Prediction and Structure of Triathlon Performance in Recreational and Elite Triathletes

Zur Erlangung des akademischen Grades eines
DOKTORS DER PHILOSOPHIE (Dr. phil.)

von der KIT-Fakultät für Geistes- und Sozialwissenschaften des
Karlsruher Instituts für Technologie (KIT)
angenommene

DISSERTATION

von
Marian Hoffmann

KIT-Dekan: Prof. Dr. Michael Schefczyk

1. Gutachter: Prof. Dr. Thorsten Stein
2. Gutachterin: apl. Prof. Dr. Ilka Seidel

Tag der mündlichen Prüfung: 12. März 2021

Acknowledgments

Die vorliegende Dissertation entstand im Rahmen meiner Tätigkeit als akademischer Mitarbeiter am BioMotion Center des Instituts für Sport und Sportwissenschaft des Karlsruher Instituts für Technologie (KIT).

In erster Linie möchte ich meinem Doktorvater Prof. Dr. Thorsten Stein für die hervorragende wissenschaftliche Unterstützung und die intensive Förderung und Begleitung meiner beruflichen Entwicklung herzlich und aufrichtig danken. Für die kompetente und wissenschaftlich anspruchsvolle Unterstützung bei den Publikationen, der Methodik und insbesondere der Möglichkeit zur Kooperation mit dem Institut für Angewandte Trainingswissenschaft Leipzig (IAT) möchte ich meiner Zweitgutachterin, Prof. Dr. Ilka Seidel, sehr herzlich danken.

Weiterhin danke ich Dr. Thomas Moeller für die gemeinsame Arbeit, insbesondere für seine wertvolle Perspektive als DTU Bundestrainer Nachwuchs, die stets lehr- und hilfreich war. Für die Ermöglichung der Promotion danke ich ebenfalls herzlich Prof. Dr. Klaus Bös.

Ein herzlicher Dank gilt meinen Kolleginnen und Kollegen des BioMotion Center für die bereichernde interdisziplinäre Zusammenarbeit, die berufliche und private Unterstützung und die jederzeit angenehme Arbeitsatmosphäre. Zudem möchte ich meinen Kolleginnen und Kollegen des Instituts für Sport und Sportwissenschaft und des Zentrums für Lehrerbildung für das angenehme Arbeitsklima und den stets abwechslungsreichen und bereichernden Arbeitsalltag danken. Für die entstandene Freundschaft und die tollen Momente, auch außerhalb der Arbeit, möchte ich Gunther Kurz, Steffen Ringhof, Frieder Krafft, Bernd Stetter und Felix Möhler danken.

Der bedingungslosen und liebevollen Unterstützung meiner Eltern und meiner Geschwister bin ich unendlich dankbar. Der Zuspruch, das Verständnis und der nötige Rückhalt, nicht nur während der Promotion, sondern in allen Phasen meines Lebens ist nicht in Worte zu fassen.

Mein größter Dank gilt meiner Ehefrau Melanie: für deine unermüdliche Geduld, dein Vertrauen in meine Fähigkeiten, deine Liebe und deine Unterstützung.

Für unsere Tochter Lorena

Summary

The sport of triathlon comprises the three classic endurance sports swimming, cycling and running, which are performed consecutively without a break and lead to an overall race time. The Olympic distance of 1.5 km swimming, 40 km cycling and 10 km running is the most common race distance, in both the amateur and professional fields. As a highly endurance-determined sport, triathlon, as with the three constituent disciplines, requires specific physiological requirements. Other fields also determine performance, such as the anthropometry of an athlete, psychological requirements and much more. In addition to the Olympic distance, there are also a shorter sprint distance and the longer half and long distance – each involving specific characteristics that can be practiced through adapted training programs.

The determination of such performance-relevant parameters of a sport or athletic performance are summarized in the field of training science as the performance structure of a sport and build up the basis for scientifically-founded statements on training programs, talent selection and more. Closely related is the prediction of performance, based on the identified parameters and their quantification by means of a current performance diagnostic. This allows prediction of the actual race performance, for example in terms of the overall race time. The combination of these two aspects – the prediction and the structure of performance in triathlon – form the core of the present thesis, whereby both amateur and professional athletes were analyzed.

The present thesis consists of eight chapters. After a short preface and a general introduction to the topic in Chapter 1, Chapter 2 provides the theoretical and methodological background. In particular, the peculiarities, boundary conditions and prerequisites of triathlon, the current state of research in the areas of prediction and structure of performance as well as the methodological approaches used in this thesis are examined in detail. Since the use of different computational methods is an important part of the thesis, their application within the three studies (Chapters 4 to 6) is discussed in more detail. Exploratory factor analysis and dominance paired comparison are applied as procedures for the preselection of performance-relevant parameters, multiple linear regression and artificial neural networks for the prediction

of individual overall race time as well as structural equation analysis as a method for building up structural models of the performance in triathlon.

After the derivation of the research questions and the description of the objectives of the present thesis (Chapter 3), the studies described in the three following chapters provide explanatory approaches.

The study in Chapter 4 provides initial explanations and demonstrates performance-relevant parameters that are used to predict the individual race performance of recreational triathletes over the sprint distance. Anthropometric, physiological and training-related parameters were recorded as part of performance diagnoses under laboratory conditions immediately before a triathlon competition, and statistical relationships were established with regard to overall race performance. Three performance prediction models were computed using linear regression and performance-relevant parameters could be identified thereby. The model based on physiological parameter blood lactate concentration after 18 min at 200 W on cycling ergometer delivers the highest explanation of variance ($R^2 = 0.71$), followed by the model based on anthropometric parameters leg length and arm span ($R^2 = 0.67$) and the model based on training-related parameter training volume in swimming ($R^2 = 0.41$). Overall, it has been shown that performance prediction is possible even with small samples and that it can provide information on the design of training programs and the individual race strategy, associated with a very limited generalizability, especially in the amateur field. A challenge in larger studies is likely to be the comparable investigation of overall race time as the dependent variable.

On this basis, the study in Chapter 5 examines the prediction of overall race times of elite triathletes over the Olympic distance. The routine performance diagnoses of triathletes, which were tested by the Institute for Applied Training Science in Leipzig in preparation for the Summer Olympics in 2012, were analyzed and used to calculate performance prediction models. The high degree of standardization of tests with a large number of recorded parameters conflicted with the need to normalize the overall race times. This was necessary because the elite triathletes participated in different triathlon races, mostly over the same race distance, but with different route profiles, starting grids, climate conditions, etc. In comparison to previous research literature, two different approaches were used for the prediction models based on anthropometric and physiological parameters: multiple regressions for linear relationships and artificial neural networks for non-linear relationships between parameters and overall race time. Both approaches yielded two prediction models. Linear regression provided $R^2 = 0.41$ in case

of anthropometric variables (predictive: pelvis width and shoulder width) and $R^2 = 0.67$ in case of physiological variables (predictive: maximum respiratory rate, running pace at 3-mmol·L^{-1} blood lactate and maximum blood lactate). The Artificial neural networks using the five most important variables after preselection yielded $R^2 = 0.43$ in case of anthropometric variables and $R^2 = 0.86$ in case of physiological variables. The advantage of neural networks over linear regressions was the possibility to take non-linear relationships into account. In contrast to the study carried out with recreational triathletes, the elite triathletes represent a very homogeneous sample that comes very close to the population of German elite national squad athletes. This is why the results and in particular the identified performance-relevant parameters are more generalizable, albeit for a very small group of athletes. In particular, to deduce important characteristics for athletes in junior squads, the results provide valuable information on potentially relevant anthropometric requirements as well as for performance-relevant physiological parameters that can be influenced by training.

The third study (Chapter 6) uses the results of the prediction models created in Chapter 5 to develop a structural model of the performance in triathlon over the Olympic distance, despite the small sample. Finally, three valid models were computed, which provide an important first step towards a scientifically-founded clarification of the performance structure in Olympic-distance triathlon. In particular, one model (that uses the experience of professional triathlon coaches in the preselection of parameters) delivers parameters that can be classified as good, and which are in accordance with the findings of the prediction models and the structural model based on theoretical considerations. Parameters classified as relevant are both anthropometric (body weight, BMI, lean body mass) and physiological (relative maximum oxygen uptake, running speed at 3 mmol/l blood lactate, maximum running speed in a specific mobilization test). While working with data from elite athletes, the use of a small sample must be mentioned as a limitation, as this can be a disadvantage when calculating structural models. The developed models are clearly defined from a mathematical and statistical point of view, but must be supplemented by further data to create more comprehensive models.

Finally, Chapter 7 provides a general discussion of the research results and an outlook for future studies. The findings of the three studies carried out are merged and compared with the current state of research for a comprehensive consideration of performance-relevant parameters of triathlon as well as of the methodological approaches used (multiple regression, artificial neural networks and structural analysis). The present thesis essentially analyzes

parameters that have already been identified as performance-relevant in research literature, but also performance parameters that have to be classified as relevant. A further major finding of the thesis is the application of the applied methods in the context of training-based performance diagnoses, as this has not yet been widely done. Due to the limitation of the small samples and thus data sets, which is unavoidable while working with elite athletes, there is clear potential for future studies and therefore an exciting and significant future research field.

The thesis concludes with a general summary (Chapter 8) of the present work.

Zusammenfassung

Die Sportart Triathlon kombiniert die drei Ausdauersportarten Schwimmen, Radfahren und Laufen, die nacheinander ohne Pause ausgeführt werden und in eine Gesamtwettkampfzeit münden. Die Olympische Distanz über 1.5 km Schwimmen, 40 km Radfahren und 10 km Laufen stellt die am meisten verbreitete Wettkampfdistanz dar, sowohl im Amateur- als auch im Profi-Bereich. Als Ausdauer determinierte Sportart erfordert Triathlon, wie auch die drei Einzeldisziplinen, spezifische physiologische Anforderungen. Zahlreiche weitere Bereiche wie die Anthropometrie der Athletinnen und Athleten, psychologische Voraussetzungen und vieles mehr können ebenfalls leistungsdeterminierend sein. Es ist davon auszugehen, dass selbst die einzelnen Streckenlängen im Triathlon – neben der Olympischen Distanz existieren noch die kürzere Sprint- sowie die längere Halb- und Langdistanz – jeweils spezifische Charakteristika mit sich bringen, die durch eine angepasste Vorbereitung und Trainingsgestaltung vorbereitet werden können.

Die Bestimmung solch leistungsrelevanter Parameter einer Sportart oder einer sportlichen Leistung werden im Bereich der Trainingswissenschaft als Struktur der sportlichen Leistung zusammengefasst und bilden die Grundlage für wissenschaftlich fundierte Aussagen zur Trainingsgestaltung, Talentauswahl und vielem mehr. Eng damit verbunden ist die Prognose sportlicher Leistung, die auf Basis der identifizierten Parameter und deren Quantifizierung mittels einer aktuellen leistungsdiagnostischen Untersuchung eine Prognose der tatsächlichen Wettkampfleistung bspw. in Form einer Gesamtwettkampfzeit ermöglichen kann. Die Verknüpfung dieser beiden Aspekte – die Prognose und die Struktur der sportlichen Leistung in der Sportart Triathlon – bilden den Kern der vorliegenden Dissertation, wobei sowohl Amateur- als auch Profi-Sportler in den Fokus genommen wurden.

Die Dissertation umfasst acht Kapitel. Nach einem kurzen Vorwort und einer allgemeinen Einführung in die Thematik (Kapitel 1) liefert Kapitel 2 den theoretischen und methodischen Hintergrund. Insbesondere werden die Besonderheiten, Rahmenbedingungen und Voraussetzungen der Sportart Triathlon, der aktuelle Forschungsstand in den Bereichen der Prognose und Struktur sportlicher Leistung sowie die in dieser Thesis verwendeten methodischen Ansätze näher beleuchtet. Da der Einsatz unterschiedlicher Methoden ein

wichtiger Bestandteil dieser Arbeit darstellt wird deren Einsatz in den drei Studien (Kapitel 4 bis 6) ausführlicher vorbereitet: die explorative Faktorenanalyse und der Dominanz-Paar-Vergleich als Verfahren zur Vorselektion leistungsrelevanter Parameter, die multiple lineare Regression und künstliche neuronale Netze zur Prognose der individuellen Gesamtwettkampfzeit sowie die Strukturgleichungsanalyse als Verfahren zur Berechnung eines Strukturgleichungsmodells der sportlichen Leistung im Triathlon.

Nach der Ableitung der Fragestellungen und der Darstellung der Ziele der vorliegenden Thesis (Kapitel 3), liefern die Forschungsarbeiten in den drei darauffolgenden Kapiteln Erklärungsansätze hierzu. Die Studie in Kapitel 4 liefert erste Erkenntnisse und weist Leistungsparameter nach, die zur Prognose der individuellen Wettkampfleistung von Amateur-Triathleten über die Sprintdistanz dienen. Hierbei wurden anthropometrische, physiologische und trainingsbezogene Parameter im Rahmen einer Leistungsdiagnostik unter Laborbedingungen unmittelbar vor einem Triathlon Wettkampf erfasst und statistische Zusammenhänge zur erbrachten Wettkampfleistung hergestellt. Drei Modelle zur Prognose der Wettkampfleistung konnten mittels linearer Regression berechnet und dabei leistungsrelevante Parameter identifiziert werden. Das auf dem physiologischen Parameter Blutlaktatkonzentration nach 18 min bei 200 W auf einem Fahrradergometer aufbauende Prognosemodel liefert die höchste Varianzaufklärung ($R^2 = 0.71$), gefolgt von den Modellen basierend auf den anthropometrischen Parametern Beinlänge und Armspannweite ($R^2 = 0.67$) und dem trainingsbezogenen Parameter Trainingsumfang im Schwimmen ($R^2 = 0.41$). Nachgewiesen werden konnte, dass dies selbst bei kleinen Stichproben möglich ist und Hinweise zur Trainingsgestaltung und zur Wettkampfeinteilung liefern kann, insbesondere im Amateur-Bereich jedoch mit einer stark eingeschränkten Generalisierbarkeit verbunden sein dürfte. Eine Herausforderung bei größeren Studien dürfte daher die vergleichbare Erfassung der Gesamtwettkampfzeit als abhängige Variable darstellen.

Die Studie in Kapitel 5 untersucht auf Basis der vorangegangenen Erfahrungen die Prognose der Gesamtwettkampfzeit von Profi-Triathleten über die olympische Distanz. Hierbei wurden die routinemäßig durchgeführten leistungsdiagnostischen Untersuchungen von Triathleten, die in der Vorbereitung auf die olympischen Sommerspiele im Jahr 2012 durch das Institut für Angewandte Trainingswissenschaft in Leipzig getestet wurden, analysiert und für die Berechnungen der Prognosemodelle verwendet. Dem hohen Maß an Standardisierung der Testungen mit einer großen Anzahl an erfassten Parametern stand die Notwendigkeit der

Normalisierung der Gesamtwettkampfzeiten gegenüber. Dies war notwendig, da die Profi-Triathleten an unterschiedlichen Wettkämpfen teilgenommen hatten, zwar überwiegend über dieselbe Streckenlänge jedoch mit unterschiedlichsten Streckenprofilen, Teilnehmerfeldern, klimatischen Bedingungen, etc. Im Vergleich zur bisherigen Literatur konnten mit zwei unterschiedlichen Ansätzen – multiple Regressionen für lineare und künstliche neuronale Netze für nichtlineare Zusammenhänge zwischen Parametern und Gesamtwettkampfzeit – gute Ergebnisse für Prognosemodelle auf Basis anthropometrischer und physiologischer Parameter erzielt werden. Beide Ansätze lieferten je zwei Prognosemodelle. Die lineare Regression führt zu $R^2 = 0.41$ auf Basis anthropometrischer Parameter (prädiktiv: Beckenbreite und Schulterbreite) und zu $R^2 = 0.67$ auf Basis physiologischer Parameter (prädiktiv: maximale Atemfrequenz, Laufgeschwindigkeit bei 3-mmol·L^{-1} Blutlaktatkonzentration und maximale Blutlaktatkonzentration). Basierend auf den jeweils fünf relevantesten Parametern einer Vorselektion führen künstliche neuronale Netze zu $R^2 = 0.43$ auf Basis anthropometrischer Parameter und $R^2 = 0.86$ auf Basis physiologischer Parameter. Der Vorteil neuronaler Netze gegenüber der linearen Regression liegt dabei in der Möglichkeit nichtlineare Zusammenhänge abzubilden. Im Gegensatz zur durchgeführten Studie mit Amateur-Triathleten stellen die Profi-Triathleten eine sehr homogene Stichprobe dar, die der Grundgesamtheit des deutschen Nationalkaders sehr nahekommt, weshalb die Ergebnisse und insbesondere die identifizierten Leistungsparameter eine höhere Generalisierbarkeit aufweisen, wenn auch für einen sehr kleinen Kreis an Athleten. Insbesondere zur Ableitung von wichtigen Merkmalen für Athletinnen und Athleten in Nachwuchskadern liefern die Ergebnisse wertvolle Hinweise auf potentiell relevante anthropometrische Voraussetzungen sowie auf leistungsrelevante und durch Training beeinflussbare physiologische Parameter.

