	0	1	4
Hughson et al. (1980)	SR	0	0	0	0	0	0
Mirabelli et al. (2005)	SR MA	1	1	1	1	1	5
Miyake (2012)	SR	1	0	1	1	1	4
Adcock et al. (2000)	SR	1	1	1	1	0	4
Morano et al. (2016)	SR MA	1	1	1	1	1	5
Na et al. (2013)	SR MA	1	1	1	1	1	5
Nakai et al. (1993)	SR	1	1	1	1	0	4
Nakai et al. (1999)		0	1	1	1	0	3
Nakai et al. (2012)		1	1	1	1	0	4
Nashold et al. (1996)	SR MA	0	0	1	1	0	2
Pillai et al. (2014)	SR	1	1	1	1	1	5
Piver (1999)	SR	1	1	0	1	1	4
Sanchez et al. (2010)	SR MA	1	0	1	1	0	3
Taylor et al. (2000)	SR	1	1	1	1	1	5
Toloo et al. (2014)	SR	1	1	1	1	1	5
Wheeler et al. (2013)	SR MA	1	1	1	1	0	4

Appendix Table 5-3. Quality and bias assessment. SR included in systematic review. SR MA: included in systematic review and meta-analysis. AFHSC: Armed Forces Health Surveillance Center

ICD9*		ICD10*	
992	Heat stroke and sunstroke	**T67.0**	Heatstroke and sunstroke
992.1	Heat syncope	**T67.1**	Heat syncope
992.2	Heat cramps	**T67.2**	Heat cramp
992.3	Heat exhaustion, anhidrotic	**T67.3**	Heat exhaustion, anhidrotic
992.4	Heat exhaustion due to salt depletion	**T67.4**	Heat exhaustion due to salt depletion
992.5	Heat exhaustion, unspecified	**T67.5**	Heat exhaustion, unspecified
992.6	Heat fatigue, transient	**T67.6**	Heat fatigue, transient
992.7	Heat oedema	**T67.7**	Heat oedema
992.8	Other specified heat effects	**T67.8**	Other effects of heat and light
992.9	Unspecified effects of heat and light	**T67.9**	Effect of heat and light, unspecified
E900	Accident caused by excessive heat due to weather condition	**X30**	Exposure to excessive natural heat
E900.1	Accidents due to excessive heat of man-made origin	**W92**	Exposure to excessive heat of man-made origin
E900.9	Accidents due to excessive heat of unspecified origin		
692.71	Sunburn		

Appendix Table 5-4. International Classification of Diseases definitions (heat illnesses) definitions. * where ICD classifications use the suffixes CM and CA, these refer to clinical modification and Canadian, respectively.

Chapter 6 Conclusions

Women's health is a niche that is becoming increasingly relevant to the military. Female participation in the military is increasing generally: in USA in 2000, 8% of military personnel were female; in 2019 this figure is 16% (18% of Officers and 15% of enlisted personnel) (Maddox, 2019). Furthermore, with ground close combat (GCC) roles now open to all genders, women will be exposed to more arduous training.

This PhD programme aimed to address concerns about the potential risks to which women in the military are exposed because of their occupation. These risks are not all well-defined, but include higher risk of stress fracture than men, potentially higher adverse risk of psychological outcomes than men and potentially higher fertility service referral rates than aged-matched civilian women. Since these conditions relate to the milieu of hormones, which regulate bone turnover, neuroendocrine axes and fertility (and which effect differences in biological sex), I aimed to explore hormone function and adaptation in women during military training. The risks to military women and the rationale for characterising hormones during training were discussed in a review article, which set out the groundwork and hypotheses for this PhD programme (**Appendix A**).

Our studies were categorised into endocrine responses to basic military training and arduous exercise in extreme cold, and sex differences in heat illness. Exploring ovulation and metabolic function during basic military training necessitated developmental work to create novel methodological approaches.

The *a priori* hypothesised was that reproductive dysfunction, associated with reduced bone mineral density (BMD) and potential subfertility, was the cumulative, negative effect of individual and environmental factors (**Figure 6-1**).

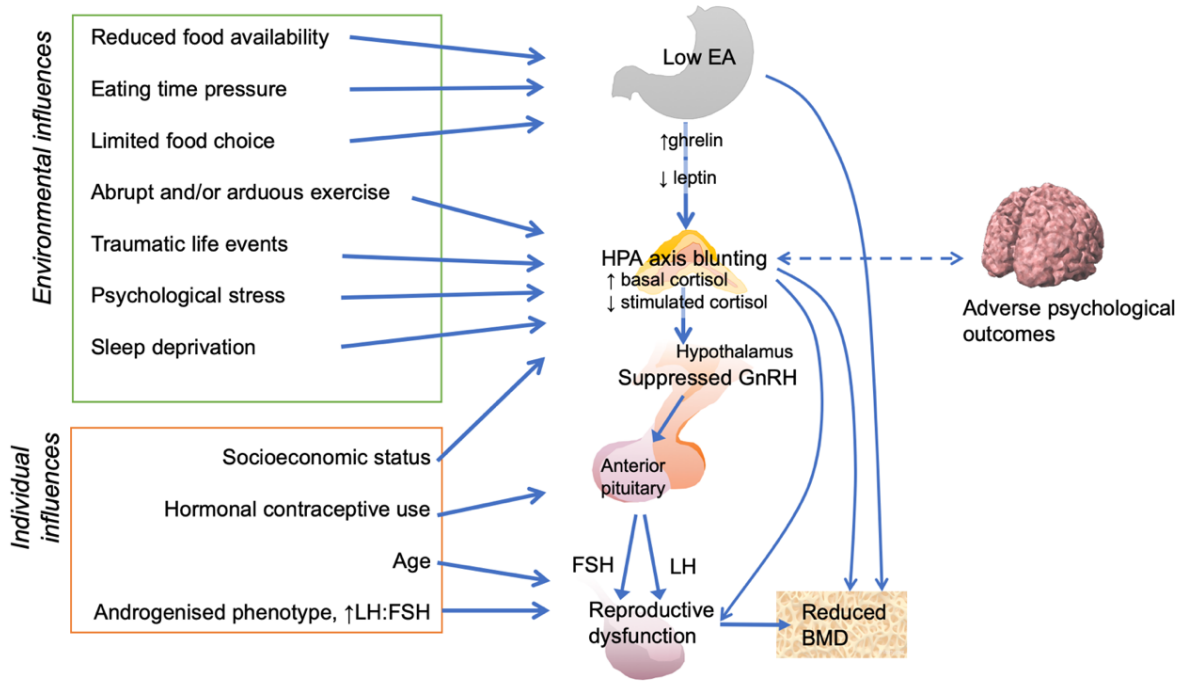


Figure 6-1. A priori hypothesis: the endocrine effects of arduous training are an accumulation of individual and environmental influences. Arrows represent direction of action. Dashed line represents indirect bidirectional relationship. EA: energy availability, HPA hypothalamic-pituitary-adrenal, GnRH gonadotrophic releasing hormone, FSH follicle stimulating hormone, LH luteinising hormone, BMD bone mineral density

6.1 Summary

6.1.1 Background

In the past, scientists have been circumspect about commending arduous physical training to women, citing concerns about the delicate balance of the female reproductive hormones (Drinkwater, 2015; James, 1998). The International Olympic Committee (IOC) only allowed the first female distance race in 1972 (1500 m) and the first women's marathon was not until 1984. However, in the last 20 years, researchers have uncovered low energy availability (EA) as the putative cause of an array of hormonal and metabolic disturbance in female athletes (EA is defined as energy intake minus exercise energy expenditure) (De Souza et al., 2019a). Citing this work, the IOC no longer restricts participation on the basis of gender. IOC consensus statements have suggested that sufficient energy intake will protect women (and men) from negative metabolic and endocrine consequences of sports participation (Mountjoy et al., 2014; Mountjoy et al., 2018). In an important subset of

athletes, high levels of androgens contribute to a different pattern of reproductive dysfunction (Javed et al., 2015), however in these athletes too low EA is also thought to be a major contributor to suppression of the hypothalamic-pituitary-gonad (HPG) axis (Rickenlund et al., 2004).

In a military context, where studies have identified the suppression of reproductive function (in men and women), this has been in association with intensive exercise and/ or dietary restriction (Anderson, 1979; Friedl et al., 1992; Henning et al., 2014; Hill et al., 2015; Hoyt and Friedl, 2006; Karl et al., 2016; Lauder et al., 1999; Nindl et al., 2007; Nindl et al., 2001; Welch, 1989). Since military training is highly distinct from training for sports, I designed a series of prospective field study measures to test multiple endocrine, metabolic and psychological domains concurrently. The central hypothesis was that low EA will be associated with suppression of reproductive function during military training (as per the IOC consensus statements on sports), while seeking to explore other facets of military training-related stressors, and phenotypic nuances of the trainees such as androgenisation.

6.1.2 Female Endocrinology In Arduous Training (FEAT) Study

6.1.2.1 Methodological innovation

In order to measure reproductive function in women, it is necessary to assess if ovulation has occurred, but studies in the military have hitherto failed to do so, (Appendix A). While measuring ovulation is achievable in a laboratory setting, using daily blood or urine samples or regular pelvic ultrasound, in a ‘real world’ setting or field environment it is challenging. Over long durations like the Commissioning Course (Course), high numbers of samples need to be taken, and the current assays for metabolites of progesterone and oestradiol are onerous and expensive. I therefore validated a novel, automated assay for urinary progesterone and creatinine, which demonstrated robust characteristics in identifying ovulation alongside gold standard methods (**Section 3.1**). This work was subsequently published (Gifford et al., 2018b).

In recent years the concept of EA has become central to the current understanding of endocrine adaptations to physical activity in women, acting via energy-conserving suppression of metabolic and endocrine systems. However, EA is challenging to measure in a field setting: it requires measurement of energy intake which is

notoriously difficult (Capling et al., 2017) and exercise energy expenditure, which presents challenges in a military field setting where load carriage increases the energy cost of walking or running (Burke et al., 2018). An assessment of an EA measurement in military training was presented in **Section 3.2**. This involved close collaboration with the engineering team who developed the GENEActiv ® triaxial accelerometer (Activinsights, Cambridge, UK), who created a syntax which robustly interpreted exercise energy expenditure from this device, compared with total activity energy expenditure based on a gold-standard measure (doubly-labelled water).

Studies of endocrine adaptation to military training have hitherto universally relied on basal measurements of hormones. However, the endocrine system is labile and dynamic, demonstrating marked within– and between–day variability. Since the finding of endocrine axis suppression was anticipated during training, this PhD programme utilised a novel means of testing the HPA and HPG axes concurrently, termed the Combined Hypothalamic-pituitary-gonadotrOph-adrenal Cortex (CHOC) test. The results of the CHOC test, described in **Sections 3.3, 3.4** and **4.1**, demonstrate its ability to detect subtle changes in gonadotroph function regardless of contraceptive use.

6.1.2.2 FEAT Study findings

After ethical approval was obtained in January 2017, the FEAT Study was commenced in July 2017. It aimed to explore changes in key hormonal, physical and metabolic traits in women during the 44-week Commissioning Course (Course) at the Royal Military Academy, Sandhurst (RMAS). Seventy seven women attended a study briefing 6 to 12 weeks before the Course, of whom 9% (12%) declined to participate, and a further 7 (9%) withdrew on or before week 1 of the Study. Fifty-two women completed the Study.

Due to underreporting of dietary intake of around 30%, it was not possible to determine EA precisely (**Section 3.2**). However, even allowing for this, both accelerometry and doubly-labelled water–based EA measures were associated with physical benefits of training (gain in lean mass, loss of fat mass, improved 1.5 mile run time) and, as expected, were inversely associated with measures of eating disorder score. The paradoxical findings in relation to body composition (that fat mass loss was greater in participants with greater EA) could be accounted for by the

energy compensation phenomenon, whereby individuals increase consumption of calorie-dense foods after periods of relative energy insufficiency (in this case, field exercises) (McNeil et al., 2017). Autonomic function demonstrated progressive improvement throughout training (i.e. greater parasympathetic activity at rest), however we did not detect any association between EA and autonomic function. This could be explained by the inaccuracy measurement of energy intake.

Section 3.3 demonstrated that despite psychological measures suggesting increased stress during basic military training, with detrimental effects on mood, basal and dynamic function of the HPA axis were not impaired. Morning saliva cortisol demonstrated an initial stress response, followed by habituation throughout the course, while fasting serum cortisol reduced from the beginning to the end of training. Dynamic cortisol response in the CHOC test was reduced marginally by the end of training in participants not using a combined contraceptive pill. Average cortisol concentration in hair rose throughout, which I interpreted as being related to ongoing physical activity. Overall, the pattern was consistent with healthy adaptation of the HPA axis to training, despite evidence of stress and sleep deprivation.

Section 3.4 aimed to detect longitudinal changes in reproductive function during training, examining the HPG axis and gonadal function, comparing changes in women alongside bone turnover and bone mineral density. Women using hormonal contraceptives were included (as far as I am aware, this was the first study to do so in a military setting), to increase relevance to the expected real-world hormonal milieu of women of childbearing age. The findings demonstrated that responsiveness of the HPG was suppressed after 28 weeks of training. Fasting blood tests after 28 weeks and 44 weeks of training demonstrated an increase in inhibin B (a marker of large ovarian follicles), oestradiol and follicle-stimulating hormone (FSH), while anti-Müllerian hormone (AMH), a marker of small follicle reserve, was unchanged. These findings suggest dysregulation of follicular development: an increase in the size and number of larger follicles under the influence of FSH and oestradiol, without ovulation, but no meaningful change in the number of smaller follicles or (it is surmised) long-term reproductive potential. Weekly measurement of urinary progesterone in women not using any hormonal contraceptives other than intrauterine device (n=22) showed marked suppression of ovulation throughout the Course. The suppression of the reproductive function was not associated with biochemical evidence of low EA from fasting metabolic blood

tests. Indeed, insulin, glucose and leptin increased significantly, while cortisol and non-esterified fatty acids decreased significantly throughout training, while the triiodothyronine was unaffected. These findings are generally the opposite of those observed in studies of low EA (Elliott-Sale et al., 2018) and suggest HPG axis disruption caused by an effect other than low EA, such as environmental stressors or androgenism. Androgen levels were unaffected. Alternatively, it is possible that reproductive dysfunction was a legacy effect of periodical low EA during field exercises, which was followed by overeating and low physical activity (energy compensation). Across the study, crude energy balance was positive, evidenced by fat mass change on DXA. An intriguing possibility is that chronic sleep disturbance and deprivation contributed to the hormonal effects we observed (discussed later in Section 6.7.5). More studies are needed to understand the metabolic maladaptation and reproductive dysfunction following basic military training.

6.1.3 Exploring the Endocrine Effects of Extreme Training (E4T) Study

The E4T Study recruited six members of Ex ICE MAIDEN, the first all-female Antarctic traverse expedition, using muscle power alone to haul 80 kg sledges over 1700km, enduring temperatures as low as –50ºC and windspeeds of over 50km/hand (**Chapter 4**). We evaluated female endocrine adaptations to highly arduous exercise in this environment, studying participants before and after the expedition which took place from November 2017 to January 2018. This study has resulted in several publications: reproductive, adrenal and metabolic adaptations (**Section 4.1** by Gifford et al. (2019a)), skeletal adaptations (O'Leary et al., 2019), autonomic adaptations (Gifford et al., 2018a), changes and sex differences in metabolic substrate utilisation (Hattersley et al., 2019a; Hattersley et al., 2019b), psychological determinants of the expedition's success (Blackadder-Weinstein et al., 2019) and practical considerations surrounding selection, training and nutrition (**Section 4.2** by Taylor et al. (2019)).

The central finding of the E4T study was that these women were resilient to the expedition, despite losing around 10kg fat mass (reflecting an energy deficit of over 1,400 kCal/d). Lean mass was maintained, and metabolic and endocrine function were preserved after the expedition compared with beforehand. Luteinising hormone responsiveness increased 10d after the expedition, suggesting it was suppressed beforehand, but not meaningfully suppressed by the expedition itself.

Cortisol responsiveness to ACTH did not change, despite an elevated free cortisol level measured in hair during the months of the expedition. Nutritional biomarkers demonstrated a pronounced decrease in leptin, glucose and non-esterified fatty acids during the expedition. Metabolic substrate utilisation was unaffected by the expedition and in 5 of the 6 participants, fitness levels at sea level and altitude increased afterwards compared with beforehand. Skeletal studies found a modest redistribution of bone mineral content, with early evidence of bone turnover uncoupling, however these findings were significantly less than would be expected during arduous military training with similar energy deficit and duration.

The biological resilience these women demonstrated was matched by expeditionary success; the team finished 15 days ahead of schedule, broke several records and took the total number of women who had ever crossed Antarctica from 4 to 10.

6.1.4 Risk of heat illness in men and women

Laboratory studies of heat strain (the combination of exercise and heat exposure) have consistently demonstrated greater heat storage capacity, and higher core temperatures with increasing heat strain, among women than men. The female sex hormones progesterone and oestradiol contribute to increased core temperature and reduced vasodilation and sweating, respectively. It has been assumed that women would be at greater risk of heat illness than men throughout much of the extant literature. However, investigators only studied changes core temperature, sweating and or vasodilation, not heat illness *per se*. It would be unethical to attempt to induce heat illness in prospective human studies, and occurrence of heat illness is idiosyncratic. Therefore, the best means of assessing sex-associated endocrine adaptation to heat was to perform a systematic review and meta-analysis of published real-world heat illness rates where sex was reported. The *a priori* hypothesis was that women would be at greater risk of heat illness than men was rejected.

The systematic review and meta-analysis in **Chapter 5** identified higher incidence heat illness among men than women, the male: female ratio of heat illness events being 2.24. Sub-analyses found the risk among men was consistently higher than women when data were stratified for age, physicality of occupation and severity of heat illness (from minor heat rash to fatal heat stroke), where these data were reported. We highlighted findings from psychology and behavioural medicine which

suggest innate sex differences in risk awareness and risk-mitigating behaviours may account for this unexpected discrepancy. While the review was limited by the nature of extant data on heat illness (e.g. retrospective, reliant on point-of-care diagnoses and database completeness), these findings represent a marker to guide the direction of future studies on the impact of sex on heat illness. Such studies should seek to understand the occurrence of heat illness *per se*, including the influence of less empiric but modifiable factors like behaviour and attitudes to risk, instead of a narrow focus on physiological adaptations to heat strain.

6.2 Discussion

In studies of extreme cold and extreme heat, women proved to be significantly more resilient than was anticipated from extant literature. Our observations in basic military training were more mixed, with benefits identified in adrenal function, but suppression of the HPG axis. Taken together, these findings do not support the notion that women are not robust enough to undertake GCC, however further research is required to delineate factors which predispose women to HPG axis dysfunction during military training and the extent to which these findings are reversible.

Ex ICE MAIDEN was associated with a significant energy deficit. Hattersley et al. (2019a) based their figure of 686 kcal/d on the lower fat loss they measured of 4 kg, rather than 10 kg we reported in **Section 4.2** and in Taylor et al. (2019). The difference is attributable to measurement timings. The post-expedition figure in Hattersley et al. was measured by air displacement plethysmography, 10 to 16 days of *ad libitum* dietary intake following the expedition, whereas the figure in **Chapter 4** was based on total body mass and estimated fat mass (four-point bioimpedance). 14 days before and 5 days after the expedition. The preserved reproductive and adrenal function of the ICE MAIDENs was in contrast to our findings in basic military trainees, in whom we observed gross energy balance and little convincing evidence of low EA.

Consistent with all literature assessing EA, it was challenging to assess EI, and made even more challenging in the field environment. Our observation of around 30% underreporting of energy intake was consistent with the systematic review of Capling et al. (2017) and was likely due largely to consumption of snacks, uncaptured regular ‘grazing’ and/or missed meals. However, we were able to

measure numerous fasting surrogate biomarkers of EA (e.g. cortisol, thyroid function, leptin, insulin, glucose, insulin-like growth factor 1 (IGF-1), free fatty acids bone turnover markers and micronutrient status) as well as body composition changes using gold-standard measurements. Estimation of EA was not attempted during Ex ICE MAIDEN; we planned *ab initio* to rely on surrogate biomarkers of EA. This meant that these datasets provided comparable assessments of EA. Moreover, in the intervening three years since the E4T and FEAT studies were planned, several studies have ratified the challenges of measuring EA directly (Burke et al., 2018), highlighting instead strong relationships between fasting metabolic surrogates and low EA, including and changes in body composition (reviewed by De Souza et al. (2019); McCall and Ackerman (2019) and Elliott-Sale et al. (2018)).

A comparison of pertinent findings from these studies is laid out in **Table 6-1**:

	E4T Study (**Chapter 4**)	FEAT Study (**Chapter 3)**
Body composition		
Fat mass	⇊ r	↓ then ↑
Lean mass	↔	↑then ↓
Bone mass	Overall ↔; redistribution from trunk, spine and ribs to limbs	Overall slight ↓; redistribution from arms and ribs to legs and pelvis
Total mass	⇊ r	↓ then ↑
Metabolic markers		
HOMA-IR	↔	↑
Leptin	⇊ pr	↑
Zinc	↓ r	↓
TSH	↔	↑[a]
Free T4	↔	↓[a]
CTX	↑	↓
Sclerostin	↔	↔
P1NP	↔	↔
BSAP	↓ r	↔
IGF-1	↔ r	↔
HPA and HPG axis markers		
Cortisol [b]	↑	↑
Basal FSH	↔	↑[c]
Basal LH	↔	↔[c]
FSH response	↔	↓
LH response	↔ pr	↓
Oestradiol	↓ r	↑
AMH	↔	↔
Inhibin B	↔	↑[c]
FAI	↓ r	↔

Table 6-1. Body composition, and reproductive and metabolic basal changes during the E4T and FEAT Studies. Arrows represent direction of change during arduous training. 'pr' and 'r' indicates partial and complete resolution 10 days after Ex ICE MAIDEN, respectively, and '↔ r' indicates no change during but increase following the expedition. [a] noted in hormone-containing contraceptive users only. [b] indicates average hair cortisol [c] noted in progesterone only and non-contraceptive users only. HOMA-IR: homeostatic modelling assessment of insulin resistance, TSH: thyroid stimulating hormone, T4: thyroxine, CTX: c-telopeptide cross-links of type 1 collagen (bone resorption marker), P1NP: amino-terminal propeptide of procollagen type 1 (bone formation marker), BSAP: bone specific isoform of alkaline phosphatase (bone formation marker), HPG: hypothalamic-pituitary-gonad, HPA: hypothalamic-pituitary-adrenal, IGF-1: insulin-like growth factor 1, FSH: follicle stimulating hormone, LH: luteinising hormone, AMH anti-Müllerian hormone, FAI: free androgen index

The profound weight loss and energy deficit noted during the E4T study (**Section 4.1**), was not associated with change in fat-free mass (FFM), REE or substrate utilisation (Hattersley et al., 2019b). Weight-adjusted VO_{2peak} did not change and anaerobic capacity improved (Thake et al., 2019). Differences in adaptation from the FEAT study could be explained by the different type of exercise or the comprehensive nutrition and physical preparation, which took place over the preceding two years (**Section 4.2**)

During basic military training, however, weight change was modest (around 1kg loss by the end of term 1, followed by 1.5 kg gain by the end of term 2, which reverted to baseline by term 3). There was a modest increase in FFM during term 1, but no change in other terms. Thus, the E4T study represented a metabolic model more akin to studies of low EA (demonstrating decreased leptin, oestradiol and increased bone resorption), whereas in the FEAT Study we largely found the opposite (increased leptin, as well as HOMA-IR, increased oestradiol and increased bone formation). Exceptions to this included the thyroid axis, which was unchanged in the E4T study, but TSH increased and free T4 decreased in the FEAT study (free T3 was unchanged), and IGF-1, which did not change during the E4T study then increased after 15 days post-expedition, but did not change during the FEAT study.

6.2.1 HPA Axis findings

Basal (fasting) cortisol has often been noted to increase following low EA (Elliott-Sale et al., 2018); however, such measurements are likely to be confounded by the stress of venepuncture early in the morning (Reynolds et al., 2001b). Fasting

plasma cortisol was observed to decrease in the FEAT study, likely representing habituation to the stress of venepuncture. In the E4T study fasting plasma cortisol did not change significantly. Hair cortisol provides a retrospective record of monthly average concentrations, 3 to 6 months before sampling. Pre-Ex ICE MAIDEN and within- basic military training hair cortisol concentrations were similar. There was a trend towards an anticipatory increase in the months leading up to both studies. This trend continued during basic military training, whereas during Ex ICE MAIDEN, a more marked increase hair cortisol was observed.

Hair cortisol is strongly associated with contemporaneous measures of stress (Stalder et al., 2017), so it was surprising that during Ex ICE MAIDEN, psychological stress levels were reduced compared with beforehand, whereas throughout basic military training, perceived psychological stress, resilience and mood were noted to deteriorate. Hair cortisol concentrations likely reflected the physically arduous nature of Ex ICE MAIDEN compared with pre-expedition training or basic military training (Gerber et al., 2012), and may have been suggestive of overtraining syndrome (Cadegiani and Kater, 2018b). Despite the elevated free hair cortisol during the expedition, the adrenal cortex response to ACTH was preserved. Dexamethasone was administered the evening before the CHOC test in the E4T study to attempt to narrow the variance of baseline cortisol, while in the FEAT study a uniform baseline was attempted by performing the test in the mid- to late-afternoon, when cortisol concentrations reach a natural nadir. However, E4T study participants were unexpectedly sensitive to dexamethasone rendering test responses between the studies incomparable (**Table 6-2**).

	E4T Study	FEAT Study
Baseline cortisol, nmol/L		
CHOC test 1	73 ±75	193 ±95
CHOC test 2	62 ±33	260 ±114
CHOC test 3	54 ±28	–
Peak cortisol, nmol/L		
CHOC test 1	128 ±33	598 ±207
CHOC test 2	140 ±24	669 ±104
CHOC test 3	137 ±27	–

Table 6-2. Comparison of Combined Hypothalamic pituitary gonadotrOpoh adrenal Cortex (CHOC) test cortisol concentration, baseline (pre-test) and peak response to 1 µg Synacthen ®. The CHOC test was undertaken three times in the E4T study (pre Expedition (Ex), Ex +5 days and Ex+16 days), and twice in the FEAT Study (weeks 1 and 28 of training). Data are for progesterone-only contraception users in the FEAT study, since these were used in the E4T study.

E4T Study pre- and post-expedition morning and evening saliva cortisol concentrations were similar to basic military trainees at the start of the Commissioning Course. The change in saliva cortisol observed after the E4T (markedly blunted, followed by resolution to baseline diurnal variation in 10 days) was likely related to the change from 24-hour sunlight and to UK time, and sleep-wake disruption.

The decrease in mood scores and modest increase in hair cortisol during basic military training could also relate to sleep deprivation. The E4T study participants slept over 6.5 hours per night, while FEAT study participants reported feeling increasingly tired with difficulty staying awake throughout the study, which may have blunted HPA axis function (Morales et al., 2019).

6.2.2 HPG Axis findings

6.2.2.1 Basal reproductive markers and ovulation

We expected suppression of the HPG axis to occur in tandem with suppression of the HPA axis, as in studies of athletes with low EA (Ackerman et al., 2013). However, the relative impact of military training on the HPG axis was more marked than on the HPA axis.

During the E4T study, AMH and inhibin B were unchanged but within normal limits. LH, FSH and oestradiol were suppressed and oestradiol tended to increase 15 days post-expedition. In the FEAT study, inhibin B and to a lesser extent oestradiol increased throughout training, suggesting follicular dysgenesis (Li et al., 2011), with AMH remaining unchanged. This may have been driven by FSH, which increased in all contraceptive groups. Oestradiol was associated with increased TSH in contraceptive users (only). In healthy ovulating women, TSH is normally slightly higher in the periovulatory phase (Benvenga et al., 2018), although it is not possible to determine whether increased pituitary thyrotroph activity was related to HPG axis changes. Exercise is normally associated with a modest decrease in oestradiol

(Ennour-Idrissi et al., 2015) and unchanged TSH (Misra, 2014), hence our observations in the FEAT study are likely related to non-exercise components of military training.

Again, sleep deprivation and stress may have contributed to the discrepancy between these studies. Sleep deprivation is associated with activation of the HPA axis, increased oestradiol, TSH and free T4, and unchanged GH (Baumgartner et al., 1993). The circadian and menstrual 'clocks' are closely related, with GnRH, LH and FSH pulsatility being sensitive to altered sleep/ wake cycle disruption (Baker and Driver, 2007; Goldstein and Smith, 2016). Activation of the HPA axis in the follicular phase, from sleep deprivation and/ or psychological stress, may prolong follicular maturation and prevent ovulation and luteal transition (Ferin, 1999). This is discussed more fully in **Section 6.7.5**.

During the E4T study, participants used progesterone-only contraceptives in an attempt to suppress menstruation. We did not attempt to assess ovulation, since anovulation is assumed following levonorgestel and it was not feasible for the participants to carry urine samples on the ice. In the FEAT study, the degree of menstrual disturbance we observed was similar to other studies from military training (Anderson, 1979; Lauder et al., 1999; Schneider et al., 2003; Schneider et al., 1999; Sonnenblick et al., 1993; Welch, 1989). The degree of menstrual disruption observed was modest compared with anovulation, which was nearly universal in non-contraceptive users.

