

Statistics Meets Sports

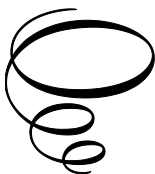
Statistics Meets Sports:

*What We Can Learn
from Sports Data*

Edited by

Yves Dominicy and Christophe Ley

**Cambridge
Scholars
Publishing**



Statistics Meets Sports: What We Can Learn from Sports Data

Edited by Yves Dominicy and Christophe Ley

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2023 by Yves Dominicy, Christophe Ley and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-9273-1

ISBN (13): 978-1-5275-9273-5

Christophe Ley dedicates this book to his biggest achievement, namely his beloved son Emilian. Yves Dominicy also dedicates this book to Emilian, his precious godchild.

Contents

Preface	xiii
----------------	-------------

Acknowledgement	xv
------------------------	-----------

1 The Female Athlete	1
-----------------------------	----------

ASTRID ZECH AND DANIEL HAMACHER

1.1	Introduction	2
1.2	Sex-Specific Differences	4
1.2.1	Female vs. Male Injury Characteristics	4
1.2.2	Female Versus Male Biomechanics	5
1.2.3	Female Versus Male Response to Resistance and Endurance Training	6
1.3	The Power of Hormones	7
1.4	Simpson’s Paradox	8
1.5	Future Research	11
1.5.1	Research Plan Development	12
1.5.2	The Optimal Sample Size	13
1.5.3	Inclusion Criteria, Randomisation, and Data Col- lection	14
1.5.4	Statistical Data Analysis and Results	14

2 National Hockey League	27
---------------------------------	-----------

MICHAEL JACK DAVIS, TIM SWARTZ, OLIVER SCHULTE, JUAN CAMILO GAMBOA HIGUERA
AND MEHRSAN JAVAN

2.1	Introduction	28
2.2	Data	31
2.3	Event-Based Glicko Ratings	31
2.4	Logistic Regression Match Prediction	33

2.4.1	Model Selection	35
2.5	Model Validation	36
2.5.1	Predictive Accuracy	36
2.5.2	Gambling Application	38
2.6	Conclusions	39
3	American Football	43
	KONSTANTINOS PELECHRINIS	
3.1	Introduction	44
3.2	Expected Points	45
3.3	Two-Point Conversion	49
3.4	Player Tracking Data	54
3.4.1	Completion Probability Above Expectation . .	54
3.4.2	In-Play Ballcarrier's Expected Rush Yards . .	62
3.5	Conclusions	71
4	Performance in Swimming	75
	ARTHUR LEROY AND ROBIN PLA	
4.1	Sports before Science	77
4.1.1	Why Does Swimming Offer an Ideal Framework for Statistical Analysis?	77
4.1.2	Performance Predictions and Talent Identification	78
4.1.3	The Influence of Morphology on Performance .	82
4.1.4	How Much Do We Progress during Competitions?	85
4.1.5	Some Practical and Visualisation Tools for Analysing Elite Swimmers	88
4.2	Science before Sports	92
4.2.1	Behind Performance Prediction: Multi-Task Learn- ing and Gaussian Processes	92
4.2.2	Behind Morphology and ISL: Bayesian Hierar- chical Modelling	95

4.3	Conclusions	98
5	Ranking tennis players	103
	MILICA MARIČIĆ AND VELJKO JEREMIĆ	
5.1	Introduction	104
5.2	Literature Review	108
5.3	Proposed Methodology	110
	5.3.1 I-distance Methodology	110
	5.3.2 Ranking Methodology	112
	5.3.3 Used Performance Metrics and Proposed Ranking	113
5.4	Results	114
	5.4.1 Clay	114
	5.4.2 Grass	117
	5.4.3 Hard	120
	5.4.4 Comparison of Rankings	123
5.5	Discussion	128
6	A Rarity in Tennis	135
	TRISTAN BARNETT AND VLADIMIR EJOV	
6.1	Introduction	136
6.2	Physics	136
6.3	Grand Slams	137
6.4	Match Statistics	137
6.5	Conclusion	140
7	Feel the Pressure	143
	TRISTAN BARNETT AND VLADIMIR EJOV	
7.1	Introduction	144
7.2	Importance	144
7.3	Grand Slams	148