Die dritte Studie (Kapitel 6) nutzt die Ergebnisse der erstellten Prognosemodelle aus Kapitel 5, um trotz des Vorhandenseins einer kleinen Stichprobe ein Strukturmodell der sportlichen Leistung im Triathlon über die olympische Distanz zu entwickeln. Hierbei konnten schlussendlich drei gültige Modelle erstellt werden, die einen ersten wichtigen Schritt zur wissenschaftlich fundierten Aufklärung der Leistungsstruktur im olympischen Triathlon liefern. Insbesondere das Modell, das die Erfahrung von professionellen Trainern in der Vorauswahl an Parametern nutzt, liefert als gut einzustufende Modellparameter, die im Einklang zu den Erkenntnissen der zuvor erstellten Prognosemodelle und des Strukturmodels basierend auf theoretischen Überlegungen und einschlägiger Literatur stehen. Als relevant einzustufende Parameter sind hier sowohl anthropometrische (Körpergewicht, BMI, fettfreie

Körpermasse) als auch physiologische (relative maximale Sauerstoffaufnahme, Laufgeschwindigkeit bei 3-mmol/l Blutlaktatkonzentration, maximale Laufgeschwindigkeit in einem spezifischen Mobilisationstest) Kenngrößen zu nennen. Als Limitation ist der Einsatz einer kleinen Stichprobe zu nennen, da dies bei der Berechnung von Strukturmodellen von Nachteil sein kann, bei der Verwendung von Daten von Profi-Athleten jedoch unvermeidbar ist. Die entwickelten Modelle sind aus mathematisch-statistischer Sicht eindeutig bestimmt, müssen jedoch durch weitere Datensätze ergänzt werden, um umfassendere Modelle zu ermöglichen.

Kapitel 7 liefert schließlich eine allgemeine Diskussion der Forschungsergebnisse und einen Ausblick auf zukünftige Studien. Die Befunde der drei durchgeführten Studien werden zusammengeführt und mit dem bisherigen Wissensstand abgeglichen, um eine umfassendere Betrachtung leistungsrelevanter Parameter der Sportart Triathlon sowie der eingesetzten methodischen Ansätze der multiplen Regression, künstlicher neuronaler Netze sowie der Strukturanalyse vorzunehmen. Die vorliegende Dissertation liefert im Wesentlichen sowohl in der Fachliteratur bereits als leistungsrelevant identifizierte Parameter aber auch bisher weniger betrachtete jedoch als potentiell relevant einzustufende Leistungsparameter. Als wesentliches Ergebnis der Dissertation muss der Einsatz der angewendeten Methoden im Kontext der trainingswissenschaftlichen Leistungsdiagnostik gesehen werden, da dies bisher wenig verbreitet ist. Wissend um die Einschränkung kleiner Stichproben, welche im Profi-Bereich unvermeidbar sind, werden die möglichen Potentiale für zukünftige Studien deutlich und zeigen somit ein spannendes und bedeutsames zukünftiges Forschungsfeld und Implikationen für sich anschließende Studien auf.

Die Dissertation schließt mit einer allgemeinen Zusammenfassung (Kapitel 8) der vorliegenden Arbeit.

Table of Contents

List of Figures

List of Tables

1 General Introduction

1.1 Preface

Triathlon is a classic endurance sport that consists of the disciplines swimming, cycling and running, with individual events that vary greatly in distance (sprint-, short-, middle- and long-distance). Over the last decade, triathlon has been a fast growing sport: around 58,000 athletes are now members of a German triathlon club (+113% from 2007 to 2017). In Germany in 2017, around 630 events took place with more than 2,000 individual races, with a total of over 270,000 participants (Deutsche Triathlon Union e.V., 2018). For those involved, high training loads are required, which incorporate 9 to 12 training sessions for more than 20 hours per week (Friel & Vance, 2013) if a respectable position in a triathlon race is to be attained. This load is independent of the athletes' performance level - recreational or elite - and mainly independent of the preferred race distance. Many triathletes therefore face (at least some of) the following questions: which training session leads to the greatest effect regarding my preferred race distance? Which contributing factors will be addressed and are therefore performance-relevant? How fast should I tackle the next triathlon competition?

A well-founded and structured training program should consider the answers to these questions and optimally support the high levels of training. Consequently, it is important to identify performance-relevant parameters, such as anthropometric, physiological or psychological parameters, as well as, for example, training extent (Hottenrott & Seidel, 2017; Landers, Blanksby, Ackland, & Smith, 2000; Schabort, Killian, St Clair Gibson, Hawley, & Noakes, 2000), as a scientific basis for a well-structured training program. The collection of data could be obtained through either laboratory (Basset & Boulay, 2000; Van Schuylenbergh, Eynde, & Hespel, 2004) or field tests (Marongiu et al., 2013), although the former allows a more standardized procedure. Both methods can provide a dataset as the fundament for the three key steps of hierarchization based on theory, internal order and prioritization (Letzelter & Letzelter, 1982). This will be the cornerstone for predicting individual performance or for clarifying the performance structure of triathlon.

To achieve the aim of prioritization of performance-related parameters, extensive statistical analyses are necessary. Commonly-used methods are the computation of correlations between single parameters and performance, and multiple regression analyses. These approaches are able to deliver scientific-based information about performance-related parameters. Nevertheless, for a deeper understanding of performance, complex modeling processes are necessary (Silva et al., 2007).

For this purpose, the present thesis investigates approaches for both performance prediction and performance structure, for both recreational and elite triathletes. Different computational approaches of multiple linear regression analyses, artificial neural networks and structural equation models are investigated. A profound understanding of the possibilities and limitations of the mentioned statistical methods in the context of performance prediction and performance structure could help future researchers in this field and coaches to develop training programs and conduct talent diagnostics in recreational and elite triathlon. Notwithstanding the above, effective statistical methods should be transferable to other athlete cohorts and for use by other sports researchers.

1.2 Outline of the thesis

The current thesis covers eight chapters, including three research studies. Chapter 2 provides the theoretical and methodological background of triathlon and the prediction and structure of performance. In particular, the current state of research about performance-relevant parameters and the computational approaches are reviewed. In Chapter 3, unresolved research issues are deduced to derive the aims and scope of the present thesis.

The three subsequent chapters (Chapter 4, Chapter 5 and Chapter 6), which address the research questions, were partially published in international peer-reviewed or applied journals.

- Chapter 4: Predictive variables of short course triathlon performance in recreational triathletes.
- Chapter 5: Predicting Elite Triathlon Performance: A Comparison of Multiple Regressions and Artificial Neural Networks
  Hoffmann, M., Moeller, T., Seidel, I., Stein, T. (2017). Predicting Elite Triathlon Performance - A Comparison of Multiple Regressions and Artificial Neural Networks.

International Journal of Computer Science in Sport, 16 (2), 101–116. https://doi.org/10.1515/ijcss-2017-0009

- Chapter 6: Modeling the Performance Structure of Elite Triathlon: A Structural Equation Approach.
  Hoffmann, M., Seidel, I., Stein, T. (2016). Aspekte der Leistungsstruktur in der Sportart Triathlon. *Leistungssport*, 46 (5), 9–13.

Finally, Chapter 7 gives an overall discussion and conclusion of the main findings of the presented work. Implications and recommendations for future research are also provided. A general conclusion closes the thesis (Chapter 8).

2 Theoretical Background

This section provides an introduction to the sport of triathlon and the field of performance diagnostics as part of sport training science. Furthermore, it will summarize research into the prediction of triathlon performance, the performance structure of a sport as well as the potential computational approaches of linear regressions, artificial neural networks and structural equation modeling. This serves as the theoretical background for the present thesis.

2.1 Triathlon

Triathlon comprises the three classic endurance sports swimming, cycling and running over four different race distances (Table 2.1). The transitions between swimming and cycling as well as between cycling and running are part of the race and are therefore included in the overall race time.

Table 2.1 Triathlon race distances (km).

Distance	Swim	Bike	Run
Sprint	0.5	20	5
Short (Olympic)	1.5	40	10
Middle	1.9	90	21.095
Long	3.8	180	42.195

Held for the first time in San Diego (USA) in 1974, triathlon has experienced a rapid development over the past decades. The short-distance triathlon has probably contributed most to this progress since it became part of the Olympic Games in 2000 (Millet, Bentley, & Vleck, 2007). In Germany, the number of members of the umbrella organization Deutsche Triathlon

Union e.V (DTU) has more than doubled since 2000 (Deutsche Triathlon Union e.V., 2018). In 2017, around 630 events took place in Germany with more than 2.000 competitions, with a total of over 270.000 participants (Deutsche Triathlon Union e.V., 2018).

This development may be due to the various motives of recreational athletes: Lembeck, Starringer, and Schönfelder (2009) identified the primary motives among recreational triathletes as the pleasure of endurance sport, a balance to work and daily life as well as the health-enhancing aspects. Beyond this recreational and health-enhancing focus, athletes could pursue the sport competitively, meaning they have to deal with different conditions. One major difference in elite triathlon over the Olympic distance is that drafting is allowed, meaning swimming or cycling in the slipstream of another athlete, which results in faster split times. This factor contributes to average overall race times in male elite Olympic-distance triathletes of 1h45 to 2h00 (Fröhlich, Klein, Pieter, Emrich, & Gießling, 2008), depending on the route profile, weather, etc. Split times in swimming of 17 to 19 minutes, 50 to 55 minutes in cycling and 30 to 32 minutes in running are necessary for a position in the leading group (Fröhlich et al., 2008; Millet & Vleck, 2000). The percentage distribution of race time between the three disciplines indicates a focus on cycling (55 % of overall race time), followed by running (29 %) and swimming (15 %); the 1 % left is needed for both transitions (Landers, Blanksby, Ackland, & Monson, 2008). This distribution gets even clearer with a look on at the annual training amounts of elite athletes: 1,000 to 1,250 km swimming (around 7 %), 10,000 to 12,500 km cycling (around 72 %) and 2,800 to 4,000 km running (around 21 %) per year (Fröhlich et al., 2008). The aforementioned drafting effect in swimming and cycling, with the consequent energy saving and other tactical possibilities (Chatard & Wilson, 2003; Hausswirth, Lehénaff, Dréano, & Savonen, 1999; Millet & Bentley, 2004), lead to a reduced significance of the bike discipline and an increase in importance of the running discipline (Bentley, Millet, Vleck, & McNaughton, 2002; Vleck, Burgi, & Bentley, 2006). Fröhlich et al. (2008) even showed that running split times under 30 minutes after the two other disciplines are common in elite Olympic-distance triathlon, which all results in a win-critical function of the run.

2.2 Performance prediction

This section investigates how performance of a sport which combines three disciplines, with many influencing factors, can be predicted and why this would be important. This should not be confused with classic antagonistic models, e.g. PerPot DoMo, in training science (Perl &

Pfeiffer, 2011), which predict the effects of training on athletic performance (Hottenrott & Seidel, 2017).

Landers et al. (2000) stated that the identification of attributes predicting performance would be important to create more specific training programs and to differentiate between talent identification programs. These two aspects are important benefits for identification of the performance structure, supporting tactical decisions and can help to develop a sport discipline:

- More specific training programs can help to improve the quality of training. Especially in a training-intensive sport such as triathlon, more specific training sessions can improve the quality of training instead of its quantity. This gets even clearer if one considers that elite triathletes normally train about 1.500 hours per year (Pfützner, 1997). Even recreational triathletes often undergo 9 to 12 training sessions per week (Friel & Vance, 2013) which means up to 15 hours per week (Rüst, Knechtle, Knechtle, Rosemann, & Lepers, 2011).
- Talent identification in triathlon often uses time trial tests in swimming and running, which is not appropriate when being used as sole criteria for selection (Bottoni, Gianfelici, Tamburri, & Faina, 2011). More relevant parameters of triathlon race performance need to be identified and combined in complex models, such as made by Bottoni et al. (2011). Such models can point out helpful variables, maybe to direct young athletes with a beneficial genotype into the sport (Landers et al., 2000) or to define the minimum requirements of certain physiological factors.
- As described by Vleck et al. (2006), tactical aspects also affect contemporary elite Olympic-distance triathlon. Besides the drafting aspect during cycling (and partially during swimming) as mentioned in section 2.1, the individual tactics and position in each discipline seem to affect the overall race result (Vleck et al., 2006). Therefore, knowledge about the predicted individual race performance based on the athlete's last performance diagnosis can help to make appropriate tactical decisions in or prior to a race.

With a view to these three aspects, a first starting point in literature is the identification of the importance of each discipline (Fröhlich et al., 2008; Landers et al., 2008; Vleck et al., 2006). It actually seems that running performance has the biggest influence on overall race performance in elite Olympic-distance triathlon, as described in section 2.1. Thereby, it has to be mentioned that this aspect cannot be generalized due to the specific regulations of this race

format. Nonetheless, endurance running or running in triathlon serve as an example where many studies emphasized the importance of maximum oxygen uptake (VO_2max) and anaerobic thresholds (Millet, Vleck, & Bentley, 2009, 2011). Moreover, these parameters show significant correlations to race performance (Bassett, 2000; McLaughlin, Howley, Bassett, Thompson, & Fitzhugh, 2010). Similar results were found for swimming and cycling (Millet et al., 2009; Sleivert & Rowlands, 1996). However, these variables normally have a prerequisite function instead of a performance predictor in homogenous samples, because of the small variation between athletes (Bassett, 2000; Sleivert & Rowlands, 1996; Stratton et al., 2009).

Besides physiological factors, it has already been shown that anthropometric variables, such as percent body fat, body mass index (BMI) or the circumferences of several parts of the body, could be important for performance in triathlon races (Knechtle, Wirth, Rüst, & Rosemann, 2011) and possibly in terms of performance prediction. In addition, blood lactate concentrations from treadmill or cycle ergometer tests have already been identified as useful parameters in predicting triathlon performance independent from athletes' performance level (Schabort et al., 2000; Van Schuylenbergh et al., 2004).

On this basis, specific prediction models have been developed in the past. Most researchers up to now used linear regression models to predict triathlon race performance (Schabort et al., 2000; Van Schuylenbergh et al., 2004). Schabort et al. (2000) used multiple linear regressions of physiological parameters to predict overall Olympic-distance triathlon race times in the South African national team, and found a highly significant correlation between predicted and actual race time ($r = 0.90$, $p < 0.001$). Multiple regression analysis ($R^2 = 0.98$; SEE = 0.95 [min]) was also used by Van Schuylenbergh et al. (2004) to predict sprint-distance triathlon performance of male physical education students. In these two studies, the subjects competed in the same triathlon competition, which most likely led to the high explanation of variance (R^2) because of the comparable conditions. Nonetheless, this kind of experimental design is rarely possible with elite triathletes due to their individual calendar.

More complex computational approaches are rarely found, even though artificial neural networks (ANNs) could be a useful alternative approach for performance prediction. Edelmann-Nusser (2005), Edelmann-Nusser, Hohmann, and Henneberg (2002) as well as Silva et al. (2007) showed that this could be a valuable method for performance modeling and a good approach without restrictions regarding distribution and independence of variables. Edelmann-Nusser et al. (2002) for example predicted the 200 m backstroke time of an elite female

swimmer in the finals of the Olympic Games 2000 very precise (error of prediction: 0.05 s) using artificial neural networks (multi-layer perceptrons) based on collected training data. The accurate results of this approach were attributed to the fact that “the adaptive behavior of the system athlete is quite a complex, non-linear problem“ (Edelmann-Nusser et al., 2002). However, multiple linear regression analyses are more often used to develop prediction models. Whereas linear regressions need linear relationships between independent variables and a dependent variable, artificial neural networks can handle non-linear relationships based on a different model architecture. Nevertheless, ANNs are rarely used to predict race performance, possibly because the network design requires substantial input concerning the number of neurons and layers, training algorithms etc. (Zhang, Eddy Patuwo, & Y. Hu, 1998).

Besides, nearly all computational approaches have a major problem while working with measurement data from elite athletes: large numbers of independent variables require many sets of data, which are not always provided while working with elite athletes. Therefore, a preselection of parameters is necessary to reduce the number of independent variables.

In summary, the prediction of overall individual race time in triathlon competition using several performance parameters, as well as different computational approaches, has not been fully investigated in all aspects.

2.3 Performance structure

With respect to Hottenrott and Seidel (2017) the term performance structure is used in a narrow sense within this thesis, which means that performance prerequisites and factors are of interest. Internal and external conditions, which are also important for triathlon performance, are part of the competition structure and therefore cannot be modeled. Also, the structure of training as a process of stress, strain and adaption (Hottenrott & Seidel, 2017) is not discussed. This is because such models mainly describe the effects of training on performance prerequisites and the necessary content, resources and methods of training - instead of the identification and prioritization of performance-relevant parameters.