The pattern of our findings from the E4T study was similar to studies of competitive athletes and soldiers with low EA (Elliott-Sale et al., 2018; Henning et al., 2014), although the effect sizes were much more modest. The HPG axis recovered in a short timeframe following cessation of arduous exercise and recuperation of energy balance (**Section 4.1**).

Free androgen index (FAI) did not change during Ex ICE MAIDEN but decreased 11 days thereafter. Sex hormone binding globulin (SHBG) was unchanged throughout the study. Conversely, FAI did not change during the FEAT study, although SHBG decreased in contraceptive users (especially combined) but was unchanged in non-users. FAI normally demonstrates an inverse relationship with insulin sensitivity, SHBG being higher and circulating androgens lower after exercise interventions compared with controls (Hakimi and Cameron, 2017; Mena et al., 2019). We did not

observe a decrease in FAI as might have been expected; it may be that the increase in insulin resistance was too modest to affect androgenisation.

6.2.2.2 Dynamic pituitary gonadotroph function

In the E4T Study, no change in gonadotroph responsiveness was demonstrated from 1 month beforehand (when it was likely to have been suppressed following pre-expedition training) to 5 days post-expedition, and a latent increase in responsiveness after a further 11 days. In the FEAT Study, both LH and FSH responses were suppressed. The absolute levels should be compared with caution due to differences in protocol (E4T study: dexamethasone suppression followed by fasting morning measurement, FEAT study: unfasted, afternoon measurement). Gonadotroph pulsatility does not vary according to the time of day (Klingman et al., 2011), although dexamethasone may suppress LH and FSH response (Frautschy and Liptrap, 1988; Rosen et al., 1988).

- In the E4T study, FSH was suppressed to a maximum fold-increase of around 1.2 throughout with no change pre- versus 5 or 16 days post- expedition.
- In the FEAT Study, the mean response of FSH was 1.5 to 2.0 fold at baseline, reducing to 1.2 after 28 weeks.
- In the E4T study, the average increase in LH responses were around 2.0-fold pre- and immediately post-expedition, rising to 3.5-fold 16 days post-expedition.
- In the FEAT Study, LH responses were higher at baseline (5- to 20-fold) but after 28 weeks had decreased around 2.5–fold.

Fold-response of both FSH and LH varied according to the negative feedback exerted by hormonal contraceptives, although the direction of change was consistent across groups of contraceptive users. Thus, LH and FSH responses in the FEAT study after 28 weeks of training were similar to the E4T study immediately post- (and pre-) expedition. The resolution in LH responsiveness after 16 days in the E4T study indicates HPG axis suppression due to low EA is reversible after correction of the energy deficit, as has been shown elsewhere (Loucks et al., 1989; Loucks and Thuma, 2003).

6.2.3 Skeletal findings

In the E4T study, we found evidence of bone uncoupling, with increased CTX and reduced BSAP, albeit the effect size was modest. DXA scanning modest decline in areal BMD from the ribs, spine and trunk, presumably to mobilise calcium to sustain remodelling in load bearing areas (legs and arms). This was evidenced by preserved BMD and microarchitecture in the tibia, where stiffness and fracture threshold were unchanged (O'Leary et al., 2019).

Skeletal imaging demonstrated a very modest decrease in total BMD in the FEAT Study (ηp2 0.078), the loss being greatest at the arms and ribs, with preserved BMD in other sites. Bone turnover demonstrated increased formation (reduced CTX, increased bone turnover ratio) at 28 weeks, but with no difference form baseline by 43 weeks. Volumetric BMD data from the tibia suggest the reduced overall BMD and increased formation may be explained as follows. Cortical area is increased from load bearing, due to increased formation of type 1 collagen matrix. However, there is a delay in secondary mineralisation, temporarily leaving weaker collagen matrix at around 28 weeks (data are held by the customer), a temporal hotspot for stress fracture development, which has been suggested previously (O'Leary et al., 2018). The cause for the delay in mineralisation is currently unclear. This would be an important area for future studies and may allow mitigation of stress fractures.

6.3 Mitigation factors

Our findings from basic military training accord with the concerns mooted in the WGCC IHR (Department of Manning (Army), 2016), that rates of infertility among military women may be higher than among age-matched civilians. We did not anticipate that HPG axis dysfunction would be associated with evidence of energy sufficiency or increased adiposity. Comparing the findings from the FEAT and E4T studies highlights a number of potential mitigation factors which are the basis for our recommendations for policymakers **(Section 6.6)**:

6.3.1 Selection and motivation

The ICE MAIDENs were selected from over 250 motivated applicants, who applied from across the Regular and Reserve Army (as described in **Section 4.2**). They appeared to be ‘outliers’ in terms of their elite physical fitness, resilience, work ethic and team dynamics (Blackadder-Weinstein et al., 2019; Thake et al., 2019) and this is likely have contributed to our findings. The ICE MAIDENs reported lower levels of

psychological stress and anxiety, and improved mood, during the expedition than beforehand, whereas in basic military training these same measures increased progressively. Women who successfully negotiate the Commissioning Course are already highly selected and motivated compared with the general population; however, continuing to emphasise the importance of rigorous selection may improve outcomes. Perhaps more importantly, training for Ex ICE MAIDEN was designed, led and operated by women. Over many decades, the training of infantry has naturally been designed for and conducted by men. There may be major challenges to women adjusting to an environment of male normative behaviours, given important sex differences in psychometrics, for example in risk awareness as discussed in **Chapter 5**.

6.3.2 Training gradient

The ICE MAIDENs' training was gradual in onset and lasted over two years, following a history of military training and employment over several years (**Section 4.2**). The E4T participants were older than FEAT study (median (range) age 33 (28 to 36) versus 24 (19 to 30) years). The first five weeks at Sandhurst are immersive, intense and unrelenting (BBC News, 2011). Hans Selye's general adaptation syndrome described activation of the HPA axis in response to abrupt stressors, leading to metabolic exhaustion, however this was mitigated when stressors were of gradual onset (Selye, 1946). While the relevance of this model has been questioned to LH disruption in the exercising female (Loucks, 2009), it remains highly plausible that upregulation of the HPA axis by non-exercise stressors during immersive military training contributes to HPG axis dysfunction (Gifford et al., 2017). Considering the general adaptation syndrome for constructing periodised military training may help mitigate against endocrine and metabolic maladaptation (Cunanan et al., 2018b).

6.3.3 Sociocultural change.

Closely linked to the gradient of training is the notion of conformation to a new sociocultural normative behaviour – 'military life'. The Services seek to instil core values during training, but additionally (and unofficially) encourage a nuanced culture, encompassing unwritten codes of behaviour, external interests, social circles, speech and dress. This change is likely to coincide with the stress of separation from normal social support networks at home. One Company Commander briefing FEAT study Officer Cadets on their first day described term 1

as ‘a competition in a submarine’, which is similar to observations of initial military training elsewhere (Bornmann, 2009; Gold and Friedman, 2000). The ICE MAIDEN team were already employed in the military, so did not have to overcome sociocultural change imposed on Officer Cadets *ab initio*. Recent efforts at Sandhurst to downplay the requirement of sociocultural uniformity, instead emphasising (constrained) individuality are interesting. The role of perceived individuality on the HPA axis and training outcomes would be a worthwhile area for future studies.

6.3.4 Locus of control.

The combination of abrupt onset training, general adaptation and sociocultural change are likely to contribute to the loss of locus of control, the perceived ability to control life circumstances. Locus of control is an important psychometric trait, which has demonstrable links with heart disease (Rosengren et al., 2004), poor diet and obesity (Neymotin and Nemzer, 2014) and therefore plausibly also insulin resistance and diabetes. The mechanism is likely related to modulation of inflammatory pathways (Rosengren et al., 2004), altered cortisol responses and reduced hippocampal volume (Bollini et al., 2004; Pruessner et al., 2005). Loss of locus of control may have contributed to the mood and metabolic outcomes we observed during military training; exploring this association would be a worthwhile area for further research.

6.3.5 Psychological support.

The ICE MAIDEN participants received training from professional psychologists, which perhaps helped build their resilience and strong sense of teamwork. We found low levels of stress and unchanged resilience and mood through the expedition, while Blackadder-Weinstein et al. (2019) also noted a strong team dynamic which overcame some disagreement during the expedition. Impaired mood and resilience during the FEAT study associated with increased stressors and tiredness (Gifford et al., 2019b). It is not possible to determine a causal relationship from these observational data, but while RMAS already employs a resident psychologist, I hypothesise that increasing support from such professionals may mitigate the psychological and biological changes we observed.

6.3.6 Nutrition.

As described in **Section 6.1.2.2,** the findings in **Sections 3.2** and **3.4** could reflect alternating energy deficit during field exercise, followed by overeating (energy compensation). This 'yo-yo effect' contrasts with the continuous intake of 24-hour rations meticulously planned for Ex ICE MAIDEN (Taylor et al., 2019). Although participants sustained a significant energy deficit throughout the expedition, it may be that the day-to-day consistency of energy balance permitted a degree of metabolic adaptation, decreasing RMR while preserving HPA axis reactivity. Alternatively, intermittent, abrupt onset energy deficit may have contributed to HPG axis maladaptation in the FEAT study (Senoo, 2000). The 32% underestimate of energy intake identified in the FEAT study likely represented the intake of snacks outside of mealtimes. Anecdotally, snacks which were popular with cadets were high in fat and may have contributed to insulin resistance. We identified an inverse relationship between EA measured during field exercises and several training benefits (fat mass loss, fat-free mass gain, 1.5 mile run time improvement). EA was also lower in participants with higher scores of disordered eating. Therefore, a focus on complete nutrition, including nutritional education, awareness of eating behaviour, regularity of energy intake and nutrient timing, especially during field exercises, might mitigate against deleterious effects of intermittent low EA and improve performance. Guidelines already exist to help guide this (Kerksick et al., 2017).

6.3.7 Sleep.

Adequacy of sleep was greater in the E4T study than the FEAT study. As well as activating the HPG axis and suppressing the HPG axis, consequences of sleep disturbance include depression, anxiety disorders, glucose dysregulation, hypertension and obesity (Goldstein and Smith, 2016; Kloss et al., 2015). Advanced military training, such as Ranger training in Norway and the USA, is associated with marked sleep deprivation and profound metabolic and endocrine maladaptation (Friedl et al., 2000; Nindl et al., 2007; Opstad, 2001). Improved sleep discipline may improve physical and academic outcomes of basic military training; however, sleep restriction appears to be an important element of 'arduousness' and has helps create a challenging training environment. Sleep was not a focus of either study. More research on sleep in military training would be important elucidate the precise mechanisms by which sleep disturbance impacts endocrine and metabolic function,

and any long-term effects of sleep deprivation in military training (including HPG axis function), before firm recommendations are made. **Section 6.7.5** contains a short discussion and recommendation for further research.

6.3.8 Contraception.

While in the FEAT study, suppression of the HPG axis was present in all contraceptive groups, in the E4T study all participants used long-acting progestogen eluting contraceptives. There is no definitive evidence of benefits of oral ethinylestradiol supplementation on bone, although in patients with amenorrhoea and low EA, transdermal oestrogen supplements appear to be beneficial (Ackerman et al., 2019b). Larger cross sectional studies building on data from the FEAT study could provide evidence to guide hormonal contraceptive use, but for the time being, user choice and concordance are of paramount importance.

6.3.9 Individual differences.

In both the E4T and FEAT studies, the level of detail of measurements allowed us to identify traits in individual outliers. For example, in the FEAT study, there were two non-hormonal contraception users who reported amenorrhoea before and during the study. Both demonstrated completely flat urinary progesterone profiles and had low BMI. Unfortunately, both withdrew from the study due to stress fractures (they were the only participants to do so) and it was not possible to complete the 28 week CHOC test in one of them. These women putatively had the ‘female athlete triad’, although neither scored highly on an eating disorder questionnaire. In the E4T study, hypercortisolism and greater HPG axis suppression were observed in two participants in whom through-expedition stress was higher than others. These findings highlight the importance of understanding individual variability in training adaptation. The first women in GCC will be likely to be biological outliers; it would be prudent to observe them very closely, considering an individualised approach towards training where unhealthy eating behaviours are observed.

6.4 Strengths and limitations

Relative strengths and weaknesses of individual components of each chapter were discussed in the relevant sections. The strengths of this thesis overall are the detailed nature of measurements to quantify changes in the HPA and HPG axes, nutrition, energy balance and bone health concurrently. We have considered the

Service women in real-world settings, where everyday stressors meet additional occupational demands, a variety of performance traits are required (not just strength and fitness) in extremes of climate, and real-world hormonal milieu, not controlled for diet or contraceptive use. The durations of training assessed were longer than previous studies, making our studies of greater real-world relevance. The studies are novel, presenting the first characterisations of female endocrine and metabolic adaptation to an extreme 2-month exercise and to 11-month initial military training.

Engaging the training establishment as a stakeholder in the research programme was also a significant strength. From the outset, we sought to maximise the association of RMAS staff with the FEAT programme. Through termly briefings, frequent face-to-face conversations and by increasing the presence of researchers within the Academy, we aimed to generate enthusiasm for the study, which could have risked being an inconvenience otherwise. Establishing trust and a perception of shared ownership with RMAS facilitated Cadets' participation, establishing the study as being of direct and wider benefit. We also sought to minimise individual withdrawal from the study by offering 'hospitality' and direct feedback on physical improvements as incentives. Our approach was successful: we were afforded an unprecedented use of facilities and support from RMAS, and no cadets withdrew from the study outright, except for reasons of injury.

The collaborative approach taken during this programme is a key strength: we have involved specialists from eight institutions in fields including biomechanics, clinical biochemistry, computer programming, diabetes, dietetics, endocrinology, nutrition, nuclear medicine, physiology, reproductive medicine, signal engineering, sports and exercise medicine and statistics. We feel these collaborations have afforded us greater breadth of expertise, variety of publications and a more balanced interpretation of extant literature than might have been obtained if a single laboratory had undertaken the programme.

The studies had several limitations. We were unable to directly measure maximal oxygen uptake or resting metabolic rate (RMR) in the E4T study, which are important variables when considering the physical and metabolic impact of training. Farenholtz et al. (2017) found a buffering effect of reduced RMR has when within-day EA was low, which is likely to confound the measurement of EA. Future studies should aim to incorporate indirect calorimetry in the field, consider non-exercise

activity thermogenesis and gold standard measurements of adaptation like VO_{2max}. Constraints of the Commissioning Course also precluded more frequent blood testing or a third CHOC test. Furthermore, the field studies did not directly measure a cortisol response to stress. Use of the Trier Psychosocial Stress Test (TSST) or cold pressor test, would have been superior. Use of dexamethasone to measure central negative feedback and to attempt to achieve a uniform baseline prior to the CHOC test in the E4T Study but not FEAT Study limits their comparison.

The observational nature of these studies naturally prevents the determination of causality. Comparison of the two studies is limited by the low participant numbers, particularly in the E4T Study. The ICE MAIDENs primarily serve as a proof of concept that women can be resilient to extreme training, but as has already been discussed, their elite physical and mental fitness meant they were not a representative sample, even arguably of future candidates for WGCC. Indeed, it could be said that women at Sandhurst are not representative of women across the military, being highly selected beforehand at the Army Officer Selection Board and motivated, by definition, to complete one of the longest and most arduous Phase 1 training programmes. Our findings are therefore primarily of relevance to motivated women who are successful in rigorous selection for military roles.

6.5 Summary

Service personnel are required to contribute to global operations in unpredictable contexts, and diverse and challenging environments. Contrary to expectations, the E4T and heat illness studies demonstrated the resilience of women to environmental extremes in terms of meaningful biological and clinical outcomes. These findings provide a proof of concept that female sex need not preclude individuals from being considered for any military role and are highly relevant to current military policy.

Military training is physically arduous; indeed, soldiers have been described as ‘tactical athletes’ (Sell et al., 2016). Our findings highlight the manifold complexity of military training compared with athletic training (whether undertaken by the military or civilians). Exercise is one of several challenges imposed alongside a steep training gradient. sociocultural change, loss of locus of control, suboptimal nutrition, sleep insufficiency, elemental exposure, conflict between leadership and subordinate roles, academic and time management pressure (Crowley et al., 2015;

Gold and Friedman, 2000). Many such stressors have been associated with or are likely to be associated with HA (Gordon et al., 2017; Pauli and Berga, 2010).

Arduousness represents an accumulation of a number of training-related and individual factors. Abrupt onset of training, sociocultural change, loss of locus of control, suboptimal nutrition or sleep restriction, and individual differences in dietary intake, eating behaviours or stress response might not individually be expected to cause meaningful changes in the HPG or HPA axes, but arduousness as a 'whole' is greater than the sum of its parts. Arduousness is dependent on an individual's response to the training environment, as illustrated in **Figure 6-2.**

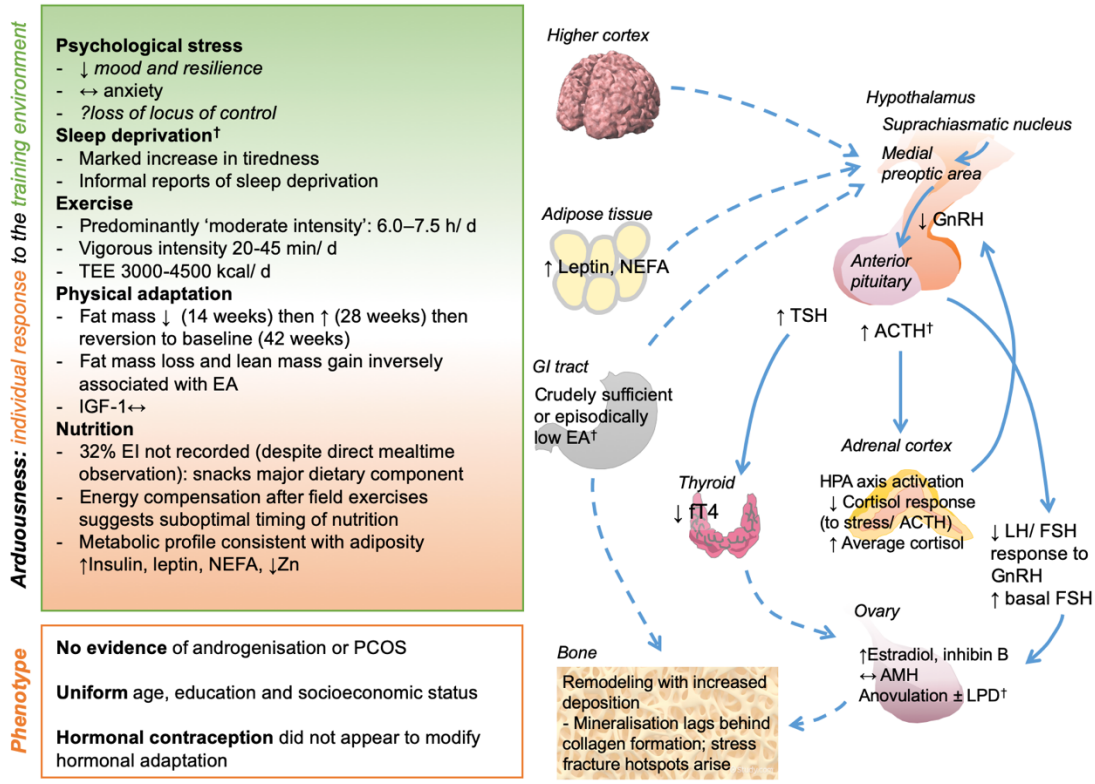


Figure 6-2. Female endocrine adaptations to arduous training. Arduousness reflects both the training environment and the subjective individual response; green and orange shading attempts to represent the interrelatedness of these factors. Individual phenotype (orange box) did not meaningfully affect outcomes. †: assessed indirectly. Relationships are represented by arrows (dashed: indirect, solid: direct). EI: energy intake, TEE: total energy expenditure, EA: energy availability, IGF-1: insulin-like growth factor 1, GI: gastrointestinal, NEFA: non-esterified fatty acids, Zn: zinc, PCOS: polycystic ovary syndrome, TSH: thyroid-stimulating hormone, fT4: free thyroxine, LH: luteinising hormone, FSH: follicle-stimulating hormone, LPD: luteal phase defect, ACTH: adrenocorticotrophic hormone, GnRH: gonadotrophin releasing hormone.

There are some important differences between our findings in **Figure 6-2** and our hypothesis (**Figure 6-1**). We no longer consider environmental and individual factors discretely (represented in **Figure 6-1** by green and orange boxes, respectively, and by a continuum between green and orange in **Figure 6-2**). Training and environmental stimuli are stressors in as much as they are subjectively perceived as stressful. Moreover, many stressors are not part of the training environment: for example, the majority of 'major life event' stressors we identified occurred outside of the training (Gifford et al., 2019b). Even stressors which are ostensibly imposed by the training environment like restricted sleep, suboptimal nutrition and episodic low EA might be exacerbated (or in all likelihood, largely caused) by individual behaviours. This was evidenced by poor sleep discipline (e.g. late-night phone use), unnecessarily missing meals, not consuming 24-hour ration packs, viewing field exercises as opportunities for weight loss and choosing fast food instead of meals in the College dining room. Thus, considering environmental or individual factors separately may be a false dichotomy.

Individual phenotype traits which we hypothesised would modulate endocrine adaptation (androgenisation, age, socioeconomic status and contraception) did not appear to be relevant to the effects observed. There was no evidence of PCOS. Age and socioeconomic status did not vary within our sample and HPG axis suppression and HPA axis activation did not differ between contraceptive groups.

Several of the seminal studies proving the importance of low EA for female athletes were undertaken under the auspices of the 1994 US Defence Women's Health Research Program (DWHRP, reported by Friedl (2005), although many important studies funded by the programme were published after this report). In the intervening 25 years, the concept whereby low EA mediates HA through leptin upregulating the HPA axis, has been substantiated by physiologists seeking to understand exercise-associated HA. The concept of RED-S has been embraced by sports and exercise physiologists as evidence that reproductive concerns about women need not be a consideration prior to participation in sports, given energy intake is sufficient (Drinkwater, 2015). However, when considering military training, this assertion must be held in equipoise with the importance of psychogenic activation of the HPA axis (Pauli and Berga, 2010) and the general adaptation syndrome (Cunanan et al., 2018a; Cunanan et al., 2018b). The culmination of individual and environmental factors we have called arduousness may cause HPG

axis suppression through all of these mechanisms. The findings in this thesis build on those of the DWHRP, and challenge the assumption of a single aetiology. Instead we advocate appropriating findings from a broad array of fields to train the 'tactical athlete' (Sell et al., 2016).

6.6 Recommendations for training

The unexpected findings from the FEAT study have created more questions than they answer. The next step would be to commence studies seeking to understand causal relationships before testing interventions to mitigate against harm. However, some tentative recommendations for training could be made based on our findings.

6.6.1 Basic military training duration, gradient and gender normative

It is important to note the difference in training duration and gradient between the E4T and FEAT cohorts, a 2-year, personalised approach as taken in the E4T study would be unlikely to be feasible for female soldiers. Some alteration of training, perhaps to reduce the training gradient, in women might reduce injury rates and could be considered an equitable approach.

A greater emphasis on pre-training (before basic military training begins) might decrease the abruptness at the start of training. This could be achieved with current trainees informing prospective recruits of their experience.

Trainers creating a gender-fair environment might consider that expected behaviours from women could accord more with their gender and native sex, since infantry training has evolved over many years around men. It may be that under extant training conditions, a female gender normative is unconsciously masculinised, and this may be associated with psychological stress. This is currently under investigation by one of a series of scoping reviews in the USA (Iobst, 2019). Studies in the UK should seek to involve experts in psychology, health behaviour and anthropology.

6.6.2 Education of staff and trainees

Physical training adaptations were less beneficial in cadets with lower EA during the FEAT Study. We recommend informing directing staff and trainees of the finding that mismatched EI and EE during field exercises is associated with doing less well

in training, to encourage regular eating and protected mealtimes and to dispel the idea that field exercises are a good opportunity for weight loss.

6.6.3 Dietician and psychologist involvement

Increasing the interface of experts in dietetics and psychology with all cadets might help educate on means of healthy nutrition and their importance. Psychologists could help with adjustment to military culture. Fat mass gains were highest where stress was greatest (term 2). Dieticians could to educate cadets about the importance of diet and exercise during times of stress.

6.6.4 Personalisation of phase 2 training for women

The FEAT study showed a different pattern of HPG axis dysfunction to RED-S, which likely represented a collective effect of several factors. We recommend monitoring the first women to enter GCC closely, including nutrition, sleep and psychological wellbeing. Identifying patterns of behaviour early might allow corrective action for these individuals and future female soldiers.

6.7 Future work

Several questions have been identified by the FEAT programme. We have identified four key areas for further research: sex differences in stress responses, development of field techniques for assessing EI and EEE, transition to military life and locus of control and sleep deprivation. Since sleep deprivation was not a focus of the FEAT study, a short review of the literature in this area provides rationale.

6.7.1 Follow-up of the FEAT Study

To establish the clinical relevance of the HPG axis suppression identified, it would be necessary to follow up the participants, particularly those in GCC roles. In addition to repeating the CHOC test and a detailed reproductive history, investigating ovarian function using imaging would be informative.

6.7.2 Male and female stress responses.

The meta-analysis by Stalder et al. (2017) found hair cortisol is normal 21% lower in women than men. In men, anticipatory cortisol responses to stressors are usually higher than women (discussed in **Section 1.3.2**). Future studies should examine sex differences in HPA axis responses to military training and to measures of whole-HPA axis response in this setting (e.g. the Trier Psychosocial Stress Test). To avoid

confounding for low EA, participants could be matched by eating behaviour and the study might take place in a metabolically neutral environment such as the interrogation element of Survival, Evasion, Resistance and Extraction training.

6.7.3 Development of field measures of energy intake and expenditure.

Our studies identified the requirement for better informed methods to quantifying energy intake. Mobile application, video and photography based techniques should be considered for validation alongside weighed food as a gold standard. A collaborative approach would need to be taken, with experts in health behaviour perhaps leading physiologists and doctors to understand dietary behaviours. An important objective would be to quantify snack consumption and explore attitudes to the diet provided by the training establishment.

We made significant progress in the use of wrist-worn triaxial accelerometers to assist estimation of EEE. Our method could be developed further to identify load carriage from wrist-worn accelerometry and use this to additional energy requirements.

Further studies should use indirect calorimetry and VO_{2max} testing to measure changes in metabolic expenditure and fitness, respectively. These techniques would be useful to validate measures of EEE, and to increase the generalisability of findings, since they are commonly used in the exercise and EA physiology literature.

6.7.4 Transition into military life and locus of control.

The psychological changes we observed during basic training were significant and could have related to adjustment to military life and culture. We recommend continuing this research theme with a focus on locus of control as a key area of interest. The intent to encourage individuality within a training environment warrants consideration for qualitative and quantitative work. Reservists are an appropriate population to explore the transition from military to civilian life, since they undergo this transition more frequently than regulars. Such work should be undertaken collaboratively by engaging experts in psychology, sociology and anthropology.

6.7.5 Impact of sleep deprivation on metabolic, HPA axis and HPG axis outcomes in military training.

The core clock is located close to the preoptic area in the suprachiasmatic nucleus of the hypothalamus and its projections modulate GnRH release via kisspeptin. Animal models demonstrate a circadian control of LH pulsatility and suprachiasmatic nucleus lesions or core clock gene knock-out result in anoestrus (Kriegsfeld and Williams III, 2012). Findings from humans are more observational, however the human core clock is influenced by core progesterone, oestradiol and temperature (Goldstein and Smith, 2016), and sleep has been found to influence various aspects of reproductive function (Goldstein and Smith, 2016). In a systematic review and meta-analysis, Stocker et al. (2014) found night shift workers experienced greater rates of spontaneous early pregnancy loss (OR 1.41 [95% CI 1.22, 1.63]), after adjustment for age and various other covariates including BMI, physical activity and smoking.

Researchers have identified associations between disturbed sleep and several of the findings we observed in the FEAT study. Oestradiol has been shown to increase with sleep deprivation (Baumgartner et al., 1993), irregularity (Merklinger-Gruchala et al., 2008) and poor quality (Sowers et al., 2008). The effect of sleep disturbance on FSH is less clear, with increase (Sowers et al., 2008) decrease (Touzet et al., 2002) and no effect (Baumgartner et al., 1993) reported in the early follicular phase. The variation in FSH findings could be explained by different patterns of reproductive dysfunction: higher FSH is associated with low ovarian reserve and reproductive ageing, while low FSH could suggest LPD or HA. AMH demonstrates a slight circadian control, being higher in the morning (Bungum et al., 2011), although a study of 1500 female nurses found no association between AMH and night shift work (Johnson et al., 2019).