7.4	Conclusion	149
8	Tournament Design	153
	LÁSZLÓ CSATÓ	
8.1	Introduction	154
8.2	The CONCACAF Qualifiers	156
8.3	Methodology	159
8.4	Results	163
8.5	Conclusions	169
9	Plus–Minus Ratings	175
	LARS MAGNUS HVATTUM, MARC-ANDRÉ KRIEGL AND MIROSLAV ČULÍK	
9.1	Introduction	176
9.2	Literature Review	177
	9.2.1 Development of Plus–Minus Ratings	177
	9.2.2 Evaluation of Plus–Minus Ratings	181
	9.2.3 Applications of Plus–Minus Ratings	182
9.3	Experimental Setup	183
	9.3.1 Data	184
	9.3.2 Framework	184
	9.3.3 Rating Model	185
9.4	Results and Discussion	188
	9.4.1 Preliminary Analysis of Players Born 1995	188
	9.4.2 Main Analysis	190
	9.4.3 Changes over Time in Player Development	193
9.5	Concluding Remarks	194
10	Competitive Balance	201
	DIMITRIS KARLIS, IOANNIS NTZOUFRAS AND VASILIS MANASIS	
10.1	Introduction	203

10.1.1	Competitive Balance	203
10.1.2	Motivation	204
10.1.3	About This Chapter	207
10.2	Background	208
10.2.1	Competitive Balance Indices	208
10.2.2	Models for Football Outcomes	211
10.3	The Proposed Methodology	213
10.4	Application	216
10.4.1	The Indices Under Complete Balance	216
10.4.2	Premier League and Bundesliga	218
10.4.3	Comparison Across Time	218
10.5	Sensitivity Analysis	220
10.6	Discussion	222
10.6.1	Main Conclusions	222
10.6.2	Future Work	224
10.6.3	Concluding Remarks	225

11 Injury Prediction 233

LUCIANA DE MICHELIS MENDONÇA, THALES REZENDE SOUZA AND SÉRGIO TEIXEIRA
FONSECA

11.1	Introduction	234
11.2	Limitations	235
11.3	Complex Phenomenon	236
11.4	Dealing with a Complex Phenomenon	239
11.5	Solutions	242
11.6	Final Considerations	245

12 World Team Championships 255

RAYMOND STEFANI

12.1	Introduction	256
------	------------------------	-----

12.2	Recognised Sports	257
12.3	Types of Rating Systems	259
12.4	Sports for This Study	261
12.5	Basic Predictability	263
12.6	Linear Regression Analysis	265
12.7	Conclusion	267

Preface

Aim of the Book

The objective of this book is to present the field of sport statistics to two very distinct target groups: on the one hand the academicians, mainly statisticians, in order to raise their interest in this growing field, and on the other hand sports fans who, even without advanced mathematical knowledge, will be able to understand the data analysis part and gain new insights into their favourite sports. The book thus offers a novel perspective on this attractive topic, by combining sports analytics, data visualization and advanced statistical procedures to extract new findings from sports data such as improved rankings or prediction methods.

This book is the follow-up of “Science Meets Sports—When Statistics are more than Numbers” and hence we consider here sports not described in our first book, as well as very timely themes such as the differences between female and male athletes, data-driven talent identification or the essential topic of injury prevention. Data scientists will also find their pleasure in the powerful analytical, statistical and visualization tools used throughout the book.

Context of the Present Book

The world of sports is currently undergoing a fundamental change thanks to the upcoming trend of sports analytics. Recent advances in data collection techniques have enabled collecting large, sometimes even massive, amounts of data in all aspects of sports, such as for instance tactics, technique, health complaints and injuries, spatiotemporal whereabouts (e.g., tracking data from GPS), but also marketing and betting. Data are by now regularly collected in almost every sport, ranging from traditional Olympic disciplines to professional football, basketball, and handball, to name but a few. Moreover, massive data from individual recreational athletes such as runners or cyclists is available. It is by far not only professional and commercially successful sports clubs that aim to analyse data, even recreational athletes and amateur clubs make use of a variety of sensors to monitor their training and performances.

This global rush towards using advanced statistics and machine learning (or, in modern terms, Data Science) methods in sports is due in large parts to the success of the Oakland Athletics baseball in the season 2002. Prior to that season, they have hired new players in a till then atypical way, namely by not relying on scouts' experience but rather on sabermetrics, the technical term for empirical/statistical analysis of baseball. This particular story has been written up in the famous book *Moneyball* in 2003, which in turn appeared as movie in 2011. The success of the Oakland Athletics team has inspired other teams in baseball, and soon after in several other sports. Since then, sports analytics as field has seen a phenomenal development, having led *inter alia* to the developments of new journals such as the Journal of Sports Analytics whose first edition appeared in 2015.