The need for scientific-based knowledge about performance-relevant parameters is undisputed for many sports (Hottenrott & Seidel, 2017, p. 67). Content-related specification and the empirical cause-and-effect relationship regarding performance prerequisites and factors on sports performance build the fundament for scientific-based recommendations for training

programs (Hottenrott & Seidel, 2017, p. 67). The necessary data for such structural analyses of sports performance can only be captured under difficult conditions (Hottenrott & Seidel, 2017), which could be the main reason for the above mentioned research gap in training science. Up to now, many correlation studies have been conducted (Knechtle et al., 2011; Miura, Kitagawa, & Ishiko, 1997; Zhou, Robson, King, & Davie, 1997) to link performance prerequisites with sports performance. These studies often used physiological or sometimes physical variables. However, the identification of performance-relevant parameters is not enough: their specific impact and effect on performance are of interest within structural analyses. Therefore, more complex models and computational approaches are needed to compensate for the drawback of correlations just linking one single variable to performance.

The term "performance structure", which was first characterized by Letzelter and Letzelter (1982) and Hohmann and Brack (1983), describes the situation where the performance prerequisites influencing a sport are identified and prioritized using statistical methods. Accordingly, uncovering the performance structure can provide a scientific basis for training programs and adjustments to them when necessary. Letzelter and Letzelter (1982) propose the idea that structuring the performance of a sport is one of the main objectives of training science, and follows three fundamental and irreversible steps of hierarchization based on theory, internal order and prioritization:

- Hierarchization means the specification of performance-relevant parameters and characteristics as well as structuring them in a model with different stages with decreasing complexity.
- Internal order means the determination of relationships and interactions within and over the stages of the model.
- Prioritization means to highlight relevant performance prerequisites and factors.

When modeling the performance structure of a sport, the sections physical condition and constitution, skills, tactical thinking and mental abilities are of special interest. Also, age- and gender-specific models should be created, even when collecting these data is difficult (Hottenrott & Seidel, 2017). It becomes clear that statistical models of the performance structure of a sport cannot be permanent because single elements or factors of a model can change over time, which requires revision of parts of the model (Hottenrott & Seidel, 2017).

Within this context, structural equation modeling (SEM) is of special interest for uncovering the performance structure, since this computational approach involves the steps of hierarchization, internal order and prioritization. Therefore, SEM delivers considerably more information about the performance structure than correlations or regression models, and makes it possible to identify indicators that explain race performance through a more complex modeling process. SEM, which was introduced in the field of social and behavioral science (Hox & Bechger, 1998), merged three historically statistical traditions: path analysis, simultaneous-equation models and factor analysis (Rosseel, 2012). Today, the areas of application of SEM are diverse, including psychology, political science, education, business-related disciplines (Jais, 2007) and sport science (Felser, Behrens, Bäumler, & Bruhn, 2015; Ostrowski & Pfeiffer, 2007). Felser et al. (2015) developed a performance structure model, based on physiological variables, of the sport short track (a discipline of speed skating) by using factor analysis, multiple regression and path analysis as methodological approaches. They identified single performance-relevant parameters as well as the starting sequence of a race as the most important partial performance. Ostrowski and Pfeiffer (2007) collected physiological parameters and race times (including split times) to develop a model of performance structure of cross-country skiing by using regression and factor analysis. They showed that single sections of a race, and especially four physiological parameters, are relevant for the athletes' overall performance. Ostrowski and Pfeiffer (2007) further stated that the commonly-used ski roller training does not seem to be adequate for the identified performance structure. The studies of Felser et al. (2015) and Ostrowski and Pfeiffer (2007) were exploratory in nature and the results need to be verified using additional datasets and further parameters.

The following section will focus on methodological approaches, which allow the prediction and structuring of performance in sport. Additional computational approaches are described, as these are necessary with regard to the data set of the three studies within this thesis and to account for the three modeling steps of Letzelter and Letzelter (1982).

2.4 Methodological approaches

Within this thesis, two different computational approaches are used to predict the overall race performance of elite triathletes, and one approach is used to analyze the performance structure of elite Olympic-distance triathlon. In both cases, preselection was mandatory to reduce the number of collected anthropometric and physiological variables, and also to identify

performance-relevant variables. The following sections will therefore give a brief overview of the applied methods.

2.4.1 Variable selection through exploratory factor analysis

In general, factor analysis as a method of multivariate data analysis can be divided into two computational approaches: confirmatory factor analysis (CFA) is designed to confirm or negate a theoretical structure, whereas exploratory factor analysis (EFA) tries to discover a structure within a set of variables (Olkin & Sampson, 2001). Both approaches normally deal with large sets of measured variables and are able to reduce the size of a data set to a smaller number of latent factors which share a common variance (Bartholomew, Knott, & Moustaki, 2011; O'Donoghue, 2010). These unobservable factors cannot be directly measured, but are linear combinations of the original variables (apart from an error term). As a contributing factor, a correlation matrix shows the relations between the measured variables, which helps to extract the factors from the data set (O'Donoghue, 2010). The ability of an EFA to uncover the structure within a set of variables was of particular use within this thesis. This allows a preselection (and finally a sort out) of parameters with high correlations and similar explanations of variance to the same underlying factor, which helps to reduce the amount of parameters with, ideally, a minimal loss of information.

A brief overview of the requirements for factor analysis is provided by Yong and Pearce (2013). The mathematical model behind EFA is built with p variables X_1, X_2,…, X_p and m latent factors F_1, F_2,…,F_m. The general formula $X_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + e_j$ means that each observable variable X_j can be expressed as a linear function of factors and a residual (Figure 2.1) (Yong & Pearce, 2013). The factor loadings a_{j1}, a_{j2}, …, a_{jm} represent the contribution of the variable to a specific factor which is similar to the weights b_i or β_i in multiple regression analysis (see section 2.4.3). Important steps within an EFA are:

- Factor extraction: The method used to extract the factors from a correlation matrix depends on the research question. Within this thesis, data reduction was the main reason to deploy an EFA. Therefore, principal component analysis was used to extract the maximum variance from the data and thereby reduce the number of variables to be considered.
- Rotation methods: The rotation step within a factor analysis helps to interpret the results because a more simple structure with variables loading on fewer factors but with higher

loadings can be achieved. The Varimax Rotation used within the thesis is a common orthogonal rotation technique.

- Interpretation of loadings: The factor loadings reflect the strength of the relationships, which makes it clear why a rotation method should be implemented before interpretation. The highest loadings identify the factors. Low loadings and cross loadings should be checked to confirm that each factor defines a unique set of interrelated variables. Cut-off values for the loadings can be used to make interpretation easier.
- Number of factors: The extracted factors should be interpretable and represent valuable common variance. In general, the eigenvalues and the so-called scree test are used to determine the number of factors. The selected criterion in this case need to be suitable to the study.

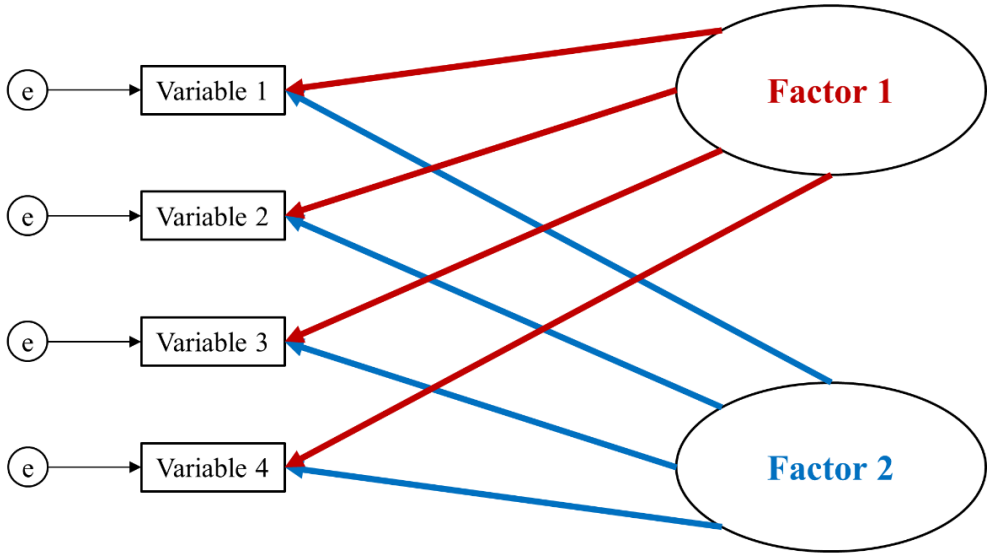


Figure 2.1 Representation of an EFA with four variables (and corresponding measurement errors) explained by two latent factors. The direction of the arrows indicates that each variable is thought to be influenced by at least one factor, not vice versa.

As mentioned before, the described steps of an EFA can be used to preselect relevant independent variables while it uncovers the structure in a large set of variables (Backhaus, Erichson, Plinke, & Weiber, 2018).

The use of factor analysis in sport science is often related to questionnaire studies, which is in general a common application. In the context of performance analysis and the use of data

from laboratory tests, the study of Pyrka, Wimmer, Fenske, Fahrmeir, and Schwirtz (2011) applied an EFA to illustrate its application to a sport science context. They identified latent factors describing parameters of performance tests in the field of ski jumping and Nordic combined (combination of ski jumping and cross-country skiing). They managed to reduce a data set of 23 measured parameters to three latent factors with a view on sport-specific arguments. These factors allow discrimination between different performance levels, allowing quicker assessment of an athlete's performance. Nevertheless, the advantage of a more efficient overview (or rather the possibility to simplify the diagnostic set-up) comes with a loss of variable-specific information (Pyrka et al., 2011).

With a view to small sample sizes, which are inevitable when working with elite triathletes, such a reduction of variables is necessary within the studies presented in this thesis. Therefore, an EFA was conducted with the principal component method. A Varimax rotation led to the final solution with variables sorted by the size of factor loadings related to a general factor. With this step, variables with similar explanation capabilities to a general factor could be excluded with a minimal loss of information.

2.4.2 Variable selection through dominance paired comparison

The method of paired comparisons as a strategy for comparative analysis has been widely used but little theorized (Tarrow, 2010). In the setting of sport science, no similar application was found. Tarrow (2010) described a wide variety of sites and settings in a political context, and found that paired comparisons have been widely used. In general, the method of pairwise comparisons "generates reliable and informative data" as stated by Farrell (2001), who used the method to quantify subjective image quality. One of the major advantages of this method is that data of subjective comparisons can be scaled or ranked (Farrell, 2001).

Through a dominance paired comparison, all variables within the analysis are compared pairwise with each other. The method adjusts for the raters' subjective criteria and the direct comparisons between each two variables is easier for raters instead of ranking a large number of variables. In general, the rater has to decide if variable A is more important than B or if B is more important than A or that both variables are equally important. After finishing all comparisons, an indirect ranking can be created based on an overall sum score, which takes into account how often a variable was preferred. In the end, the relative importance of a variable is

given. This could also be done with more than one rater (Bortz & Döring, 2006) by adding up the overall sum scores of all raters to build an indirect ranking.

The dominance paired comparisons within this thesis were conducted to identify performance-relevant parameters based on the expertise of professional German triathlon coaches and therefore reduce the amount of variables within the dataset. This comparison helps coaches to prioritize the influencing variables in a more systematic and objective way. Therefore, personal preferences and subjective influences could be avoided. Each coach had to rate the significance of each single variable against all others. The three possible options are: 1 if variable A is less important than variable B, 2 for equal priority and 3 if A is more important than B to overall race performance in triathlon. The overall sum score led to the final prioritization. The final number of variables rated as relevant must be specified manually. The dominance paired comparisons were conducted separately for anthropometric and physiological variables due to the large number of variables the coaches need to compare pairwise.

2.4.3 Prediction through multiple regression

A multiple linear regression analysis can have two main objectives: 1) a quantitative description and explanation of the relationships and 2) the prediction of a dependent variable by a given set of explanatory variables (Backhaus et al., 2018). There is widespread use of regression analysis in education and research (Backhaus et al., 2018), presumably because common statistics software (SPSS, Stata, SAS, R, etc.) have these analyses built-in. The overall aim of a multiple linear regression is to model the linear relationship between the explanatory (independent) variables and the response (dependent) variable by fitting a linear equation to the observed data. The classic model for multiple linear regression with n observations is $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_j x_{ij} + \varepsilon$ for $i = 1, 2, \ldots n$ with x_1, x_2, $\ldots x_j$ independent variables, n observed values for y and a random error ε. By minimizing the sum of the squares of the vertical deviations from each data point to the line, the best-fitting line is calculated for the given observed data. The least-square fitted values b_0, b_1,... b_j estimate the parameters $\beta_0, \beta_0, \ldots \beta_j$ of the overall regression line (Figure 2.2). The coefficient of determination (R-squared or R^2) is the most commonly-used indicator to denote how much of the variation of the dependent variable can be explained by the variation of the independent variables. It is important to

remember that R^2 increases with a larger number of independent variables within the model even though these variables are not related to the dependent variable.

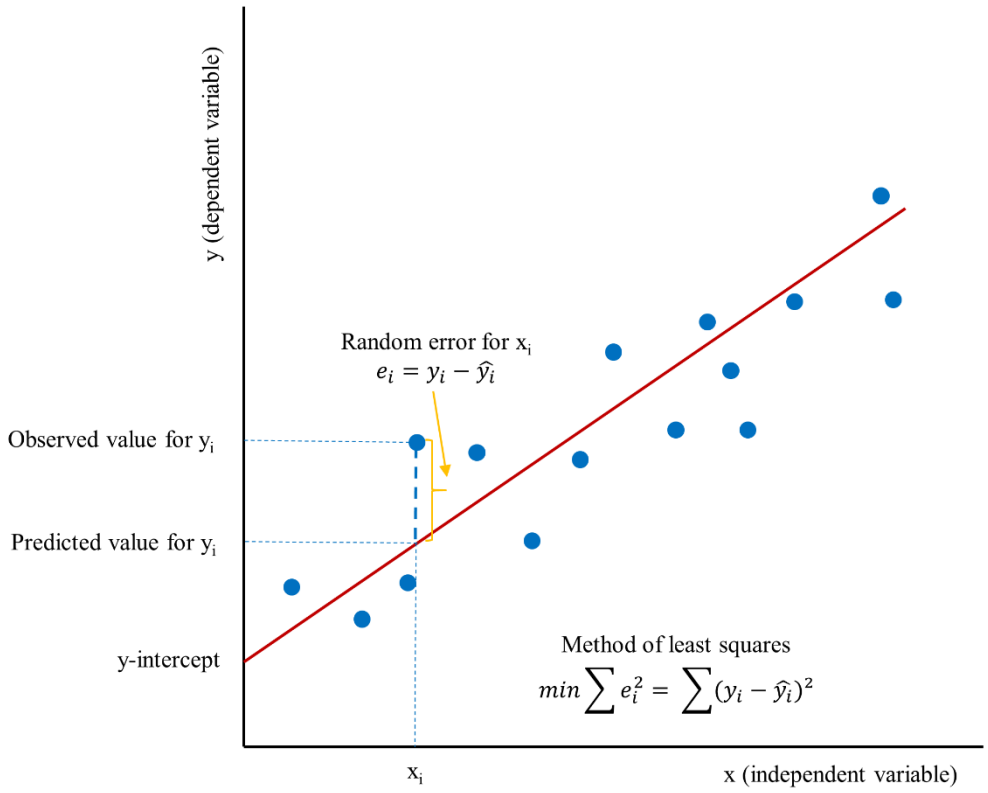


Figure 2.2 Representation of a linear regression.

In sport science, regression analyses are widely used (Atkinson & Nevill, 2001), even to predict race performance (Schabort et al., 2000; Van Schuylenbergh et al., 2004). The main research focus of the mentioned investigations was to predict overall race times of triathletes based on individual physiological parameters measured through laboratory tests.

This thesis uses the basic idea of investigating laboratory-obtained parameters to generate new approaches to performance prediction and structuring. To generate new knowledge, anthropometric and physiological parameters from laboratory tests of elite and recreational triathletes were used to calculate regression models based on previously-conducted preselection. In all multiple regression analyses within this thesis, the most important assumptions of normality, linearity, reliability of measurement and homoscedasticity (Osborne

J. & Waters E., 2002) were checked because multiple regressions are not highly robust to violations of these assumptions.

2.4.4 Prediction through artificial neural networks

Artificial neural networks (ANNs) are a strong and commonly-used machine learning approach inspired by biological neural networks, and are used to approximate functions that are generally unknown. The variety of network types can be classified by their structure, data flow, number of neurons, layers used, etc. Some examples are feedforward, radial basis function, Kohonen self-organizing and recurrent neural networks (Aggarwal, 2018; Haykin, 2009).