Lack of sleep adversely affects mood as well as and menstrual function (Nam et al., 2017). Sleep loss is associated with upregulation of the thyroid axis, with increased average TSH and thyroid gland activity (Baumgartner et al., 1993; Spiegel et al., 1999) and loss of normal diurnal variation in TSH (Brabant et al., 1990). Sleep studies predominantly including healthy, young men have found acute sleep deprivation is associated with increased HOMA-IR (Spiegel et al., 1999) as observed in the FEAT study, but reduced leptin (Spiegel et al., 2004) and increased NEFA (Broussard et al., 2015), contrary to findings in the FEAT Study. Population-

wide studies have identified associations between sleep disturbance is associated with obesity and diabetes (reviewed in (McHill and Wright Jr, 2017) and (Grandner et al., 2016)).

Sleep disturbance could cause metabolic complications through chronic activation of neuroendocrine stress responses. Sympathetic adrenal medulla (SAM), but not HPA axis, activation was noted following poor sleep quality in a study of middle-aged men, although more robust measures of the SAM system were used than of the HPA axis (24-h metanephrines and single morning cortisol, respectively) (Louis et al., 2011). Indeed, sleep loss is associated with elevated evening cortisol levels (Spiegel et al., 1999) and cortisol and ACTH is elevated in insomniacs between 9 p.m. and 12.30 a.m. (Vgontzas et al., 2001). Reproductive and metabolic disturbances due to sleep loss could be mediated via activation of the HPA axis or SAM system (Kloss et al., 2015; Louis et al., 2011; Spiegel et al., 2004).

Another possible mechanism linking sleep disturbance with reproductive dysfunction is a direct ovarian effect of melatonin. Melatonin is released by the pineal gland diurnally and is suppressed by light. Melatonin is abundant in ovarian tissue, reflecting production in the ovary as well as the pineal gland (Goldstein and Smith, 2016). Light supplementation during the day restores suppressed FSH, oestradiol and TSH (Danilenko and Samoilova, 2007) and improved LH pulsatility (Kripke et al., 2010), but these studies did not measure melatonin directly. Fertility treatment improvements have been seen with melatonin supplementation, especially in women with PCOS (Goldstein and Smith, 2016). Melatonin improves insulin sensitivity in rats, providing a possible peripheral pathway (Dantas-Ferreira et al., 2018). A randomised control trial of melatonin supplementation to improve outcomes of fertility treatments showed no improvement (Fernando et al., 2018a; 2018b).

In summary, sleep disturbance has several negative effects on reproductive function. Both sleep deprivation and impaired sleep quality have been reported in military training (Crowley et al., 2012) but anecdotal evidence suggests insufficient quantity was the greater factor during the Commissioning Course. Studies assessing sleep quality and quantity indices alongside HPA and HPG axis responsiveness would be valuable in military training. In future investigators might

consider a randomised control trial of melatonin supplementation on metabolic outcomes and HPG axis function.

6.8 Wider implications

The discussions in this thesis are naturally oriented around application to a military niche, since the studies were conducted to address problems in Servicewomen. However, stressors observed in the military, like assessment, physical exercise, psychosocial upheaval, conflict between leadership and peer roles, metabolic stressors, energy compensation and sleep deprivation, occur in civilian life. Military training is unusual only in that it prepares individuals for the use of violent force (but even in this it is not unique). Clinicians might therefore consider military training a useful model of real-world stressors in aspects of civilian life.

Since there is a paucity of studies involving military women, **Chapters 1 to 4** drew on data from civilian athletes to form hypotheses and interpret findings. The necessary assumption made was that endocrine adaptations to military training primarily relate to military training's physicality. The American College of Sports Medicine, International Olympic Committee (IOC) and British Association of Sports and Exercise Medicine all emphasise that reproductive dysfunction and impaired bone turnover in sports and exercise are caused by low EA (Nattiv et al., 2007; Keay and Rankin, 2019; Mountjoy et al., 2018). Some (notably the Female Athlete Triad Coalition) emphasise that the clinically relevant manifestation of this cause-effect paradigm is the female athlete triad (De Souza et al., 2014a). These authors assert that in most exercising women, low EA is a reversible phenomenon and should not lead to concerns which might preclude women from physical activity (De Souza et al., 2014b). The Triad Coalition have emphasised that the stress *of exercise* does not contribute to reproductive dysfunction (Loucks, 2009). Some disagree, noting that that in settings outside the laboratory or clinical trial environment, the stress of life apart from exercise is an important cause of HA (Pauli and Berga, 2010). Others, including the IOC, have ascribed even greater importance to low EA, stating that chronic energy deficiency among athletes and sportspeople manifests in a wide spectrum of pathology (the relative energy deficiency in sports syndrome, RED-S (Mountjoy et al., 2014; Mountjoy et al., 2018). Evidence for conditions other than the female athlete triad is lacking, and the RED-S concept may cause confusion, exaggerating the scale of the problem in women (De Souza et al., 2014b).

Leadership, not physical activity, is the primary occupational focus of training in RMAS. This thesis has demonstrated that the single aetiology paradigm of low EA (or RED-S) does not appear to be as relevant to reproductive dysfunction in this setting as it is for civilian athletes. The findings presented here might be generalisable to other areas of work where similar stressors are encountered. For example, financiers working long hours under significant pressure in male-dominated environments experience high levels of stress (Bourbonnais et al., 1996). Externalisation of locus of control among city workers (i.e. loss of control over one's one life circumstances) affects women more than men and junior workers more than seniors (Bernardi, 1997). Female air crew can experience high levels of sleep disturbance, which may be associated with job stress (Hwang and Kim, 2015; MacDonald et al., 2003). Post-traumatic stress and sleep disturbance have been reported among firefighters (Corneil et al., 1999; Eastlake et al., 2015) among whom women may experience greater work-related stress and symptoms of PTSD than men (Murphy et al., 1999; Noor et al., 2019). This thesis may serve to emphasise the importance of such stressors for women, the cumulative effect of many concurrent stressors, and the possibility of their being associated with HA (Gordon et al., 2017). More studies are needed to explore sex differences in endocrine stress responses, and to help policymakers mitigate against reproductive dysfunction.

The reproductive dysfunction observed in this thesis occurred in tandem with metabolic maladaptation, similar to that observed following studies of exercise interventions (energy compensation) (McNeil et al., 2017; Riou et al., 2015). Individuals undertaking exercise for weight loss (and clinicians recommending it) should be aware of energy compensation and perhaps plan accordingly with gradual onset training and additional dietetic support. The chronic catabolic effects of elevated cortisol and psychological stress are similar to the effects of energy compensation and include central adiposity, insulin resistance and increased cardiovascular risk (Abell et al., 2016a; Manenschijn et al., 2013; Ulrich-Lai et al., 2015). The importance of chronic work-related stress and metabolic maladaptation should be underlined.

The E4T and heat illness studies found women to be resilient, thriving in extremes of physical endeavour and environmental exposure. The FEAT study suggests that in the less extreme setting of general training, female reproductive dysfunction may nevertheless occur. This latter setting is probably more generalisable to civilians,

and the FEAT study model might be applied in other employment settings, for example financial services, police forces or lawcourts. It would be necessary to assess the prevalence of reproductive dysfunction and associated pathology as a pretext for such studies.

Overall, the findings of this thesis do not preclude women from ground close combat roles. Gender specific training might enable greater equity for women, recognising sex differences and gender complementarity. Further studies are needed to delineate the mechanisms of the reproductive dysfunction observed and to determine the precise scale of the problem of reproductive dysfunction in military women.

APPENDIX A: Narrative Review: Reproductive dysfunction and associated pathology in women undergoing military training

Sqn Ldr Robert M Gifford, R M Reynolds, J Greeves, R A Anderson, D R Woods

J Roy Army Med Corps, 2017 doi:10.1136/jramc- 2016-000727

Introduction: Evidence from civilian athletes raises the question of whether reproductive dysfunction may be seen in female soldiers as a result of military training. Such reproductive dysfunction consists of impaired ovulation with or without long-term subfertility. **Methods** A critical review of pertinent evidence following an extensive literature search. **Results** The evidence points towards reduced energy availability as the most likely explanation for exercise-induced reproductive dysfunction. Evidence also suggests that reproductive dysfunction is mediated by activation of the hypothalamic–pituitary–adrenal axis and suppression of the hypothalamic–pituitary–gonadal axis, with elevated ghrelin and reduced leptin likely to play an import- ant role. The observed reproductive dysfunction exists as part of a female athletic triad, together with osteopenia and disordered eating. If this phenomenon was shown to exist with UK military training, this would be of significant concern. We hypothesise that the nature of military training and possibly field exercises may contribute to greater risk of reproductive dysfunction among female military trainees compared with exercising civilian controls. We discuss the features of military training and its participants, such as energy availability, age at recruitment, body phenotype, type of physical training, psycho- genic stressors, altered sleep pattern and elemental exposure as contributors to reproductive dysfunction. **Conclusions** We identify lines of future research to more fully characterise reproductive dysfunction in military women and suggest possible interventions that, if indicated, could improve their future well-being.

This manuscript outlined the argument for the work described in this Thesis, highlighting to policymakers important gaps in current knowledge and areas of concern for military women. The article formed the basis for **Chapter 1.**

APPENDIX B: Materials

Equipment

Item	Supplier
S-monovette serum gel 7.5 mL and 9 mL, EDTA 4.9 mL, 7.5 mL and 9 mL, fluoride oxalate 2.7 mL	Sarstedt, Leicester, UK
Salivette ® saliva collection tube	
Conical skirted base aliquot tube with cap	
Safety multifly 21G	
10 mL universal container	
2.5 mL, 5 mL and 10 mL luerlock syringe	
Cannula membrane and luer adapters	
20G Cannula	B Braun, Sheffield, UK
Tegaderm ® 3M cannula dressing	esuppliesmedical.co.uk
Bench top centrifuge 5810R	Eppendorf Ltd., Stevenage, UK
Lunar Prodigy iDXA	GE Healthcare, Chalfront, Amersham, UK
Xtreme eCT II	Scanco Medical, Brüttisellen, Switzerland
CheckMyHeart HRV Device	DailyCare Biomedical, Taoyuan City, Taiwan
Cobas e411 ® analyser	Roche Ltd., Welwyn Garden City, UK
Architect ® i2000sr analyser	Abbott , Maidenhead, UK
Cobas Fara centrifugal analyser	Roche Ltd., Welwyn Garden City, UK
Cobas Mira turbimetric analyser	Roche Ltd., Welwyn Garden City, UK

Consumables

Item	Supplier
Anti rabbit PDG IgG	Scottish Antibody Production Unit, Carluke, UK
Rabbit anti-PDG antibody, PDG-HRP	In house reagents

Brady 163-499 Aliquot labels and ribbon	Scientific Laboratory Supplies Ltd., Glasgow, UK
0.9% NaCl solution, 250 mL	Baxter, Newbury, Berks, UK
Water for injection, 10 mL	
Gonadorelin ®	Intrapharm Ltd., Maidenhead, Berks, UK
Synacthen ®	Mallinckrodt, Staines-upon-Thames, UK
Reagents for Cobas Fara: Glucose, Lactate, Creatinine, Albumin, Transferrin, Calcium, Zinc, Iron, Magnesium, NEFA	Alpha Laboratories, Eastleigh, Sussex, UK
Reagents for Cobas e411: Progesterone, Insulin, C-peptide, IGF-1, Ferritin, Anti-Müllerian Hormone, P1NP, Beta-CTX, TSH free T4 total T3, prolactin	Roche Ltd., Welwyn Garden City, UK
Reagents for Architect ®: Progesterone Oestradiol, LH, FSH	Abbott, Maidenhead, UK
Leptin ELISA	R&D Systems, Inc., Abingdon, UK
Inhibin B ELISA	Beckmann Coulter, High Wycombe, UK
Sclerostin ELISA	Biomedica Medizinprodukte GmbH, Vienna, Austria
BSAP ELISA	Quidel Corporation, San Diego, CA, USA

Software

Programme	**Supplier**
SPSS for Mac v 24	IBM, Armonk, NY, USA
Prism 8.2.0	GraphPad, graphpad.com
MasterPlex Readerfit	MairiBio Hitachi Solutions, London EC2N, UK
Brady® IdentiLab	Brady PLC, London EC3R, UK
Microsoft Office 365	Microsoft, Redmond, WA, USA
CheckMyHeart version 3.0	DailyCare Biomedical, Taoyuan City, Taiwan
Kubios HRV Premium 3.20	Kubios Oy. https://www.kubios.com/

APPENDIX C: High resolution peripheral quantitative computerised tomography (HR-pQCT) from Ex ICE MAIDEN

Skeletal responses to an all-female unassisted Antarctic traverse

O'Leary, Thomas J, Gifford, Robert M, Double, Rebecca L, Reynolds, Rebecca M, Woods, David R, Wardle, Sophie L, Greeves, Julie P, Bone. 2019 Apr;121:267-276. doi: 10.1016/j.bone.2019.02.002. Epub 2019 Feb 5.

PURPOSE: To investigate the skeletal effects of the first all-female trans-Antarctic traverse. **METHODS**: Six women (mean+/-SD, age 32+/-3years, height 1.72+/-0.07m, body mass 72.8+/-4.0kg) hauled 80kg sledges over 1700km in 61days from coast-to-coast across the Antarctic. Whole-body areal bone mineral density (aBMD) (dual-energy X-ray absorptiometry) and tibial volumetric BMD (vBMD), geometry, microarchitecture and estimated mechanical properties (high-resolution peripheral quantitative computed tomography) were assessed 39days before (pre-expedition) and 15days after the expedition (post-expedition). Serum and plasma markers of bone turnover were assessed pre-expedition, and 4 and 15days after the expedition. **RESULTS**: There were reductions in trunk (-2.6%), ribs (-5.0%) and spine (-3.4%) aBMD from pre- to post-expedition (all P</=0.046); arms, legs, pelvis and total body aBMD were not different (all P>/=0.075). Tibial vBMD, geometry, microarchitecture and estimated mechanical properties at the metaphysis (4% site) and diaphysis (30% site) were not different between pre- and post-expedition (all P>/=0.082). Bone-specific alkaline phosphatase was higher 15days post- than 4days post-expedition (1.7mugl(-1), P=0.028). Total 25(OH)D decreased from pre- to 4days post-expedition (-36nmoll(-1), P=0.008). Sclerostin, procollagen 1N-terminal propeptide, C-telopeptide cross-links of type 1 collagen and adjusted calcium were unchanged (all P>/=0.154). **CONCLUSION**: A decline in aBMD of the axial skeleton may be due to indirect and direct effects of prolonged energy deficit. We propose that weight-bearing exercise was protective against the effects of energy deficit on tibial vBMD, geometry, microarchitecture and strength.

References

Abdulla, S.H., Bouchard, T.P., Leiva, R.A., Boyle, P., Iwaz, J., Ecochard, R., 2018. Hormonal Predictors of Abnormal Luteal Phases in Normally Cycling Women. Front Public Health 6, 144.

Abell, J.G., Shipley, M.J., Ferrie, J.E., Kivimaki, M., Kumari, M., 2016a. Recurrent short sleep, chronic insomnia symptoms and salivary cortisol: A 10-year follow-up in the Whitehall II study. Psychoneuroendocrinology 68, 91-99.

Abell, J.G., Stalder, T., Ferrie, J.E., Shipley, M.J., Kirschbaum, C., Kivimaki, M., Kumari, M., 2016b. Assessing cortisol from hair samples in a large observational cohort: The Whitehall II study. Psychoneuroendocrinology 73, 148-156.

Abraham, G., 1978. The normal menstrual cycle, in: Givens, J.R., (Ed.,) Endocrine Causes of Menstrual Disroders. Chicago: Year Book Medical Publishers, pp. 15-44.

Acheson, K.J., 1975. A longitudinal study to assess the current field techniques for measuring energy balance in man. Br. Antarct. Surv. Bull. 41 & 42, 87-97.

Ackerman, K.E., Cano Sokoloff, N., De Nardo Maffazioli, G., Clarke, H.M., Lee, H., Misra, M., 2015. Fractures in Relation to Menstrual Status and Bone Parameters in Young Athletes. Medicine and science in sports and exercise 47, 1577-1586.

Ackerman, K.E., Holtzman, B., Cooper, K.M., Flynn, E.F., Bruinvels, G., Tenforde, A.S., Popp, K.L., Simpkin, A.J., Parziale, A.L., 2019a. Low energy availability surrogates correlate with health and performance consequences of Relative Energy Deficiency in Sport. British journal of sports medicine 53, 628-633.

Ackerman, K.E., Misra, M., 2018. Amenorrhoea in adolescent female athletes. Lancet Child Adolesc Health, 2, 677-688.

Ackerman, K.E., Patel, K.T., Guereca, G., Pierce, L., Herzog, D.B., Misra, M., 2013. Cortisol secretory parameters in young exercisers in relation to LH secretion and bone parameters. Clin Endocrinol (Oxf) 78, 114-119.

Ackerman, K.E., Singhal, V., Baskaran, C., Slattery, M., Campoverde Reyes, K.J., Toth, A., Eddy, K.T., Bouxsein, M.L., Lee, H., Klibanski, A., Misra, M., 2019b. Oestrogen replacement improves bone mineral density in oligo-amenorrhoeic athletes: a randomised clinical trial. British journal of sports medicine 53, 229-236.

Ackerman, K.E., Slusarz, K., Guereca, G., Pierce, L., Slattery, M., Mendes, N., Herzog, D.B., Misra, M., 2012. Higher ghrelin and lower leptin secretion are associated with lower LH secretion in young amenorrheic athletes compared with eumenorrheic athletes and controls. American journal of physiology. Endocrinology and metabolism 302, E800-806.

Activinsights, 2019. Downloads and software. Available at: https://www.activinsights.com/resources-support/geneactiv/downloads-software/ (Accessed: 19 December 2019).

Adam, E.K., Quinn, M.E., Tavernier, R., McQuillan, M.T., Dahlke, K.A., Gilbert, K.E., 2017. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. Psychoneuroendocrinology 83, 25-41.

Adams, J.D., Ganio, M.S., Burchfield, J.M., Matthews, A.C., Werner, R.N., Chokbengboun, A.J., Dougherty, E.K., LaChance, A.A., 2015. Effects of obesity on body temperature in otherwise-healthy females when controlling hydration and heat production during exercise in the heat. European journal of applied physiology 115, 167-176.

Adcock, M.P., Bines WH., Smith FW., 2000. Heat-related illnesses, deaths, and risk factors--Cincinnati and Dayton, Ohio, 1999, and United States, 1979-1997. (Reprinted) Journal of the American Medical Association 284, 34-35.

Adler, A.B., Williams, J., McGurk, D., Moss, A., Bliese, P.D., 2015. Resilience training with soldiers during basic combat training: randomisation by platoon. Applied Psychology: Health and Well-Being 7, 85-107.

Ahrens, K.A., Vladutiu, C.J., Mumford, S.L., Schliep, K.C., Perkins, N.J., Wactawski-Wende, J., Schisterman, E.F., 2014. The effect of physical activity across the menstrual cycle on reproductive function. Annals of epidemiology 24, 127-134.

Alemany, J.A., Nindl, B.C., Kellogg, M.D., Tharion, W.J., Young, A.J., Montain, S.J., 2008. Effects of dietary protein content on IGF-I, testosterone, and body composition during 8 days of severe energy deficit and arduous physical activity. Journal of Applied Physiology 105, 58-64.

Alfonso, A.J., Ryan, D.H., Williamson, D.A., Martin, P.D., 2006. Psychological Adjustment during Army Basic Training. Military medicine 171, 157-160.

Allen, A.P., Kennedy, P.J., Dockray, S., Cryan, J.F., Dinan, T.G., Clarke, G., 2017. The Trier Social Stress Test: Principles and practice. Neurobiology of stress 6, 113-126.

Andersen, N.E., Karl, J.P., Cable, S.J., Williams, K.W., Rood, J.C., Young, A.J., Lieberman, H.R., McClung, J.P., 2010. Vitamin D status in female military personnel during combat training. Journal of the International Society of Sports Nutrition 7, 38.

Anderson, J., 1979. Women's sports and fitness programs at the US Military Academy. The Physician and sportsmedicine 7, 72-82.

Anderson, R.J., Reed, G., Knochel, J., 1983. Heatstroke. Advances in internal medicine 28, 115-140.

Andrews, M.A., Schliep, K.C., Wactawski-Wende, J., Stanford, J.B., Zarek, S.M., Radin, R.G., Sjaarda, L.A., Perkins, N.J., Kalwerisky, R.A., Hammoud, A.O., Mumford, S.L., 2015. Dietary factors and luteal phase deficiency in healthy eumenorrheic women. Human reproduction (Oxford, England) 30, 1942-1951.

Apter, D., Gemzell-Danielsson, K., Hauck, B., Rosen, K., Zurth, C., 2014. Pharmacokinetics of two low-dose levonorgestrel-releasing intrauterine systems and effects on ovulation rate and cervical function: pooled analyses of phase II and III studies. Fertility and sterility 101, 1656-1662.e1651-1654.

Armed Forces Health Surveillance Center, 2011. Update: heat injuries, active component, U.S. Armed Forces, 2010. MSMR 18, 6-8.

Armed Forces Health Surveillance Center, 2012. Update: heat injuries, active component, U.S. Armed Forces, 2011. MSMR 19, 14-16.

Armed Forces Health Surveillance Center, 2013. Update: Heat injuries, active component, U.S. Armed Forces, 2012. MSMR 20, 17-20.

Armed Forces Health Surveillance Center, 2014. Update: Heat injuries, active component, U.S. Armed Forces, 2013. MSMR 21, 10-13.

Armed Forces Health Surveillance Center, 2015. Update: Heat injuries, active component, U.S. Armed Forces, 2014. MSMR 22, 17-20.

Armed Forces Health Surveillance Center, 2016. Update: Heat injuries, active component, U.S. Army, Navy, Air Force, and Marine Corps, 2015. MSMR 23, 16-19.

Armed Forces Health Surveillance Center, 2017. Update: Heat illness, active component, U.S. Armed Forces, 2016. MSMR 24, 9-13.

Armstrong, L.E., Casa, D.J., Millard-Stafford, M., Moran, D.S., Pyne, S.W., Roberts, W.O., 2007. American College of Sports Medicine position stand. Exertional heat illness during training and competition. Medicine and science in sports and exercise 39, 556-572.

Armstrong, L.E., Maresh, C.M., Keith, N.R., Elliott, T.A., Vanheest, J.L., Scheett, T.P., Stoppani, J., Judelson, D.A., De Souza, M.J., 2005. Heat acclimation and physical training adaptations of young women using different contraceptive hormones. American journal of physiology. Endocrinology and metabolism 288, E868-875.

Armstrong, R.G., Ahmad, S., Seely, A.J., Kenny, G.P., 2012. Heart rate variability and baroreceptor sensitivity following exercise-induced hyperthermia in endurance trained men. European journal of applied physiology 112, 501-511.

Atlis, M.M., Leon, G.R., Sandal, G.M., Infante, M.G., 2004. Decision Processes and Interactions During a Two-Woman Traverse of Antarctica. Environment and Behavior 36, 402-423.

Avendano, M.S., Vazquez, M.J., Tena-Sempere, M., 2017. Disentangling puberty: novel neuroendocrine pathways and mechanisms for the control of mammalian puberty. Human reproduction update 23, 737-763.

Bae, Y.J., Kratzsch, J., 2015. Corticosteroid-binding globulin: modulating mechanisms of bioavailability of cortisol and its clinical implications. Best Practice & Research Clinical Endocrinology & Metabolism 29, 761-772.

Baevsky, R., Berseneva, A., 2008. Use of Kardivar System for Determination of the Stress Level and Estimation of the Body Adaptability.

Bai, L., Ding, G., Gu, S., Bi, P., Su, B., Qin, D., Xu, G., Liu, Q., 2014. The effects of summer temperature and heat waves on heat-related illness in a coastal city of China, 2011-2013. Environmental Research 132, 212-219.

Baird, D.D., Weinberg, C.R., Wilcox, A.J., McConnaughey, D.R., Musey, P.I., 1991. Using the ratio of urinary oestrogen and progesterone metabolites to estimate day of ovulation. Statistics in medicine 10, 255-266.

Baker, F.C., Driver, H.S., 2007. Circadian rhythms, sleep, and the menstrual cycle. Sleep Medicine 8, 613-622.

Bandelow, B., Baldwin, D., Abelli, M., Bolea-Alamanac, B., Bourin, M., Chamberlain, S.R., Cinosi, E., Davies, S., Domschke, K., Fineberg, N., Grunblatt, E., Jarema, M., Kim, Y.K., Maron, E., Masdrakis, V., Mikova, O., Nutt, D., Pallanti, S., Pini, S., Strohle, A., Thibaut, F., Vaghi, M.M., Won, E., Wedekind, D., Wichniak, A., Woolley, J., Zwanzger, P., Riederer, P., 2017. Biological markers for anxiety disorders, OCD and PTSD: A consensus statement. Part II: Neurochemistry, neurophysiology and neurocognition. The world journal of biological psychiatry 18, 162-214.

Banning, E.J., 2010. The effect of homesickness on Air Force Academy cadets. Doctor of Psychology thesis. Philadelphia College of Osteopathic Medicine. Available at: https://pdfs.semanticscholar.org/316e/ac04f0599f82248d26aea5aa40f0f0a641c9.pdf (Accessed: 19 December 2019).

Barrack, M.T., Ackerman, K.E., Gibbs, J.C., 2013. Update on the female athlete triad. Curr Rev Musculoskelet Med 6, 195-204.

Bartlett, B.A., Mitchell, K.S., 2015. Eating disorders in military and veteran men and women: A systematic review. International Journal of Eating Disorders 48, 1057-1069.

Bassett, S.M., Lupis, S.B., Gianferante, D., Rohleder, N., Wolf, J.M., 2015. Sleep quality but not sleep quantity effects on cortisol responses to acute psychosocial stress. Stress (Amsterdam, Netherlands) 18, 638-644.

Batig, A.L., 2017. The Incidence of Intrauterine Device Provision in a Military Tertiary Care Facility From 2008-2014 Correlated to Contemporary Contraception Guidelines. Military medicine 182, e1869-e1873.

Baumgartner, A., Dietzel, M., Saletu, B., Wolf, R., Campos-Barros, A., Gräf, K.-J., Kürten, I., Mannsmann, U., 1993. Influence of partial sleep deprivation on the secretion of thyrotropin, thyroid hormones, growth hormone, prolactin, luteinizing hormone, follicle stimulating hormone, and estradiol in healthy young women. Psychiatry Research 48, 153-178.

BBC News, 2011. Sandhurst Military Academy is like 'Hogwarts with guns'. Available at: https://www.bbc.co.uk/news/uk-14929441 (Accessed: 7 May 2019).

BBC News, 2016. Women to serve in close combat roles in the British military. Available at: https://www.bbc.co.uk/news/uk-36746917 (Accessed: 10 May 2019).

Beals, K., Darnell, M.E., Lovalekar, M., Baker, R.A., Nagai, T., San-Adams, T., Wirt, M.D., 2015. Suboptimal Nutritional Characteristics in Male and Female Soldiers Compared to Sports Nutrition Guidelines. Military medicine 180, 1239-1246.

Beck, A.T., Epstein, N., Brown, G., Steer, R.A., 1988. An inventory for measuring clinical anxiety: psychometric properties. Journal of consulting and clinical psychology 56, 893-897.

Bedno, S.A., Li, Y., Han, W., Cowan, D.N., Scott, C.T., Cavicchia, M.A., Niebuhr, D.W., 2010. Exertional heat illness among overweight U.S. Army recruits in basic training. Aviation Space & Environmental Medicine 81, 107-111.

Bedno, S.A., Urban, N., Boivin, M.R., Cowan, D.N., 2014. Fitness, obesity and risk of heat illness among army trainees. Occupational Medicine (Oxford) 64, 461-467.

Beitins, I.Z., McArthur, J.W., Turnbull, B.A., Skrinar, G.S., Bullen, B.A., 1991. Exercise Induces Two Types of Human Luteal Dysfunction: Confirmation by Urinary Free Progesterone*. The Journal of Clinical Endocrinology & Metabolism 72, 1350-1358.

Beksinska, M.E., Kleinschmidt, I., Smit, J.A., Farley, T.M., Rees, H.V., 2009. Bone mineral density in young women aged 19-24 after 4-5 years of exclusive and mixed use of hormonal contraception. Contraception 80, 128-132.

Belanger, D., Gosselin, P., Valois, P., Abdous, B., 2014. Perceived adverse health effects of heat and their determinants in deprived neighbourhoods: a cross-sectional survey of nine cities in Canada. Int J Environ Res Public Health. 11, 11028-11053. doi: 11010.13390/ijerph111111028.

Bellenger, C.R., Fuller, J.T., Thomson, R.L., Davison, K., Robertson, E.Y., Buckley, J.D., 2016. Monitoring athletic training status through autonomic heart rate regulation: a systematic review and meta-analysis. Sports medicine 46, 1461-1486.

Benvenga, S., Di Bari, F., Granese, R., Borrielli, I., Giorgianni, G., Grasso, L., Le Donne, M., Vita, R., Antonelli, A., 2018. Circulating thyrotropin is upregulated by estradiol. Journal of clinical & translational endocrinology 11, 11-17.