The present book inscribes itself in this context and aims at further contributing to this stimulating research area thanks to its unique feature of targeting academicians and sports fans.

Content of the Present Book

The book starts with a chapter focussing on female athletes and the need for a differentiated analysis of both sexes (Chapter 1 by Zech and Hamacher), then continues with sports not treated in our first book such as hockey (Chapter 2 by Davis, Swartz, Schulte, Higuera and Javan), American football (Chapter 3 by Pelechrinis), and swimming (Chapter 4 by Leroy and Pla), before returning on other facets of tennis (Chapter 5 by Maričić and Jeremić, and Chapters 6 and 7 by Barnett and Ejoy). Next the book considers topics of interest for many sports such as the role of tournament design in sporting success (Chapter 8 by Csató), talent identification via the plus-minus ratings (Chapter 9 by Hvattum, Kriegl and Čulík), uncertainty in competitive balance (Chapter 10 by Karlis, Ntzoufras and Manasis), risk profile identification and injury prediction (Chapter 11 by De Michelis Mendonça, Rezende Souza and Teixeira Fonseca), and finally rating systems and the predictability of World Team Championships (Chapter 12 by Stefani).

Acknowledgement

We wish to thank all contributors to the present book which, we hope, will please the reader. We also thank Yvonne Fromme for her professional proof-reading of the entire book, and Max Sinner for his Latex typesetting help with some chapters initially written in Word. All remaining mistakes are ours.

Chapter 1

The Female Athlete—Exploring the Need For a Differentiated Analysis of Both Sexes

ASTRID ZECH
INSTITUTE OF SPORTS SCIENCE, FRIEDRICH SCHILLER
UNIVERSITY JENA

DANIEL HAMACHER
INSTITUTE OF SPORTS SCIENCE, FRIEDRICH SCHILLER
UNIVERSITY JENA

Abstract

Although male and female athletes differ regarding their physiology, anatomy, biomechanics, injury risk, and performances, research studies in exercise and sports medicine pay little attention on sex-specific analyses. These often lack a well-balanced ratio of women and men in mixed-sex studies or entirely disregard female athletes. This chapter summarises the current evidence of sex-specific differences in sports in order to emphasise the need for a differentiated analysis of male and female athletes. Thereafter, we demonstrate that including sex as a factor in the statistical model could aid in achieving valid results and avoiding wrong conclusions when both sexes are recruited. Finally, we discuss some methodological and statistical issues, and we give some considerations for further studies planning to include male and female participants.

1.1 Introduction

Sports science is currently facing the same problem that has already been acknowledged for other fields of human research: an under representation and underestimation of women (Costello et al., 2014; Hutchins et al., 2021). Considerably more men than women are examined in research studies.

Considerably more men than women are examined in research studies in the field of exercise and sports medicine (Costello et al., 2014; Hutchins et al., 2021). Consequently, most evidence-based recommendations for female athletes are derived from studies with male participants (Elliott-Sale et al., 2021). This problem does not only apply to the ratio between single-sex studies. Mixed-sex studies also often have an unequal number of female and male participants and, in combination with the lack of sex-different data analysis, can result in wrong conclusions (Hagstrom et al., 2021). While this can lead to fatal results in medicine, it remains unclear how it affects our understanding of female motor control, the development of female performance in elite sports, or the prevention and treatment of sports injuries. The major question is whether the biological and non-biological differences between men and women are truly neglectable, for example, exercise-related adaptations or if our current understanding of evidence in sports science should be modified towards a better consideration of sex-specific differences.

Distinguishing sex and gender

In this chapter the term “sex” is mainly used to describe movement and sports-related differences between male and female athletes. Sex and gender differences are caused by a variety of factors including genetic, endocrine, environmental, social, economic, and behavioural differences (Wainer et al., 2020). Sex refers to the biological attributes that distinguish organisms as male, female, intersex, and hermaphrodite (Tannenbaum et al., 2019). Gender refers to psychological, social, and cultural factors that shape attitudes, behaviours, stereotypes, technologies, and knowledge (Organization, 2002; Tannenbaum et al., 2019; Wainer et al., 2020). Participation and performance in sports might be moderated by gender roles (Chalabaev et al., 2013).