In general, ANNs used for (performance) prediction typically contain a feedforward design (Bunker & Thabtah, 2019) where input variables predict an output variable, and the information moves in only one direction through the network. The main characteristic of a feedforward ANN therefore is that the connections within the network do not form cycles or loops unlike in recurrent neural networks.

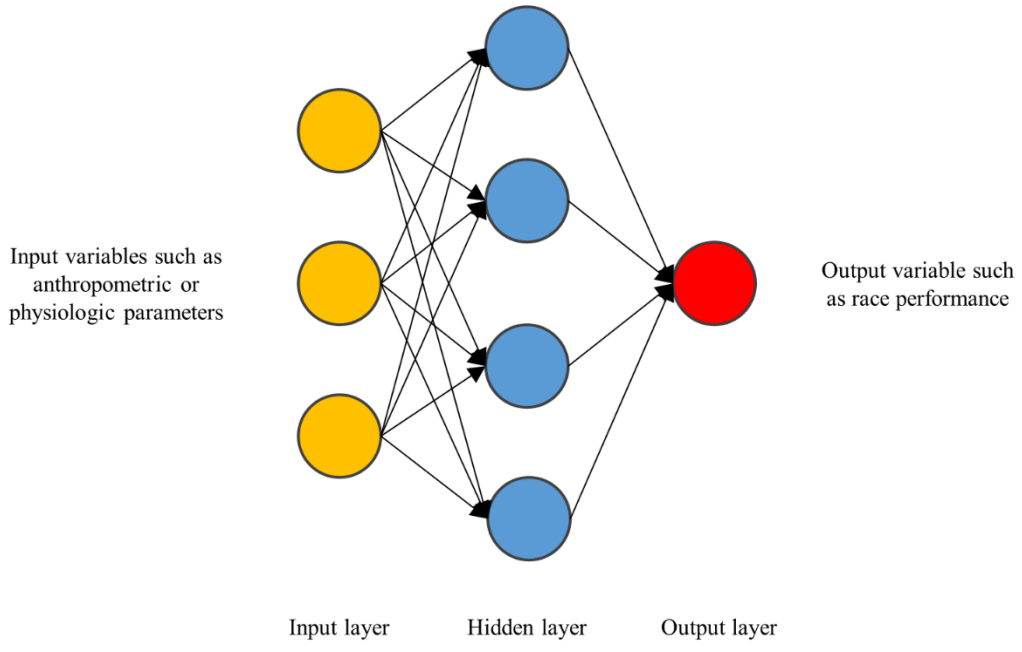


Figure 2.3 Architecture of an ANN: An input layer with the initial data, a hidden layer as an intermediate layer where all computations take place, and an output layer that produces the result for the given data set.

But how does such a neural network learn to predict? In this thesis, multilayer perceptrons using a backpropagation learning algorithm are used. This class of feedforward neural network consists of an input layer, an output layer and at least one hidden layer (Figure 2.3). Each variable (also called a neuron) within the network has directed weighted connections to all neurons of the subsequent layer. These weights need to be adjusted within the training process of an ANN and therefore a supervised learning algorithm such as the backpropagation algorithm is necessary. This kind of algorithm "repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector" (Rumelhart, Hinton, & Williams, 1986). In simple words, after each forward pass through an ANN, the backpropagation algorithm performs a backward pass while adjusting the model's parameters (weights and biases).

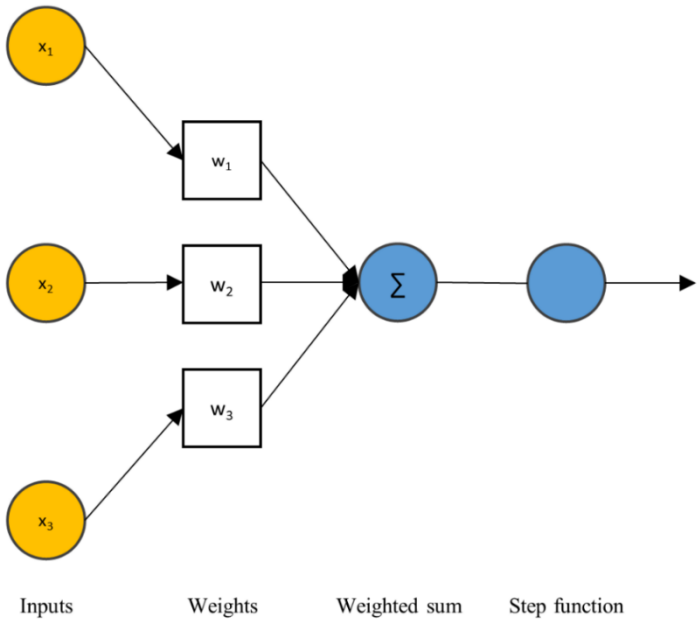


Figure 2.4 Example of a neuron in the hidden layer within an artificial neural network.

Figure 2.4 illustrates how a training data set is processed through an ANN: nodes x_1, x_2 and x_3 of the input layer have connections with specific weights w_1, w_2 and w_3 to each node within the next (hidden) layer. Computationally, each input value is multiplied by the specific weight and the resulting values are summarized and processed through an activation function which defines how active this node will be based on the summarized value (Zhang et al., 1998).

A sigmoid function is the most widely-used activation function in ANNs (Haykin, 2009; Zhang et al., 1998).

This procedure runs for each node in parallel and, after each iteration, a loss function as the difference between the output predicted by the network and the real output value is calculated as the sum of the error. This loss functions needs to be minimized to minimize the error of the neural network, which can be achieved by adapting the weights optimal. After several iterations, the network should be able to predict outputs based on new input data. This procedure is called supervised learning and (in this case) can be seen as an optimization problem (Haykin, 2009).

For more detail about computation using matrices, minimizing the loss function (especially with regard to local and global minimums) and the backpropagation algorithm see Haykin (2009).

ANNs are widely used, even in the field of performance prediction in sport a few applications could be found: for example Edelmann-Nusser et al. (2002) and Silva et al. (2007) showed that they could be a valuable method for performance modelling, without the restrictions of distribution and independence of variables. Edelmann-Nusser et al. (2002) predicted the 200 m backstroke time of an elite female swimmer in the finals of the Olympic Games using ANNs (multilayer perceptrons) based on collected training data. The accuracy of the results of this approach were attributed to the fact that performance of an athlete is quite a complex, non-linear problem (Edelmann-Nusser et al., 2002). Maszczyk et al. (2014) stated that neural networks are especially useful to model complex input-output relationships no matter if they are linear or non-linear. Nevertheless, ANNs have rarely been used to predict race performance, possibly because the network design of an ANN requires substantial input concerning the number of neurons, layers, training algorithm etc. (Zhang et al., 1998).

Within this thesis, ANNs were used after preselection through a dominance paired comparison to reduce the number of variables obtained from laboratory tests of elite triathletes. Anthropometric and physiological parameters are used both separately and combined within the models. Conceptually, multilayer perceptrons using backpropagation learning algorithm are used as already mentioned.

2.4.5 Structure through structural equation modeling

The main reason for using structural equation models (SEM) is often the proof of theory-based models or constructs including latent variables with available data. Thereby, SEM "provides a very general and convenient framework for statistical analysis" (Hox & Bechger, 1998), combining classical multivariate analyses such as factor analysis and regression. SEM, which was introduced in the field of social and behavioral science (Hox & Bechger, 1998), merged three historically statistical traditions: path analysis, simultaneous-equation models and factor analysis (Rosseel, 2012). While the computation itself is realized through matrix equations, the visualization can be done by path diagrams (Hox & Bechger, 1998). Since the 1970s, a lot of SEM programs have become available such as Lisrel (Jöreskog & Sörbom, 1993, 2001), MPlus (Muthén & Muthén, 1998-2011), AMOS (Arbuckle, 2014) and also R (R Development Core Team, 2008), which is used within this thesis.

The conceptual framework of SEM consists of a measurement model and a structural model (Figure 2.5). The measurement model consists of observed or measured variables, traditionally depicted as rectangles, and the structural model consists of latent or unobserved variables, traditionally depicted as ovals. A line between two variables symbolizes the causal effect of a latent variable on an observed or another latent variable (Schreiber, Nora, Stage, Barlow, & King, 2006).

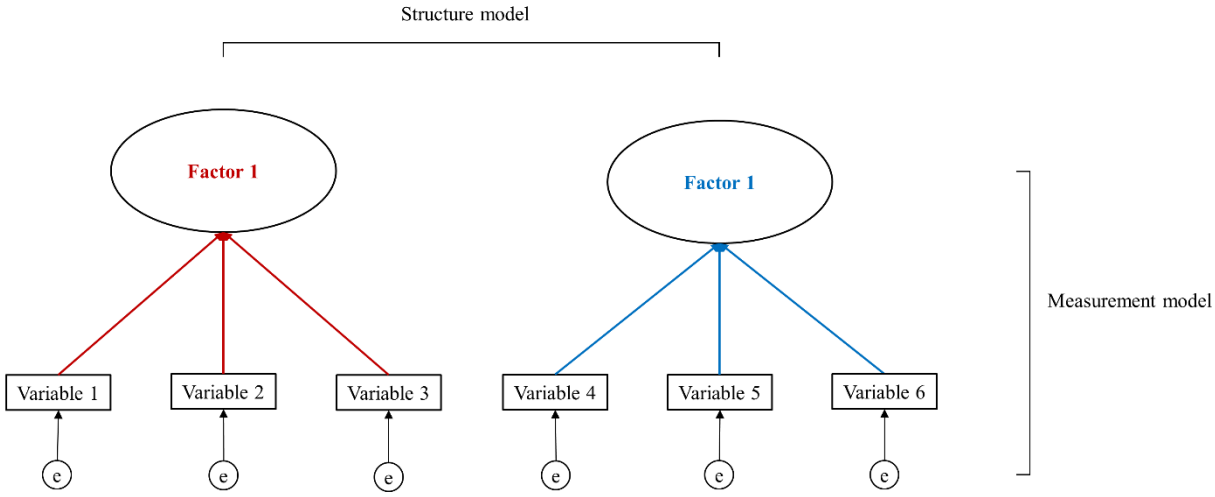


Figure 2.5 Conceptual framework of a structural equation model consisting of a measurement and a structural model.

After a theory-based model is created, it has to be fitted to the available data, which means estimating the model parameters by solving a set of equations (Hox & Bechger, 1998). A number of estimation procedures exist, taking into account the different conditions for application (e.g. normal or non-normal data distribution, ordinal or interval scaled data, etc.)

After computing a model, one of many available goodness-of-fit indicators can be used to assess the model fit (Schreiber et al., 2006). Given the available data within this thesis, the Comparative Fit Index (CFI), the Tucker-Lewis-Index (TLI), also known as the Non-Normed Fit Index (NNFI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) were used and described within the study in Chapter 6 (for details, see: Hooper, Coughlan, & Mullen, 2008; Hox & Bechger, 1998; Hu & Bentler, 1999).

Nowadays, the areas of application of SEM are diverse, including psychology, political science, education, business-related disciplines (Jais, 2007) and even sport science (Felser et al., 2015; Ostrowski & Pfeiffer, 2007). Felser et al. (2015) developed a performance structure model based on physiological variables found in short track, and Ostrowski and Pfeiffer (2007) collected physiological parameters and race times to develop a model of performance structure of cross-country skiing. Both studies were exploratory in nature and the results need to be verified using additional datasets and further parameters.

In general, performance structure is an important modeling approach in the field of training science in sport, with a focus on the identification of performance-relevant variables. SEM could be of special interest for uncovering the performance structure of triathlon, since this computational approach involves the mentioned steps of hierarchization, internal order and prioritization.

Working with elite athletes within this thesis means that the sample size is small for using SEM. Therefore, the application is exploratory in nature. As stated by MacCallum, Widaman, Zhang, and Hong (1999), a small sample size is not an obstacle to SEM, but it requires the factors to be well determined and the computations of the factor analysis or the SEM need to converge on an appropriate solution.

3 Aims and Scope of the Thesis

The present thesis aims to investigate the field of performance prediction and performance structure of Olympic-distance triathlon. For this purpose, the previously mentioned theoretical background was fundamental to deducing outstanding issues in training science research. The computational approaches described in section 2.4 have previously been used in many different research contexts because they can be widely applied. The strength of these methods was useful within this thesis to find which performance-relevant parameters determine triathlon success, with a focus on the two shorter distances. Thereby, the research gap can be filled by investigating two groups - recreational and elite triathletes - competing under different race conditions.

Accordingly, this work encompasses three main issues:

(1) Individual performance prediction of recreational triathletes
(2) Individual performance prediction of elite triathletes
(3) Performance structure of elite Olympic-distance triathlon

The following chapters 4 to 6 comprise three research studies, each considering one of the main issues. Chapters 4 and 5 encompass studies investigating the performance prediction of overall triathlon race time, whereas Chapter 6 assesses the performance structure of elite triathlon. Figure 3.1 provides a schema of the scientific work founding the thesis and of the overall project, especially the applied computational methods in each case.

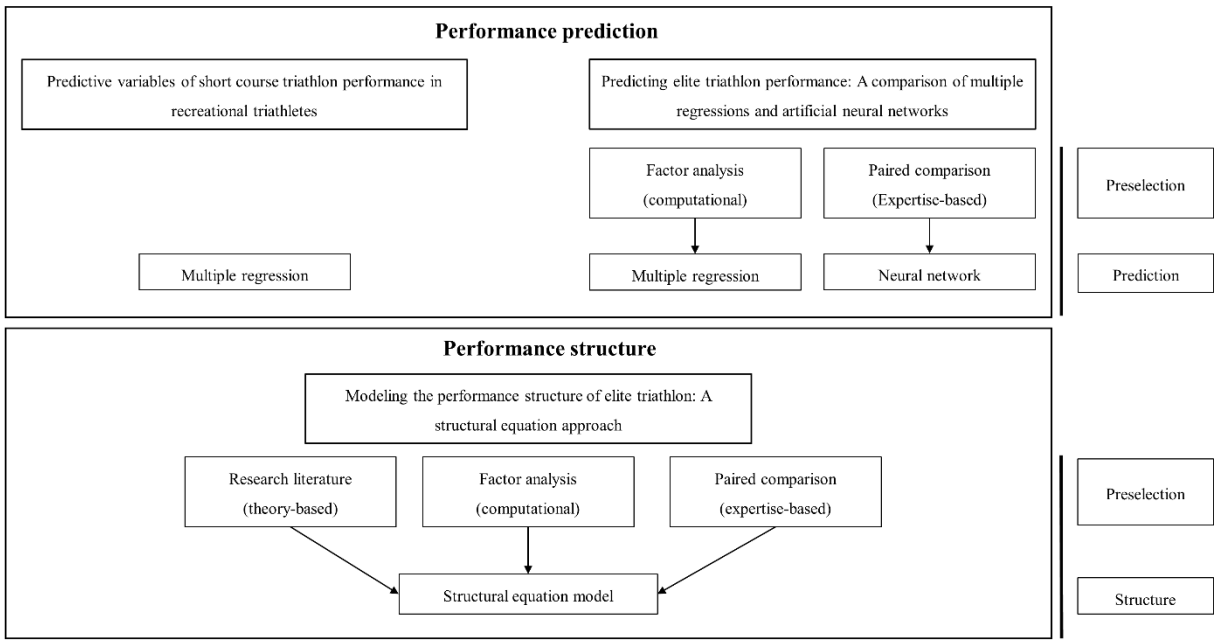


Figure 3.1 Schematic of the scientific work performed.

The studies were conducted at the BioMotion Center of the Institute of Sports and Sports Science at the Karlsruhe Institute of Technology. The data acquisition took place at the BioMotion Center (Chapter 4) and as part of a cooperative project at the Institute of Applied Training Science (Leipzig, Germany; Chapters 5 and 6).

3.1 Individual performance prediction of recreational triathletes

In a first step, previous knowledge about recreational triathletes was used to conduct a study with a heterogeneous group of recreational, competitive triathletes. Previous studies mostly focused on physiological requirements (e.g. Schneider & Pollack, 1991; Sleivert & Rowlands, 1996) and conducted simulated performance tests (Miura et al., 1997). Also, previous studies investigated the relationships between such physiological parameters and race performance in the context of recreational triathlon (Sleivert & Rowlands, 1996; Van Schuylenbergh et al., 2004). To gain more knowledge about performance-relevant parameters, the study in Chapter 4 was conducted with a heterogeneous group of recreational triathletes to collect physiological and anthropometric parameters as well as information about individual training volume. All triathletes competed in a sprint distance triathlon race within a German regional league. Their overall race times were normalized to make them comparable and used as performance criteria.

To detect relationships between the collected individual parameters and overall race time, multiple linear regression was used instead of only calculating correlations.