Berenson, A.B., Breitkopf, C.R., Grady, J.J., Rickert, V.I., Thomas, A., 2004. Effects of hormonal contraception on bone mineral density after 24 months of use. Obstetrics & Gynecology 103, 899-906.

Berga, S.L., 2008. Stress and reprodution: a tale of false dichotomy? Endocrinology 149, 867-868.

Berger, P.K., Pollock, N.K., Laing, E.M., Chertin, V., Bernard, P.J., Grider, A., Shapses, S.A., Ding, K.-H., Isales, C.M., Lewis, R.D., 2015. Zinc supplementation increases procollagen type 1 amino-terminal propeptide in premenarcheal girls: a randomized controlled trial. The Journal of nutrition 145, 2699-2704.

Bernardi, R.A., 1997. The relationships among locus of control, perceptions of stress and performance. Journal of Applied Business Research (JABR) 13, 1-8.

Black, A.E., Prentice, A.M., Coward, W.A., 1986. Use of food quotients to predict respiratory quotients for the doubly-labelled water method of measuring energy expenditure. Human nutrition. Clinical nutrition 40, 381-391.

Blackadder-Weinstein, J., Leon, G.R., Norris, R.C., Venables, N.C., Smith, M., 2019. Individual Attributes, Values, and Goals of an All-Military Women Antarctic Expedition. Aerospace medicine and human performance 90, 18-25.

Blackwell, L.F., Cooke, D.G., Brown, S., 2018. The Use of Estrone-3-Glucuronide and Pregnanediol-3-Glucuronide Excretion Rates to Navigate the Continuum of Ovarian Activity. Frontiers in Public Health 6, 153.

Bland, J.M., Altman, D.G., 1999. Measuring agreement in method comparison studies. Statistical methods in medical research 8, 135-160.

Blundell, J.E., King, N.A., 1999. Physical activity and regulation of food intake: current evidence. Medicine and science in sports and exercise 31, S573-583.

Blundell, J.E., Stubbs, R.J., Hughes, D.A., Whybrow, S., King, N.A., 2003. Cross talk between physical activity and appetite control: does physical activity stimulate appetite? The Proceedings of the Nutrition Society 62, 651-661.

Boeckmann, M., Rohn, I., 2014. Is planned adaptation to heat reducing heat-related mortality and illness? A systematic review. BMC public health 14, 1112.

Boesch, M., Sefidan, S., Annen, H., Ehlert, U., Roos, L., Van Uum, S., Russell, E., Koren, G., La Marca, R., 2015. Hair cortisol concentration is unaffected by basic military training, but related to sociodemographic and environmental factors. Stress (Amsterdam, Netherlands) 18, 35-41.

Boesch, M., Sefidan, S., Ehlert, U., Annen, H., Wyss, T., Steptoe, A., La Marca, R., 2014. Mood and autonomic responses to repeated exposure to the Trier Social Stress Test for Groups (TSST-G). Psychoneuroendocrinology 43, 41-51.

Bollini, A.M., Walker, E.F., Hamann, S., Kestler, L., 2004. The influence of perceived control and locus of control on the cortisol and subjective responses to stress. Biological psychology 67, 245-260.

Boos, C., Vincent, E., Mellor, A., O Hara, J., Newman, C., Cruttenden, R., Scott, P., Cooke, M., Matu, J., Woods, D.R., 2017. The effect of sex on heart rate variability at high altitude. Medicine and science in sports and exercise 49, 2562-2569.

Booth, C.K., Probert, B., Forbes-Ewan, C., Coad, R.A., 2006. Australian army recruits in training display symptoms of overtraining. Military medicine 171, 1059-1064.

Borner, U., Szasz, G., Bablok, W., Busch, E.W., 1979. [A specific fully enzymatic method for creatinine: reference values in serum (author's transl)]. Journal of clinical chemistry and clinical biochemistry. Zeitschrift fur klinische Chemie und klinische Biochemie 17, 679-682.

Bornmann, J.W., 2009. Becoming soldiers: Army Basic Training and the negotiation of identity. PhD Thesis. George Washington University. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.615.9700&rep=rep1&type=pdf (Accessed 19 December 2019)

Bosy-Westphal, A., Kahlhofer, J., Lagerpusch, M., Skurk, T., Muller, M.J., 2015. Deep body composition phenotyping during weight cycling: relevance to metabolic efficiency and metabolic risk. Obes Rev 16 Suppl 1, 36-44.

Bouassida, A., Zalleg, D., Bouassida, S., Zaouali, M., Feki, Y., Zbidi, A., Tabka, Z., 2006. Leptin, its implication in physical exercise and training: a short review. J Sports Sci Med 5, 172.

Bouchama, A., Knochel, J.P., 2002. Heat stroke. The New England journal of medicine 346, 1978-1988.

Bourbonnais, R., Brisson, C., Moisan, J., Vézina, M., 1996. Job strain and psychological distress in white-collar workers. Scandinavian journal of work, environment & health 22, 139-145.

Brabant, G., Prank, K., Ranft, U., Schuermeyer, T., Wagner, T.O.F., Hauser, H., Kummer, B., Feistner, H., Hesch, R.D., MüHlen, A.V.Z., 1990. Physiological Regulation of Circadian and Pulsatile Thyrotropin Secretion in Normal Man and Woman*. The Journal of Clinical Endocrinology & Metabolism 70, 403-409.

Braun, B., Newman, A., 2019. Accounting for the Nutritional Context to Correctly Interpret Results from Studies of Exercise and Sedentary Behavior. Nutrients 11, 2230.

Breland, J.Y., Donalson, R., Nevedal, A., Dinh, J.V., Maguen, S., 2017. Military experience can influence Women's eating habits. Appetite 118, 161-167.

Briffa, J., 2005. What's in your basket? Available at: https://www.theguardian.com/lifeandstyle/2005/may/15/foodanddrink (Accessed: 19 December 2019).

Broussard, J.L., Chapotot, F., Abraham, V., Day, A., Delebecque, F., Whitmore, H.R., Tasali, E., 2015. Sleep restriction increases free fatty acids in healthy men. Diabetologia 58, 791-798.

Brown, M.A., Howatson, G., Quin, E., Redding, E., Stevenson, E.J., 2017. Energy intake and energy expenditure of pre-professional female contemporary dancers. PloS one 12, e0171998.

Brundu, B., Loucks, T.L., Adler, L.J., Cameron, J.L., Berga, S.L., 2006. Increased cortisol in the cerebrospinal fluid of women with functional hypothalamic amenorrhea. The Journal of clinical endocrinology and metabolism 91, 1561-1565.

Bullen, B.A., Skrinar, G.S., Beitins, I.Z., von Mering, G., Turnbull, B.A., McArthur, J.W., 1985. Induction of menstrual disorders by strenuous exercise in untrained women. The New England journal of medicine 312, 1349-1353.

Bungum, L., Jacobsson, A.-K., Rosen, F., Becker, C., Yding Andersen, C., Güner, N., Giwercman, A., 2011. Circadian variation in concentration of anti-Müllerian hormone in regularly menstruating females: relation to age, gonadotrophin and sex steroid levels. Human reproduction 26, 678-684.

Burkart, K., Breitner, S., Schneider, A., Khan, M.M., Kramer, A., Endlicher, W., 2014. An analysis of heat effects in different subpopulations of Bangladesh. International Journal of Biometeorology 58, 227-237.

Burke, L.M., Lundy, B., Fahrenholtz, I.L., Melin, A.K., 2018. Pitfalls of Conducting and Interpreting Estimates of Energy Availability in Free-Living Athletes. International journal of sport nutrition and exercise metabolism 28, 350-363.

Byrnes, J.P., Miller, D.C., Schafer, W.D., 1999. Gender Differences in Risk Taking: A Meta-Analysis. Psychological bulletin 125, 367-383.

Cadegiani, F.A., Kater, C.E., 2017a. Hormonal aspects of overtraining syndrome: a systematic review. BMC sports science, medicine and rehabilitation 9, 14.

Cadegiani, F.A., Kater, C.E., 2017b. Hypothalamic-Pituitary-Adrenal (HPA) Axis Functioning in Overtraining Syndrome: Findings from Endocrine and Metabolic Responses on Overtraining Syndrome (EROS)—EROS-HPA Axis. Sports Med Open 3, 45

Cadegiani, F.A., Kater, C.E., 2018a. Body composition, metabolism, sleep, psychological and eating patterns of overtraining syndrome: Results of the EROS study (EROS-PROFILE). Journal of sports sciences 36, 1902-1910.

Cadegiani, F.A., Kater, C.E., 2018b. Hormonal response to a non-exercise stress test in athletes with overtraining syndrome: results from the Endocrine and metabolic Responses on Overtraining Syndrome (EROS) - EROS-STRESS. J Sci Med Sport 21, 648-653.

Calbet, J.A., Ponce-González, J.G., Calle-Herrero, J.d.L., Perez-Suarez, I., Martin-Rincon, M., Santana, A., Morales-Alamo, D., Holmberg, H.-C., 2017. Exercise preserves lean mass and performance during severe energy deficit: the role of exercise volume and dietary protein content. Frontiers in physiology 8, 483.

Campbell-Sills, L., Forde, D.R., Stein, M.B., 2009. Demographic and childhood environmental predictors of resilience in a community sample. J Psychiatr Res 43, 1007-1012.

Campbell-Sills, L., Stein, M.B., 2007. Psychometric analysis and refinement of the connor–davidson resilience scale (CD-RISC): Validation of a 10-item measure of resilience. Journal of traumatic stress 20, 1019-1028.

Cano Sokoloff, N., Misra, M., Ackerman, K.E., 2016. Exercise, Training, and the Hypothalamic-Pituitary-Gonadal Axis in Men and Women. Frontiers of hormone research 47, 27-43.

Capling, L., Beck, K., Gifford, J., Slater, G., Flood, V., O'Connor, H., 2017. Validity of dietary assessment in athletes: A systematic review. Nutrients 9, 1313.

Carbone, E.G., Cigrang, J.A., Todd, S., Fiedler, E., 1998. Mental Health Attrition from Air Force Basic Military Training. Military medicine 163, 834-838.

Carson, J.A., Manolagas, S.C., 2015. Effects of sex steroids on bones and muscles: Similarities, parallels, and putative interactions in health and disease. Bone 80, 67-78.

Carter, R., 3rd, Cheuvront, S.N., Williams, J.O., Kolka, M.A., Stephenson, L.A., Sawka, M.N., Amoroso, P.J., 2005. Epidemiology of hospitalizations and deaths from heat illness in soldiers. Medicine & Science in Sports & Exercise 37, 1338-1344.

Castellani, J.W., Delany, J.P., O'Brien, C., Hoyt, R.W., Santee, W.R., Young, A.J., 2006. Energy expenditure in men and women during 54 h of exercise and caloric deprivation. Medicine and science in sports and exercise 38, 894-900.

Castrission, J., 2013. Extreme South. Adelaide: Hachette.

Cea-Soriano, L., Garcia Rodriguez, L.A., Machlitt, A., Wallander, M.A., 2014. Use of prescription contraceptive methods in the UK general population: a primary care study. BJOG : an international journal of obstetrics and gynaecology 121, 53-60; discussion 60-51.

Centers for Disease, Control and Prevention, 2006. Heat-related deaths--United States, 1999-2003. MMWR - Morbidity & Mortality Weekly Report 55, 796-798.

Charkoudian, N., Johnson, J.M., 1997. Modification of active cutaneous vasodilation by oral contraceptive hormones. Journal of applied physiology (Bethesda, Md. : 1985) 83, 2012-2018.

Charkoudian, N., Joyner, M.J., 2004. Physiologic considerations for exercise performance in women. Clinics in chest medicine 25, 247-255.

Charkoudian, N., Stachenfeld, N., 2016. Sex hormone effects on autonomic mechanisms of thermoregulation in humans. Autonomic Neuroscience 196, 75-80.

Charkoudian, N., Stephens, D.P., Pirkle, K.C., Kosiba, W.A., Johnson, J.M., 1999. Influence of female reproductive hormones on local thermal control of skin blood flow. Journal of applied physiology (Bethesda, Md. : 1985) 87, 1719-1723.

Che Muhamed, A.M., Atkins, K., Stannard, S.R., Mündel, T., Thompson, M.W., 2016. The effects of a systematic increase in relative humidity on thermoregulatory and circulatory responses during prolonged running exercise in the heat. Temperature (Austin, Tex.) 3, 455-464.

Chen, T., Liu, H.-x., Yan, H.-y., Wu, D.-m., Ping, J., 2016. Developmental origins of inflammatory and immune diseases. Mol Hum Reprod 22, pp. 858-865.

Cheung, A.M., Adachi, J.D., Hanley, D.A., Kendler, D.L., Davison, K.S., Josse, R., Brown, J.P., Ste-Marie, L.G., Kremer, R., Erlandson, M.C., Dian, L., Burghardt, A.J., Boyd, S.K., 2013. High-resolution peripheral quantitative computed tomography for the assessment of bone strength and structure: a review by the Canadian Bone Strength Working Group. Current osteoporosis reports 11, 136-146.

Chida, Y., Steptoe, A., 2009. Cortisol awakening response and psychosocial factors: A systematic review and meta-analysis. Biological psychology 80, 265-278.

Cho, G.J., Han, S.W., Shin, J.-H., Kim, T., 2017. Effects of intensive training on menstrual function and certain serum hormones and peptides related to the female reproductive system. Medicine 96, e6876-e6876.

Chou, S.H., Chamberland, J.P., Liu, X., Matarese, G., Gao, C., Stefanakis, R., Brinkoetter, M.T., Gong, H., Arampatzi, K., Mantzoros, C.S., 2011. Leptin is an effective treatment for hypothalamic amenorrhea. Proceedings of the National Academy of Sciences of the United States of America 108, 6585-6590.

Cialdella-Kam, L., Guebels, C.P., Maddalozzo, G.F., Manore, M.M., 2014. Dietary intervention restored menses in female athletes with exercise-associated menstrual dysfunction with limited impact on bone and muscle health. Nutrients 6, 3018-3039.

Clark, H.L., Heileson, J., DeMay, J., Cole, R.E., 2017. Misperceptions of Weight Status in Military Men and Women. Military medicine 182, e1792-e1798.

Clow, A., Edwards, S., Owen, G., Evans, G., Evans, P., Hucklebridge, F., Casey, A., 2006. Post-awakening cortisol secretion during basic military training. Int J Psychophysiol 60, 88-94.

Cohen, S., Doyle, W.J., Baum, A., 2006. Socioeconomic status is associated with stress hormones. Psychosom Med 68, 414-420.

Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. Journal of health and social behavior, 385-396.

Compher, C., Frankenfield, D., Keim, N., Roth-Yousey, L., Group, E.A.W., 2006. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. Journal of the American Dietetic Association 106, 881-903.

Connor, K.M., Davidson, J.R., 2003. Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). Depression and anxiety 18, 76-82.

Constantini, N.W., Warren, M.P., 1995. Menstrual dysfunction in swimmers: a distinct entity. The Journal of clinical endocrinology and metabolism 80, 2740-2744.

Coombs, G.B., Cramer, M.N., Ravanelli, N.M., Morris, N.B., Jay, O., 2015. Acute acetaminophen ingestion does not alter core temperature or sweating during exercise in hot-humid conditions. Scandinavian journal of medicine & science in sports 25 Suppl 1, 96-103.

Cooper, Z., Cooper, P.J., Fairburn, C.G., 1989. The validity of the eating disorder examination and its subscales. The British journal of psychiatry : the journal of mental science 154, 807-812.

Corneil, W., Beaton, R., Murphy, S., Johnson, C., Pike, K., 1999. Exposure to traumatic incidents and prevalence of posttraumatic stress symptomatology in urban firefighters in two countries. Journal of occupational health psychology 4, 131-141.

Corr, M., De Souza, M., Toombs, R., Williams, N., 2011. Circulating leptin concentrations do not distinguish menstrual status in exercising women. Human reproduction 26, 685-694.

Coward, W., 1988. The doubly-labelled-water (2 H 2 18 O) method: principles and practice. Proceedings of the Nutrition Society 47, 209-218.

Cox, A.T., Gillingham, S., Johnson, S.A., Sharma, S., Wilson, D., 2016. An assessment of the accuracy and utility of the primary care electronic medical record as used by the British Army. Journal of the Royal Army Medical Corps 162, 361-365.

Craig, H., 1957. Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. Geochimica et Cosmochimica Acta 12, 133-149.

Cramer, M.N., Jay, O., 2015. Explained variance in the thermoregulatory responses to exercise: the independent roles of biophysical and fitness/fatness-related factors. Journal of applied physiology (Bethesda, Md. : 1985) 119, 982-989.

Cramer, M.N., Jay, O., 2016. Biophysical aspects of human thermoregulation during heat stress. Autonomic neuroscience : basic & clinical 196, 3-13.

Crowley, S.K., Wilkinson, L.L., Burroughs, E.L., Muraca, S.T., Wigfall, L.T., Louis-Nance, T., Williams, E.M., Glover, S.H., Youngstedt, S.D., 2012. Sleep during basic combat training: a qualitative study. Military medicine 177, 823-828.

Crowley, S.K., Wilkinson, L.L., Wigfall, L.T., Reynolds, A.M., Muraca, S.T., Glover, S.H., Wooten, N.R., Sui, X., Beets, M.W., Durstine, J.L., 2015. Physical fitness and depressive symptoms during army basic combat training. Medicine and science in sports and exercise 47, 151.

Cunanan, A.J., DeWeese, B.H., Wagle, J.P., Carroll, K.M., Sausaman, R., Hornsby, W.G., Haff, G.G., Triplett, N.T., Pierce, K.C., Stone, M.H., 2018a. Authors' reply to buckner et al.:'Comment on:"The general adaptation syndrome: A foundation for the concept of periodization". Sports Medicine 48, 1755-1757.

Cunningham, J.J., 1991. Body composition as a determinant of energy expenditure: a synthetic review and a proposed general prediction equation. The American journal of clinical nutrition 54, 963-969.

Danilenko, K.V., Samoilova, E.A., 2007. Stimulatory effect of morning bright light on reproductive hormones and ovulation: results of a controlled crossover trial. PLoS clinical trials 2, e7.

Dantas-Ferreira, R.F., Raingard, H., Dumont, S., Schuster-Klein, C., Guardiola-Lemaitre, B., Pevet, P., Challet, E., 2018. Melatonin potentiates the effects of metformin on glucose metabolism and food intake in high-fat-fed rats. Endocrinology, Diabetes & Metabolism 1, e00039.

Davidson, J.R., 2018. Connor-Davidson Resilience Scale (CDRISC) Manual. Available at: www.cd-risc.com (Accessed 19 December 2019).

Davis, R.E., Knappenberger, P.C., Michaels, P.J., Novicoff, W.M., 2003. Changing heat-related mortality in the United States. Environmental Health Perspectives 111, 1712-1718.

De Souza, M.J., Nattiv, A., Joy, E., Misra, M., Williams, N.I., Mallinson, R.J., Gibbs, J.C., Olmsted, M., Goolsby, M., Matheson, G., 2014. 2014 Female Athlete Triad Coalition Consensus Statement on Treatment and Return to Play of the Female Athlete Triad: 1st International Conference held in San

Francisco, California, May 2012 and 2nd International Conference held in Indianapolis, Indiana, May 2013. British journal of sports medicine 48, 289.

De Souza, M.J., Koltun, K.J., Etter, C.V., Southmayd, E.A., 2017. Current Status of the Female Athlete Triad: Update and Future Directions. Current osteoporosis reports 15, 577-587.

De Souza, M.J., Koltun, K.J., Strock, N.C.A., Williams, N.I., 2019a. Rethinking the concept of an energy availability threshold and its role in the Female Athlete Triad. Current Opinion in Physiology 10, 35-42.

De Souza, M.J., Koltun, K.J., Williams, N.I., 2019b. What is the evidence for a Triad-like syndrome in exercising men? Current Opinion in Physiology 10, 27-34.

De Souza, M.J., Leidy, H.J., O'Donnell, E., Lasley, B., Williams, N.I., 2004. Fasting ghrelin levels in physically active women: relationship with menstrual disturbances and metabolic hormones. The Journal of clinical endocrinology and metabolism 89, 3536-3542.

De Souza, M.J., Miller, B.E., Loucks, A.B., Luciano, A.A., Pescatello, L.S., Campbell, C.G., Lasley, B.L., 1998. High frequency of luteal phase deficiency and anovulation in recreational women runners: blunted elevation in follicle-stimulating hormone observed during luteal-follicular transition. The Journal of clinical endocrinology and metabolism 83, 4220-4232.

De Souza, M.J., Nattiv, A., Joy, E., Misra, M., Williams, N.I., Mallinson, R.J., Gibbs, J.C., Olmsted, M., Goolsby, M., Matheson, G., 2014a. 2014 Female Athlete Triad Coalition Consensus Statement on Treatment and Return to Play of the Female Athlete Triad: 1st International Conference held in San Francisco, California, May 2012 and 2nd International Conference held in Indianapolis, Indiana, May 2013. Br J Sports Med 48, 289.

De Souza, M.J., Toombs, R.J., Scheid, J.L., O'Donnell, E., West, S.L., Williams, N.I., 2010. High prevalence of subtle and severe menstrual disturbances in exercising women: confirmation using daily hormone measures. Human reproduction 25, 491-503.

De Souza, M.J., Williams, N.I., Nattiv, A., Joy, E., Misra, M., Loucks, A.B., Matheson, G., Olmsted, M.P., Barrack, M., Mallinson, R.J., 2014b. Misunderstanding the female athlete triad: refuting the IOC consensus statement on Relative Energy Deficiency in Sport (RED-S). Br J Sports Med 48, 1461-1465.

Defence Statistics (Health), 2018. UK Armed Forces Mental Health: Annual Summary & Trends Over Time, 2007/08 - 2017/18. Ministry of Defence, London

Department of Manning (Army), 2016. Interim report on the health risks to women in ground close combat roles 2016. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/536381/20160706_ADR006101_Report_Women_in_Combat_WEB-FINAL.PDF (Accessed 19 December 2019).

Dewailly, D., Andersen, C.Y., Balen, A., Broekmans, F., Dilaver, N., Fanchin, R., Griesinger, G., Kelsey, T.W., La Marca, A., Lambalk, C., Mason, H., Nelson, S.M., Visser, J.A., Wallace, W.H., Anderson, R.A.,

2014. The physiology and clinical utility of anti-Mullerian hormone in women. Human reproduction update 20, 370-385.

Dhalwani, N.N., Fiaschi, L., West, J., Tata, L.J., 2013. Occurrence of fertility problems presenting to primary care: population-level estimates of clinical burden and socioeconomic inequalities across the UK. Human reproduction (Oxford, England) 28, 960-968.

Dhillo, W.S., Liang, S., Kinghorn, A., Voliotis, M., Prague, J.K., Veldhuis, J.D., Tsaneva-Atanasova, K., McArdle, C.A., Li, H.W.R., Cass, T., Tanner, J., 2019. SAT-LB040 Measuring LH Pulsatility in Patients with Reproductive Disorders Using a Novel Robotic Aptamer-Enabled Electrochemical Reader (RAPTER))From ENDO March 23-26 2019, New Orleans) Journal of the endocrine society, 2019-SAT-LB040

DiPietro, L., Stachenfeld, N.S., 2006. The myth of the female athlete triad, Br J Sports Med, 40, 490-493.

Doyle-Lucas, A.F., Akers, J.D., Davy, B.M., 2010. Energetic efficiency, menstrual irregularity, and bone mineral density in elite professional female ballet dancers. Journal of Dance Medicine & Science 14, 146-154.

Drain, J.R., Groeller, H., Burley, S.D., Nindl, B.C., 2017. Hormonal response patterns are differentially influenced by physical conditioning programs during basic military training. J Sci Med Sport 20 Suppl 4, S98-s103.

Drinkwater, B., 2015. Active Voice: ACSM's Role in Promoting Physical Activity for Women, Sports Medicine Bulletin. Available at: http://www.multibriefs.com/briefs/acsm/active3-10.htm (Accessed 19 December 2019).

Druyan, A., Makranz, C., Moran, D., Yanovich, R., Epstein, Y., Heled, Y., 2012. Heat tolerance in women--reconsidering the criteria. Aviat Space Environ Med. 83, 58-60.

Du Preez, E.J., Graham, K.S., Gan, T.Y., Moses, B., Ball, C., Kuah, D.E., 2017. Depression, Anxiety, and Alcohol Use in Elite Rugby League Players Over a Competitive Season. Clin J Sport Med 27, 530-535

Duckworth, A.L., Eskreis-Winkler, L., 2013. True grit. Available at: https://www.psychologicalscience.org/observer/true-grit (Accessed: 16 December 2019).

Dulac, S., Quirion, A., Brisson, G.R., Decarufel, D., 1986. Sex difference in serum testosterone response to long-distance swimming. Hormone and metabolic research 18, 420-421.

Dulloo, A., Jacquet, J., 1999. The control of partitioning between protein and fat during human starvation: its internal determinants and biological significance. British Journal of Nutrition 82, 339-356.

Durnin, J.V., Womersley, J., 1974. Body fat assessed from total body density and its estimation from skinfold thickness: measurements on 481 men and women aged from 16 to 72 years. British journal of nutrition 32, 77-97.

Eastlake, A.C., Knipper, B.S., He, X., Alexander, B.M., Davis, K.G., 2015. Lifestyle and safety practices of firefighters and their relation to cardiovascular risk factors. Work 50, 285-294.

Ebbeling, C.B., Feldman, H.A., Klein, G.L., Wong, J.M.W., Bielak, L., Steltz, S.K., Luoto, P.K., Wolfe, R.R., Wong, W.W., Ludwig, D.S., 2018. Effects of a low carbohydrate diet on energy expenditure during weight loss maintenance: randomized trial. Bmj 363, k4583.

Ebisch, I., Thomas, C., Peters, W., Braat, D., Steegers-Theunissen, R., 2006. The importance of folate, zinc and antioxidants in the pathogenesis and prevention of subfertility. Human reproduction update 13, 163-174.

Ecochard, R., Bouchard, T., Leiva, R., Abdulla, S., Dupuis, O., Duterque, O., Garmier Billard, M., Boehringer, H., Genolini, C., 2017. Characterization of hormonal profiles during the luteal phase in regularly menstruating women. Fertil Steril 108, 175-182.e171.

Ecochard, R., Leiva, R., Bouchard, T., Boehringer, H., Direito, A., Mariani, A., Fehring, R., 2013. Use of urinary pregnanediol 3-glucuronide to confirm ovulation. Steroids 78, 1035-1040.

Elia, M., Livesey, G., 1988. Theory and validity of indirect calorimetry during net lipid synthesis. The American journal of clinical nutrition 47, 591-607.

Elliott-Sale, K.J., Tenforde, A.S., Allyson, L.P., Holtzman, B., Ackerman, K.E., 2018. Endocrine Effects of Relative Energy Deficiency in Sport. International journal of sport nutrition and exercise metabolism 28, 335-349.

Ellis, F.P., 1972. Mortality from heat illness and heat-aggravated illness in the United States. Environmental Research 5, 1-58.

Ellis, F.P., Prince, H.P., Lovatt, G., Whittington, R.M., 1980. Mortality and Morbidity in Birmingham during the 1976 Heatwave. QJM: An International Journal of Medicine 49, 1-8.

Ennour-Idrissi, K., Maunsell, E., Diorio, C., 2015. Effect of physical activity on sex hormones in women: a systematic review and meta-analysis of randomized controlled trials. Breast Cancer Research 17, 139.

Epstein, Y., 1990. Heat intolerance: predisposing factor or residual injury? Medicine and science in sports and exercise 22, 29-35.

Esliger, D.W., Rowlands, A.V., Hurst, T.L., Catt, M., Murray, P., Eston, R.G., 2011. Validation of the GENEA Accelerometer. Medicine and science in sports and exercise 43, 1085-1093.

Fahrenholtz, I.L., Sjodin, A., Benardot, D., Tornberg, A.B., Skouby, S., Faber, J., Sundgot-Borgen, J.K., Melin, A.K., 2018. Within-day energy deficiency and reproductive function in female endurance athletes. Scandinavian journal of medicine & science in sports 28, 1139-1146.

Faix, J.D., 2013. Principles and pitfalls of free hormone measurements. Best practice & research clinical endocrinology & metabolism 27, 631-645.

Fallowfield, J.L., Delves, S.K., Hill, N.E., Cobley, R., Brown, P., Lanham-New, S.A., Frost, G., Brett, S.J., Murphy, K.G., Montain, S.J., Nicholson, C., Stacey, M., Ardley, C., Shaw, A., Bentley, C., Wilson, D.R., Allsopp, A.J., 2014. Energy expenditure, nutritional status, body composition and physical fitness of

Royal Marines during a 6-month operational deployment in Afghanistan. The British journal of nutrition 112, 821-829.

Falsetti, L., Gambera, A., Barbetti, L., Specchia, C., 2002. Long-term follow-up of functional hypothalamic amenorrhea and prognostic factors. The Journal of clinical endocrinology and metabolism 87, 500-505.