In most professional and recreational sports, women and men compete separately from each other and are rarely compared when it comes to peak performances. The only exceptions are sports in which the sex-specific biological characteristics are considered to be playing a subordinate role (e.g., equestrian sports, sailing) or in which both males and females compete with an equal number in the same team (e.g., pair figure skating, mixed tennis). For most of the other sports, it is generally assumed that males would outperform females due to their higher muscle strength or aerobic capacities (Thibault et al., 2010). The development of world records in running, swimming, cycling, and jumping sports supports this (Cheuvront et al., 2005; Nevill et al., 2007; Nevill and Whyte, 2005; Tønnessen et al., 2015). Elite male and female sports performances have constantly improved over the last century, but the gap between sexes has remained nearly unchanged (Thibault et al., 2010).

Considering the noticeable differences between male and female physical capacities (Cureton et al., 1986; Handelsman, 2017; Perez-Gomez et al., 2008) and their influence on the main outcome performance, the limited consideration of sex-specific aspects in human biomechanics, training, and testing as well as in the development of evidence-based recommendations for sports practice is astonishing. Sports-specific training is often based on a mixture of personal experience or beliefs of athletes or coaches and scientific evidence. Although there is wide acceptance for the need for individualised training in order to address the different baseline conditions and responses to training, hardly any of the existing recommendations mention the specific circumstances of female versus male athletes. Therefore, there is still a widespread assumption among coaches and athletic trainers that sex-specific aspects play a subordinate role in the planning of sports-specific training. However, in research there is an increasing body of literature emphasising the need for a more selective approach for both sexes.

1.2 Examples for Sex–Specific Differences in Sports and Biomechanics

1.2.1 Female vs. Male Injury Characteristics

Male and female athletes have different sports–specific injury profiles (Hollander et al., 2021; Montalvo et al., 2019; Zech et al., 2022). For example, in running sports women tend to have more frequent bone stress injuries while men have a higher rate of Achilles tendinopathies (Hollander et al., 2021). In major team sports, such as football (soccer), rugby, hockey, volleyball or handball, male players have a greater risk for injuries in general especially hip, thigh, and foot injuries (Zech et al., 2022). According to the literature, male football players are also more susceptible to hamstring (Cross et al., 2013) and groin injuries (Waldén et al., 2015) than female players. For female athletes, however, the knee seems to be the weak point when it comes to sports injuries (Montalvo et al., 2019). Compared to men they have a higher risk for anterior cruciate ligament (ACL) injuries (Arendt and Dick, 1995; Montalvo et al., 2019; Zech et al., 2022) as well as more than a twofold increased rate of patellofemoral pain (Boling et al., 2010) and iliotibial band friction syndromes (Taunton, 2002).

Based on these findings, sex has often been introduced as a risk factor for sports injuries, such as ACL injuries (Bittencourt et al., 2016) or ankle sprains (Delahunt and Remus, 2019). Future studies on the prevalence of risk factors of sports injuries should therefore either have a balanced ratio of men and women with comparable sports–specific backgrounds or should make clear whether the observed data apply to male, female, or both.

However, injury surveillance and risk factor studies are not the only ones that need sex–specific considerations. It is unknown if injury prevention measures should take the differences between male and female risk profiles into account. Based on the evidence, it is recommended to regularly participate in specific injury prevention training programmes (Hübscher et al., 2010; Steib et al., 2017). These recommendations do not differentiate between the sexes. This, however, is mostly due to the lack of comparative data between male and female participants. The systematic reviews of Thorborg et al. (2017) and Al Attar et al. (2017) showed that previous randomised controlled

studies on injury prevention training had a considerably lower number of female participants compared to male participants.

1.2.2 Female Versus Male Biomechanics

Although the reasons contributing to the different injury risks between the sexes are not fully known, sex-specific musculoskeletal characteristics likely play a considerable role. Women tend to have a greater joint laxity and lower joint resistance to translation and rotation movements (Eriksen and Gribble, 2012; Renstrom et al., 2008). They have a wider pelvis leading to a lower extremity alignment than men as well as a higher likelihood of increased knee valgus (Nguyen and Shultz, 2007). The increased knee valgus in women has often been linked to a greater strain of the iliotibial band (Kim et al., 2021) and greater ACL risk (Hewett et al., 2005).