3.2 Individual performance prediction of elite triathletes

Section 2.2 briefly introduced the possible benefits of individual performance prediction models: more specific training programs based on general performance-relevant parameters for the preferred triathlon race distance, more differentiated talent identification programs and - especially for contemporary elite triathlon - individual race strategies based on actual performance diagnostics (Landers et al., 2000; Vleck et al., 2006). Unlike in the individual sports of swimming, cycling and running, the relationship between one or especially the combination of anthropometric and physiological parameters and overall race time with regard to performance prediction has rarely been investigated in elite triathlon. Schabort et al. (2000) and Van Schuylenbergh et al. (2004) therefore used multiple linear regressions and standardized diagnostics, especially participation in the same triathlon competition as a performance variable, to predict triathlon race times. This kind of experimental design is rarely possible with elite triathletes, due to their individual season calendar, and necessitates race time normalization to produce an equivalent dependent variable.

An alternative computational approach for performance prediction without the restrictions of distribution and independence of variables of linear regressions could be ANNs (Edelmann-Nusser et al., 2002; Silva et al., 2007). While multiple linear regression analyses have been widely used to develop prediction models, ANNs can also handle non-linear relationships based on a different model architecture. Both computational approaches can deliver meaningful results and both have the same limitation when working with measurement data from elite athletes: large numbers of independent variables require many sets of data, which are rarely available from these athletes. Therefore, a preselection of parameters is necessary to reduce the number of independent variables. As described in section 2.4, a purely statistical approach like exploratory factor analysis or an expertise-based approach like dominance paired comparison based on the experience of professional triathlon coaches could therefore be beneficial.

The study in Chapter 5 therefore investigated whether the overall Olympic-distance triathlon race time of elite athletes could be predicted using regular performance diagnostics,

which do not interrupt individual training programs and can handle the different season calendars of elite athletes. Therefore, two computational approaches were compared: 1) a purely statistical approach consisting of an exploratory factor analysis to preselect variables in combination with a multiple linear regression to predict overall race time and 2) an expertise-based non-linear approach consisting of a dominance paired comparison as a preselection method in combination with an ANN to predict overall race time.

3.3 Performance structure of elite Olympic-distance triathlon

Section 2.3 stated that the link between performance prerequisites and triathlon performance has so far mainly been provided by correlation studies (Knechtle et al., 2011; Miura et al., 1997; Zhou et al., 1997). The drawback of correlations linking just one variable to a key performance indicator must be eliminated by investigating the effect of performance variables on performance. The acquisition of data sets for structural analyses of sports performance is difficult, especially when dealing with elite athletes, which is possibly why only a few studies exist in this research area (Hottenrott & Seidel, 2017).

Performance-relevant parameters have also been identified in other sports. Felser et al. (2015) investigated short track using a step-by-step approach up to path analysis which led to a structural model. For cross-country skiing, Ostrowski and Pfeiffer (2007) followed the steps of hierarchization, internal order and prioritization using factor analysis and regression analysis to identify a structural model . They could identify influencing variables and generated thought-provoking conclusions for athletes and trainers.

Following Letzelter and Letzelter (1982) as well as Hohmann and Brack (1983) and based on the results of the performance prediction studies conducted within this thesis (Chapters 4 and 5), the main part of the present work is to give an initial approach to uncover the performance structure in elite triathlon. The results have the potential to provide new scientific indications for training programs and adjustments to them. The mentioned drawback of correlations between performance indicators and overall race performance is avoided by using structural equation modeling (section 2.4.5) as a more complex modeling process. This is of special interest for uncovering the performance structure, since this computational approach involves the steps of hierarchization, internal order and prioritization.

It must be taken into account that the present results of such statistical models are limited to the target group of elite triathletes, and especially to the triathlon race distance studied. Therefore, the developed models must be seen as exploratory in nature.

4 Study I: Predictive Variables of Short Course Triathlon Performance in Recreational Triathletes

4.1 Abstract

The purpose of this study was to statistically analyze laboratory tests of male recreational triathletes to predict sprint distance (0.5 km swim - 20 km bike - 5 km run) triathlon race time and to identify performance-relevant variables. In training for intensive and multidisciplinary sports like triathlon, performance prediction could be beneficial for optimizing training protocols and to identify an optimal race strategy, which is important for recreational athletes in particular. Therefore, 11 ambitious triathletes underwent anthropometric measurements and a cycle-run test under laboratory conditions. The athletes' race times in official triathlon races were used to compute multiple linear regression models to determine the best predictors of overall triathlon race time. Three different performance prediction models were computed. The anthropometric parameters leg length and arm span led to an adjusted R^2 of 56.9% ($R^2 = 0.665$) and the equation *predicted race time [s] = 8,386.30 + 45.65 × (leg length [cm]) - 42.62 × (arm span [cm])* to predict overall race time. The physiological parameter blood lactate concentration after 18 minutes at 200 W on a cycling ergometer (BLC_C_18) led to an adjusted R^2 of 67.9% ($R^2 = 0.711$) and the equation *predicted race time [s] = 3,773.99 + 416.40 × (BLC_C_18 [mmol·L^{-1}])*. The training parameter swimming volume led to an adjusted R^2 of 33.6% ($R^2 = 0.410$) and the equation *predicted race time [s] = 5,005.98 - 22.40 × (training volume in swimming [km])*. The performance of triathletes with heterogeneous performance levels is demonstrably dependent on anthropometric, physiological and training parameters. Overall race time was best predicted with the physiological model. The computed prediction equations, combined with an actual performance diagnostic of an individual triathlete, could be a possibility to determine individual race tactics.

4.2 Introduction

The physiological characteristics of recreational triathletes have been investigated in several previous studies (Basset & Boulay, 2000; Hue, Le Gallais, Chollet, & Préfaut, 2000; Sleivert & Wenger, 1993). The importance of maximum oxygen uptake (VO_2max) and anaerobic thresholds (Millet et al., 2009, 2011) as performance-limiting factors has been verified for triathletes with different performance levels (Butts, Henry, & Mclean, 1991), and especially in endurance running or running in triathlon (McLaughlin et al., 2010). Moreover, these parameters show significant correlations to race performance (Miura et al., 1997; Schabort et al., 2000). Similar results were found for swimming and cycling (Millet et al., 2009; Sleivert & Rowlands, 1996). Furthermore, blood lactate concentrations from treadmill or cycle ergometer tests have been identified as possible parameters for predicting triathlon performance, independent from athletes' performance level (Schabort et al., 2000; Van Schuylenbergh et al., 2004). Besides these physiological factors, anthropometric variables such as percent body fat, body mass index or the circumferences of several parts of the body could also be important for performance in triathlon races (Knechtle et al., 2011) and possibly in terms of performance prediction.

In training for intensive and multidisciplinary sports like triathlon, performance prediction could be beneficial for optimizing training protocols and to identify an optimal race strategy. For recreational triathletes, assessment of their running and overall pace might be a major problem, because an ambitious initial speed often results in a rapid decrease of performance. Therefore, prediction of their individual race performance could be useful to reduce uncertainty or anxiety. As an example of quite general advice, Hausswirth et al. (2010) identified that a 5% slower running pace than the average 10-km running pace on the first kilometer of an Olympic-distance triathlon run phase was an optimal pacing strategy. To provide more background knowledge and more precise prediction models, it is important to identify relevant parameters of triathlon race performance (Landers et al., 2000).

In contrast to the individual sports swimming, cycling and running, the relationship between single anthropometric or physiological parameters and overall race time is rarely investigated with regard to performance prediction. Schabort et al. (2000) identified the physiological parameters peak treadmill running speed [$km \cdot h^{-1}$] and blood lactate value at 4 $W \cdot kg^{-1}$ body mass on a cycle ergometer as predictive parameters for elite triathletes. The

equation to predict their overall Olympic-distance triathlon race time computed by multiple linear regression analysis provided a highly significant correlation between predicted and actual race time ($r = 0.90$, $p < 0.001$). Van Schuylenbergh et al. (2004) also predicted sprint distance triathlon performance of male physical education students with multiple regression analysis ($R^2 = 0.98$; SEE = 0.95 [min]). Running speed at maximal lactate steady state (MLSS) during laboratory testing as well as blood lactate concentration in running at MLSS were identified as predictive variables. The study of Hue (2003) also used multiple regression analysis ($r = 0.96$, $p < 0.02$) in elite triathletes and Olympic-distance triathlon performance and found two relevant parameters: 1) lactate concentration at the end of the cycle phase in a simulated field test, and 2) the distance covered during a submaximal run. The high explanation of variance (R^2) in these studies could be a consequence of the comparable conditions, since all subjects competed in the same triathlon.

The purpose of this study, therefore, is to predict individual overall race time of recreational triathletes to identify performance-relevant parameters for this specific triathlon cohort and to give potentially relevant indications for optimizing training protocols and to identify optimal race strategies. Following previous studies that determined mainly the physiological requirements of recreational triathletes (Kohrt, Morgan, Bates, & Skinner, 1987; Millet et al., 2011; Miura et al., 1997; Sleivert & Wenger, 1993), this study analyze selected anthropometric and physiological variables measured during laboratory tests as well as training volumes of recreational triathletes with different performance levels to predict their overall triathlon race time by multiple linear regression analyses. As measure of performance, normalized race time of an official triathlon race the triathletes took part in shortly after the laboratory test were used.

4.3 Methods

Experimental Approach to the Problem

The present study used multiple linear regression analysis to predict overall race times in short course triathlon and to identify performance-relevant variables. The following independent anthropometric measurements were taken: age, body height, body weight, BMI, arm span, shoulder width, chest width, hip width, hand circumferences, trunk length and leg length. Furthermore, several lactate values from a cycle-run test were taken as physiological variables,

and distance / time for swimming, cycling, running and overall were taken as training variables (Table 4.1). The dependent variable was the normalized overall race time of each triathlete.

Subjects

Eleven male recreational triathletes (age: 35.09 ± 12.49 years) were included in this study. All triathletes were competing in national amateur league races and underwent at least 10 hours of training a week. The test protocol was approved by the Institutional Review Board and written informed consent was obtained from all participants prior to testing. Table 4.1 shows the descriptive characteristics (mean value and standard deviation (SD)) as well as the coefficient of variation (CV = (SD/Mean)*100)) of the sample.

Procedures

The study consisted of two laboratory tests at the BioMotion Center of the Institute of Sports and Sports Science at the Karlsruhe Institute of Technology (Germany). At the beginning, the triathletes had to perform a classic step test on a treadmill (saturn 300/100rs, h/p/cosmos sports & medical GmbH, Germany). Starting with 8 $km \cdot h^{-1}$ and a gradient of 1 %, the speed increased by about 2 $km \cdot h^{-1}$ every 3 minutes until exhaustion. After each step, a break of 20 seconds was necessary to collect the blood sample. The results of the first test were used to produce an optimal test procedure for the second test after four weeks of training. Due to the performance of the triathletes and to simulate a more race-specific situation, the initial speed of the treadmill was adapted to 10 $km \cdot h^{-1}$ for the second test.

About four weeks later, the same triathletes performed a cycle-run test consisting of a 26 min cycling protocol on a cycling ergometer (SRM Ergometer, SRM GmbH, Germany) followed by a fast transition and a classic step test on the same treadmill as above. The standardized cycling protocol consisted of a 5 min warm-up phase (power: 150 W), 20 min constant load (power: 200 W) and 1 min to prepare themselves for transition (power: 150 W). The transition to running step test was prepared before starting the test to ensure comparable conditions to the transition in a triathlon race. The running step test began at 10 $km \cdot h^{-1}$ and increased by 2 $km \cdot h^{-1}$ every 3 min. The results of the first running step test ensured that all triathletes could perform at least four complete steps.

Meanwhile, the following blood lactate measurements (Biosen, EKF Diagnostics, Germany) were conducted:

- value prior to testing
- value after 18 min constant load (200 W) on a cycling ergometer
- value after transition and prior to treadmill running test
- values at the end of each step of the treadmill running test

Anthropometric parameters (Table 4.1) were collected separately in accordance with the methods of Knussmann and Barlett (1988) and Tittel and Wutscherk (1972).

The whole dataset consisted of 11 anthropometric, eight physiological (lactate) parameters and eight variables concerning the training volume, which were used for computations (Table 4.1).

Table 4.1 Descriptive variables of male recreational triathletes (N = 11).

	Mean ± SD	**CV (%)**
Anthropometric		
Age [yrs]	35.09 ± 12.49	35.59
Body height [cm]	182.64 ± 6.55	3.59
Body weight [kg]	78.42 ± 4.75	6.06
BMI [$kg \cdot m^{-2}$]	23.51 ± 0.81	3.45
Arm span [cm]	188.28 ± 7.17	3.81
Shoulder width [cm]	44.73 ± 2.14	4.78
Chest width [cm]	31.16 ± 1.76	5.65
Hip width [cm]	29.85 ±1.19	3.99
Hand circumferences [cm]	21.84 ± 1.01	4.62
Trunk length [cm]	63.38 ± 3.24	5.11
Leg length [cm]	86.84 ± 4.12	4.74
Training volume		
Swim ($km \cdot week^{-1}$)	7.17 ± 2.17	30.26
Bike ($km \cdot week^{-1}$)	271.04 ± 95.19	35.12
Run ($km \cdot week^{-1}$)	42.88 ± 27.85	64.95
Overall race time for sprint distance [min]	72.10 ± 3.99	5.53

Training Recordings

All triathletes had to record their training volume (kilometers and hours) for the three disciplines as well as additional training within the period between the two laboratory tests to provide a

transparent view on the overall training volume. The triathletes were instructed to complete their usual training routine. Because of slightly different periods between the two tests, the training volume was normalized to a representative 28-day period. Due to missing values of one triathlete, the sample used in training volume-based regression consisted of ten triathletes.

Race Time and Normalization

To obtain comparable overall race times, all triathletes competed in an official triathlon race within two weeks after their second test. Race time normalization was necessary because three triathletes participated in different events, compared to eight triathletes who took part in the same race of the LBS Cup triathlon league (Germany). Unfortunately, this is unavoidable if triathlon race times are collected under real conditions rather than a simulated laboratory test.

The winner's time of the LBS Cup triathlon race was set as the baseline, and reference factors were calculated for each winner's time of the other three races. Then, the individual race times of each triathlete participating in the study were normalized to obtain comparable race times as a dependent variable in multiple regression analysis.

reference factor = winner's time of triathlon race / winner's time of LBS Cup triathlon race

normalized individual race time = individual race time / related reference factor

Statistical Analyses

IBM SPSS Statistics (Version 22, IBM) was used for all statistical analyses. The level of significance for the prediction models and for each parameter was set to $p < 0.05$.

Stepwise multiple linear regression analyses (MLR; backward method, default exclusion criteria: probability of F to remove ≥ 0.1) were used to detect potential performance prediction models and to investigate the relationships between independent variables and the dependent variable. Due to the small sample size, the parameters had to be divided into three clusters to compute MLR analyses. Therefore, anthropometric, physiological and training parameters could only be used separately and not in combination. Variance Inflation Factor (VIF) was checked to avoid multicollinearity (Hair, 1995). The normality of residuals as well as residual independence and homoscedasticity were analyzed by corresponding plots. To

identify and remove influential cases in the case of homoscedasticity, Cook's Distance was used with a cut-off of ≥ 1 (Heiberger & Holland, 2004).

To evaluate the models, coefficients of determination (percentage of variance explained; R^2) and the standard error of the estimate (SEE) were used. The adjusted R^2, which considers the number of variables used in each regression, allows comparison between the different models.

4.4 Results

Normalization of Race Time

The normalization of race times led to a mean (± SD) overall race time of 4,326.29 ± 239.19 s (approximately 1:12 h) for male recreational triathletes.

Performance Prediction Models

MLR analyses provided three different performance prediction models (Table 4.2). Statistical assumptions (normal distribution of regression residuals, homoscedasticity) were validated by assessment and testing of residuals. The best anthropometric predictors of overall race time in sprint distance triathlon were leg length and arm span. The adjusted R^2 showed an explanation of variance of 56.9% ($R^2 = 0.665$) of overall race time by the anthropometric-based model. The best physiological predictor of overall race time was the blood lactate concentration after 18 minutes at 200 W on a cycling ergometer. The adjusted R^2 showed an explanation of variance of 67.9% ($R^2 = 0.711$) of overall race time by the physiological-based model. The best predictor out of the training parameters was swimming training volume. The adjusted R^2 showed an explanation of variance of 33.6% ($R^2 = 0.410$) of overall race time by this model.

Table 4.2 Parameter and model estimates of MLR analyses for male recreational athletes (Model 1 = anthropometric parameters; Model 2 = physiological parameters; Model 3 = training parameters).