Ferin, M., 1999. Stress and the Reproductive Cycle. The Journal of Clinical Endocrinology & Metabolism 84, 1768-1774.

Fernando, S., Biggs, S.N., Horne, R.S.C., Vollenhoven, B., Lolatgis, N., Hope, N., Wong, M., Lawrence, M., Lawrence, A., Russell, C., Leong, K., Thomas, P., Rombauts, L., Wallace, E.M., 2018a. The impact of melatonin on the sleep patterns of women undergoing IVF: a double blind RCT. Human Reproduction Open 2017.

Fernando, S., Wallace, E.M., Vollenhoven, B., Lolatgis, N., Hope, N., Wong, M., Lawrence, M., Lawrence, A., Russell, C., Leong, K., Thomas, P., Rombauts, L., 2018b. Melatonin in Assisted Reproductive Technology: A Pilot Double-Blind Randomized Placebo-Controlled Clinical Trial. Frontiers in endocrinology 9, 545-545.

Firman, N., Palmer, M.J., Timæus, I.M., Wellings, K., 2018. Contraceptive method use among women and its association with age, relationship status and duration: findings from the third British National Survey of Sexual Attitudes and Lifestyles (Natsal-3). BMJ Sexual & Reproductive Health 44, 165-174.

Flatt, A.A., Howells, D., Williams, S., 2018. Effects of consecutive domestic and international tournaments on heart rate variability in an elite rugby sevens team. J Sci Med Sport, 22, 616-621

Fortes, M.B., Diment, B.C., Greeves, J.P., Casey, A., Izard, R., Walsh, N.P., 2011. Effects of a daily mixed nutritional supplement on physical performance, body composition, and circulating anabolic hormones during 8 weeks of arduous military training. Applied physiology, nutrition, and metabolism = Physiologie appliquee, nutrition et metabolisme 36, 967-975.

Fortune, M.K., Mustard, C.A., Etches, J.J., Chambers, A.G., 2013. Work-attributed illness arising from excess heat exposure in Ontario, 2004-2010. Canadian Journal of Public Health. Revue Canadienne de Sante Publique 104, e420-426.

Fowler, D.R., Mitchell, C.S., Brown, A., Paulson, J., Noller, A.C., Oscanyan, K., Vaidyanathan, A., Wolkinm A, Taylor, E., 2013. Heat-related deaths after an extreme heat event--four states, 2012, and United States, 1999-2009. MMWR Morb Mortal Wkly Rep. 62, 433-436.

Frautschy, S.A., Liptrap, R.M., 1988. Anovulation and plasma hormone concentrations after administration of dexamethasone during the middle of the luteal phase in sows undergoing estrous cycles. American journal of veterinary research 49, 1270-1275.

Friedl, K.E., 2005. Biomedical research on health and performance of military women: accomplishments of the Defense Women's Health Research Program (DWHRP). Journal of women's health (2002) 14, 764-802.

Friedl, K.E., Moore, R.J., Hoyt, R.W., Marchitelli, L.J., Martinez-Lopez, L.E., Askew, E.W., 2000. Endocrine markers of semistarvation in healthy lean men in a multistressor environment. Journal of applied physiology (Bethesda, Md. : 1985) 88, 1820-1830.

Friedl, K.E., Nuovo, J.A., Patience, T.H., Dettori, J.R., 1992. Factors associated with stress fracture in young army women: indications for further research. Military medicine 157, 334-338.

Friedl, K.E., Westphal, K.A., Marchitelli, L.J., Patton, J.F., Chumlea, W.C., Guo, S.S., 2001. Evaluation of anthropometric equations to assess body-composition changes in young women. The American journal of clinical nutrition 73, 268-275.

Frisch, R.E., Hall, G.M., Aoki, T.T., Birnholz, J., Jacob, R., Landsberg, L., Munro, H., Parker-Jones, K., Tulchinsky, D., Young, J., 1984. Metabolic, endocrine, and reproductive changes of a woman channel swimmer. Metabolism: clinical and experimental 33, 1106-1111.

Frisch, R.E., McArthur, J.W., 1974. Menstrual cycles: fatness as a determinant of minimum weight for height necessary for their maintenance or onset. Science 185, 949-951.

Frye, A.J., Kamon, E., 1981. Responses to dry heat of men and women with similar aerobic capacities. Journal of Applied Physiology: Respiratory, Environmental & Exercise Physiology 50, 65-70.

Gaffney-Stomberg, E., Nakayama, A.T., Guerriere, K.I., Lutz, L.J., Walker, L.A., Staab, J.S., Scott, J.M., Gasier, H.G., McClung, J.P., 2019. Calcium and vitamin D supplementation and bone health in Marine recruits: Effect of season. Bone 123, 224-233.

Gagnon, D., Crandall, C.G., Kenny, G.P., 2013. Sex differences in postsynaptic sweating and cutaneous vasodilation. Journal of applied physiology (Bethesda, Md. : 1985) 114, 394-401.

Gagnon, D., Kenny, G.P., 2011. Sex modulates whole-body sudomotor thermosensitivity during exercise. The Journal of physiology 589, 6205-6217.

Gagnon, D., Kenny, G.P., 2012a. Does sex have an independent effect on thermoeffector responses during exercise in the heat? The Journal of physiology 590, 5963-5973.

Gagnon, D., Kenny, G.P., 2012b. Sex differences in thermoeffector responses during exercise at fixed requirements for heat loss. Journal of applied physiology (Bethesda, Md. : 1985) 113, 746-757.

Garner, D.M., Garfinkel, P.E., 1979. The Eating Attitudes Test: an index of the symptoms of anorexia nervosa. Psychological medicine 9, 273-279.

Garrett, J.R., 1999. Effects of autonomic nerve stimulations on salivary parenchyma and protein secretion. Neural mechanisms of salivary gland secretion 11, 59-79.

Garthe, I., Raastad, T., Sundgot-Borgen, J., 2011. Long-term effect of nutritional counselling on desired gain in body mass and lean body mass in elite athletes. Applied Physiology, Nutrition, and Metabolism 36, 547-554.

Gasparrini, A., Guo, Y., Sera, F., Vicedo-Cabrera, A.M., Huber, V., Tong, S., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P.H.N., Lavigne, E., Correa, P.M., Ortega, N.V., Kan, H., Osorio, S., Kyselý, J.,

Urban, A., Jaakkola, J.J.K., Ryti, N.R.I., Pascal, M., Goodman, P.G., Zeka, A., Michelozzi, P., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Diaz, M., Cruz, J.C., Seposo, X., Kim, H., Tobias, A., Iñiguez, C., Forsberg, B., Åström, D.O., Ragettli, M.S., Guo, Y.L., Wu, C.-f., Zanobetti, A., Schwartz, J., Bell, M.L., Dang, T.N., Van, D.D., Heaviside, C., Vardoulakis, S., Hajat, S., Haines, A., Armstrong, B., 2017. Projections of temperature-related excess mortality under climate change scenarios. The Lancet. Planetary health 1, e360-e367.

Gay, J.L., Buchner, D.M., Smith, J., He, C., 2017. An examination of compensation effects in accelerometer-measured occupational and non-occupational physical activity. Preventive medicine reports 8, 55-59.

Gemming, L., Utter, J., Ni Mhurchu, C., 2015. Image-assisted dietary assessment: a systematic review of the evidence. Journal of the Academy of Nutrition and Dietetics 115, 64-77.

George, J.T., Veldhuis, J.D., Roseweir, A.K., Newton, C.L., Faccenda, E., Millar, R.P., Anderson, R.A., 2011. Kisspeptin-10 is a potent stimulator of LH and increases pulse frequency in men. The Journal of clinical endocrinology and metabolism 96, E1228-1236.

Gerber, M., Brand, S., Lindwall, M., Elliot, C., Kalak, N., Herrmann, C., Pühse, U., Jonsdottir, I.H., 2012. Concerns regarding hair cortisol as a biomarker of chronic stress in exercise and sport science. Journal of sports science & medicine 11, 571.

Gibbs, J.C., Nattiv, A., Barrack, M.T., Williams, N.I., Rauh, M.J., Nichols, J.F., De Souza, M.J., 2014. Low bone density risk is higher in exercising women with multiple triad risk factors. Medicine and science in sports and exercise 46, 167-176.

Gibbs, J.C., Williams, N.I., De Souza, M.J., 2013. Prevalence of individual and combined components of the female athlete triad. Medicine and science in sports and exercise 45, 985-996.

Gifford, R.M., Boos, C.J., Reynolds, R.M., Woods, D.R., 2018a. Recovery time and heart rate variability following extreme endurance exercise in healthy women. Physiological reports 6, e13905-e13905.

Gifford, R.M., Howie, F., Wilson, K., Johnston, N., Todisco, T., Crane, M., Greeves, J.P., Skorupskaite, K., Woods, D.R., Reynolds, R.M., Anderson, R.A., 2018b. Confirmation of ovulation from urinary progesterone analysis: assessment of two automated assay platforms. Scientific reports 8, 17621.

Gifford, R.M., O'Leary, T., Cobb, R., Blackadder-Weinstein, J., Double, R., Wardle, S.L., Anderson, R.A., Thake, C.D., Hattersley, J., Imray, C.H.E., Wilson, A., Greeves, J.P., Reynolds, R.M., Woods, D.R., 2019a. Female Reproductive, Adrenal, and Metabolic Changes during an Antarctic Traverse. Medicine and science in sports and exercise 51, 556-567.

Gifford, R.M., O'Leary, T.J., Double, R.L., Wardle, S.L., Wilson, K., Boyle, L.D., Homer, N.Z.M., Kirschbaum, C., Greeves, J.P., Woods, D.R., Reynolds, R.M., 2019b. Positive adaptation of HPA axis function in women during 44 weeks of infantry-based military training. Psychoneuroendocrinology 110, 104432.

Gifford, R.M., Reynolds, R.M., 2017. Sex differences in early-life programming of the hypothalamic-pituitary-adrenal axis in humans. Early human development 114, 7-10.

Gifford, R.M., Reynolds, R.M., Greeves, J., Anderson, R.A., Woods, D.R., 2017. Reproductive dysfunction and associated pathology in women undergoing military training. Journal of the Royal Army Medical Corps 163, 301-310.

Gifford, R.M., Todisco, T., Stacey, M., Fujisawa, T., Allerhand, M., Woods, D.R., Reynolds, R.M., 2019c. Risk of heat illness in men and women: A systematic review and meta-analysis. Environ Res 171, 24-35.

Gold, M.A., Friedman, S.B., 2000. Cadet basic training: an ethnographic study of stress and coping. Military medicine 165, 147-152.

Goldstein, C.A., Smith, Y.R., 2016. Sleep, Circadian Rhythms, and Fertility. Current Sleep Medicine Reports 2, 206-217.

Golier, J.A., Caramanica, K., Makotkine, I., Sher, L., Yehuda, R., 2014. Cortisol response to cosyntropin administration in military veterans with or without posttraumatic stress disorder. Psychoneuroendocrinology 40, 151-158.

Goodpaster, B.H., Kelley, D.E., Wing, R.R., Meier, A., Thaete, F.L., 1999. Effects of weight loss on regional fat distribution and insulin sensitivity in obesity. Diabetes 48, 839-847.

Gordley, L.B., Lemasters, G., Simpson, S.R., Yiin, J.H., 2000. Menstrual disorders and occupational, stress, and racial factors among military personnel. Journal of Occupational and Environmental Medicine 42, 871-881.

Gordon, C.M., Ackerman, K.E., Berga, S.L., Kaplan, J.R., Mastorakos, G., Misra, M., Murad, M.H., Santoro, N.F., Warren, M.P., 2017. Functional Hypothalamic Amenorrhea: An Endocrine Society Clinical Practice Guideline. The Journal of clinical endocrinology and metabolism 102, 1413-1439.

Graham, T., 1988. Thermal, metabolic, and cardiovascular changes in men and women during cold stress. Medicine and science in sports and exercise 20, S185-192.

Grandner, M.A., Seixas, A., Shetty, S., Shenoy, S., 2016. Sleep duration and diabetes risk: population trends and potential mechanisms. Current diabetes reports 16, 106.

Green, K.T., Beckham, J.C., Youssef, N., Elbogen, E.B., 2014. Alcohol Misuse and Psychological Resilience among U.S. Iraq and Afghanistan Era Veteran Military Personnel. Addictive behaviors 39, 406-413.

Greeves, J.P., 2015. Physiological Implications, Performance Assessment and Risk Mitigation Strategies of Women in Combat-Centric Occupations. Journal of strength and conditioning research / National Strength & Conditioning Association 29 Suppl 11, S94-100.

Gu, S., Huang, C., Bai, L., Chu, C., Liu, Q., 2016. Heat-related illness in China, summer of 2013. International Journal of Biometeorology 60, 131-137.

Gubernot, D.M., Anderson, G.B., Hunting, K.L., 2014. The epidemiology of occupational heat exposure in the United States: a review of the literature and assessment of research needs in a changing climate. Int J Biometeorol 58, 1779-1788.

Gubernot, D.M., Anderson, G.B., Hunting, K.L., 2015. Characterizing occupational heat-related mortality in the United States, 2000-2010: an analysis using the Census of Fatal Occupational Injuries database. American Journal of Industrial Medicine 58, 203-211.

Guermandi, E., Vegetti, W., Bianchi, M.M., Uglietti, A., Ragni, G., Crosignani, P., 2001. Reliability of ovulation tests in infertile women. Obstetrics and gynecology 97, 92-96.

Guyon, A., Balbo, M., Morselli, L.L., Tasali, E., Leproult, R., L'Hermite-Balériaux, M., Van Cauter, E., Spiegel, K., 2014. Adverse Effects of Two Nights of Sleep Restriction on the Hypothalamic-Pituitary-Adrenal Axis in Healthy Men. The Journal of Clinical Endocrinology & Metabolism 99, 2861-2868.

Hackney, A., 2008. Effects of endurance exercise on the reproductive system of men: the “exercise-hypogonadal male condition”. Journal of endocrinological investigation 31, 932-938.

Hagmar, M., Berglund, B., Brismar, K., Hirschberg, A.L., 2009. Hyperandrogenism may explain reproductive dysfunction in olympic athletes. Medicine and science in sports and exercise 41, 1241-1248.

Hakimi, O., Cameron, L.C., 2017. Effect of Exercise on Ovulation: A Systematic Review. Sports medicine (Auckland, N.Z.) 47, 1555-1567.

Hale, R.W., Kosasa, T., Krieger, J., Pepper, S., 1983. A marathon: the immediate effect on female runners' luteinizing hormone, follicle-stimulating hormone, prolactin, testosterone, and cortisol levels. Am J Obstet Gynecol 146, 550-556.

Hallal, P.C., Andersen, L.B., Bull, F.C., Guthold, R., Haskell, W., Ekelund, U., 2012. Global physical activity levels: surveillance progress, pitfalls, and prospects. The Lancet 380, 247-257.

Halsey, L.G., Stroud, M.A., 2011. Could Scott have survived with today's physiological knowledge? Current Biology 21, R457-R461.

Halsey, L.G., Stroud, M.A., 2012. 100 years since Scott reached the pole: a century of learning about the physiological demands of Antarctica. Physiological reviews 92, 521-536.

Hamilton, C., 2000. South Pole 2000. London: Virgin Books.

Hardoff, D., Halevy, A., 2006. Health perspectives regarding adolescents in military service. Curr Opin Pediatr 18, 371-375.

Harduar Morano, L., Bunn, T.L., Lackovic, M., Lavender, A., Dang, G.T., Chalmers, J.J., Li, Y., Zhang, L., Flammia, D.D., 2015. Occupational heat-related illness emergency department visits and inpatient hospitalizations in the southeast region, 2007-2011. American Journal of Industrial Medicine 58, 1114-1125.

Harduar Morano, L., Watkins, S., Kintziger, K., 2016. A Comprehensive Evaluation of the Burden of Heat-Related Illness and Death within the Florida Population. Int J Environ Res Public Health 13.

Hattersley, J., Wilson, A., Gifford, R.M., Facer Childs, J., Stoten, O., Cobb, R., Thake, C.D., Reynolds, R.M., Woods, D.R., Imray, C., 2019a. A comparison of the metabolic effects of sustained strenuous activity in polar environments on men and women. In Submission.

Hattersley, J., Wilson, A.J., Gifford, R.M., Cobb, R., Thake, C.D., Reynolds, R.M., Woods, D.R., Imray, C.H.E., 2019b. Pre- to postexpedition changes in the energy usage of women undertaking sustained expeditionary polar travel. Journal of Applied Physiology 126, 681-690.

Hayashi, T., Yamada, K., Esaki, T., Kuzuya, M., Satake, S., Ishikawa, T., Hidaka, H., Iguchi, A., 1995. Estrogen increases endothelial nitric oxide by a receptor-mediated system. Biochem Biophys Res Commun 214, 847-855.

Headquarters of the Surgeon General, 2017. Joint Service Publication 539 version 3.0. Climatic illness and injury in the Armed Forces: Force Protection and initial medical treatment. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793094/JSP_539_Part_2_V3.1__Updated_04-19_.pdf (Accessed 15 December 2019).

Heil, K., Thomas, R., Robertson, G., Porter, A., Milner, R., Wood, A., 2016. Freezing and non-freezing cold weather injuries: a systematic review. British Medical Bulletin 117, 79-93.

Hellhammer, D.H., Wüst, S., Kudielka, B.M., 2009. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology 34, 163-171.

Henning, P.C., Scofield, D.E., Spiering, B.A., Staab, J.S., Matheny, R.W., Jr., Smith, M.A., Bhasin, S., Nindl, B.C., 2014. Recovery of endocrine and inflammatory mediators following an extended energy deficit. The Journal of clinical endocrinology and metabolism 99, 956-964.

Herbst, J., Mason, K., Byard, R.W., Gilbert, J.D., Charlwood, C., Heath, K.J., Winskog, C., Langlois, N.E., 2014. Heat-related deaths in Adelaide, South Australia: review of the literature and case findings - an Australian perspective. Journal of Forensic & Legal Medicine 22, 73-78.

Hertel, J., Konig, J., Homuth, G., Van der Auwera, S., Wittfeld, K., Pietzner, M., Kacprowski, T., Pfeiffer, L., Kretschmer, A., Waldenberger, M., Kastenmuller, G., Artati, A., Suhre, K., Adamski, J., Langner, S., Volker, U., Volzke, H., Nauck, M., Friedrich, N., Grabe, H.J., 2017. Evidence for Stress-like Alterations in the HPA-Axis in Women Taking Oral Contraceptives. Scientific reports 7, 14111.

Hess, J.J., Saha, S., Luber, G., 2014. Summertime acute heat illness in U.S. emergency departments from 2006 through 2010: analysis of a nationally representative sample.[Erratum appears in Environ Health Perspect. 2014 Nov;122(11):A293]. Environmental Health Perspectives 122, 1209-1215.

Hill, N.E., Deighton, K., Matu, J., Misra, S., Oliver, N.S., Newman, C., Mellor, A., O'Hara, J., Woods, D., 2018. Continuous Glucose Monitoring at High Altitude-Effects on Glucose Homeostasis. Medicine and science in sports and exercise 50, 1679-1686.

Hill, N.E., Woods, D.R., Delves, S.K., Murphy, K.G., Davison, A.S., Brett, S.J., Quinton, R., Turner, S., Stacey, M., Allsopp, A.J., Fallowfield, J.L., 2015. The gonadotrophic response of Royal Marines during an operational deployment in Afghanistan. Andrology 3, 293-297.

Hill, R.J., Davies, P.S., 2001. The validity of self-reported energy intake as determined using the doubly labelled water technique. The British journal of nutrition 85, 415-430.

Hoch, A.Z., Pajewski, N.M., Moraski, L., Carrera, G.F., Wilson, C.R., Hoffmann, R.G., Schimke, J.E., Gutterman, D.D., 2009. Prevalence of the female athlete triad in high school athletes and sedentary students. Clinical journal of sport medicine: official journal of the Canadian Academy of Sport Medicine 19, 421.

Hoch, A.Z., Papanek, P., Szabo, A., Widlansky, M.E., Schimke, J.E., Gutterman, D.D., 2011. Association between the female athlete triad and endothelial dysfunction in dancers. Clinical journal of sport medicine: official journal of the Canadian Academy of Sport Medicine 21, 119.

Hodge, D., Jones, D., Martinez, R., Buono, M.J., 2013. Time course of the attenuation of sympathetic nervous activity during active heat acclimation. Autonomic Neuroscience 177, 101-103.

Hofmann, T., Elbelt, U., Haas, V., Ahnis, A., Klapp, B.F., Rose, M., Stengel, A., 2017. Plasma kisspeptin and ghrelin levels are independently correlated with physical activity in patients with anorexia nervosa. Appetite 108, 141-150.

Hopkins, M., Finlayson, G., Duarte, C., Whybrow, S., Ritz, P., Horgan, G.W., Blundell, J.E., Stubbs, R.J., 2016. Modelling the associations between fat-free mass, resting metabolic rate and energy intake in the context of total energy balance. International journal of obesity (2005) 40, 312-318.

Horn, J., 2013. Shackleton's Ad - Men Wanted for Hazardous Journey. Available at: http://discerninghistory.com/2013/05/shackletons-ad-men-wanted-for-hazerdous-journey/ (Accessed: 19 December 2019).

Houghton, B.L., Holowatz, L.A., Minson, C.T., 2005. Influence of progestin bioactivity on cutaneous vascular responses to passive heating. Medicine and science in sports and exercise 37, 45-51; discussion 52.

Hoyt, R.W., Friedl, K.E., 2006. Field studies of exercise and food deprivation. Current Opinion in Clinical Nutrition & Metabolic Care 9, 685-690.

Hoyt, R.W., Opstad, P.K., Haugen, A.-H., DeLany, J.P., Cymerman, A., Friedl, K.E., 2006. Negative energy balance in male and female rangers: effects of 7 d of sustained exercise and food deprivation. The American journal of clinical nutrition 83, 1068-1075.

Huang, W., Kan, H., Kovats, S., 2010. The impact of the 2003 heat wave on mortality in Shanghai, China. Science of the Total Environment 408, 2418-2420.

Huffman, E.A., Yard, E.E., Fields, S.K., Collins, C.L., Comstock, R.D., 2008. Epidemiology of rare injuries and conditions among United States high school athletes during the 2005-2006 and 2006-2007 school years. Journal of athletic training 43, 624-630.

Hughes, J.M., Foulis, S.A., Taylor, K.M., Guerriere, K.I., Walker, L.A., Hand, A.F., Popp, K.L., Gaffney-Stomberg, E., Heaton, K.J., Sharp, M.A., Grier, T.L., Hauret, K.G., Jones, B.H., Bouxsein, M.L., McClung, J.P., Matheny, R.W., Proctor, S.P., 2019. A prospective field study of U.S. Army trainees to identify the physiological bases and key factors influencing musculoskeletal injuries: a study protocol. BMC musculoskeletal disorders 20, 282.

Hughes, J.M., Smith, M.A., Henning, P.C., Scofield, D.E., Spiering, B.A., Staab, J.S., Hydren, J.R., Nindl, B.C., Matheny, R.W., Jr., 2014. Bone formation is suppressed with multi-stressor military training. European journal of applied physiology 114, 2251-2259.

Hughson, R.L., Green, H.J., Houston, M.E., Thomson, J.A., MacLean, D.R., Sutton, J.R., 1980. Heat injuries in Canadian mass participation runs. Canadian Medical Association Journal 122, 1141-1144.

Hulmi, J.J., Isola, V., Suonpää, M., Järvinen, N.J., Kokkonen, M., Wennerström, A., Nyman, K., Perola, M., Ahtiainen, J.P., Häkkinen, K., 2017. The effects of intensive weight reduction on body composition and serum hormones in female fitness competitors. Frontiers in physiology 7, 689.

Hwang, H.M., Kim, M.R., 2015. The relationship of quality of sleep, fatigue and job stress of airlines international flight attendants in full service carriers. Journal of the Korea Academia-Industrial cooperation Society 16, 7013-7020.

Ichinose-Kuwahara, T., Inoue, Y., Iseki, Y., Hara, S., Ogura, Y., Kondo, N., 2010. Sex differences in the effects of physical training on sweat gland responses during a graded exercise. Exp Physiol 95, 1026-1032.

Ihle, R., Loucks, A.B., 2004. Dose-response relationships between energy availability and bone turnover in young exercising women. Journal of Bone and Mineral Research 19, 1231-1240.

Informed-Sport, 2018. Certified Products. Available at: https://www.informed-sport.com/certified-product-brands (Accessed 19 December 2019).

Iob, E., Kirschbaum, C., Steptoe, A., 2018. Positive and negative social support and HPA-axis hyperactivity: Evidence from glucocorticoids in human hair. Psychoneuroendocrinology 96, 100-108.

Iobst, S., 2019. The Social Ecological Model As A Framework For Scoping Reviews Of Active Duty U.S. Military Women's Health, Medical Health System Resaerch Symposium, Orlando, FL, USA.

Israel, S.L., Schneller, O., 1950. The Thermogenic Property of Progesterone. Fertility & Sterility 1, 53-65.

Ivarsson, A., Johnson, U., Andersen, M.B., Tranaeus, U., Stenling, A., Lindwall, M., 2017. Psychosocial factors and sport injuries: meta-analyses for prediction and prevention. Sports medicine 47, 353-365.

Iwen, K.A., Wenzel, E.T., Ott, V., Perwitz, N., Wellhöner, P., Lehnert, H., Dodt, C., Klein, J., 2011. Cold-induced alteration of adipokine profile in humans. Metabolism: clinical and experimental 60, 430-437.

Iyoho, A.E., Ng, L.J., MacFadden, L., 2017. Modeling of Gender Differences in Thermoregulation. Military medicine 182, 295-303.

Jackowski, S.A., Baxter-Jones, A.D.G., McLardy, A.J., Pierson, R.A., Rodgers, C.D., 2016. The associations of exposure to combined hormonal contraceptive use on bone mineral content and areal bone mineral density accrual from adolescence to young adulthood: A longitudinal study. Bone reports 5, e333-e341.

Jain, A., Polotsky, A.J., Rochester, D., Berga, S.L., Loucks, T., Zeitlian, G., Gibbs, K., Polotsky, H.N., Feng, S., Isaac, B., Santoro, N., 2007. Pulsatile luteinizing hormone amplitude and progesterone metabolite

excretion are reduced in obese women. The Journal of clinical endocrinology and metabolism 92, 2468-2473.

James, W.H., 1998. The Health and Hormone Status of Female Cadets and Active-Duty Women in the US Armed Forces. Archives of pediatrics & adolescent medicine 152, 101.

Jardine, C., 2009. The Big Chill: the inside story of James Cracknell and Ben Fogle's Antarctic epic. The Daily Telegraph, London, UK. Available at: https://www.telegraph.co.uk/sport/othersports/south-pole-race/4435059/The-Big-Chill-the-inside-story-of-James-Cracknell-and-Ben-Fogles-Antarctic-epic.html (Accessed: 19 December 2019).

Javed, A., Kashyap, R., Lteif, A.N., 2015. Hyperandrogenism in female athletes with functional hypothalamic amenorrhea: a distinct phenotype. International journal of women's health 7, 103.

Jayasena, C.N., Abbara, A., Veldhuis, J.D., Comninos, A.N., Ratnasabapathy, R., De Silva, A., Nijher, G.M., Ganiyu-Dada, Z., Mehta, A., Todd, C., Ghatei, M.A., Bloom, S.R., Dhillo, W.S., 2014. Increasing LH pulsatility in women with hypothalamic amenorrhoea using intravenous infusion of Kisspeptin-54. The Journal of clinical endocrinology and metabolism 99, E953-961.

Jette, M., Sidney, K., Blumchen, G., 1990. Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. Clin Cardiol 13, 555-565.

Johnson, C.Y., Tanz, L.J., Lawson, C.C., Howards, P.P., Bertone-Johnson, E.R., Eliassen, A.H., Schernhammer, E.S., Rich-Edwards, J.W., 2019. Anti-Müllerian hormone levels in nurses working night shifts. Archives of Environmental & Occupational Health, 1-8.

Johnson, D.C., Polusny, M.A., Erbes, C.R., King, D., King, L., Litz, B.T., Schnurr, P.P., Friedman, M., Pietrzak, R.H., Southwick, S.M., 2011. Development and initial validation of the Response to Stressful Experiences Scale. Military medicine 176, 161-169.

Johnson, J.M., Kellogg, D.L., Jr., 2010. Local thermal control of the human cutaneous circulation. Journal of applied physiology (Bethesda, Md. : 1985) 109, 1229-1238.

Johnston, L.W., Harris, S.B., Retnakaran, R., Giacca, A., Liu, Z., Bazinet, R.P., Hanley, A.J., 2018. Association of NEFA composition with insulin sensitivity and beta cell function in the Prospective Metabolism and Islet Cell Evaluation (PROMISE) cohort. Diabetologia 61, 821-830.

Joiner, T.E., Schmidt, N.B., Lerew, D.R., Cook, J.H., Gencoz, T., Gencoz, F., 2000. Differential roles of depressive and anxious symptoms and gender in defensiveness. Journal of personality assessment 75, 200-211.