Considering the anatomical and physiological differences between the sexes, it seems unsurprising they also differ regarding their biomechanics in running. This includes hip and knee kinematics (Ferber et al., 2003; Vannatta et al., 2020) and the gluteus and hamstring muscle activation (Vannatta and Kernozek, 2021). Sex differences in biomechanics and, in particular, knee joint mechanics were also found for jumping (Fuchs et al., 2019) and landing (Jacobs et al., 2007; Jenkins et al., 2017; Quatman et al., 2006) as well as changing of direction (Thomas et al., 2020), squatting, and side-stepping (Mendiguchia et al., 2011; Schmitz et al., 2008).

Only a few studies already linked the sex-specific differences in skeletal morphology and mechanics to the risk of specific types of injuries (e.g., Willy et al. (2012)). Previous studies on movement-related mechanisms contributing to sports injuries often neglected or underestimated the influence of female and male characteristics. However, taking the above-mentioned characteristics of basic movements into account, there is a strong need to consider sex-specific factors in sports biomechanics.

1.2.3 Female Versus Male Response to Resistance and Endurance Training

Current evidence also suggests that male and female athletes respond differently to endurance and resistance training. Moderate to vigorous resistance training generally leads to immediate decreased strength and delayed onset of muscle soreness. This has been observed in both sexes, but in some studies women were more prone to exercise-induced muscle damage (Clarkson and Hubal, 2002; Markus et al., 2021; Morawetz et al., 2020). Also, women are usually less fatigable than men for similar intensity isometric fatiguing contractions, although this can change when the requirements of the task are altered (Hunter, 2014). The magnitude and interaction of mechanisms of sex-based differences in neuromuscular physiology and fatigability are still mostly unknown, but the range of potential influencing factors include physiological (e.g., muscle mass, blood flow, fibre types, hormones) and non-physiological (e.g., bias of studying and reporting or physical activity) mechanisms (Hunter, 2014). Nonetheless, strength improvements and hypertrophy following resistance training are comparable between women and men emphasising the beneficial effects on health and performance in both (Roberts et al., 2020).

Some studies reported sex differences regarding the cardiorespiratory response to long-term endurance training (Howden et al., 2015; Joyner, 2017). While men showed a progressive (almost linear) increase of maximum oxygen uptake throughout the three-month stages of one-year high-load endurance training in the study of (Howden et al., 2015) there was a different response in women: After an immediate increase, measured three months after the start of the endurance training, no further changes were observed during the subsequent stages. Potential underlying mechanisms for the sex-related differences in response to endurance training include hormonal regulations, a suboptimal energy intake, cardiac and pulmonary adaptations as well insufficient recovery or other life-related stressors in female athletes (Archiza et al., 2021; Dominelli et al., 2019; Harms, 2006; Howden et al., 2015; Tipton and Witard, 2007). Nevertheless, most studies showing sex-related differences in the response to resistance and endurance training had a small sample size and/or an unequal distribution of female and male athletes. The prevalence of male bias in studies may mask our current understanding of the magnitude and mechanisms of sex-based adaptations to exercising.

1.3 The Power of Hormones. How Estradiol, Progesterone and Testosterone Regulate Athletic Performance and Injury Risk

Many of the aforementioned sex differences are triggered by changes in hormonal status during adolescence (Handelsman, 2017; Quatman et al., 2008). Although there is already a small gender gap in basic motor skills between young boys and girls, the differences become palpable at the onset of male puberty (Handelsman, 2017; Tomkinson et al., 2018). The swimming, running, jumping, agility, and aerobic fitness abilities rapidly improve in male adolescents while they change less in female adolescents (Handelsman, 2017; Ortega et al., 2011). At the end of puberty, the sex divergence reaches a plateau and remains there subsequently (Handelsman, 2017). This suggests that sex hormones are important for the biology and physiology of the human collagen, muscles, and bones and their contribution to physical performance and movement characteristics (Beynnon et al., 2005; Renstrom et al., 2008).

In female athletes the menstrual cycle-related changes in estradiol and progesterone concentration are associated with changes in the ACL injury risk (Herzberg et al., 2017; Hewett et al., 2007), modifications of the hip, knee, and ankle mechanics as well as changes in aerobic performance (Balachandar et al., 2017; Bell et al., 2014; Hohmann et al., 2015). In their systematic review, Hewett et al. (2007) demonstrated an increased rate of cruciate ligament injuries in the pre-ovulatory phase (high estrogen level) of the menstrual cycle. Another systematic review by Herzberg et al. (2017) confirmed the close relationship between menstrual hormone fluctuations and the incidence of ACL injuries with an increased risk in the pre-ovulatory and ovulatory phases and occasionally in the follicular phase, while the luteal phase is not associated with a higher injury rate. The authors also showed that oral contraceptives tend to have a positive influence on the risk of injury.