	Value	**β-coefficient**	**R^2**	**Adjusted R^2**	**SEE [s]**	**p-value**	**VIF**
Model 1 (anthropometric)			0.655	0.569	156.97	0.014	
Constant	8,386.30						
LL	45.65	0.786				0.05	2.699
AS	-42.62	-1.277				0.006	2.699
Model 2 (physiological)			0.711	0.679	135.51	0.001	
Constant	3,773.99						
BLC_C_18	416.40	0.843				0.001	1.000
Model 3 (training)			0.410	0.336	205.43	0.046	
Constant	5,005.98						
SW_KM	-22.40	-0.640				0.046	1.000

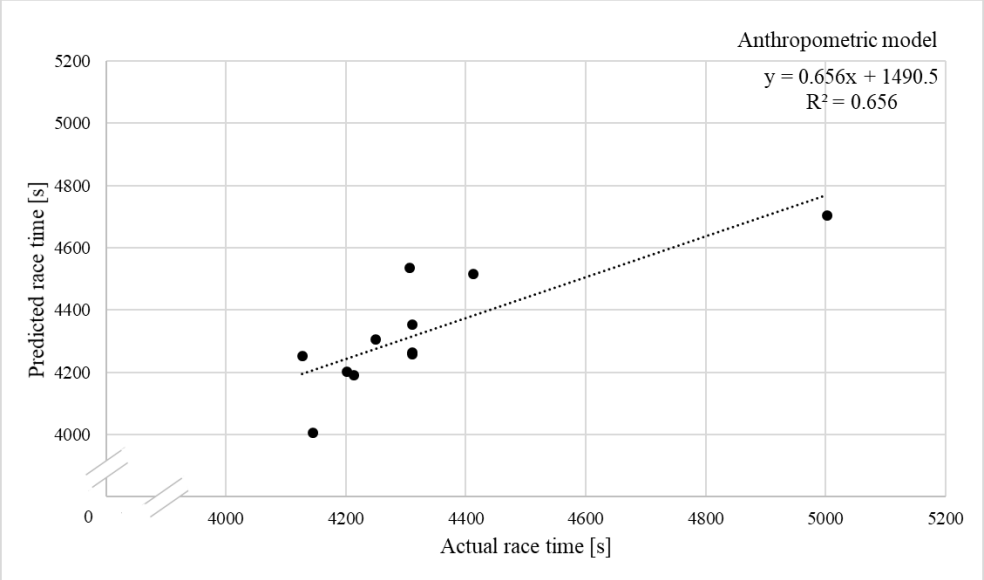
Notes: SEE = standard error of the estimate, VIF = Variance Inflation Factor, LL = leg length; AS = arm span; BLC_C_18 = blood lactate concentration after 18 min at 200 W on a cycling ergometer; SW_KM = training volume in swimming. General format for multiple regression equation: y = constant + value1 × variable1 + value2 × variable2 + …

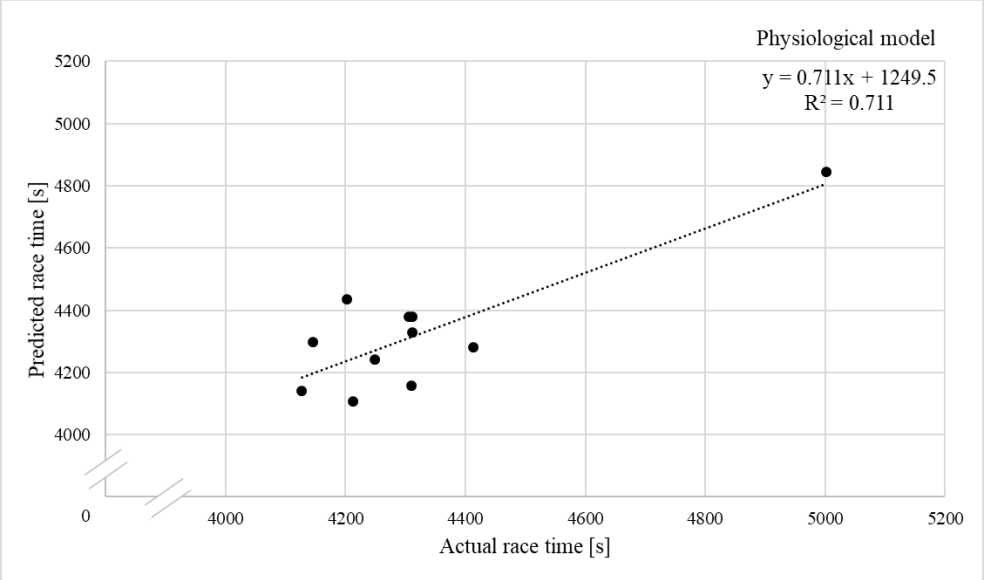
Prediction Equations

These results led to three different equations predicting overall triathlon race time for the sample:

- Predicted race time [s] = 8,386.30 + 45.65 × (leg length [cm]) - 42.62 × (arm span [cm])
- Predicted race time [s] = 3,773.99 + 416.40 × (blood lactate concentration after 18 min at 200 W on cycling ergometer [mmol·L-1])
- Predicted race time [s] = 5,005.98 - 22.40 × (training volume in swimming [km])

Figure 4.1 shows the predicted overall race time plotted against the actual overall race time in sprint distance triathlon of each recreational athlete used for computation.





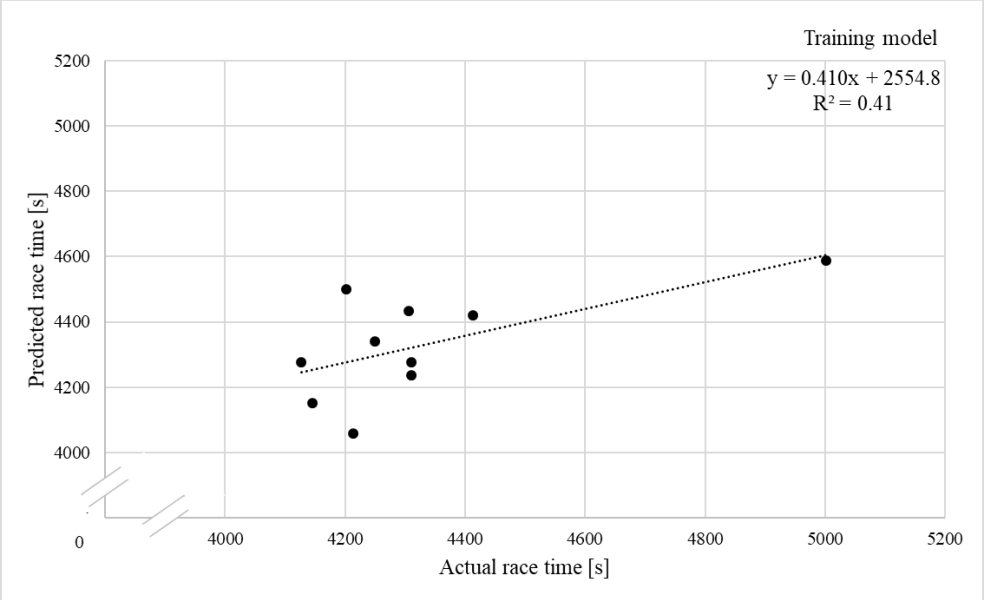


Figure 4.1 Predicted and actual overall race time in sprint distance triathlon of the anthropometric-, physiological- and training-based models.

4.5 Discussion

The aim of the current study was to reveal relationships between anthropometric, physiological and training parameters with the overall sprint distance triathlon race time of recreational triathletes. The best performance prediction was obtained with the physiological model (adjusted R^2 showed an explanation of variance of 67.9%; $R^2 = 0.711$) followed by the

anthropometric model (adjusted R^2 showed an explanation of variance of 56.9%; $R^2 = 0.665$). The training-based model provided poorer prediction results (adjusted R^2 showed an explanation of variance of 33.6%; $R^2 = 0.410$).

Selection of Parameters and Sample Composition

In this study, only separated blocks of parameters (anthropometric-, physiological- and training-based) were used to compute the MLR due to the small sample size. This should be borne in mind, because the composition of parameters affects the prediction results. To determine the relevant parameters prior to testing, the results of a previous investigation were used (Hoffmann, Moeller, Seidel, & Stein, 2015). Although the sample showed a relatively large spread concerning the age of the triathletes, performance level was comparable which is shown by the fact that they compete in the same league and the small CV in overall race time of 5.53% (Table 4.1). The anthropometric parameters (Table 4.1) of body height, body weight and resulting BMI are in accordance with Sleivert and Wenger (1993) (slightly taller and higher body weight which results in a comparable BMI), and slightly taller and heavier than reported by Kohrt et al. (1987) and Hue et al. (1998). Knechtle et al. (2011) showed similar values for body height, body weight, BMI and leg length for long distance triathletes.

Interestingly, the triathletes of the present study had high training volumes considering their recreational background: on average, 7.17 km swimming, 271.04 km cycling and 42.88 km running per week. The slightly higher training volumes compared to Hue et al. (1998), Kohrt et al. (1987) and Kohrt et al. (1989) had two main reasons: the triathletes of the present study were more experienced on average (Kohrt et al., 1987; Kohrt et al., 1989) and some completed privately-organized training camps. These trainings camps are also responsible for the high CVs in training volume parameters. Compared to other studies using recreational triathletes, the sample showed similar anthropometric and training volume parameters.

Normalization of Race Time

To obtain comparable overall race times independent of the individually-selected triathlon races, race time normalization was necessary. Mean and standard deviation of overall race time (4,326.29 ± 239.19 s) are comparable to another study (Taylor & Smith, 2014). The slightly faster mean race times and lower SD in this study could be due to different conditions compared

to the previous study. The lower SD indicates a more homogenous group of triathletes, which could be because they were members of the same league.

Performance Prediction

Based on performance parameters identified through prediction models, new training programs with more specific priorities could be created. Also, in the field of talent diagnoses the relevant parameters allow (with a certain degree of caution) a more objective selection and a more precisely aligned training process for young athletes (Landers et al., 2000). The number of previous races and personal best times were previously identified as predictive variables for middle- and long-distance triathlon race times (Knechtle, Zingg, Rosemann, & Rüst, 2015; Rüst et al., 2012). Unfortunately, no sufficient and usable information was available for the present sample.

The training volume-based prediction model provided the poorest results and is difficult to interpret due to very individual race preparation protocols before season opening, even if the weekly training volume is comparable to other studies.

The best anthropometric predictors were leg length and arm span. Previous studies already showed the relationship between arm span and swim performance (Lätt et al., 2010), which should also be valid for swimming in triathlon even if it is only one discipline. So far, no consensus concerning the importance of leg length in running exists, although previous studies characterized long-distance runners as long-legged compared to sprinters. This is an interesting indication with regard to the results of the present study (Barnes & Kilding, 2015).

The general importance of physiological parameters in triathlon, especially for performance prediction, is indisputable (Schabort et al., 2000; Suriano & Bishop, 2010). Blood lactate concentrations are common in research, and have previously been used to describe or predict overall triathlon performance (Schabort et al., 2000; Van Schuylenbergh et al., 2004). Hue (2003) also demonstrated that lactate concentration measured at the end of the cycling part in a simulated cycle-run test appears to be a performance predictor in triathlon, which confirms the present findings. With regard to different race tactics in races with or without drafting, Hue (2003) - in reference to Hausswirth et al. (1999) - highlights the importance of specific test protocols and their influence on the results of performance prediction. Therefore, the use of a constant workload of 200 W on a cycling ergometer to standardize conditions should be

modified in future studies to a normalized workload based on each triathlete's ability (such as % VO_2max or % lactate threshold). This should lead to a better linear regression fit in the case of the physiological model due to a more personalized cycling part.

The SEEs of the three performance prediction models vary in precision: the anthropometric (156.97 s) and physiological (135.51 s) models are close to performance variation of the top five triathletes in overall 2015 LBS Cup triathlon league ranking (their average overall ranking in four triathlon races was 7^{th} position; average time span between 1^{st} and 10^{th} place: 105 s). With this in mind, the SEE of 205.43 s for the model based on training volume indicates too much inaccuracy.

In summary, three different prediction models of overall race time of recreational triathletes in sprint distance triathlon were computed with variables measured in laboratory tests and reported individual training volume. This succeeded, although there are confounding variables like the environment, the conditions of the race, the terrain, etc. which add variance to the modeled equations. Arm span and leg length as anthropometric variables were identified as important parameters of individual performance, accounting for a variance explanation of 65.5% of overall race time. Blood lactate concentration after 18 min at 200 W on a cycling ergometer in a simulated cycle-run test was identified as an important parameter concerning the physiological model, with a variance explanation of 71.1%. Swimming training volume during a short training period was also identified as fairly important parameter, accounting for a variance explanation of 41% of overall race time.

Practical Applications

The present study shows that overall race time in short course triathlon performance could be predicted with variables measured through laboratory tests. For recreational triathletes, assessment of their running and overall pace might be a major problem when determining an individual race tactic. The choice of an optimum individual pace in each discipline, especially in swimming and running, can prevent a decline in performance or even a dropout. Therefore, prediction of their individual race performance could be useful. The computed prediction equations, combined with an actual performance diagnostic of an individual triathlete, could be a possibility to reduce their concerns. In the field of talent diagnostics, the identified parameters of performance prediction models allow, with a certain degree of caution, a more objective

basis of selection and a more precisely aligned training process for young athletes (Landers et al., 2000).

Limitations and Outlook

Potential for future studies can arise from the application of a normalized workload on cycling ergometer based on each triathlete's ability (such as % VO_2max or % lactate threshold) instead of using a constant workload. This could lead to a better fit in linear regression.

Another limitation appears as the result of using a small sample of heterogeneous triathletes. As shown in Figure 4.1, the actual race time of one triathlete was about 10 minutes slower compared to all other athletes, which led to an outlier that maybe influence the results of linear regression.

Further research should therefore focus on collecting larger samples and the application of more specific laboratory tests, preferably combining the three different disciplines, to determine more extensive and specific performance prediction models. Based on these models, new training programs with more specific training priorities could also be created. Information about previous races such as overall or split times could lead to a better prediction (Gilinsky, Hawkins, Tokar, & Cooper, 2014), in particular in samples of recreational triathletes because of their heterogeneous characteristics compared to elite triathletes.

5 Study II: Predicting Elite Triathlon Performance: A Comparison of Multiple Regressions and Artificial Neural Networks

5.1 Abstract

Two different computational approaches were used to predict Olympic-distance triathlon race time of German male elite triathletes. Anthropometric measurements and two treadmill-running tests to collect physiological variables were repeatedly conducted on eleven male elite triathletes between 2008 and 2012. After race time normalization, exploratory factor analysis (EFA), as a mathematical preselection method, followed by multiple linear regression (MLR) and dominance paired comparison (DPC), as a preselection method considering professional expertise, followed by nonlinear artificial neural network (ANN) were conducted to predict overall race time. Both computational approaches yielded two prediction models. MLR provided $R^2 = 0.41$ in case of anthropometric variables (predictive: pelvis width and shoulder width) and $R^2 = 0.67$ in case of physiological variables (predictive: maximum respiratory rate, running pace at 3-mmol·L^{-1} blood lactate and maximum blood lactate). ANNs using the five most important variables after DPC yielded $R^2 = 0.43$ in case of anthropometric variables and $R^2 = 0.86$ in case of physiological variables. The advantage of ANNs over MLRs was the possibility to take non-linear relationships into account. Overall, race time of male elite triathletes could be well predicted without interfering with individual training programs and season calendars.

5.2 Introduction

Performance prediction in training-intensive sports like triathlon (a combination of swimming, cycling, and running) could be beneficial for optimizing training protocols and identifying talent. It is therefore important to identify performance parameters predicting triathlon race performance (Landers et al., 2000), such as anthropometric and physiological parameters based on laboratory tests (Schabort et al., 2000).

Several studies have shown the importance of maximum oxygen uptake (VO_2max) and anaerobic thresholds (Millet et al., 2009, 2011) in endurance running or running in triathlon. These parameters showed significant correlations to race performance (Bassett, 2000; McLaughlin et al., 2010). Similar results were found for swimming and cycling (Millet et al., 2009; Sleivert & Rowlands, 1996). However, these variables only have a prerequisite function, and are not performance predictors in homogenous samples, because of the small variation between athletes (Bassett, 2000; Sleivert & Rowlands, 1996; Stratton et al., 2009). Nonetheless, blood lactate concentrations from treadmill or cycle ergometer tests were useful parameters in predicting triathlon performance independent of athletes' performance level (Schabort et al., 2000; Van Schuylenbergh et al., 2004). Besides such physiological factors, anthropometric variables such as percent body fat, body mass index (BMI), and the circumferences of several parts of the body could also be important for performance in triathlon races (Knechtle et al., 2011) and therefore for performance prediction.