Kahn, P.M., Leon, G.R., 1994. Group climate and individual functioning in an all-women Antarctic expedition team. Environment and Behavior 26, 669-697.

Kajantie, E., Eriksson, J., Barker, D.J., Forsen, T., Osmond, C., Wood, P.J., Andersson, S., Dunkel, L., Phillips, D.I., 2003. Birthsize, gestational age and adrenal function in adult life: studies of dexamethasone suppression and ACTH1-24 stimulation. Eur J Endocrinol 149, 569-575.

Kamel, R.M., 2010. Management of the infertile couple: an evidence-based protocol. Reprod Biol Endocrinol 8, 21.

Karl, J.P., Smith, T.J., Wilson, M.A., Bukhari, A.S., Pasiakos, S.M., McClung, H.L., McClung, J.P., Lieberman, H.R., 2016. Altered metabolic homeostasis is associated with appetite regulation during and following 48-h of severe energy deprivation in adults. Metabolism: clinical and experimental 65, 416-427.

Kazman, J.B., Purvis, D.L., Heled, Y., Lisman, P., Atias, D., Van Arsdale, S., Deuster, P.A., 2015. Women and exertional heat illness: identification of gender specific risk factors. US Army Medical Department Journal, 58-66.

Keay, N., Rankin, A., 2019. Infographic. Relative energy deficiency in sport: an infographic guide. British journal of sports medicine, bjsports-2018-100354.

Kenny, G.P., Jay, O., 2013. Thermometry, calorimetry, and mean body temperature during heat stress. Compr Physiol 3, 1689-1719.

Kenny, G.P., Sigal, R.J., McGinn, R., 2016. Body temperature regulation in diabetes. Temperature (Austin, Tex.) 3, 119-145.

Kenttamies, A., Karkola, K., 2008. Death in sauna. Journal of Forensic Sciences 53, 724-729.

Kerksick, C.M., Arent, S., Schoenfeld, B.J., Stout, J.R., Campbell, B., Wilborn, C.D., Taylor, L., Kalman, D., Smith-Ryan, A.E., Kreider, R.B., Willoughby, D., Arciero, P.J., VanDusseldorp, T.A., Ormsbee, M.J., Wildman, R., Greenwood, M., Ziegenfuss, T.N., Aragon, A.A., Antonio, J., 2017. International society of sports nutrition position stand: nutrient timing. Journal of the International Society of Sports Nutrition 14, 33.

Kerksick, C.M., Wilborn, C.D., Roberts, M.D., Smith-Ryan, A., Kleiner, S.M., Jäger, R., Collins, R., Cooke, M., Davis, J.N., Galvan, E., 2018. ISSN exercise & sports nutrition review update: research & recommendations. Journal of the International Society of Sports Nutrition 15, 38.

Khare, S., Hajat, S., Kovats, S., Lefevre, C.E., de Bruin, W.B., Dessai, S., Bone, A., 2015. Heat protection behaviour in the UK: results of an online survey after the 2013 heatwave. BMC public health 15, 878.

Kim, H.G., Cheon, E.J., Bai, D.S., Lee, Y.H., Koo, B.H., 2018. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry investigation 15, 235-245.

King, J.C., Brown, K.H., Gibson, R.S., Krebs, N.F., Lowe, N.M., Siekmann, J.H., Raiten, D.J., 2016. Biomarkers of Nutrition for Development (BOND)-Zinc Review. The Journal of nutrition.

Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28, 76-81.

Kirschbaum, C., Tietze, A., Skoluda, N., Dettenborn, L., 2009. Hair as a retrospective calendar of cortisol production—increased cortisol incorporation into hair in the third trimester of pregnancy. Psychoneuroendocrinology 34, 32-37.

Kirschbaum, C., Wust, S., Hellhammer, D., 1992. Consistent sex differences in cortisol responses to psychological stress. Psychosom Med 54, 648-657.

Kiviniemi, A.M., Tulppo, M.P., Hautala, A.J., Vanninen, E., Uusitalo, A.L., 2014. Altered relationship between R-R interval and R-R interval variability in endurance athletes with overtraining syndrome. Scandinavian journal of medicine & science in sports 24, e77-85.

Klingman, K.M., Marsh, E.E., Klerman, E.B., Anderson, E.J., Hall, J.E., 2011. Absence of circadian rhythms of gonadotropin secretion in women. The Journal of clinical endocrinology and metabolism 96, 1456-1461.

Kloss, J.D., Perlis, M.L., Zamzow, J.A., Culnan, E.J., Gracia, C.R., 2015. Sleep, sleep disturbance, and fertility in women. Sleep Medicine Reviews 22, 78-87.

Knapik, J., Montain, S.J., McGraw, S., Grier, T., Ely, M., Jones, B.H., 2012. Stress fracture risk factors in basic combat training. Int J Sports Med 33, 940-946.

Knapik, J.J., Sharp, M.A., Montain, S.J., 2018. Association between stress fracture incidence and predicted body fat in United States Army Basic Combat Training recruits. BMC musculoskeletal disorders 19, 161.

Knapp, J., Aerni, G., Anderson, J., 2014. Eating disorders in female athletes: use of screening tools. Current sports medicine reports 13, 214-218.

Knauff, E.A., Eijkemans, M.J., Lambalk, C.B., ten Kate-Booij, M.J., Hoek, A., Beerendonk, C.C., Laven, J.S., Goverde, A.J., Broekmans, F.J., Themmen, A.P., de Jong, F.H., Fauser, B.C., 2009. Anti-Mullerian hormone, inhibin B, and antral follicle count in young women with ovarian failure. The Journal of clinical endocrinology and metabolism 94, 786-792.

Kocalevent, R.D., Hinz, A., Brahler, E., 2013. Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. General hospital psychiatry 35, 551-555.

Koehler, K., Williams, N.I., Mallinson, R.J., Southmayd, E.A., Allaway, H.C., De Souza, M.J., 2016. Low resting metabolic rate in exercise-associated amenorrhea is not due to a reduced proportion of highly active metabolic tissue compartments. American Journal of Physiology-Endocrinology and Metabolism 311, E480-E487.

Korbi, M., Youssef, M., Ben Brahim, H., Chaabane, A., Mohamed, M., Akkari, H., Belhadjali, H., Zili, J., 2015. [Allopurinol-induced DRESS complicated by hemophagocytic lymphohistiocytosis]. Annales de Dermatologie et de Venereologie 142, 767-770.

Kowalski, W., Chatterton, R.T., Jr., Kazer, R.R., Wentz, A.C., 1993. The impact of subchronic hypercortisolemia on progesterone metabolism and the luteinizing hormone-progesterone axis in the cynomolgus monkey. The Journal of clinical endocrinology and metabolism 77, 1597-1604.

Krief, S., Lönnqvist, F., Raimbault, S., Baude, B., Van Spronsen, A., Arner, P., Strosberg, A.D., Ricquier, D., Emorine, L.J., 1993. Tissue distribution of beta 3-adrenergic receptor mRNA in man. Journal of Clinical Investigation 91, 344-349.

Kriegsfeld, L.J., Williams III, W.P., 2012. Circadian control of neuroendocrine circuits regulating female reproductive function. Frontiers in endocrinology 3, 60.

Kripke, D.F., Elliott, J.A., Youngstedt, S.D., Parry, B.L., Hauger, R.L., Rex, K.M., 2010. Weak evidence of bright light effects on human LH and FSH. Journal of circadian rhythms 8, 5.

Kroenke, K., Spitzer, R.L., Williams, J.B., 2001. The PHQ-9: validity of a brief depression severity measure. Journal of general internal medicine 16, 606-613.

Kudielka, B.M., Hellhammer, D., Wüst, S., 2009. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. Psychoneuroendocrinology 34, 2-18.

Kudielka, B.M., Kirschbaum, C., 2005. Sex differences in HPA axis responses to stress: a review. Biological psychology 69, pp. 113-132.

Kuehl, L.K., Hinkelmann, K., Muhtz, C., Dettenborn, L., Wingenfeld, K., Spitzer, C., Kirschbaum, C., Wiedemann, K., Otte, C., 2015. Hair cortisol and cortisol awakening response are associated with criteria of the metabolic syndrome in opposite directions. Psychoneuroendocrinology 51, 365-370.

Kuht, J.A., Woods, D., Hollis, S., 2018. Case series of non-freezing cold injury: epidemiology and risk factors. Journal of the Royal Army Medical Corps 165, 400-404.

Kumari, M., Shipley, M., Stafford, M., Kivimaki, M., 2011. Association of Diurnal Patterns in Salivary Cortisol with All-Cause and Cardiovascular Mortality: Findings from the Whitehall II Study. The Journal of Clinical Endocrinology & Metabolism 96, 1478-1485.

Kyriakidis, M., Caetano, L., Anastasiadou, N., Karasu, T., Lashen, H., 2016. Functional hypothalamic amenorrhoea: leptin treatment, dietary intervention and counselling as alternatives to traditional practice - systematic review. European journal of obstetrics, gynecology, and reproductive biology 198, 131-137.

Lagowska, K., Kapczuk, K., Friebe, Z., Bajerska, J., 2014. Effects of dietary intervention in young female athletes with menstrual disorders. Journal of the International Society of Sports Nutrition 11, 21.

Lauder, T.D., 1997. The Female Athlete Triad: Prevalence in Military Women, Defense Women's Research Program. Technical Report for U.S. Army Medical Research and Materiel Command. Available at: https://apps.dtic.mil/dtic/tr/fulltext/u2/a330021.pdf (Accessed: 19 December 2019).

Lauder, T.D., Williams, M.V., Campbell, C.S., Davis, G., Sherman, R., Pulos, E., 1999. The female athlete triad: prevalence in military women. Military medicine 164, 630-635.

Lawson, E.A., Donoho, D., Miller, K.K., Misra, M., Meenaghan, E., Lydecker, J., Wexler, T., Herzog, D.B., Klibanski, A., 2009. Hypercortisolemia is associated with severity of bone loss and depression in hypothalamic amenorrhea and anorexia nervosa. The Journal of clinical endocrinology and metabolism 94, 4710-4716.

LeBlanc, J., Ducharme, M.B., Pasto, L., Thompson, M., 2003. Response to thermal stress and personality. Physiology & behavior 80, 69-74.

Lebrun, C.M., Joyce S. M., Constantini, N. W., 2013. Effects of Female Reproductive Hormones on Sports Performance, in: Constantini, N.H., Hackney, A.C. (Eds.), Endocrinology of Physical Activity and Sport, 2nd ed. New York: Springer.

Lee, E.A.D., Theus, S.A., 2012. Lower heart rate variability associated with military sexual trauma rape and posttraumatic stress disorder. Biological research for nursing 14, 412-418.

Lehrner, A., Daskalakis, N., Yehuda, R., 2016. Cortisol and the hypothalamic–pituitary–adrenal axis in PTSD, in: Bremner, D.J. (Ed.), Posttraumatic stress disorder: From neurobiology to treatment. John WIley & Sons, Hoboken, NJ, USA, pp. 265-274.

Lehrner, A., Yehuda, R., 2014. Biomarkers of PTSD: military applications and considerations. European journal of psychotraumatology 5: 10.3402/ejpt.v5.23797. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4138702/ (Accessed: 19 December 2019).

Leiva, R., Bouchard, T., Boehringer, H., Abulla, S., Ecochard, R., 2015. Random serum progesterone threshold to confirm ovulation. Steroids 101, 125-129.

Leiva, R., McNamara-Kilian, M., Niezgoda, H., Ecochard, R., Bouchard, T., 2019. Pilot observational prospective cohort study on the use of a novel home-based urinary pregnanediol 3-glucuronide (PDG) test to confirm ovulation when used as adjunct to fertility awareness methods (FAMs) stage 1. BMJ Open 9, e028496.

Leon, L.R., Bouchama, A., 2015. Heat stroke. Compr Physiol 5, 611-647.

Leproult, R., Van Cauter, E., 2011. Effect of 1 week of sleep restriction on testosterone levels in young healthy men. Jama 305, 2173-2174.

Levy, J.C., Matthews, D.R., Hermans, M.P., 1998. Correct homeostasis model assessment (HOMA) evaluation uses the computer program. Diabetes Care 21, 2191-2192.

Lewis, G.F., Hourani, L., Tueller, S., Kizakevich, P., Bryant, S., Weimer, B., Strange, L., 2015. Relaxation training assisted by heart rate variability biofeedback: Implication for a military predeployment stress inoculation protocol. Psychophysiology 52, 1167-1174.

Lewis, J.G., Elder, P.A., 2011. Corticosteroid-binding globulin reactive centre loop antibodies recognise only the intact natured protein: elastase cleaved and uncleaved CBG may coexist in circulation. The Journal of steroid biochemistry and molecular biology 127, 289-294.

Lewis, J.G., Elder, P.A., 2013. Intact or "active" corticosteroid-binding globulin (CBG) and total CBG in plasma: determination by parallel ELISAs using monoclonal antibodies. Clinica chimica acta; international journal of clinical chemistry 416, 26-30.

Li, H.W., Anderson, R.A., Yeung, W.S., Ho, P.C., Ng, E.H., 2011. Evaluation of serum antimullerian hormone and inhibin B concentrations in the differential diagnosis of secondary oligoamenorrhea. Fertil Steril 96, 774-779.

Li, H.W., Wong, B.P., Ip, W.K., Yeung, W.S., Ho, P.C., Ng, E.H., 2016. Comparative evaluation of three new commercial immunoassays for anti-Mullerian hormone measurement. Human reproduction (Oxford, England) 31, 2796-2802.

Li, J., Xu, X., Ding, G., Zhao, Y., Zhao, R., Xue, F., Li, J., Gao, J., Yang, J., Jiang, B., Liu, Q., 2016. A Cross-Sectional Study of Heat Wave-Related Knowledge, Attitude, and Practice among the Public in the Licheng District of Jinan City, China. International journal of environmental research and public health 13 (7). Available at: https://pubmed.ncbi.nlm.nih.gov/27367715-a-cross-sectional-study-of-heat-wave-related-knowledge-attitude-and-practice-among-the-public-in-the-licheng-district-of-jinan-city-china/ (Accessed: 19 December 2019).

Lie Fong, S., Visser, J.A., Welt, C.K., de Rijke, Y.B., Eijkemans, M.J., Broekmans, F.J., Roes, E.M., Peters, W.H., Hokken-Koelega, A.C., Fauser, B.C., Themmen, A.P., de Jong, F.H., Schipper, I., Laven, J.S., 2012. Serum anti-mullerian hormone levels in healthy females: a nomogram ranging from infancy to adulthood. The Journal of clinical endocrinology and metabolism 97, 4650-4655.

Lieberman, H.R., Farina, E.K., Caldwell, J., Williams, K.W., Thompson, L.A., Niro, P.J., Grohmann, K.A., McClung, J.P., 2016a. Cognitive function, stress hormones, heart rate and nutritional status during simulated captivity in military survival training. Physiology & behavior 165, 86-97.

Lieberman, H.R., Karl, J.P., McClung, J.P., Williams, K.W., Cable, S., 2016b. Improved mood state and absence of sex differences in response to the stress of Army Basic Combat Training. Applied Psychology: Health and Well-Being 8, 351-363.

Lieberman, H.R., Kellogg, M.D., Bathalon, G.P., 2008. Female marine recruit training: mood, body composition, and biochemical changes. Medicine and science in sports and exercise 40, S671-676.

Lieberman, J.L., De Souza, M.J., Wagstaff, D.A., Williams, N.I., 2018. Menstrual Disruption with Exercise Is Not Linked to an Energy Availability Threshold. Medicine and science in sports and exercise 50, 551-561.

Lindqvist, P.G., Marsal, K., Merlo, J., Pirhonen, J.P., 2003. Thermal response to submaximal exercise before, during and after pregnancy: a longitudinal study. Journal of Maternal-Fetal & Neonatal Medicine 13, 152-156.

Lindsay, T.J., Vitrikas, K.R., 2015. Evaluation and treatment of infertility. Am Fam Physician 91, 308-314.

Liu, T., Xu, Y.J., Zhang, Y.H., Yan, Q.H., Song, X.L., Xie, H.Y., Luo, Y., Rutherford, S., Chu, C., Lin, H.L., Ma, W.J., 2013. Associations between risk perception, spontaneous adaptation behavior to heat waves and heatstroke in Guangdong province, China. BMC public health 13, 913.

Loewen, N., Bancroft, A., 2001. Four to the Pole!: The American Women's Expedition to Antarctica, 1992-1993. Hamden, CT: Linnet Books.

Logue, D., Madigan, S.M., Delahunt, E., Heinen, M., Mc Donnell, S.-J., Corish, C.A., 2018. Low energy availability in athletes: a review of prevalence, dietary patterns, physiological health, and sports performance. Sports Medicine 48, 73-96.

Lopez, R.M., Casa, D.J., 2009. The influence of nutritional ergogenic aids on exercise heat tolerance and hydration status. Current sports medicine reports 8, 192-199.

Loucks, A.B., 2004. Energy balance and body composition in sports and exercise. Journal of sports sciences 22, 1-14.

Loucks, A.B., 2006. The response of luteinizing hormone pulsatility to 5 days of low energy availability disappears by 14 years of gynecological age. The Journal of clinical endocrinology and metabolism 91, 3158-3164.

Loucks, A.B., 2009. Is stress measured in joules? Military Psychology 21, S101.

Loucks, A.B., 2013. Exercise Training in the Normal Female: Effects of Low Energy Availability on Reproductive Function, in: Constantini, N.H., Hackney, A.C. (Eds.), Endocrinology of Physical Activity and Sport, 2nd ed. New York: Springer,

Loucks, A.B., Horvath, S.M., 1984. Exercise-induced stress responses of amenorrheic and eumenorrheic runners. The Journal of clinical endocrinology and metabolism 59, 1109-1120.

Loucks, A.B., Horvath, S.M., Freedson, P.S., 1984. Menstrual status and validation of body fat prediction in athletes. Human biology 56, 383-392.

Loucks, A.B., Kiens, B., Wright, H.H., 2011. Energy availability in athletes. Journal of sports sciences 29 Suppl 1, S7-15.

Loucks, A.B., Laughlin, G.A., Mortola, J.F., Girton, L., Nelson, J.C., Yen, S.S., 1992. Hypothalamic-pituitary-thyroidal function in eumenorrheic and amenorrheic athletes. The Journal of clinical endocrinology and metabolism 75, 514-518.

Loucks, A.B., Mortola, J.F., Girton, L., Yen, S.S.C., 1989. Alterations in the Hypothalamic-Pituitary-Ovarian and the Hypothalamic-Pituitary-Adrenal Axes in Athletic Women. The Journal of clinical endocrinology and metabolism 68, 402-411.

Loucks, A.B., Redman, L.M., 2004. The effect of stress on menstrual function. Trends in endocrinology and metabolism 15, 466-471.

Loucks, A.B., Thuma, J.R., 2003. Luteinizing hormone pulsatility is disrupted at a threshold of energy availability in regularly menstruating women. The Journal of clinical endocrinology and metabolism 88, 297-311.

Loucks, A.B., Verdun, M., Heath, E.M., 1998. Low energy availability, not stress of exercise, alters LH pulsatility in exercising women. Journal of applied physiology (Bethesda, Md. : 1985) 84, 37-46.

Louis, G.M.B., Lum, K.J., Sundaram, R., Chen, Z., Kim, S., Lynch, C.D., Schisterman, E.F., Pyper, C., 2011. Stress reduces conception probabilities across the fertile window: evidence in support of relaxation. Fertility and sterility 95, 2184-2189.

Lundgren-Kownacki, K., Hornyanszky, E.D., Chu, T.A., Olsson, J.A., Becker, P., 2018. Challenges of using air conditioning in an increasingly hot climate. International Journal of Biometeorology 62, 401-412.

Lutz, L.J., Gaffney-Stomberg, E., Karl, J.P., Hughes, J.M., Guerriere, K.I., McClung, J.P., 2019. Dietary Intake in Relation to Military Dietary Reference Values During Army Basic Combat Training; a Multi-center, Cross-sectional Study. Military medicine 184, e223-e230.

Luxton, D.D., Greenburg, D., Ryan, J., Niven, A., Wheeler, G., Mysliwiec, V., 2011. Prevalence and Impact of Short Sleep Duration in Redeployed OIF Soldiers. Sleep 34, 1189-1195.

Lynch, K.E., Mumford, S.L., Schliep, K.C., Whitcomb, B.W., Zarek, S.M., Pollack, A.Z., Bertone-Johnson, E.R., Danaher, M., Wactawski-Wende, J., Gaskins, A.J., Schisterman, E.F., 2014. Assessment of anovulation in eumenorrheic women: comparison of ovulation detection algorithms. Fertility and Sterility 102, 511-518.e512.

MacDonald, L.A., Deddens, J.A., Grajewski, B.A., Whelan, E.A., Hurrell, J.J., 2003. Job stress among female flight attendants. Journal of Occupational and Environmental Medicine 45, 703-714.

Maddox, Y., 2019. Uniformed Services University Military Women's Health Consortium updates, Military Health System Research Symposium, Orlando, FL, USA.

Magkos, F., Yannakoulia, M., 2003. Methodology of dietary assessment in athletes: concepts and pitfalls. Current opinion in clinical nutrition and metabolic care 6, 539-549.

Maguen, S., Turcotte, D.M., Peterson, A.L., Dremsa, T.L., Garb, H.N., McNally, R.J., Litz, B.T., 2008. Description of risk and resilience factors among military medical personnel before deployment to Iraq. Military medicine 173, 1-9.

Maïmoun, L., Coste, O., Georgopoulos, N.A., Roupas, N.D., Mahadea, K.K., Tsouka, A., Mura, T., Philibert, P., Gaspari, L., Mariano-Goulart, D., Leglise, M., Sultan, C., 2013. Despite a high prevalence of menstrual disorders, bone health is improved at a weight-bearing bone site in world-class female rhythmic gymnasts. The Journal of clinical endocrinology and metabolism 98, 4961-4969.

Maïmoun, L., Georgopoulos, N.A., Sultan, C., 2014. Endocrine disorders in adolescent and young female athletes: impact on growth, menstrual cycles, and bone mass acquisition. The Journal of clinical endocrinology and metabolism 99, 4037-4050.

Makras, P., Koukoulis, G.N., Bourikas, G., Papatheodorou, G., Bedevis, K., Menounos, P., Pappas, D., Kartalis, G., 2005. Effect of 4 weeks of basic military training on peripheral blood leucocytes and urinary excretion of catecholamines and cortisol. Journal of sports sciences 23, 825-834.

Mallinson, R.J., De Souza, M.J., 2014. Current perspectives on the etiology and manifestation of the "silent" component of the Female Athlete Triad. International journal of women's health 6, 451-467.

Mancini, A., Vergani, E., Bruno, C., Barini, A., Silvestrini, A., Meucci, E., Messana, C., Romualdi, D., Apa, R., Lanzone, A., 2019. Relationships Between Thyroid Hormones, Insulin-Like Growth Factor-1 and Antioxidant Levels in Hypothalamic Amenorrhea and Impact on Bone Metabolism. Hormone and metabolic research 51, 302-308.

Manenschijn, L., Schaap, L., van Schoor, N.M., van der Pas, S., Peeters, G.M., Lips, P., Koper, J.W., van Rossum, E.F., 2013. High long-term cortisol levels, measured in scalp hair, are associated with a history of cardiovascular disease. The Journal of clinical endocrinology and metabolism 98, 2078-2083.

Marcora, S.M., Staiano, W., Manning, V., 2009. Mental fatigue impairs physical performance in humans. Journal of applied physiology (Bethesda, Md. : 1985) 106, 857-864.

Marcus, M.D., Loucks, T.L., Berga, S.L., 2001. Psychological correlates of functional hypothalamic amenorrhea. Fertil Steril 76, 310-316.

Marek, R.J., Coulon, S.M., Brown, J.D., Lydecker, J.A., Marek, S., Malcolm, R., O'Neil, P.M., 2017. Characteristics of Weight Loss Trajectories in a Comprehensive Lifestyle Intervention. Obesity 25, 2062-2067.

Margolis, L.M., Murphy, N.E., Martini, S., Gundersen, Y., Castellani, J.W., Karl, J.P., Carrigan, C.T., Teien, H.K., Madslien, E.H., Montain, S.J., Pasiakos, S.M., 2016. Effects of Supplemental Energy on Protein Balance during 4-d Arctic Military Training. Medicine and science in sports and exercise 48, 1604-1612.

Marmar, C.R., Weiss, D.S., Metzler, T.J., Ronfeldt, H.M., Foreman, C., 1996. Stress responses of emergency services personnel to the Loma Prieta earthquake Interstate 880 freeway collapse and control traumatic incidents. Journal of traumatic stress 9, 63-85.

Marsh, S.A., Jenkins, D.G., 2002. Physiological responses to the menstrual cycle: implications for the development of heat illness in female athletes. Sports medicine (Auckland, N.Z.) 32, 601-614.

Martin, A., Rief, W., Klaiberg, A., Braehler, E., 2006. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. General hospital psychiatry 28, 71-77.

Martin, C.K., Nicklas, T., Gunturk, B., Correa, J.B., Allen, H.R., Champagne, C., 2014. Measuring food intake with digital photography. Journal of human nutrition and dietetics : the official journal of the British Dietetic Association 27 Suppl 1, 72-81.

Martin, N.M., Conlon, C.A., Smeele, R.J.M., Mugridge, O.A.R., von Hurst, P.R., McClung, J.P., Beck, K.L., 2019. Iron status and associations with physical performance during basic combat training in female New Zealand Army recruits. The British journal of nutrition 121, 887-893.

Martinez Rodriguez, A., Vicente Salar, N., Montero Carretero, C., Cervello Gimeno, E., Roche Collado, E., 2015. Eating disorders and diet management in contact sports; EAT-26 questionnaire does not seem appropriate to evaluate eating disorders in sports. Nutricion hospitalaria 32, 1708-1714.

Martinsen, M., Holme, I., Pensgaard, A.M., Torstveit, M.K., Sundgot-Borgen, J., 2014. The development of the brief eating disorder in athletes questionnaire. Medicine and science in sports and exercise 46, 1666-1675.

Matzkin, E., Curry, E.J., Whitlock, K., 2015. Female Athlete Triad: Past, Present, and Future. The Journal of the American Academy of Orthopaedic Surgeons 23, 424-432.

Maughan, R.J., Burke, L.M., 2011. Practical nutritional recommendations for the athlete. Nestle Nutrition Institute workshop series 69, 131-149.

Mayorga-Vega, D., Bocanegra-Parrilla, R., Ornelas, M., Viciana, J., 2016. Criterion-Related Validity of the Distance- and Time-Based Walk/Run Field Tests for Estimating Cardiorespiratory Fitness: A Systematic Review and Meta-Analysis. PLoS One 11, e0151671.

McCall, L.M., Ackerman, K.E., 2019. Endocrine and Metabolic Repercussions of Relative Energy Deficiency in Sport (RED-S). Current Opinion in Endocrine and Metabolic Research 9, 56-65.

McClung, J.P., Gaffney-Stomberg, E., Lee, J.J., 2014. Female athletes: a population at risk of vitamin and mineral deficiencies affecting health and performance. Journal of trace elements in medicine and biology 28, 388-392.

McCraty, R., 2017. New frontiers in heart rate variability and social coherence research: techniques, technologies, and implications for improving group dynamics and outcomes. Frontiers in public health 5, 267.

McHill, A., Wright Jr, K., 2017. Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. Obesity Reviews 18, 15-24.

McLennan, S.N., Ihle, A., Steudte-Schmiedgen, S., Kirschbaum, C., Kliegel, M., 2016. Hair cortisol and cognitive performance in working age adults. Psychoneuroendocrinology 67, 100-103.

McNeil, J., Brenner, D.R., Courneya, K.S., Friedenreich, C.M., 2017. Dose-response effects of aerobic exercise on energy compensation in postmenopausal women: combined results from two randomized controlled trials. International journal of obesity (2005) 41, 1196-1202.

McNeilly, A.S., 2012. Diagnostic applications for inhibin and activins. Molecular and cellular endocrinology 359, 121-125.

Meeusen, R., Duclos, M., Foster, C., Fry, A., Gleeson, M., Nieman, D., Raglin, J., Rietjens, G., Steinacker, J., Urhausen, A., 2013. Prevention, diagnosis, and treatment of the overtraining syndrome: joint consensus statement of the European College of Sport Science and the American College of Sports Medicine. Medicine and science in sports and exercise 45, 186-205.

Melin, A., Tornberg, A.B., Skouby, S., Faber, J., Ritz, C., Sjodin, A., Sundgot-Borgen, J., 2014. The LEAF questionnaire: a screening tool for the identification of female athletes at risk for the female athlete triad. British journal of sports medicine 48, 540-545.

Mena, G.P., Mielke, G.I., Brown, W.J., 2019. The effect of physical activity on reproductive health outcomes in young women: a systematic review and meta-analysis. Human reproduction update.

Merklinger-Gruchala, A., Ellison, P.T., Lipson, S.F., Thune, I., Jasienska, G., 2008. Low estradiol levels in women of reproductive age having low sleep variation. European Journal of Cancer Prevention 17, 467-472.