Cycle-related hormonal fluctuations also seem to be associated with changes in athletic performance (Meignié et al., 2021). Female football players' endurance is lower in the luteal phase (second half of the menstrual cycle) than in the early follicular phase (starts on the first day of

menstruation) (Janse de Jonge, 2003; Julian et al., 2017). In addition, the ability to run worsens in the luteal phase compared to the early and late follicular phase (Goldsmith and Glaister, 2020). This was explained by an increased core temperature and minute ventilation in the luteal phase (Goldsmith and Glaister, 2020). Female athletes also respond differently to strength training interventions throughout the menstrual cycle (Thompson et al., 2020). Greater effects were observed in the follicular phase compared to the luteal phase emphasising the positive effect of an increased estrogen level on strength training adaptations. The recovery after muscle damage also seemed to be influenced by the different phases of the menstrual cycle as estrogen appears to reduce the extent of muscle damage (Enns and Tiidus, 2010).

However, jump and sprint performances seem to be mostly unaffected by hormonal changes throughout the menstrual cycle (Julian et al., 2017; Meignié et al., 2021; Wiecek et al., 2016).

1.4 An Example of How It Can Go Horribly Wrong: Simpson’s Paradox Using Sex as a Confounder in a Simulated Data Set

As outlined, there are various sex-related differences in sports and movement science. Therefore, including men and women in the study should be a “guiding principle in biomedicine” (Clayton, 2016), and there are many statements encouraging researchers to do so (Springer et al., 2012). If we continue to predominantly include men in our studies, we will frequently disregard half of the relevant population (Wainer et al., 2020), and thereby greatly limit the external validity of the results. Since we do know that men and women respond differently to some treatments or conditions, the application of conclusions derived from male-based studies could lead to non-optimal recommendations for women, preventing optimal performance (Elliott-Sale et al., 2021). On the other hand, as we will demonstrate, including men and women in a study and statistically analysing the aggregated data of both can potentially lead to entirely wrong conclusions. We

will use the well-known (Yule-) Simpson's paradox ¹ (Simpson, 1951; Yule, 1903) in a simulated sports-related data set. We do know that the relative risk of sports injuries can be different in women and men. For example, ACL injuries happen more often in female football players (Arendt and Dick, 1995; Carter et al., 2018). The amount of sports injuries can be reduced with neuromuscular training (Hübscher et al., 2010). Our data set simulates an evaluation study for a neuromuscular training intervention to reduce the amounts of ACL injuries. 10,000 participants are allocated into a group that receives no neuromuscular training (control group: CG) and another 10,000 into a group that does receive neuromuscular training (intervention group: IG). Our example study is badly randomised resulting in an unequal allocation of men and women into the two groups. For the primary outcome it is documented if our virtual participants developed at least one ACL injury. In studies on injury prevention, the incidence rate (usually the number of injuries in 1,000 exposure hours) is frequently used. In our example, for simplicity, we just use the number of participants reporting an injury. The results are displayed in Table 1.1. According to the results of the aggregated data (all participants, not distinguishing between women and men), 3.6% of the participants in CG (360 out of 10,000) and 5.1% in the IG (512 out of 10,000) reported an ACL injury. Since more injuries happened in the IG, we would conclude that the intervention is harmful for our participants. However, if we analyse men only, 3.0% in the CG (270 out of 9,000) and 1.8% in the IG (18 out of 1,000 men) got injured. The intervention reduced the number of ACL injuries for men. The results are quite comparable for women: in the CG 9.0% (90 out of 1,000) and in the IG 5.5% (494 out of 9,000) reported an ACL injury. Taken together, and to quote the title of Baker and Kramer (2001) on the Simpson's paradox: The intervention is "good for women, good for men, bad for people". The Simpson's paradox demonstrates a particular type of confounding (Norton and Divine, 2015). To distort the results, a confounder²

¹The Simpson's paradox is a statistical phenomenon demonstrating the effects of a particular type of confounding. Not accounting for such a confounder might distort the results and might even flip the conclusions completely.