Unlike in the individual sports of swimming, cycling, and running, which constitute triathlon, the relationship between one or a combination of anthropometric and physiological parameters and overall race time with regard to performance prediction have rarely been investigated in triathlon. Schabort et al. (2000) used multiple linear regressions to predict overall Olympic-distance triathlon race time of the South African national team by using physiological parameters such as peak treadmill running speed [$km \cdot h^{-1}$] and blood lactate value at 4 $W \cdot kg^{-1}$ body mass on a cycle ergometer. The correlation between predicted and actual race time was highly significant ($r = 0.90$, $p < 0.001$). Multiple regression analysis ($R^2 = 0.98$; SEE = 0.95 [min]) was also used by Van Schuylenbergh et al. (2004) to predict sprint distance triathlon performance of male physical education students. In each of these two studies, subjects competed in the same triathlon competition, which likely caused the high explanation of variance (R^2) because of comparable conditions. Nonetheless, this kind of experimental design

is rarely possible with elite triathletes due to their individual season calendar. Artificial neural networks (ANNs) are an alternative computational approach for performance prediction. Edelmann-Nusser et al. (2002) as well as Silva et al. (2007) showed that artificial neural networks could be a valuable method for performance modelling, without the restrictions of distribution and independence of variables. Edelmann-Nusser et al. (2002) predicted the 200 m backstroke time of an elite female swimmer in the finals of the Olympic Games by using artificial neural networks (multi-layer perceptrons) based on collected training data. The accuracy of the results of this approach were attributed to the fact that "the adaptive behavior of the system athlete is quite a complex, non-linear problem" (Edelmann-Nusser et al., 2002). However, multiple linear regression analyses have been more widely used to develop prediction models. Linear regressions require linear relationships between independent variables and a dependent variable, whereas artificial neural networks could also handle non-linear relationships based on a different model architecture. Nevertheless, ANNs have rarely been used to predict race-performance, possibly because the network design of an ANN requires substantial input concerning the number of neurons, layers, training algorithm etc. (Zhang et al., 1998).

Both computational approaches have a major limitation while working with measurement data from elite athletes: large numbers of independent variables require many sets of data, which are rarely available while working with elite athletes. Therefore, a preselection of parameters is necessary to reduce the number of independent variables. If there are only a few variables with non-linear relationships, a purely statistical approach like an exploratory factor analysis can be used to preselect variables before computing a prediction model. In case of ANNs, which could also handle non-linear relationships, a dominance paired comparison based on the expertise of professional triathlon coaches could be beneficial, since this method utilizes a more subjective point of view and practical experiences.

In summary, the prediction of individual overall race time in elite Olympic-distance triathlon competition, by using several anthropometric and physiological parameters as well as different computational approaches, has not been investigated to date. Previous studies mostly tested recreational triathletes (Kohrt et al., 1987; Millet et al., 2011; Miura et al., 1997; Sleivert & Wenger, 1993) because of the availability of a larger number of potential athletes. National squads normally consist of 4–5 athletes, which makes it very difficult to get a sufficient sized sample. Moreover, elite athletes are often reluctant to participate in experiments. In addition,

individual training programs and different season calendars complicate experimental laboratory studies with elite athletes. Therefore, the first aim of this study was to assess whether overall Olympic-distance triathlon race time of elite athletes could be predicted using regular performance diagnostics. The second aim was to compare two computational approaches and determine whether one is better than the other. A purely statistical approach consisting of an exploratory factor analysis to preselect variables in combination with a multiple linear regression to predict overall race time was compared to an expertise-based non-linear approach consisting of a dominance paired comparison as a preselection method in combination with an artificial neural network to predict overall race time. In both cases, several anthropometric and physiological variables measured during laboratory tests over a period of four years in German male elite Olympic-distance triathletes were used.

5.3 Methods

Subjects

Eleven male German elite triathletes (age: 23.38 ± 2.79 years) competing in national or international championships were included in this study. Written informed consent in the form of an athlete agreement between each national squad triathlete and the German national triathlon association (DTU), as well as a cooperation agreement with the Institute for Applied Training Science (Leipzig, Germany), which is responsible for classic performance diagnostics of elite athletes in Germany, were mandatory. Participation in the performance diagnostics was voluntary and the triathletes could opt out at any time. After data acquisition, all statistical analyses were conducted anonymously. Table 5.1 shows descriptive characteristics (mean value and standard deviation (SD) as well as the coefficient of variation (CV = (SD/Mean)*100)) of the triathletes.

Table 5.1 Descriptive variables of German elite triathletes (N = 11).

	Mean ± SD	**CV (%)**
Anthropometric		
Age [yrs]	23.38 ± 2.79	11.93
Body height [cm]	187.0 ± 2.90	1.55
Body weight [kg]	74.46 ± 4.28	5.75
Seat height [cm]	96.38 ± 1.59	1.65
Shoulder width [cm]	40.16 ± 2.24	5.58
Pelvis width [cm]	28.65 ± 1.61	5.62
Thorax width [cm]	28.27 ± 1.29	4.56
Thorax depth [cm]	21.06 ± 1.41	6.70
Quetelet Index [$g \cdot cm^{-1}$]	398.15 ± 21.32	5.35
BMI [$kg \cdot m^{-2}$]	21.29 ± 1.17	5.50
Body fat [%]	10.70 ± 1.36	12.71
Body fat [kg]	8.00 ± 1.35	16.88
Lean body mass [kg]	66.46 ± 3.27	4.92
Physiological		
VO_2max [$mL \cdot min^{-1}$]	5457.67 ± 292.56	5.36
VO_2max [$mL \cdot min^{-1} \cdot kg^{-1}$]	72.02 ± 4.29	5.96
PL3 [$m \cdot s^{-1}$]	5.08 ± 0.23	4.53
Max running pace [$m \cdot s^{-1}$]	5.22 ± 0.27	5.17

Max running pace mobi [$m \cdot s^{-1}$]	6.92 ± 0.17	2.46
LA max mobi [$mmol \cdot L^{-1}$]	9.18 ± 1.30	14.16
VCO_2 max mobi [mL]	6472.75 ± 431.74	6.67
Max distance mobi [m]	1762.69 ± 136.70	7.76
RMV max mobi [$mL \cdot min^{-1}$]	187.73 ± 12.40	6.61
RR max mobi [$breaths \cdot min^{-1}$]	63.18 ± 10.10	15.99
BLC 3 min [$mmol \cdot L^{-1}$]	8.08 ± 1.31	16.21
BLC 6 min [$mmol \cdot L^{-1}$]	9.13 ± 1.29	14.13
BLC 10 min [$mmol \cdot L^{-1}$]	8.62 ± 1.38	16.01
Normalized overall race time Olympic distance [min]	113.79 ± 3.21	2.82

Notes: PL3 = running pace at 3-$mmol \cdot L^{-1}$ blood lactate; mobi = mobilization test; LA max mobi = maximum blood lactate in mobilization test; RMV max Mo = maximum respiratory minute volume in mobilization test; RR max Mo = maximum respiratory rate in mobilization test; BLC 3, 6, 10 min = blood lactate concentration 3, 6, 10 min after load in mobilization test

Experimental Procedure

The data in this study were derived from laboratory tests performed between 2008 and 2012 at the Institute for Applied Training Science (Leipzig, Germany) within the frame of national squad investigations. Because elite triathletes were tested at various time slots based on their competition calendar, the distribution of tests was not consistent. Overall, 23 men completed 58 laboratory tests between 2008 and 2012. The iterative approach to select valid sets of variables was based on the following requirements: (1) complete sets of variables of the laboratory tests and (2) finished Olympic-distance triathlon races within 8 weeks after each

single performance diagnostic. Twenty-five sets of variables from eleven triathletes fulfilled these criteria and were eventually used.

The anthropometric characteristics of each triathlete were selected and determined based on the information provided by Tittel and Wutscherk (1972) and Knussmann and Barlett (1988). Body height and segment lengths and widths were measured using precise measuring instruments and valid measurement regulations, and provided the basis to calculate the various indices (Tittel & Wutscherk, 1972). Body fat was determined by measuring skin fold thickness of ten skin folds with a caliper (Tittel & Wutscherk, 1972); lean body mass could then be calculated. The anthropometric variables mentioned in Table 5.1 (except age and body weight, which are only listed for a better characterization of the sample) were used for computation.

For the physiological parameters, the triathletes had to perform two different motorized treadmill running tests under laboratory conditions (gradient of 0°). First, a classic step test with an individual initial speed between 4 and 4.5 $m \cdot s^{-1}$ depending on general performance was conducted. The step length was 4 km, with an increasing rate of 0.25 $m \cdot s^{-1}$ between two consecutive steps. The test was stopped after a maximum of four steps. One day later, a maximum mobilization test with an initial speed of 5 $m \cdot s^{-1}$, an increasing rate of 0.25 $m \cdot s^{-1}$ per step, and a step length of 30 s until voluntary exhaustion was performed. In both tests, blood lactate measurements and spirometry were conducted. Pulmonary and respiratory gas-exchange parameters were measured using a calibrated breath-by-breath gas-analyzer (Cortex METAMAX 3B and Cortex METALYZER 3B). The physiological variables considered for computation are shown in Table 5.1 (except relative VO_2max, which is only listed for a better characterization of the sample).

Data analysis

Normalization of race time

Normalization was necessary to obtain comparable individual race times independent of the various triathlon races in which the subjects participated. These normalized race times were fundamental to all following analyses, since they accounted for the slightly different competition calendars of each elite triathlete. Races with a maximum time lag of 8 weeks to each single performance diagnostic were selected from official results (www.triathlondata.org). To guarantee a similar race progress, the minimum requirement was participation in the

German, European, or World championships as well as in races within the World Triathlon Series (WTS).

The reference factor was calculated as the mean value of overall race times of the Top 10 athletes in WTS between 2009 and 2012. All finished races within each year were considered. The resulting mean value for Olympic-distance triathlon race time was used to normalize each individual race time.

reference factor = mean (overall race times of Top 10 WTS athletes of all races within the WTS 2009, 2010, 2011, 2012)

Up to two races of the WTS are sprint distance triathlons. To use these race times, the same approach was applied to determine a factor transforming sprint distance triathlon race time into an Olympic distance equivalent.

Statistical methods

The statistical analyses applied after race time normalization could be divided into two computational approaches to identify performance-relevant parameters and predict overall race times of German male elite triathletes:

- A purely statistical approach consisting of an exploratory factor analysis, to preselect important anthropometric and physiological parameters, and multiple linear regressions to identify performance-relevant parameters and predict overall race times of German male elite triathletes.

- An expertise-based non-linear approach consisting of a dominance paired comparison with four professional German triathlon coaches, to preselect important anthropometric and physiological parameters, and the application of artificial neural networks to predict overall race times of German male elite triathletes.

The converse implementation of the preselection and prediction methods (e.g. dominance paired comparison and multiple linear regressions) were deemed unsuitable because of their different fields of application, based on the linear and non-linear relationships between the independent variables and the dependent variable, normalized overall race time.

The purely statistical approach could be divided into two consecutive steps: An exploratory factor analysis (EFA) was first applied to preselect relevant independent variables, followed by a multiple linear regression (MLR) to determine potential prediction models and the priority of the used parameters. An EFA helps to uncover structures in large sets of variables. This allows a preselection of parameters with high correlations among themselves and similar explanation of variance to the same underlying factor. For small sample sizes, which are inevitable while working with elite triathletes, a reduction of variables can improve the results in MLR, and prevent multicollinearity. Therefore, an EFA was conducted using the 'principal component' method. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy of 0.726 (based on anthropometric variables) and 0.697 (based on physiological variables) show a "middling" suitability (Kaiser & Rice, 1974) for both EFAs. A Varimax rotation led to the final solution, with variables sorted by the size of factor loadings related to a general factor. With this step, variables such as relative VO_2max (less descriptive than absolute VO_2max) and the maximum blood lactate concentration in classic step test (less descriptive than maximum blood lactate concentration in mobilization test or blood lactate concentrations after load) could be excluded with a minimal loss of information. Based on these results, a stepwise multiple linear regression analysis (backward method, default exclusion criteria: probability of F to remove ≥ 0.1) was used to detect the relationships between independent variables and overall race time in Olympic-distance triathlon. Each parameter had to be significant ($p < 0.05$). To avoid multicollinearity, Variance Inflation Factor (VIF) was checked with a cut-off of 10 (Hair, 1995). Additionally, the normality of residuals was examined via normal distribution plots, and residual independence and homoscedasticity were determined by plotting the residuals against the estimated data. Furthermore, Cook's Distance was used with a cut-off ≥ 1 to identify and remove influential cases in case of homoscedasticity (Heiberger & Holland, 2004). The coefficient of determination (percentage of variance explained; R^2) and the standard error of the estimate (SEE) were used to evaluate the models. The adjusted R^2, in particular, allows a comparison between several MLR models, considering the number of variables used in each case.

The expertise-based non-linear approach also consisted of two consecutive steps: A dominance paired comparison was first conducted to identify performance-relevant parameters, based on the expertise of four professional German triathlon coaches, followed by the computation of artificial neural networks (ANNs) to determine potential prediction models. A dominance paired comparison helped raters to prioritize influencing variables in a systematic

and objective way. Thus, personal preferences and subjective influences could be avoided with regard to prioritization of the independent variables. Each national triathlon coach had to rate the significance of each variable against all others. The overall sum score was used for the final prioritization. To ensure solvability of the numerous connections in the artificial neural networks with regard to the sample size, the five most relevant variables were finally selected. Two dominance paired comparisons were conducted (for anthropometric and physiological variables separately). The selected relevant parameters were used to compute two-layer feedforward artificial neural networks as a non-linear approach to predict overall race time in Olympic-distance triathlon of elite triathletes. In general, ANNs have the ability to learn relationships between variables in complex, non-linear contexts. A multi-layer perceptron with one input layer (one input neuron for each independent variable), one hidden layer (two neurons), and one output layer (one neuron for the dependent variable, normalized overall race time), as shown in Figure 5.1, was selected as a universal approach (Hornik, Stinchcombe, & White, 1989). To minimize mean squared error, Levenberg-Marquardt algorithm was used as a training algorithm due to its attribute of robustness (Marquardt, 1963). The dataset was randomly divided into datasets for training (80% of the sample), validation (10% of the sample), and testing (10% of the sample). In the training process, a set of input-output patterns was used to adjust the weights of all interconnections between the neurons in an ANN. The validation set is mainly used to avoid over fitting in the learning process. The test data is finally used to predict an output, which should be within an acceptable margin compared to the actually given output. The presented results below involving the entire sample. The coefficient of determination (R^2) and the standard error of the estimate (SEE) were used to evaluate the models. The SEE was calculated to ensure comparability between both computational approaches, even though it is not common in ANNs.

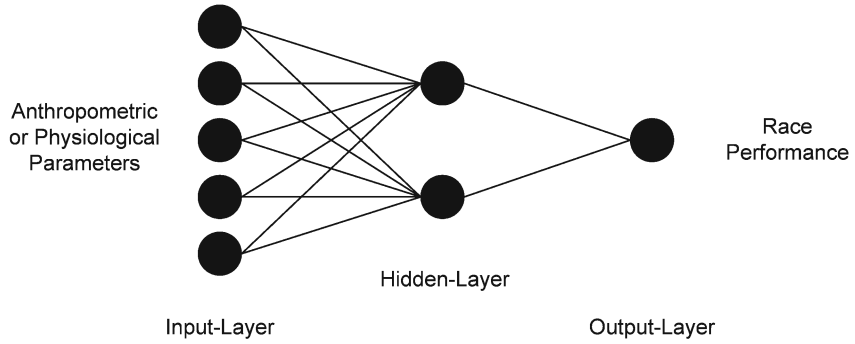


Figure 5.1 Internal characteristics of an Artificial Neural Network consisting of five Input-Neurons, two Hidden-Neurons and one Output-Neuron.

SPSS Statistics (Version 22, IBM) and MATLAB (Version R2015b, MathWorks) with Neural Network Toolbox were used for statistical analyses. The level of significance was set to $p < 0.05$.

5.4 Results

Normalization of race time

The normalization of race times yielded a mean ± standard deviation of overall race time in Olympic-distance triathlon of 6,827.57 ± 192.56 [s] (approximately 1:54 h) for male elite triathletes. The conversion factor for sprint distance race times into an Olympic distance equivalent is 2.08 ± 0.03.

Preselection of variables

Exploratory factor analysis

EFA yielded four factors in case of anthropometric variables and three factors in case of physiological variables. Tables 5.2 and 5.3 show the variables sorted by the size of factor loadings related to the general factor, and after Varimax rotation. A suppression level of 0.5 was used to point out decisive variables (Hair, 1995) and to exclude variables with poorer explanation to one general factor (e.g. relative VO_2max).

Table 5.2 Varimax rotated factor loadings of exploratory factor analysis for anthropometric variables.

	Factor 1	Factor 2	Factor 3	Factor 4
BMI	.908			
Quetelet Index	.863			
Lean body mass	.777		.551	
Thorax depth	.613			
Body fat %		.968		
Body fat kg		.839		
Pelvis width		.583		
Body height			.930	
Seat height			.802	
Shoulder width				.888
Thorax width				.852
Possible factor interpretation	„body composition“	„body fat“	„height“	„segment width“

Table 5.3 Varimax rotated factor loadings of exploratory factor analysis for physiological variables.