Mesen, T.B., Young, S.L., 2015. Progesterone and the luteal phase: a requisite to reproduction. Obstetrics and gynecology clinics of North America 42, 135-151.

Michael, S., Jay, O., Halaki, M., Graham, K., Davis, G.M., 2016. Submaximal exercise intensity modulates acute post-exercise heart rate variability. European journal of applied physiology 116, 697-706.

Michelozzi, P., de Donato, F., Bisanti, L., Russo, A., Cadum, E., DeMaria, M., D'Ovidio, M., Costa, G., Perucci, C.A., 2005. The impact of the summer 2003 heat waves on mortality in four Italian cities. Euro Surveillance: Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin 10, 161-165.

Michopoulos, V., Mancini, F., Loucks, T.L., Berga, S.L., 2013. Neuroendocrine recovery initiated by cognitive behavioral therapy in women with functional hypothalamic amenorrhea: a randomized, controlled trial. Fertil Steril 99, 2084-2091.e2081.

Miller, R., Plessow, F., Rauh, M., Gröschl, M., Kirschbaum, C., 2013. Comparison of salivary cortisol as measured by different immunoassays and tandem mass spectrometry. Psychoneuroendocrinology 38, 50-57.

Miller, R.C., Brindle, E., Holman, D.J., Shofer, J., Klein, N.A., Soules, M.R., O'Connor, K.A., 2004. Comparison of specific gravity and creatinine for normalizing urinary reproductive hormone concentrations. Clinical chemistry 50, 924-932.

Ministry of Defence, 2016. News story: ban on women in ground close combat roles lifted. Available at: https://www.gov.uk/government/news/ban-on-women-in-ground-close-combat-roles-lifted (accessed 19 December 2019).

Ministry of Defence, 2018. Joint Service Publication 950, London, UK, Lft 6-7-7.

Mirabelli, M.C., Richardson, D.B., 2005. Heat-related fatalities in North Carolina. American Journal of Public Health 95, 635-637.

Misra, M., 2014. Neuroendocrine mechanisms in athletes, Handbook of clinical neurology. Elsevier, pp. 373-386.

Misra, M., Klibanski, A., 2014. Endocrine consequences of anorexia nervosa. The Lancet Diabetes & Endocrinology 2, 581-592.

Miyake, Y., 2012. [Characteristics of non-exertional heat-related illness in Japan]. Nippon Rinsho - Japanese Journal of Clinical Medicine 70, 997-1004.

Miyake, Y., 2013. [Characteristics of elderly heat illness patients in Japan--analysis from Heatstroke STUDY 2010]. Nihon rinsho. Japanese journal of clinical medicine 71, 1065-1073.

Moore, A.C., Stacey, M.J., Bailey, K.G.H., Bunn, R.J., Woods, D.R., Haworth, K.J., Brett, S.J., Folkes, S.E.F., 2016. Risk factors for heat illness among British soldiers in the hot Collective Training Environment. Journal of the Royal Army Medical Corps 162, 434-439.

Morales, J., Yanez, A., Fernandez-Gonzalez, L., Montesinos-Magraner, L., Marco-Ahullo, A., Solana-Tramunt, M., Calvete, E., 2019. Stress and autonomic response to sleep deprivation in medical residents: A comparative cross-sectional study. PLoS One 14, e0214858.

Moran, D., Evans, R., Arbel, Y., Luria, O., Hadid, A., Yanovich, R., Milgrom, C., Finestone, A., 2013. Physical and psychological stressors linked with stress fractures in recruit training. Scandinavian journal of medicine & science in sports 23, 443-450.

Moran, D.S., Shapiro, Y., Laor, A., Izraeli, S., Pandolf, K.B., 1999. Can gender differences during exercise-heat stress be assessed by the physiological strain index? American Journal of Physiology 276, R1798-1804.

Morgan, J.F., Reid, F., Lacey, J.H., 1999. The SCOFF questionnaire: assessment of a new screening tool for eating disorders. BMJ 319, 1467-1468.

Morosini, P.P., Sarzani, R., Arnaldi, G., Taccaliti, A., 1989. [Hypothalamic amenorrhea. Different patterns in the pulsatile secretion of LH during 24 hours and different responses to the stimulation test with GnRH]. Minerva endocrinologica 14, 153-158.

Mountjoy, M., Sundgot-Borgen, J., Burke, L., Carter, S., Constantini, N., Lebrun, C., Meyer, N., Sherman, R., Steffen, K., Budgett, R., Ljungqvist, A., 2014. The IOC consensus statement: beyond the Female Athlete Triad--Relative Energy Deficiency in Sport (RED-S). British journal of sports medicine 48, 491-497.

Mountjoy, M., Sundgot-Borgen, J.K., Burke, L.M., Ackerman, K.E., Blauwet, C., Constantini, N., Lebrun, C., Lundy, B., Melin, A.K., Meyer, N.L., Sherman, R.T., Tenforde, A.S., Klungland Torstveit, M., Budgett, R., 2018. IOC consensus statement on relative energy deficiency in sport (RED-S): 2018 update. British journal of sports medicine 52, 687-697.

Moyce, S., Joseph, J., Tancredi, D., Mitchell, D., Schenker, M., 2016. Cumulative Incidence of Acute Kidney Injury in California's Agricultural Workers. J Occup Environ Med 58, 391-397.

Munro, C.J., Stabenfeldt, G.H., Cragun, J.R., Addiego, L.A., Overstreet, J.W., Lasley, B.L., 1991. Relationship of serum estradiol and progesterone concentrations to the excretion profiles of their major urinary metabolites as measured by enzyme immunoassay and radioimmunoassay. Clinical chemistry 37, 838-844.

Murakami, H., Kawakami, R., Nakae, S., Nakata, Y., Ishikawa-Takata, K., Tanaka, S., Miyachi, M., 2016. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. JAMA internal medicine 176, 702-703.

Murphy, N.E., Carrigan, C.T., Karl, J.P., Pasiakos, S.M., Margolis, L.M., 2018. Threshold of Energy Deficit and Lower-Body Performance Declines in Military Personnel: A Meta-Regression. Sports Medicine, 1-10.

Murphy, S.A., Beaton, R.D., Pike, K.C., Johnson, L., 1999. Occupational stressors, stress responses, and alcohol consumption among professional firefighters: A prospective, longitudinal analysis. International Journal of Stress Management 6, 179-196.

Na, W., Jang, J.Y., Lee, K.E., Kim, H., Jun, B., Kwon, J.W., Jo, S.N., 2013. The effects of temperature on heat-related illness according to the characteristics of patients during the summer of 2012 in the Republic of Korea. Journal of Preventive Medicine & Public Health / Yebang Uihakhoe Chi 46, 19-27.

Nadel, E.R., Pandolf, K.B., Roberts, M.F., Stolwijk, J.A., 1974. Mechanisms of thermal acclimation to exercise and heat. J Appl Physiol. 37, 515-520.

Nagata, I., Kato, K., Seki, K., Furuya, K., 1986. Ovulatory disturbances. Causative factors among Japanese student nurses in a dormitory. Journal of adolescent health care : official publication of the Society for Adolescent Medicine 7, 1-5.

Nakai, S., 2012. [The epidemiology of heat disorders]. Nippon Rinsho - Japanese Journal of Clinical Medicine 70, 934-939.

Nakai, S., 2015. [Historical changes and recent trends in heat stroke]. Koshu Eisei 79, 366-372.

Nakai, S., Itoh, T., Morimoto, T., 1999. Deaths from heat-stroke in Japan: 1968-1994. International Journal of Biometeorology 43, 124-127.

Nakkas, C., Annen, H., Brand, S., 2016. Psychological distress and coping in military cadre candidates. Neuropsychiatric disease and treatment 12, 2237.

Nam, G.E., Han, K., Lee, G., 2017. Association between sleep duration and menstrual cycle irregularity in Korean female adolescents. Sleep Medicine 35, 62-66.

Nashold, R.D., Jentzen, J.M., Peterson, P.L., Remington, P.L., 1996. Heat-related deaths during the summer of 1995, Wisconsin. Clin Neuropharmacol. 19, 252-258.

Nater, U.M., Rohleder, N., 2009. Salivary alpha-amylase as a non-invasive biomarker for the sympathetic nervous system: current state of research. Psychoneuroendocrinology 34, 486-496.

Nathan, R.S., Mary Alice, M., Kimberly, A., Crystal, M., Adam, M.B., Patricia, A.R., Brett, T.L., Consortium, S.S., 2012. A Scheme for Categorizing Traumatic Military Events. Behavior Modification 36, 787-807.

Nattiv, A., Loucks, A.B., Manore, M.M., Sanborn, C.F., Sundgot-Borgen, J., Warren, M.P., 2007. American College of Sports Medicine position stand. The female athlete triad. Medicine and science in sports and exercise 39, 1867-1882.

Nelson, N.G., Collins, C.L., Comstock, R.D., McKenzie, L.B., 2011. Exertional heat-related injuries treated in emergency departments in the U.S., 1997-2006. American Journal of Preventive Medicine 40, 54-60.

Neymotin, F., Nemzer, L.R., 2014. Locus of Control and Obesity. Frontiers in Endocrinology 5, 159

NHS Digital, 2018. Statistics on Sexual and Reproductive Health Services (Contraception) England: 2017/2018. NHS England, London, UK.

NICE, 2017. Clinical guideline [CG156] Fertility problems: assessment and treatment, in: Excellence, N.I.o.C. (Ed.), p. Section 1.3.4.

Nindl, B.C., Barnes, B.R., Alemany, J.A., Frykman, P.N., Shippee, R.L., Friedl, K.E., 2007. Physiological consequences of U.S. Army Ranger training. Medicine and science in sports and exercise 39, 1380-1387.

Nindl, B.C., Billing, D.C., Drain, J.R., Beckner, M.E., Greeves, J., Groeller, H., Teien, H.K., Marcora, S., Moffitt, A., Reilly, T., 2018. Perspectives on resilience for military readiness and preparedness: Report of an international military physiology roundtable. Journal of science and medicine in sport 21, 1116-1124.

Nindl, B.C., Kraemer, W.J., Deaver, D.R., Peters, J.L., Marx, J.O., Heckman, J.T., Loomis, G.A., 2001. LH secretion and testosterone concentrations are blunted after resistance exercise in men. Journal of applied physiology (Bethesda, Md. : 1985) 91, 1251-1258.

Nindl, B.C., McClung, J.P., Miller, J.P., Karl, J.P., Pierce, J.R., Scofield, D.E., Young, A.J., Lieberman, H.R., 2011. Bioavailable IGF-I Is Associated with Fat-Free Mass Gains after Physical Training in Women. Medicine & Science in Sports & Exercise 43, 793-799.

Nitschke, M., Hansen, A., Bi, P., Pisaniello, D., Newbury, J., Kitson, A., Tucker, G., Avery, J., Dal Grande, E., 2013. Risk Factors, Health Effects and Behaviour in Older People during Extreme Heat: A Survey in South Australia. International journal of environmental research and public health 10, 6721-6733.

Noor, N., Pao, C., Dragomir-Davis, M., Tran, J., Arbona, C., 2019. PTSD symptoms and suicidal ideation in US female firefighters. Occupational medicine (Oxford) [Online ahead of print]

Notley, S.R., Park, J., Tagami, K., Ohnishi, N., Taylor, N.A.S., 2017. Variations in body morphology explain sex differences in thermoeffector function during compensable heat stress. Exp Physiol 102, 545-562.

Nunan, D., Sandercock, G.R., Brodie, D.A., 2010. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. Pacing and clinical electrophysiology : PACE 33, 1407-1417.

O'Leary, T.J., Gifford, R.M., Double, R.L., Reynolds, R.M., Woods, D.R., Wardle, S.L., Greeves, J.P., 2019. Skeletal responses to an all-female unassisted Antarctic traverse. Bone 121, 267-276.

O'Connor, K.A., Brindle, E., Holman, D.J., Klein, N.A., Soules, M.R., Campbell, K.L., Kohen, F., Munro, C.J., Shofer, J.B., Lasley, B.L., Wood, J.W., 2003. Urinary Estrone Conjugate and Pregnanediol 3-Glucuronide Enzyme Immunoassays for Population Research. Clinical chemistry 49, 1139-1148.

Opstad, P., 2001. Endocrine and metabolic changes during exhaustive multifactorial military stress. Results from studies during the Ranger training course of the Norwegian Military Academy. Defense Technical Information Center Compliation Part Notice ADP010649. Available at:

https://pdfs.semanticscholar.org/ac2a/184e71a275e43780c1206926fb6b343739e7.pdf (Accessed 19 December 2019).

Opstad, P.K., Falch, D., Oktedalen, O., Fonnum, F., Wergeland, R., 1984. The thyroid function in young men during prolonged exercise and the effect of energy and sleep deprivation. Clin Endocrinol (Oxf) 20, 657-669.

Otis, C.L., Drinkwater, B., Johnson, M., Loucks, A., Wilmore, J., 1997. American College of Sports Medicine position stand. The Female Athlete Triad. Medicine and science in sports and exercise 29, i-ix.

Palm-Fischbacher, S., Ehlert, U., 2014. Dispositional resilience as a moderator of the relationship between chronic stress and irregular menstrual cycle. Journal of Psychosomatic Obstetrics and Gynaecolology 35, 42-50.

Papageorgiou, M., Elliott-Sale, K.J., Parsons, A., Tang, J.C.Y., Greeves, J.P., Fraser, W.D., Sale, C., 2017. Effects of reduced energy availability on bone metabolism in women and men. Bone 105, 191-199.

Paravidino, V.B., Mediano, M.F.F., Sichieri, R., 2017. Does a single bout of exercise influence subsequent physical activity and sedentary time in overweight boys? Physiology & behavior 173, 231-235.

Pauli, S.A., Berga, S.L., 2010. Athletic amenorrhea: energy deficit or psychogenic challenge? Ann N Y Acad Sci 1205, 33-38.

Paulin, S., Roberts, J., Roberts, M., Davis, I., 2015. A case study evaluation of competitors undertaking an antarctic ultra-endurance event: nutrition, hydration and body composition variables. Extreme physiology & medicine 4, 3.

Payne, S., Macintosh, A., Stock, J., 2018. Body size and body composition effects on heat loss from the hands during severe cold exposure. American Journal of Physical Anthropology 166, 313-322.

Periard, J.D., Racinais, S., Sawka, M.N., 2015. Adaptations and mechanisms of human heat acclimation: Applications for competitive athletes and sports. Scandinavian journal of medicine & science in sports 25 Suppl 1, 20-38.

Peric, M., Zenic, N., Sekulic, D., Kondric, M., Zaletel, P., 2016. Disordered eating, amenorrhea, and substance use and misuse among professional ballet dancers: Preliminary analysis. Medycyna pracy 67, 21-27.

Petit, M.A., Prior, J.C., 2013. Exercise and the Hypothalamus. Ovulatory Adaptations, in: Constantini, N.H., Hackney, A.C. (Eds.), Endocrinology of Physical Activity and Sport, 2nd ed. New York: Springer

Petitti, D.B., Piaggio, G., Mehta, S., Cravioto, M.C., Meirik, O., 2000. Steroid hormone contraception and bone mineral density: a cross-sectional study in an international population. The WHO Study of Hormonal Contraception and Bone Health. Obstetrics and gynecology 95, 736-744.

Petraglia, F., Sutton, S., Vale, W., Plotsky, P., 1987. Corticotropin-releasing factor decreases plasma luteinizing hormone levels in female rats by inhibiting gonadotropin-releasing hormone release into hypophysial-portal circulation. Endocrinology 120, 1083-1088.

Pierce, M.E., Pritchard, L.M., 2016. Lower stress-reactive cortisol in female veterans associated with military status but not PTSD. Stress (Amsterdam, Netherlands) 19, 486-491.

Pillai, S.K., Noe, R.S., Murphy, M.W., Vaidyanathan, A., Young, R., Kieszak, S., Freymann, G., Smith, W., Drenzek, C., Lewis, L., Wolkin, A.F., 2014. Heat illness: predictors of hospital admissions among emergency department visits-Georgia, 2002-2008. Journal of Community Health 39, 90-98.

Pires, V., 2018. Amenorrhea and high intensity training. Presented at the 20th European Congress of Endocrinology 2018, Barcelona, Spain. Endocrine Abstracts 56, P922.

Piver, W.T., Ando, M., Ye, F., Portier, C.J., 1999. Temperature and air pollution as risk factors for heat stroke in Tokyo, July and August 1980-1995. Environmental Health Perspectives 107, 911-916.

Pope, Z., Gao, Y., Bolter, N., Pritchard, M., 2015. Validity and reliability of eating disorder assessments used with athletes: A review. Journal of Sport and Health Science 4, 211-221.

Portnoy, G.A., Relyea, M.R., Decker, S., Shamaskin-Garroway, A., Driscoll, M., Brandt, C.A., Haskell, S.G., 2018. Understanding Gender Differences in Resilience Among Veterans: Trauma History and Social Ecology. Journal of Traumatic Stress 31, 845-855.

Potter, G.D.M., Skene, D.J., Arendt, J., Cade, J.E., Grant, P.J., Hardie, L.J., 2016. Circadian Rhythm and Sleep Disruption: Causes, Metabolic Consequences, and Countermeasures. Endocrine Reviews 37, 584-608.

Pruessner, J.C., Baldwin, M.W., Dedovic, K., Renwick, R., Mahani, N.K., Lord, C., Meaney, M., Lupien, S., 2005. Self-esteem, locus of control, hippocampal volume, and cortisol regulation in young and old adulthood. Neuroimage 28, 815-826.

Public Health England, 2016. Government Dietary Recommendations, London, UK, p. 7. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/618167/government_dietary_recommendations.pdf (Accessed 19 December 2019).

Rangan, A.M., Tieleman, L., Louie, J.C., Tang, L.M., Hebden, L., Roy, R., Kay, J., Allman-Farinelli, M., 2016. Electronic Dietary Intake Assessment (e-DIA): relative validity of a mobile phone application to measure intake of food groups. The British journal of nutrition 115, 2219-2226.

Rauh, M.J., Macera, C.A., Trone, D.W., Shaffer, R.A., Brodine, S.K., 2006. Epidemiology of stress fracture and lower-extremity overuse injury in female recruits. Medicine and science in sports and exercise 38, 1571-1577.

Raukar, N., Lemieux, R., Finn, G., Stearns, R., Casa, D.J., 2015. Heat Illness - A Practical Primer. Rhode Island Medicine 98, 28-31.

Reed, J.L., De Souza, M.J., Mallinson, R.J., Scheid, J.L., Williams, N.I., 2015. Energy availability discriminates clinical menstrual status in exercising women. Journal of the International Society of Sports Nutrition 12, 11.

Reid, H.A.B., Dennison, N.C., Quayle, J., Preston, T., 2017. Starvation Resulting From Inadequate Dietary Planning for a 50-Day Rowing Expedition. International journal of sports physiology and performance 12, 1012-1016.

Remer, T., Maser-Gluth, C., Wudy, S.A., 2008. Glucocorticoid measurements in health and disease--metabolic implications and the potential of 24-h urine analyses. Mini reviews in medicinal chemistry 8, 153-170.

Reyna, R., Traynor, K.D., Hines, G., Boots, L.R., Azziz, R., 2001. Repeated freezing and thawing does not generally alter assay results for several commonly studied reproductive hormones. Fertil Steril 76, 823-825.

Reynolds, R.M., Walker, B.R., Syddall, H.E., Andrew, R., Wood, P.J., Whorwood, C.B., Phillips, D.I., 2001a. Altered control of cortisol secretion in adult men with low birth weight and cardiovascular risk factors. The Journal of clinical endocrinology and metabolism 86, pp. 245-250.

Reynolds, R.M., Walker, B.R., Syddall, H.E., Whorwood, C.B., Wood, P.J., Phillips, D.I., 2001b. Elevated plasma cortisol in glucose-intolerant men: differences in responses to glucose and habituation to venepuncture. The Journal of clinical endocrinology and metabolism 86, 1149-1153.

Rhodes, M., Sorenson, H., 2017. Polar Medicine In:Auerbach's Wilderness Medicine. 7th ed. Boston: Elsevier.

Ricci, M.R., Fried, S.K., Mittleman, K.D., 2000. Acute cold exposure decreases plasma leptin in women. Metabolism: clinical and experimental 49, 421-423.

Richmond, V.L., Carter, J.M., Wilkinson, D.M., Homer, F.E., Rayson, M.P., Wright, A., Bilzon, J.L., 2012. Comparison of the physical demands of single-sex training for male and female recruits in the British Army. Military medicine 177, 709-715.

Richmond, V.L., Wilkinson, D.M., Rayson, M.P., 2010. Physical Demands of the Commissioning Course for Officer Cadets at the Royal Military Academy Sandhurst (CC042: May 2004 to March 2005). Report OPL 10/07. Bristol: Optimal Performance.

Rickenlund, A., Thorén, M., Carlström, K., Von Schoultz, B., Hirschberg, A.L., 2004. Diurnal profiles of testosterone and pituitary hormones suggest different mechanisms for menstrual disturbances in endurance athletes. The Journal of clinical endocrinology and metabolism 89, 702-707.

Riou, M.E., Jomphe-Tremblay, S., Lamothe, G., Stacey, D., Szczotka, A., Doucet, E., 2015. Predictors of Energy Compensation during Exercise Interventions: A Systematic Review. Nutrients 7, 3677-3704.

Rivier, C., Rivest, S., 1991. Effect of stress on the activity of the hypothalamic-pituitary-gonadal axis: peripheral and central mechanisms. Biology of reproduction 45, 523-532.

Robinson, N., Saudan, C., Sottas, P.-E., Mangin, P., Saugy, M., 2007. Performance characteristics of two immunoassays for the measurement of urinary luteinizing hormone. Journal of pharmaceutical and biomedical analysis 43, 270-276.

Roether, W., 1970. Water-CO2 exchange set-up for the routine 18oxygen assay of natural waters. The International journal of applied radiation and isotopes 21, 379-387.

Romero-Franco, N., Jiménez-Reyes, P., Castaño-Zambudio, A., Capelo-Ramírez, F., Rodríguez-Juan, J.J., González-Hernández, J., Toscano-Bendala, F.J., Cuadrado-Peñafiel, V., Balsalobre-Fernández, C., 2017. Sprint performance and mechanical outputs computed with an iPhone app: Comparison with existing reference methods. European journal of sport science 17, 386-392.

Roos, J., Johnson, S., Weddell, S., Godehardt, E., Schiffner, J., Freundl, G., Gnoth, C., 2015. Monitoring the menstrual cycle: Comparison of urinary and serum reproductive hormones referenced to true ovulation. The European Journal of Contraception & Reproductive Health Care 20, 438-450.

Rosen, H., Jameel, M.L., Barkan, A.L., 1988. Dexamethasone suppresses gonadotropin-releasing hormone (GnRH) secretion and has direct pituitary effects in male rats: differential regulation of GnRH receptor and gonadotropin responses to GnRH. Endocrinology 122, 2873-2880.

Rosengren, A., Hawken, S., Ounpuu, S., Sliwa, K., Zubaid, M., Almahmeed, W.A., Blackett, K.N., Sitthi-amorn, C., Sato, H., Yusuf, S., 2004. Association of psychosocial risk factors with risk of acute myocardial infarction in 11119 cases and 13648 controls from 52 countries (the INTERHEART study): case-control study. Lancet 364, 953-962.

Ross, T., 2015. British Army's women soldiers to go into combat. Available at: https://www.telegraph.co.uk/news/uknews/defence/12060225/British-Armys-women-soldiers-to-go-into-combat.html (Accessed: 19 December 2019).

Roupas, N.D., Georgopoulos, N.A., 2011. Menstrual function in sports. Hormones (Athens) 10, 104-116.

Rumball, J.S., Lebrun, C.M., 2004. Preparticipation physical examination: selected issues for the female athlete. Clinical journal of sport medicine : official journal of the Canadian Academy of Sport Medicine 14, 153-160.

Sanchez, C.A., Thomas, K.E., Malilay, J., Annest, J.L., 2010. Nonfatal natural and environmental injuries treated in emergency departments, United States, 2001-2004. Fam Community Health. 33, 3-10.

Saunders, B., Maddison, J., 2015. Following Scott. Available at: https://www.sidetracked.com/the-scott-expedition (Accessed: 19 December 2019).

Schaal, K., Van Loan, M.D., Casazza, G.A., 2011. by Martin Krause. Medicine & Science in Sports & Exercise 43, 34-43.

Scheid, J.L., De Souza, M.J., 2010. Menstrual irregularities and energy deficiency in physically active women: the role of ghrelin, PYY and adipocytokines. Medicine and sport science 55, 82-102.

Scheid, J.L., De Souza, M.J., Hill, B.R., Leidy, H.J., Williams, N.I., 2013. Decreased luteinizing hormone pulse frequency is associated with elevated 24-hour ghrelin after calorie restriction and exercise in premenopausal women. American journal of physiology. Endocrinology and metabolism 304, E109-116.

Schnakenberg, D., Hill, T., Kretsch, M., Morris, B., 1981. Diary—interview technique to assess food consumption patterns of individual military personnel. Assessing changing food consumption patterns, 187-197.

Schneider, M.B., Bijur, P.E., Fisher, M., Friedman, S.B., Toffler, C.P.A., 2003. Menstrual irregularity in female military cadets: comparison of data utilizing short-term and long-term recall. Journal of pediatric and adolescent gynecology 16, 89-93.

Schneider, M.B., Fisher, M., Friedman, S.B., Bijur, P.E., Toffler, A.P., 1999. Menstrual and premenstrual issues in female military cadets: a unique population with significant concerns. Journal of pediatric and adolescent gynecology 12, 195-201.

Scholes, D., Ichikawa, L., LaCroix, A.Z., Spangler, L., Beasley, J.M., Reed, S., Ott, S.M., 2010. Oral contraceptive use and bone density in adolescent and young adult women. Contraception 81, 35-40.

Schubert, M.M., Palumbo, E., Seay, R.F., Spain, K.K., Clarke, H.E., 2017. Energy compensation after sprint- and high-intensity interval training. PLoS One 12, e0189590.

Seelig, A.D., Jacobson, I.G., Smith, B., Hooper, T.I., Boyko, E.J., Gackstetter, G.D., Gehrman, P., Macera, C.A., Smith, T.C., Team, M.C.S., 2010. Sleep Patterns Before, During, and After Deployment to Iraq and Afghanistan. Sleep 33, 1615-1622.

Sell, T.C., Abt, J.P., Nagai, T., Deluzio, J.B., Lovalekar, M., Wirt, M.D., Lephart, S.M., 2016. The Eagle Tactical Athlete Program Reduces Musculoskeletal Injuries in the 101st Airborne Division (Air Assault). Military medicine 181, 250-257.

Selye, H., 1939. The effect of adaptation to various damaging agents on the female sex organs in the rat. Endocrinology 25, 615-624.

Selye, H., 1946. The general adaptation syndrome and the diseases of adaptation. The Journal of clinical endocrinology and metabolism 6, pp. 117-230.

Senoo, H., 2000. Physiology of stress and starvation-like conditions, The laboratory rat. Elsevier, pp. 447-460.

Shaffer, F., Ginsberg, J.P., 2017. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health 5, 258.

Sharma, R., Biedenharn, K.R., Fedor, J.M., Agarwal, A., 2013. Lifestyle factors and reproductive health: taking control of your fertility. Reprod Biol Endocrinol 11, 1-15.

Shattuck, G., Hilferty, M., 1933. Causes of Deaths from Heat in Massachusetts. New England Journal of Medicine 209, 319-329.

Shay, L.E., Seibert, D., Watts, D., Sbrocco, T., Pagliara, C., 2009. Adherence and weight loss outcomes associated with food-exercise diary preference in a military weight management program. Eating behaviors 10, 220-227.

Shen, T., Howe, H.L., Alo, C., Moolenaar, R.L., 1998. Toward a broader definition of heat-related death: comparison of mortality estimates from medical examiners' classification with those from total death differentials during the July 1995 heat wave in Chicago, Illinois. The American journal of forensic medicine and pathology 19, 113-118.

Siddall, A.G., Powell, S.D., Needham-Beck, S.C., Edwards, V.C., Thompson, J.E., Kefyalew, S.S., Singh, P.A., Orford, E.R., Venables, M.C., Jackson, S., 2019. Validity of energy expenditure estimation methods during 10 days of military training. Scand J Med Sci Sports, 29, 1313-1321

Sidossis, L., Kajimura, S., 2015. Brown and beige fat in humans: thermogenic adipocytes that control energy and glucose homeostasis. The Journal of clinical investigation 125, 478-486.

Silva, A.M., Matias, C.N., Santos, D.A., Thomas, D., Bosy-Westphal, A., Mu, L.M., Heymsfield, S.B., Sardinha, L.B., 2017. Compensatory Changes in Energy Balance Regulation over One Athletic Season. Medicine and science in sports and exercise 49, 1229-1235.

Silva, M.-R.G., Silva, H.-H., Paiva, T., 2018. Sleep duration, body composition, dietary profile and eating behaviours among children and adolescents: a comparison between Portuguese acrobatic gymnasts. European journal of pediatrics 177, 815-825.