²A confounder is a third variable that is correlated with the independent variable. In our example, sex is the confounder. Men and women are not equally allocated into the intervention and control group (sex is correlated with the independent variable). A confounder also affects the dependent variable. In our

Table 1.1: Simulated data set on Simpson’s paradox: Frequencies of participants depicting an ACL injury (dependent variable, binary) within the intervention (neuromuscular training to reduce the likelihood of an ACL injury) group compared to a control group (no neuromuscular training).

	Injured	Not injured	Injured (%)	p-value (X2-test)	Total N for each group
Aggregated data (Men and women)					
Control group	360	9640	3.6%	<.001	10000
Intervention group	512	9488	5.1%		10000
Men only					
Control group	270	8730	3.0%	0.031	9000
Intervention group	18	982	1.8%		1000
Women only					
Control group	90	910	9.0%	<.001	1000
Intervention group	494	8506	5.5%		9000
Note: In our example, the X ² test is used as a statistical hypothesis test to analyse frequencies in the contingency tables.					

needs to differ between the groups and affect the dependent variable (Norton and Divine, 2015). This is the case in our example. Men and women were not equally allocated to the study groups (e.g., significantly more women in the intervention group) and sex co-varies with the outcome measure (women depict more ACL injuries in both groups). Therefore, sex is a confounder in our simulated data set. In this case it flips the conclusions completely if not accounted for. Such confounding might not only flip the results entirely but also increase or decrease the effects.

Of course, using (simple) randomisation, such an unequal allocation of men and women into the groups is unlikely for big samples. However, in studies with small samples, men and women are fre-

example, ACL injuries occur more often in females. If not accounted for a confounder in a study, spurious statistical results might occur, probably leading to wrong conclusion.

Table 1.2: Effects of neuromuscular training and participants depicting an injury; simulated and aggregated data (men and women together). Notes: injury level “injured” coded as class 1; confidence intervals were omitted to archive better readability.

	Estimate	Standard Error	Odds Ratio	z	Wald statistic	df	p-value
(Intercept)	3.288	0.054	26.778	61.244	3750.856	1	<.001
group (IG)	-0.368	0.070	0.692	-5.237	27.431	1	<.001
Note : An Odds Ratio is an effect size measure. Values less than 1 indicate better effects in the control group.							

quently unevenly allocated into the different intervention/condition groups, probably confounding the results and therefore diminishing the internal validity of the study. Analysing pooled data always bears the potential of misinterpretation even if men and women are equally allocated into the groups. If there is a sex or interaction effect that is not accounted for, the effects might be masked, or the estimated effects might be false (in the size or direction) (Tannenbaum et al., 2019). One possibility is to include the confounder “sex” as a moderating variable in the statistical model. The results of the logistic regression without considering sex as a confounder and accounting for sex are shown in Table 1.2 and Table 1.3, respectively. If the intervention effects differ for men and women, they need to be modelled, too. Including sex in the model also tells us if the effect of the intervention differed between men and women (the interaction effect “group (IG) * sex (m)” in Table 1.3), at least if the study was adequately powered³.

1.5 Considerations For Future Research in Sports and Motor Control

In the previous sections, we concluded that, if possible, we should include men and women in our study, but we also need to take care not to confound our results (to maintain internal validity). Therefore, in

³A study is adequately powered if the sample size is sufficiently large, enabling the study to detect relevant effects if there are any in the population (to avoid false negatives).

Table 1.3: Effects of neuromuscular training and participants depicting an injury; simulated data set; analysis “accounting” for sex differences. Notes: injury level “injured” coded as class 1; confidence intervals were omitted to archive better readability.

	Estimate	Standard Error	Odds Ratio	z	Wald statistic	df	p-value
(Intercept)	2.314	0.110	10.111	20.938	438.403	1	<.001
group (IG)	0.532	0.120	1.703	4.444	19.747	1	<.001
sex (m)	1.162	0.127	3.198	9.182	84309	1	<.001
group (IG) * sex(m)	-0.009	0.273	0.991	-0.034	0.001	1	0.973
Note: An Odds Ratio is an effect size measure. For the independent variable group, values exceeding 1 indicate better effects in the intervention group.							

the following paragraphs, we want to give some considerations for further studies. Since there is no gold standard that fits all cases, we will focus on factorial designs⁴. For a more general view on methodological and statistical issues, we refer to Springer et al. (2012), Beltz et al. (2019) and Rich-Edwards et al. (2018). However, all these examples and recommendations only apply to studies that

- had men and women recruited into the planned study,
- had sex as a subordinate or no factor of interest,
- had outcomes and constructs measured in both men and women.