Simoes, T.M., Zapata, C.L., Donangelo, C.M., 2015. [Influence of hormonal contraceptives on indices of zinc homeostasis and bone remodeling in young adult women]. Revista brasileira de ginecologia e obstetricia : revista da Federacao Brasileira das Sociedades de Ginecologia e Obstetricia 37, 402-410.

Sinclair, W.H., Crowe, M.J., Spinks, W.L., Leicht, A.S., 2007. Pre-Pubertal Children and Exercise in Hot and Humid Environments: A Brief Review. Journal of Sports Science & Medicine 6, 385-392.

Sinreih, M., Zukunft, S., Sosic, I., Cesar, J., Gobec, S., Adamski, J., Lanisnik Rizner, T., 2015. Combined liquid chromatography-tandem mass spectrometry analysis of progesterone metabolites. PLoS One 10, e0117984.

Skog, C., 2011. Antarktis. 71 dager,1800 kilometer og én million tanker [Antartica. 71 days, 1800 kilometers and one million thoughts]. Glydendal, Copenhagen, Denmark.

Skoluda, N., Dettenborn, L., Stalder, T., Kirschbaum, C., 2012. Elevated hair cortisol concentrations in endurance athletes. Psychoneuroendocrinology 37, 611-617.

Skoluda, N., Strahler, J., Schlotz, W., Niederberger, L., Marques, S., Fischer, S., Thoma, M.V., Spoerri, C., Ehlert, U., Nater, U.M., 2015. Intra-individual psychological and physiological responses to acute laboratory stressors of different intensity. Psychoneuroendocrinology 51, 227-236.

Skorupskaite, K., George, J.T., Veldhuis, J.D., Anderson, R.A., 2017. Neurokinin B regulates gonadotropin secretion, ovarian follicle growth and the timing of ovulation in healthy women. The Journal of clinical endocrinology and metabolism.

Slater, J., Brown, R., McLay-Cooke, R., Black, K., 2017. Low Energy Availability in Exercising Women: Historical Perspectives and Future Directions. Sports medicine. 24, 207-220

Smith, G.I., Yoshino, J., Kelly, S.C., Reeds, D.N., Okunade, A., Patterson, B.W., Klein, S., Mittendorfer, B., 2016. High-protein intake during weight loss therapy eliminates the weight-loss-induced improvement in insulin action in obese postmenopausal women. Cell reports 17, 849-861.

Smith, S.K., Lenton, E.A., Landgren, B.M., Cooke, I.D., 1984. The short luteal phase and infertility. British journal of obstetrics and gynaecology 91, 1120-1122.

Smith, S.M., Vale, W.W., 2006. The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. Dialogues Clin Neurosci 8, pp. 383-395.

Sonnenblick, M., Friedlander, Y., Rosin, A.J., 1993. Diuretic-induced severe hyponatremia. Review and analysis of 129 reported patients. Chest 103, 601-606.

Southmayd, E.A., Williams, N.I., Mallinson, R.J., De Souza, M.J., 2019. Energy Deficiency Suppresses Bone Turnover in Exercising Women With Menstrual Disturbances. The Journal of Clinical Endocrinology & Metabolism 104, 3131-3145.

Sowers, M.F., Zheng, H., Kravitz, H.M., Matthews, K., Bromberger, J.T., Gold, E.B., Owens, J., Consens, F., Hall, M., 2008. Sex steroid hormone profiles are related to sleep measures from polysomnography and the Pittsburgh Sleep Quality Index. Sleep 31, 1339-1349.

Sowers, M.R., Eyvazzadeh, A.D., McConnell, D., Yosef, M., Jannausch, M.L., Zhang, D., Harlow, S., Randolph, J.F., Jr., 2008. Anti-mullerian hormone and inhibin B in the definition of ovarian aging and the menopause transition. The Journal of clinical endocrinology and metabolism 93, 3478-3483.

Spaanderman, D.C.E., Nixon, M., Buurstede, J.C., Sips, H.C., Schilperoort, M., Kuipers, E.N., Backer, E.A., Kooijman, S., Rensen, P.C.N., Homer, N.Z.M., Walker, B.R., Meijer, O.C., Kroon, J., 2018. Androgens modulate glucocorticoid receptor activity in adipose tissue and liver. The Journal of endocrinology. [Online ahead of print]

Speakman, J., 1997. Doubly labelled water: theory and practice. Springer Science & Business Media.

Spiegel, K., Leproult, R., L'Hermite-Balériaux, M., Copinschi, G., Penev, P.D., Van Cauter, E., 2004. Leptin Levels Are Dependent on Sleep Duration: Relationships with Sympathovagal Balance, Carbohydrate Regulation, Cortisol, and Thyrotropin. The Journal of Clinical Endocrinology & Metabolism 89, 5762-5771.

Spiegel, K., Leproult, R., Van Cauter, E., 1999. Impact of sleep debt on metabolic and endocrine function. The Lancet 354, 1435-1439.

Stacey, M.J., Delves, S.K., Britland, S.E., Allsopp, A.J., Brett, S.J., Fallowfield, J.L., Woods, D.R., 2018b. Copeptin reflects physiological strain during thermal stress. European journal of applied physiology 118, 75-84.

Stacey, M.J., Hill, N., Woods, D. R., 2018a. Physiological monitoring for healthy military personnel. J Roy Army Med Corps 164, 290-292

Stacey, M.J., Parsons, I.T., Woods, D.R., Taylor, P.N., Ross, D., J Brett, S., 2015. Susceptibility to exertional heat illness and hospitalisation risk in UK military personnel. BMJ Open Sport & Exercise Medicine 1, e000055

Stacey, M.J., Woods, D.R., Ross, D., Wilson, D., 2014. Heat illness in military populations: asking the right questions for research. J Roy Army Med Corps 160, 121-124.

Stachenfeld, N.S., Silva, C., Keefe, D.L., 2000. Estrogen modifies the temperature effects of progesterone. Journal of applied physiology (Bethesda, Md. : 1985) 88, 1643-1649.

Stalder, T., Kirschbaum, C., 2012. Analysis of cortisol in hair--state of the art and future directions. Brain Behav Immun 26, 1019-1029.

Stalder, T., Kirschbaum, C., Kudielka, B.M., Adam, E.K., Pruessner, J.C., Wust, S., Dockray, S., Smyth, N., Evans, P., Hellhammer, D.H., Miller, R., Wetherell, M.A., Lupien, S.J., Clow, A., 2016. Assessment of the cortisol awakening response: Expert consensus guidelines. Psychoneuroendocrinology 63, 414-432.

Stalder, T., Steudte-Schmiedgen, S., Alexander, N., Klucken, T., Vater, A., Wichmann, S., Kirschbaum, C., Miller, R., 2017. Stress-related and basic determinants of hair cortisol in humans: A meta-analysis. Psychoneuroendocrinology 77, 261-274.

Stanczyk, F.Z., Gentzschein, E., Ary, B.A., Kojima, T., Ziogas, A., Lobo, R.A., 1997. Urinary progesterone and pregnanediol. Use for monitoring progesterone treatment. The Journal of reproductive medicine 42, 216-222.

Stephens, M.A., Mahon, P.B., McCaul, M.E., Wand, G.S., 2016. Hypothalamic-pituitary-adrenal axis response to acute psychosocial stress: Effects of biological sex and circulating sex hormones. Psychoneuroendocrinology 66, 47-55.

Stephenson, L.A., Kolka, M.A., 1985. Menstrual cycle phase and time of day alter reference signal controlling arm blood flow and sweating. The American journal of physiology 249, R186-191.

Stephenson, L.A., Kolka, M.A., 1993. Thermoregulation in women. Exercise and Sport Sciences Reviews 21, 231-262.

Stephenson, L.A., Kolka, M.A., 1999. Esophageal temperature threshold for sweating decreases before ovulation in premenopausal women. Journal of applied physiology (Bethesda, Md. : 1985) 86, 22-28.

Steudte-Schmiedgen, S., Stalder, T., Schonfeld, S., Wittchen, H.U., Trautmann, S., Alexander, N., Miller, R., Kirschbaum, C., 2015. Hair cortisol concentrations and cortisol stress reactivity predict PTSD symptom increase after trauma exposure during military deployment. Psychoneuroendocrinology 59, 123-133.

Stewart A., Marfell-Jones M., Olds T., for the Interational Society for Advancement of Kinanthropometry, 2011. International standards for anthropometric assessment. Lower Hutt, New Zealand.

Stieg, M.R., Sievers, C., Farr, O., Stalla, G.K., Mantzoros, C.S., 2015. Leptin: A hormone linking activation of neuroendocrine axes with neuropathology. Psychoneuroendocrinology 51, 47-57.

Stirrat, L.I., Walker, J.J., Stryjakowska, K., Jones, N., Homer, N.Z.M., Andrew, R., Norman, J.E., Lightman, S.L., Reynolds, R.M., 2018. Pulsatility of glucocorticoid hormones in pregnancy: Changes with gestation and obesity. Clin Endocrinol (Oxf) 88, 592-600.

Stocker, L.J., Macklon, N.S., Cheong, Y.C., Bewley, S.J., 2014. Influence of shift work on early reproductive outcomes: a systematic review and meta-analysis. Obstetrics & Gynecology 124, 99-110.

Street, A.E., Dardis, C.M., 2018. Using a social construction of gender lens to understand gender differences in posttraumatic stress disorder. Clinical psychology review 66, 97-105.

Stroud, M., Jackson, A., Waterlow, J., 1996. Protein turnover rates of two human subjects during an unassisted crossing of Antarctica. British Journal of Nutrition 76, 165-174.

Stroup, D.F., Berlin, J.A., Morton, S.C., Olkin, I., Williamson, G.D., Rennie, D., Moher, D., Becker, B.J., Sipe, T.A., Thacker, S.B., 2000. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. Jama 283, 2008-2012.

Stubbs, R.J., Hopkins, M., Finlayson, G., Duarte, C., Gibbons, C., Blundell, J., 2018. Potential effects of fat mass and fat-free mass on energy intake in different states of energy balance. European journal of clinical nutrition 72, 698.

Su, H.-W., Yi, Y.-C., Wei, T.-Y., Chang, T.-C., Cheng, C.-M., 2017. Detection of ovulation, a review of currently available methods. Bioengineering & Translational Medicine 2, 238-246.

Tarnopolsky, M.A., 2008. Sex differences in exercise metabolism and the role of 17-beta estradiol. Medicine and science in sports and exercise 40, 648-654.

Task Force, 1996. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Eur Heart J 17, 354-381.

Tawatsupa, B., Yiengprugsawan, V., Kjellstrom, T., Berecki-Gisolf, J., Seubsman, S.A., Sleigh, A., 2013. Association between heat stress and occupational injury among Thai workers: findings of the Thai Cohort Study. Industrial Health 51, 34-46.

Taylor, A.J., McGwin, G., Jr., 2000. Temperature-related deaths in Alabama. Southern Medical Journal 93, 787-792.

Taylor, M.K., Kviatkovsky, S.A., Hernández, L.M., Sargent, P., Segal, S., Granger, D.A., 2016. Anabolic hormone profiles in elite military men. Steroids 110, 41-48.

Taylor, M.K., Larson, G.E., Hiller Lauby, M.D., Padilla, G.A., Wilson, I.E., Schmied, E.A., Highfill-McRoy, R.M., Morgan, C.A., 3rd, 2014. Sex differences in cardiovascular and subjective stress reactions: prospective evidence in a realistic military setting. Stress (Amsterdam, Netherlands) 17, 70-78.

Taylor, M.K., Sausen, K.P., Mujica-Parodi, L.R., Potterat, E.G., Yanagi, M.A., Kim, H., 2007a. Neurophysiologic methods to measure stress during survival, evasion, resistance, and escape training. Aviation, space, and environmental medicine 78, B224-B230.

Taylor, M.K., Sausen, K.P., Potterat, E.G., Mujica-Parodi, L.R., Reis, J.P., Markham, A.E., Padilla, G.A., Taylor, D.L., 2007b. Stressful military training: endocrine reactivity, performance, and psychological impact. Aviation, space, and environmental medicine 78, 1143-1149.

Taylor, N., Gifford, R.M., Cobb, R., Wardle, S.L., Jones, S., Blackadder-Weinstein, J., Hattersley, J., Wilson, A., Imray, C., Greeves, J.P., Reynolds, R., Woods, D.R., 2019. Experience from the selection and nutritional preparation for Expedition ICE MAIDEN: the first successful all-female unassisted Antarctic traverse. Journal of the Royal Army Medical Corps. Published Online First: 15 May 2019, doi: 10.1136/jramc-2019-001175

Taylor, N.A., 2014. Human heat adaptation. Comprehensive Physiology 4, 325-365

Telfer, E.E., McLaughlin, M., 2007. Natural history of the mammalian oocyte. Reproductive biomedicine online 15, 288-295.

Thacher, T.D., Clarke, B.L., 2011. Vitamin D Insufficiency. Mayo Clinic proceedings 86, 50-60.

Thake, C.D., Hattersley, J., Wilson, A., Gifford, R.M., Cobb, R., Reynolds, R.M., Woods, D., Imray, C., 2019. Exercise ICE MAIDEN: Physiological responses to incremental exercise in normoxia and hypoxia before and after an Antarctic expedition, International Conference on Environmental Ergonomics, Amsterdam, NL.

Tharion, W.J., Lieberman, H.R., Montain, S.J., Young, A.J., Baker-Fulco, C.J., Delany, J.P., Hoyt, R.W., 2005. Energy requirements of military personnel. Appetite 44, 47-65.

Thayer, J.F., Sternberg, E., 2006. Beyond heart rate variability: vagal regulation of allostatic systems. Annals of the New York Academy of Sciences 1088, 361-372.

The Comprehensive R Archive Network, 2019. The Comprehensive R Archive Network. Available at: https://cran.r-project.org/ (Accessed: 19 December 2019).

Thompson, F.E., Subar, A.F., Loria, C.M., Reedy, J.L., Baranowski, T., 2010. Need for technological innovation in dietary assessment. Journal of the American Dietetic Association 110, 48-51.

Tiedeken, M., Westhoff, C.L., Cohen, A., Cremers, S., Sitruk-Ware, R., Blithe, D.L., 2019. Bone turnover markers in women participating in a dose-finding trial of a contraceptive vaginal ring releasing Nestorone and estradiol. Contraception 99, 329-334

Tolin, D.F., Foa, E.B., 2006. Sex differences in trauma and posttraumatic stress disorder: a quantitative review of 25 years of research. Psychological bulletin 132, 959-992.

Toloo, G.S., Yu, W., Aitken, P., FitzGerald, G., Tong, S., 2014. The impact of heatwaves on emergency department visits in Brisbane, Australia: a time series study. Crit Care. 18, R69. doi: 10.1186/cc13826.

Torstveit, M.K., Fahrenholtz, I., Stenqvist, T.B., Sylta, Ø., Melin, A., 2018. Within-day energy deficiency and metabolic perturbation in male endurance athletes. International journal of sport nutrition and exercise metabolism 28, 419-427.

Torstveit, M.K., Sundgot-Borgen, J., 2005. The female athlete triad exists in both elite athletes and controls. Medicine and science in sports and exercise 37, 1449-1459.

Toulemon, L., Barbieri, M., 2008. The mortality impact of the August 2003 heat wave in France: investigating the 'harvesting' effect and other long-term consequences. Population Studies 62, 39-53.

Touzet, S., Rabilloud, M., Boehringer, H., Barranco, E., Ecochard, R., 2002. Relationship between sleep and secretion of gonadotropin and ovarian hormones in women with normal cycles. Fertility and sterility 77, 738-744.

Tran, K.V., Azhar, G.S., Nair, R., Knowlton, K., Jaiswal, A., Sheffield, P., Mavalankar, D., Hess, J., 2013. A cross-sectional, randomized cluster sample survey of household vulnerability to extreme heat among slum dwellers in ahmedabad, india. International Journal of Environmental Research & Public Health [Electronic Resource] 10, 2515-2543.

Turrell, G., Lynch, J.W., Leite, C., Raghunathan, T., Kaplan, G.A., 2007. Socioeconomic disadvantage in childhood and across the life course and all-cause mortality and physical function in adulthood: evidence from the Alameda County Study. Journal of epidemiology and community health 61, 723-730.

Ulrich-Lai, Y.M., Fulton, S., Wilson, M., Petrovich, G., Rinaman, L., 2015. Stress exposure, food intake and emotional state. Stress (Amsterdam, Netherlands) 18, 381-399.

Vaha-Eskeli, K., Erkkola, R., Seppanen, A., 1991. Is the heat dissipating ability enhanced during pregnancy? European Journal of Obstetrics, Gynecology, & Reproductive Biology 39, 169-174.

Välimäki, V.-V., Alfthan, H., Lehmuskallio, E., Löyttyniemi, E., Sahi, T., Suominen, H., Välimäki, M.J., 2005. Risk factors for clinical stress fractures in male military recruits: a prospective cohort study. Bone 37, 267-273.

van der Voorn, B., Hollanders, J.J., Ket, J.C., Rotteveel, J., Finken, M.J., 2017. Gender-specific differences in hypothalamus-pituitary-adrenal axis activity during childhood: a systematic review and meta-analysis. Biol Sex Differ 8, 3

van Hees, V.T., Fang, Z., Langford, J., Assah, F., Mohammad, A., da Silva, I.C., Trenell, M.I., White, T., Wareham, N.J., Brage, S., 2014. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. Journal of applied physiology (Bethesda, Md. : 1985) 117, 738-744.

van Hees, V.T., Renstrom, F., Wright, A., Gradmark, A., Catt, M., Chen, K.Y., Lof, M., Bluck, L., Pomeroy, J., Wareham, N.J., Ekelund, U., Brage, S., Franks, P.W., 2011. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. PLoS One 6, e22922.

VanHeest, J.L., Rodgers, C.D., Mahoney, C.E., De, M.S., 2014. Ovarian suppression impairs sport performance in junior elite female swimmers. Medicine and science in sports and exercise 46, 156-166.

Varnes, J.R., Stellefson, M.L., Janelle, C.M., Dorman, S.M., Dodd, V., Miller, M.D., 2013. A systematic review of studies comparing body image concerns among female college athletes and non-athletes, 1997–2012. Body image 10, 421-432.

Veldhuis, J.D., Keenan, D.M., Pincus, S.M., 2008. Motivations and methods for analyzing pulsatile hormone secretion. Endocrine reviews 29, 823-864.

Vgontzas, A.N., Bixler, E.O., Lin, H.M., Prolo, P., Mastorakos, G., Vela-Bueno, A., Kales, A., Chrousos, G.P., 2001. Chronic insomnia is associated with nyctohemeral activation of the hypothalamic-pituitary-adrenal axis: clinical implications. The Journal of clinical endocrinology and metabolism 86, 3787-3794.

Viechtbauer, W., 2007. Hypothesis tests for population heterogeneity in meta-analysis. British Journal of Mathematical and Statistical Psychology, 29–60.

Viechtbauer, W., 2019. The Metafor Package: A Meta-Analysis Package for R. Available at: http://www.metafor-project.org/doku.php (Last accessed: 19 December 2019).

Vilayphiou, N., Boutroy, S., Szulc, P., van Rietbergen, B., Munoz, F., Delmas, P.D., Chapurlat, R., 2011. Finite element analysis performed on radius and tibia HR-pQCT images and fragility fractures at all sites in men. Journal of Bone and Mineral Research 26, 965-973.

Viner, R.T., Harris, M., Berning, J.R., Meyer, N.L., 2015. Energy availability and dietary patterns of adult male and female competitive cyclists with lower than expected bone mineral density. International journal of sport nutrition and exercise metabolism 25, 594-602.

Viola, A.S., Castro, S., Marchi, N.M., Bahamondes, M.V., Viola, C.F., Bahamondes, L., 2011. Long-term assessment of forearm bone mineral density in postmenopausal former users of depot medroxyprogesterone acetate. Contraception 84, 122-127.

Vogel, J.A., Friedl, K.E., 1992. Body fat assessment in women. Special considerations. Sports medicine (Auckland, N.Z.) 13, 245-269.

Vogt, D., Vaughn, R., Glickman, M.E., Schultz, M., Drainoni, M.-L., Elwy, R., Eisen, S., 2011. Gender differences in combat-related stressors and their association with postdeployment mental health in a nationally representative sample of U.S. OEF/OIF veterans. Journal of Abnormal Psychology 120, 797-806.

Volpe, S. L., Stanzione, J. R., 2017. Balancing Nutrition: Recommended Dietary Guidelines, in: Bushman, B., (Ed.). American College of Sports Medicine's Complete Guide to Fitness & Health, 2e. Champaign: Human Kinetics.

Vulliemoz, N.R., Xiao, E., Xia-Zhang, L., Rivier, J., Ferin, M., 2008. Astressin B, a nonselective corticotropin-releasing hormone receptor antagonist, prevents the inhibitory effect of ghrelin on luteinizing hormone pulse frequency in the ovariectomized rhesus monkey. Endocrinology 149, 869-874.

Wade, G.N., Schneider, J.E., 1992. Metabolic fuels and reproduction in female mammals. Neuroscience and biobehavioral reviews 16, 235-272.

Wagner, J.A., Horvath, S.M., 1985. Influences of age and gender on human thermoregulatory responses to cold exposures. Journal of Applied Physiology 58, 180-186.

Wallace, R.F., Kriebel, D., Punnett, L., Wegman, D.H., Wenger, C.B., Gardner, J.W., Gonzalez, R.R., 2005. The effects of continuous hot weather training on risk of exertional heat illness. Medicine & Science in Sports & Exercise 37, 84-90.

Wallace, R.F., Kriebel, D., Punnett, L., Wegman, D.H., Wenger, C.B., Gardner, J.W., Kark, J.A., 2006. Risk factors for recruit exertional heat illness by gender and training period. Aviation Space & Environmental Medicine 77, 415-421.

Walsh, J.S., Eastell, R., Peel, N.F., 2008. Effects of depot medroxyprogesterone acetate on bone density and bone metabolism before and after peak bone mass: a case-control study. The Journal of Clinical Endocrinology & Metabolism 93, 1317-1323.

Wandrag, L., Siervo, M., Riley, H.L., Khosravi, M., Fernandez, B.O., Leckstrom, C.A., Martin, D.S., Mitchell, K., Levett, D.Z.H., Montgomery, H.E., Mythen, M.G., Stroud, M.A., Grocott, M.P.W., Feelisch, M., 2017. Does hypoxia play a role in the development of sarcopenia in humans? Mechanistic insights from the Caudwell Xtreme Everest Expedition. Redox biology 13, 60-68.

Warner, C.M., Warner, C.H., Breitbach, J., Rachal, J., Matuszak, T., Grieger, T.A., 2007. Depression in entry-level military personnel. Military medicine 172, 795-799.

Warren, M., Perlroth, N., 2001. The effects of intense exercise on the female reproductive system. The Journal of endocrinology 170, 3-11.

Warren, M.P., 1980. The effects of exercise on pubertal progression and reproductive function in girls. The Journal of clinical endocrinology and metabolism 51, 1150-1157.

Wathen, N.C., Perry, L., Lilford, R.J., Chard, T., 1984. Interpretation of single progesterone measurement in diagnosis of anovulation and defective luteal phase: observations on analysis of the normal range. British medical journal (Clinical research ed.) 288, 7-9.

Weeke, J., Hansen, A.P., 1975. Serum tsh and serum t3 levels during normal menstrual cycles and during cycles on oral contraceptives. Acta endocrinologica 79, 431-438.

Weiss, D.S., 1997. The Impact of Event Scale-Revised., in: Wilson, J.P., Keane, T.M. (Eds.), Assessing psychological trauma and PTSD: a practitioner's handbook, 2nd ed. Guildford Press, New York, pp. 168-189.

Welch, M.J., 1989. Women in the Military Academies: US Army (Part 3 of 3). The Physician and sportsmedicine 17, 89-96.

Wells, T.S., Horton, J.L., LeardMann, C.A., Jacobson, I.G., Boyko, E.J., 2013. A comparison of the PRIME-MD PHQ-9 and PHQ-8 in a large military prospective study, the Millennium Cohort Study. Journal of affective disorders 148, 77-83.

Welt, C.K., Chan, J.L., Bullen, J., Murphy, R., Smith, P., DePaoli, A.M., Karalis, A., Mantzoros, C.S., 2004. Recombinant human leptin in women with hypothalamic amenorrhea. The New England journal of medicine 351, 987-997.

Wentz, L., Liu, P.-Y., Haymes, E., Ilich, J.Z., 2011. Females Have a Greater Incidence of Stress Fractures Than Males in Both Military and Athletic Populations: A Systemic Review. Military medicine 176, 420-430.

Wheeler, K., Lane, K., Walters, S., Matte, T., 2013. Heat illness and deaths--New York City, 2000-2011. MMWR Morb Mortal Wkly Rep. 62, 617-621.

Whybrow, S., Hughes, D.A., Ritz, P., Johnstone, A.M., Horgan, G.W., King, N., Blundell, J.E., Stubbs, R.J., 2008. The effect of an incremental increase in exercise on appetite, eating behaviour and energy balance in lean men and women feeding ad libitum. The British journal of nutrition 100, 1109-1115.

Williams, J.M., Andersen, M.B., 1998. Psychosocial antecedents of sport injury: Review and critique of the stress and injury model'. Journal of applied sport psychology 10, 5-25.

Williams, N.I., Helmreich, D.L., Parfitt, D.B., Caston-Balderrama, A., Cameron, J.L., 2001. Evidence for a causal role of low energy availability in the induction of menstrual cycle disturbances during strenuous exercise training. The Journal of clinical endocrinology and metabolism 86, 5184-5193.

Williams, N.I., Leidy, H.J., Hill, B.R., Lieberman, J.L., Legro, R.S., De Souza, M.J., 2015. Magnitude of daily energy deficit predicts frequency but not severity of menstrual disturbances associated with exercise and caloric restriction. American journal of physiology. Endocrinology and metabolism 308, E29-39.

Williams, N.I., Statuta, S.M., Austin, A., 2017. Female Athlete Triad: Future Directions for Energy Availability and Eating Disorder Research and Practice. Clinics in sports medicine 36, 671-686.

Williams, N.I., Young, J.C., McArthur, J.W., Bullen, B., Skrinar, G.S., Turnbull, B., 1995. Strenuous exercise with caloric restriction: effect on luteinizing hormone secretion. Medicine and science in sports and exercise 27, 1390-1398.

Williamson, D.A., Allen, H.R., Martin, P.D., Alfonso, A.J., Gerald, B., Hunt, A., 2003. Comparison of digital photography to weighed and visual estimation of portion sizes. Journal of the American Dietetic Association 103, 1139-1145.

Winterer, J., Cutler, G.B., Jr., Loriaux, D.L., 1984. Caloric balance, brain to body ratio, and the timing of menarche. Medical hypotheses 15, 87-91.

Wong, B.J., Hollowed, C.G., 2017. Current concepts of active vasodilation in human skin. Temperature (Austin) 4, 41-59.

Woodruff, S.J., Meloche, R.D., 2013. Energy availability of female varsity volleyball players. International journal of sport nutrition and exercise metabolism 23, 24-30.

Woods, D.R., O'Hara, J.P., Boos, C.J., Hodkinson, P.D., Tsakirides, C., Hill, N.E., Jose, D., Hawkins, A., Phillipson, K., Hazlerigg, A., Arjomandkhah, N., Gallagher, L., Holdsworth, D., Cooke, M., Green, N.D.C., Mellor, A., 2017. Markers of physiological stress during exercise under conditions of normoxia, normobaric hypoxia, hypobaric hypoxia, and genuine high altitude. European journal of applied physiology 117, 893-900.

Yamamoto, T., Todani, M., Oda, Y., Kaneko, T., Kaneda, K., Fujita, M., Miyauchi, T., Tsuruta, R., 2015. Predictive Factors for Hospitalization of Patients with Heat Illness in Yamaguchi, Japan. International Journal of Environmental Research & Public Health [Electronic Resource] 12, 11770-11780.

Yaqub, B., Al Deeb, S., 1998. Heat strokes: aetiopathogenesis, neurological characteristics, treatment and outcome. Journal of the Neurological Sciences 156, 144-151.

Yehuda, R., 2002. Current status of cortisol findings in post-traumatic stress disorder. The Psychiatric clinics of North America 25, 341-368.

Yehuda, R., Bierer, L.M., Pratchett, L.C., Lehrner, A., Koch, E.C., Van Manen, J.A., Flory, J.D., Makotkine, I., Hildebrandt, T., 2015. Cortisol augmentation of a psychological treatment for warfighters with posttraumatic stress disorder: Randomized trial showing improved treatment retention and outcome. Psychoneuroendocrinology 51, 589-597.

Zhao, J., Dong, X., Hu, X., Long, Z., Wang, L., Liu, Q., Sun, B., Wang, Q., Wu, Q., Li, L., 2016. Zinc levels in seminal plasma and their correlation with male infertility: A systematic review and meta-analysis. Scientific reports 6, 22386.

Zschucke, E., Renneberg, B., Dimeo, F., Wüstenberg, T., Ströhle, A., 2015. The stress-buffering effect of acute exercise: Evidence for HPA axis negative feedback. Psychoneuroendocrinology 51, 414-425.