1.5.1 Research Plan Development

The inclusion of the variable sex should be considered at all stages of a research project, starting with the development of the research plan (Mannon et al., 2020; Springer et al., 2012; Wainer et al., 2020). As discussed in the previous sections, some outcome measures are moderated by sex. Therefore, even if sex differences are not part of the original research question, the researcher should undertake a literature

⁴A factorial design is an experiment with one or more factors (e.g., group or condition variables) as independent variables. Here, factors are categorical variables.

review, searching for possible sound hypotheses concerning sex as a confounder in their study (Mannon et al., 2020; Rich-Edwards et al., 2018; Springer et al., 2012). If there is a theory-based sound hypothesis, this provides a rationale for subgroup analyses (McGregor et al., 2016), and sex should be included as an additional independent factor in the study. Since sex does not explain any underlying mechanisms itself (Springer et al., 2012; Wainer et al., 2020), it would be even better to operationalise the biological and or biosocial mechanisms directly instead of using the proxy sex (Springer et al., 2012). Other third variables that are strongly co-varying with sex (e.g., strength, height, weight) might better explain the underlying mechanisms. If so, using sex as a proxy could result in a loss of information and statistical power. Taken together, if based on the theory that sex (or even better another variable addressing the underlying mechanism directly) might be a confounder, a corresponding hypothesis should be formulated a priori. This gives a rationale for performing subgroup analyses (Aulakh and Anand, 2007; Springer et al., 2012). As a next step plan a control group/condition for your original research question but also include control groups/conditions for possible underlying biological or biosocial mechanisms (Springer et al., 2012). Even more complex, the biological mechanism might be affected by the hormonal milieu especially in women (Elliott-Sale et al., 2021). Hence, gathering data on reproductive stages and cycles could be considered (Rich-Edwards et al., 2018). If no sound hypotheses on sex as a confounder could be deduced from the scientific literature, men and women should be included nevertheless, preferably in equal subsample sizes (Hagstrom et al., 2021) to increase the external validity of the study.

1.5.2 The Optimal Sample Size

If sex is considered as a factor in the research plan, it must also be considered in the a priori power analysis to avoid underpowered designs. If sex-related differences are included in the hypotheses, the sample size needs to be large enough to test the interaction effect of sex with the main intervention/condition with adequate statistical power (Rich-Edwards et al., 2018). Examining the interaction effects is the correct method to determine if the intervention/condition effect differs between men and women (Brookes et al., 2004). While the required sample size will be considerably larger compared to the sample size

needed to detect the main or treatment effect alone (Brookes et al., 2004), this needs to be done if we want to test for sex differences in order to avoid an underpowered design, to avoid false negatives, and therefore to avoid false conclusions on sex differences (Rich-Edwards et al., 2018). If the analyses of sex differences are not a priori planned but conducted post hoc without a sound theory-based hypothesis or with an underpowered study design, this may even “create more noise than light” (Rich-Edwards et al., 2018).

1.5.3 Inclusion Criteria, Randomisation, and Data Collection

As seen in our simulated example on injury prevention (Section 1.4), it might be important to allocate the sexes equally into the study groups/conditions. If sex is a possible confounder, this is needed to distribute the sex effect equally into the groups/conditions to avoid confounding. This can be achieved with a sex-stratified (block) randomisation (Rich-Edwards et al., 2018). Furthermore, the chances of revealing an interaction effect of sex and the intervention/condition are more likely if the subgroups (men and women) are equal in size (Brookes et al., 2004). If the effect of treatment/condition is also moderated by the phases of the menstrual cycle, the researcher might consider standardisation of this aspect using the inclusion criteria or balancing the factor during the randomisation process (Rich-Edwards et al., 2018). In the study report, the research should mention if sex or gender was measured and how this information was assessed (Rich-Edwards et al., 2018).

1.5.4 Statistical Data Analysis and Results

There are different ways to conduct the statistical data analysis (Beltz et al., 2019). We will outline just one possibility for a factorial design. Other statistical tools might be needed if the designs are more complicated. In the best-case scenario, there is an a priori sound hypothesis for sex-related differences, and the study is appropriately powered (sample is large enough to provide sufficient statistical power) to identify relevant interaction effects of the main outcome with sex. Examining the interaction effects is the correct method to determine if the intervention/condition effect differs between men and women