	Factor 1	Factor 2	Factor 3
LA max mobi	.966		
BLC 6 min	.959		
BLC 10 min	.921		
BLC 3 min	.824		
VCO_2 max mobi		.870	
VO_2 max		.834	
PL3		.708	.591
Max running pace		.682	.523
RR max mobi			.788
Max distance mobi			.689
Max running pace mobi			.684
RMV max mobi			.655
Possible factor interpretation	„lactate“	„respiration and velocity“	„respiration and velocity mobi“

Most of the variables showed a strong relationship to one single factor. Lean body mass was related to body composition and height. Running pace at 3-mmol·L^{-1} blood lactate and maximum running pace in classic step test were both related to respiration and velocity as well as respiration and velocity in the mobilization test. The variables in Table 5.2 and Table 5.3 were therefore used to compute the following multiple linear regression analyses.

Dominance paired comparison

The dominance paired comparisons as a second preselection approach yielded five parameters concerning anthropometric and physiological parameters, shown in Table 5.4. Anthropometric parameters mostly described the body composition of the athletes. The selection of physiological parameters consisted of respiratory, lactate, and velocity-related variables.

Table 5.4 Results of dominance paired comparisons with national triathlon coaches for anthropometric and physiological variables.

Five most important parameters each	
Anthropometric	**Physiological**
Body weight [kg]	Absolute VO_2max [$mL \cdot min^{-1}$]
BMI [$kg \cdot m^{-2}$]	Relative VO_2max [$mL \cdot min^{-1} \cdot kg^{-1}$]
Body fat [%]	Running pace at 3-$mmol \cdot L^{-1}$ blood lactate [$m \cdot s^{-1}$]
Body fat [kg]	Max running pace [$m \cdot s^{-1}$]
Lean body mass [kg]	Max running pace in mobilization test [$m \cdot s^{-1}$]

Performance prediction models

Multiple linear regression

Statistical assumptions of multiple linear regression (normal distribution of regression residuals, homoscedasticity) were validated by assessment and testing of residuals. Multiple linear regression analysis after EFA revealed that, among anthropometric parameters, pelvis width and shoulder width were the best predictors of overall race time in Olympic-distance triathlon. The R^2 showed an explanation of variance of 40.5% of overall race time by the anthropometric based model. The multiple linear regression model after EFA based on

physiological parameters included running pace at 3-mmol·L^{-1} blood lactate, maximum lactate, and maximum respiratory rate in the mobilization test. The physiological model showed a higher R^2 of 66.5, with a lower SEE in comparison (Table 5.5). The results led to two equations predicting overall Olympic-distance triathlon race time for male elite triathletes:

Predicted race time [s] based on anthropometric variables = 7643.56 - 80.889 × (pelvis width [cm]) + 37.388 × (shoulder width [cm])

Predicted race time [s] based on physiological variables = 8521.03 + 8.556 × (maximum respiratory rate [breaths·min^{-1}]) - 332.80 × (running pace at 3-mmol·L^{-1} blood lactate [m·s^{-1}]) - 61.658 × (maximum blood lactate [mmol·L^{-1}])

Table 5.5 Parameter and model estimates of multiple linear regression analyses for male elite triathletes.

	Value	β-coefficient	R^2	Adjusted R^2	SEE [s]	p-value	VIF
EFA + MLR (anthropometric)			0.405	0.351	155.14	0.003	
Constant	7643.56						
SW	37.39	0.434				0.025	1.199
PW	-80.89	-0.674				0.001	1.199
EFA + MLR (physiological)			0.665	0.582	117.27	0.003	
Constant	8521.03						
PL3	-332.80	-0.474				0.018	1.065
LA max Mo	-61.66	-0.450				0.028	1.161
RR max Mo	8.56	0.505				0.014	1.103

Notes: SEE = standard error of the estimate, VIF = Variance Inflation Factor, SW = shoulder width; PW = pelvis width; PL3 = running pace at 3-mmol·L^{-1} blood lactate; LA max Mo = maximum blood lactate in mobilization test; RR max Mo = maximum respiratory rate in mobilization test; general format for multiple regression equation: y = constant + value1 × variable1 + value2 × variable2 + …

Artificial neural networks

The artificial neural network computed after dominance paired comparison using the anthropometric variables body weight, BMI, lean body mass, and absolute and relative body fat explained 43.4% of the variance in overall race time (R^2 = 0.43; SEE = 144.56 [s]). The

artificial neural network after dominance paired comparison using the physiological variables maximum running pace, running pace at 3-mmol·L^{-1} blood lactate, absolute and relative VO_2max, and maximum running pace in mobilization test explained 86.2% of the variance in overall race time (R^2 = 0.86; SEE = 91.82 [s]). Both artificial neural networks, with their specific characteristics, could be used to predict overall Olympic-distance triathlon race time based on a single input pattern.

5.5 Discussion and Conclusion

The aim of the current study was to assess whether overall Olympic-distance triathlon race time of male elite athletes could be predicted using regular performance diagnostics, and to compare two different computational approaches. Anthropometric and physiological variables measured during routine laboratory tests provided a database for the prediction, without interfering with individual training programs and season calendars of the elite triathletes. Both the combinations assessed (an exploratory factor analysis and multiple linear regression, and a dominance paired comparison and artificial neural network), yielded prediction models of overall triathlon race time.

Assessment of parameters

Table 5.1 shows the homogeneous appearance of elite triathletes within the sample. Anthropometric characteristics had only small variations, except for body fat [% and kg], which became obvious because of a larger CV. The physiological variables showed a partially similar distribution: VO_2max [mL·min^{-1} and mL·min^{-1}·kg^{-1}] showed a small variation because of its premising function in samples consisting of elite triathletes. Maximum lactate value and maximum respiratory rate in mobilization tests as well as the lactate values after load showed higher CVs. Different individual strengths in the three disciplines likely affected the results of running-specific step tests.

The selection of parameters has an important effect on the prediction results. Body height, body weight, and resulting BMI, as well as age were in accordance to the reports of Hue (2003), Schabort et al. (2000) and Hue, Le Gallais, Boussana, Chollet, and Prefaut (2000) (slightly lower body height and weight) as well as Ackland, Blanksby, Landers, and Smith (1998) (slightly older, smaller, and lighter). VO_2max [mL·min^{-1}·kg^{-1}] as gross criterion of endurance performance was slightly lower than that reported by Hue, Le Gallais, Chollet, and

Préfaut (2000) and Schabort et al. (2000), and similar to that reported by Hue (2003). Lactate values could not be compared because of various specifications such as defined bounds, running paces, or power outputs while cycling. McLaughlin et al. (2010) showed a considerably slower running pace at 3-mmol·L^{-1} blood lactate (4.41 m·s^{-1}), which is likely because their sample consisted of well-trained but non-elite triathletes. In summary, our set of variables seemed to be accurately selected and showed values similar to those reported in other studies using male elite triathletes.

Normalization of race time

The mean and standard deviation of normalized overall race time in Olympic-distance triathlon for male triathletes (6,827.57 ± 192.56 [s]; approximately 1:54 h) calculated in this study were comparable to those reported by Landers et al. (2000). A closer look at the Top 10 ranked athletes in the WTS from 2009 until 2012 showed that the mean ± SD of overall race time were consistent with the values used, considering that German elite triathletes commonly have a Top 20 position in WTS races.

Preselection of variables

The purely statistical approach using exploratory factor analysis is devoid of subjective influences, and prioritizes variables based on their influence to a general factor. Therefore, variables with a small variance will typically be sorted out. This could be why only two anthropometric and three physiological variables provided a significant contribution in the computed linear regression models, which is unfavorable because it could result in a lack of explanation of variance. Additionally, a sufficient sized sample must be available. In contrast, the dominance paired comparison does not have high demands regarding the number of coaches consulted. An objective prioritization based on professional expertise seems to be a plausible preprocessing step, if combined with ANNs to model complex and non-linear patterns. A reduction of variables similar to an exploratory factor analysis could therefore not be achieved and the maximum number of variables used in the computational model must be specified manually. As a prime example, VO_2max is a common parameter characterizing the endurance of heterogeneous groups and predicting performance (Butts et al., 1991; Miura et al., 1997). This could be why national triathlon coaches select absolute and relative VO_2max as predictive parameters. In homogenous groups, VO_2max normally has only premising instead of predicting character, because of the small variation (Sleivert & Rowlands, 1996). This could possibly be

a drawback of subjective assessments compared to the exploratory factor analysis as a purely statistical approach, which sorted out VO_2max.

Performance prediction

Landers et al. (2000) underlined the importance of identifying parameters predicting race performance. Besides potentially supporting the creation of new training programs, the information provided by performance prediction models could also be used in the field of talent diagnostics. Considering that the small and homogenous sample limits generalizability, the reported performance prediction models showed that specific influencing parameters generally exist. These parameters could allow more objective talent selection by defining minimum physical requirements (e.g. for specific age groups). Talent identification programs could also use information on advantageous anthropometric requirements to direct young athletes to the sport of triathlon. The design of training programs could be influenced by focusing on optimal training levels (e.g. to improve identified lactate levels).

The combination of a professional triathlon coach survey and ANNs provided two performance prediction models with medium and large explained percentages of variance, respectively (anthropometric: $R^2 = 0.43$; physiological: $R^2 = 0.86$). In comparison, the MLR showed clearly poorer results (anthropometric: $R^2 = 0.41$; physiological: $R^2 = 0.67$). Therefore, the predictions using ANNs outperformed those from the purely statistical approach comprising factor analysis and multiple regressions. Furthermore, a closer look at the SEE (based on MLR: anthropometric: 155.14 [s]; physiological: 117.27 [s]; based on ANN: anthropometric: 144.56 [s]; physiological: 92.82 [s]) revealed that these are smaller than the performance variation of individual athletes (e.g. SD of race time of Javier Gomez during WTS 2014: 200.93 [s]) and of the Top 10 athletes in WTS 2014 (SD of race time: 217.83 [s]), which confirms the results of the performance prediction models.

The first MLR model yielded the anthropometric parameters pelvis width and shoulder width as significant predictors of overall race time in elite Olympic-distance triathlon. These two variables could theoretically have an impact on running economy (Barnes & Kilding, 2015). Shoulder width seems to be a predictor for swimming performance, which is necessary to be in the first group getting out of the water. In contrast, pelvis width should be smaller, which was already shown for distance runners (Anderson, 1996; Williams, Cavanagh, & Ziff, 1987), and is therefore plausible in connection with the importance of the run part in elite

Olympic-distance triathlon. The ANN model used the five anthropometric parameters, body weight, BMI, lean body mass, and absolute as well as relative body fat, which were identified through dominance paired comparison as most important for overall race time in elite Olympic-distance triathlon. Parameters such as body height or BMI normally show too small variations to get significant results in small and homogenous samples (Table 5.1). In the present study, the triathlon coaches were partially responsible for young athletes in national squads, where the mentioned variables have a higher influence and a greater variance than in elite triathletes.

The second MLR model yielded the physiological parameters running pace at 3-mmol·L^{-1} blood lactate, maximum lactate, and maximum respiratory rate in mobilization test as significant predictors of overall race time in elite Olympic-distance triathlon. The ANN model used the five physiological parameters, maximum running pace, running pace at 3-mmol·L^{-1} blood lactate, maximum running pace in mobilization test, and absolute as well as relative VO_2max identified by a dominance paired comparison as most important for overall race time in elite Olympic-distance triathlon. Both approaches identified running pace at 3 mmol·L^{-1} blood lactate as important for overall race time. This variable describes the possibility of an athlete to realize a higher pace with the same utilization of metabolic processes. The mentioned lactate interval is mainly used while competing in Olympic-distance triathlon, and therefore leads directly to a faster race time. Some studies identified VO_2max or ventilatory thresholds as important for performance prediction in heterogeneous groups (Butts et al., 1991; Miura et al., 1997). This could be why national triathlon coaches select absolute and relative VO_2max as predictive parameters, particularly for young athletes. In homogenous groups, VO_2max normally has only premising instead of predicting character because of only small variation in VO_2max (Sleivert & Rowlands, 1996). In contrast, maximum values such as maximum lactate and maximum respiratory rate in mobilization as well as maximum running pace allow a valid assessment of anaerobic capacities. EFA and MLR as well as DPC and ANN used these kind of variables, which seems to be plausible: nearly all races of the WTS were actually won during the running discipline, especially in the final spurt. High lactate values and high running paces could therefore be important factors for overall race time. The maximum respiratory rate could also influence this kind of race situation, because a selectively high oxygen uptake is required to prevent the formation of lactate.

Limitations

The sample in this study was elite, small, and homogenous, which limits generalizability to other triathlete cohorts. However, generalizability of results to other triathlete cohorts was not the aim of this study; we focused on elite athletes. National squads for triathlon are generally small; compared to other sports, only 4 - 5 athletes are included in the elite Olympic-distance triathlon squad each year, and elite athletes are often reluctant to participate in experiments. Additionally, individual training schedules and differences in season calendars complicate experimental laboratory studies with this special population. Therefore, one of our aims was to assess whether overall Olympic-distance triathlon race time of elite athletes could be predicted using regular performance diagnostics. To overcome the drawback of having a small number of available athletes, we developed an algorithm that helped us to increase the number of datasets used in the statistical analyses, by collecting performance diagnostics over a period of four years. We only included data sets if two requirements were fulfilled: (1) availability of a complete set of variables from the laboratory tests and (2) a finished Olympic-distance triathlon race within 8 weeks after each single performance diagnostic. However, despite this improvement, no prediction models could be determined by combining anthropometric and physiological variables due to the sample size.

Further, we did not take the results of laboratory tests in swimming and cycling into account. This was because the protocols slightly changed over the period of interest (2008-2012), which led to inconsistent data sets. Therefore, we decided to exclude these sources of information to avoid a further reduction of the sample size. Nevertheless, the general tactical behavior in elite Olympic-distance triathlon races allows the use of running diagnostics alone to generate meaningful results. The swimming and cycling disciplines in elite Olympic-distance triathlon only have premising function whereas the running discipline is normally the critical factor for success (Fröhlich et al., 2008; Vleck et al., 2006). Therefore, the results of the present prediction models, with only running diagnostics as physiological parameter, could be considered appropriate. However, we are planning to incorporate more comprehensive data sets (swimming, cycling, and running diagnostics) in future studies, since the test protocols for swimming and cycling have now been standardized.

Conclusion

Two different approaches to determine performance prediction models of overall race time in elite Olympic-distance triathlon were developed without interfering with individual training programs, through triathlete participation in a standardized experimental study and the identification of important parameters collected through laboratory tests. According to these models, the combination of an exploratory factor analysis and multiple linear regression provided appropriate explanations of variance in case of anthropometric (R^2 = 0.41) and physiological (R^2 = 0.67) variables. These were selected with a strong analytical procedure, using variables with a greater variance. The corresponding SEEs of 155.14 [s] (anthropometric variables) and 117.27 [s] (physiological variables) showed acceptable results when compared to performance variations of individual athletes (e.g. SD of race time of Javier Gomez during WTS 2014: 200.93 [s]) and of the Top 10 athletes in WTS 2014 (SD of race time: 217.83 [s]), and therefore confirmed the results of the performance prediction models.

The advantage of ANNs compared to MLRs is the possibility to take non-linear relationships into account and to model more complex patterns. Therefore, the trained ANNs considering expertise of professional triathlon coaches through dominance paired comparison as preselection method could preferably be used to predict individual race time based on the values of an actual performance diagnostic. The explanations of variance and the standard errors of the estimate in case of anthropometric (R^2 = 0.43; SEE = 144.56 [s]) and physiological (R^2 = 0.86; SEE = 91.82 [s]) variables were an improvement over those of the purely statistical approach.

Finally, the results of the present study show that future research should focus on collecting larger samples, and on the developmental process of young triathletes, with a focus on the influence on performance prediction models. Information from previous races, such as overall or split times and training indicators, could also enhance prediction (Gilinsky et al., 2014).

6 Study III: Modeling the Performance Structure of Elite Triathlon: A Structural Equation Approach

6.1 Abstract

This study analyzed the performance structure of the Olympic-distance triathlon using a scientifically based approach to identify performance relevant parameters that could be beneficial for optimizing training programs, especially in sports with high amounts of training, such as triathlon. The computational approach of structural equation modeling includes the necessary steps of hierarchization, internal order and prioritization needed for the performance modeling process. Valid structural equation models were developed based on anthropometric and physiological parameters collected through the regular performance diagnostics of male elite triathletes. The model derived from the research literature and theoretical considerations, consisted of two anthropometric variables (body weight and BMI) and three physiological variables (running pace at 3-mmol·L^{-1} blood lactate, maximum running pace, relative VO_2max), had a good model fit (CFI = 0.99, TLI = 0.99, RMSEA = 0.03), and yielded the best result of the three models tested in the study. It showed the effects of anthropometric and physiological parameters on overall race performance were very similar. The results demonstrate that structural equation modeling can be a powerful analytical procedure for use in the field of training science and that it is able to identify performance relevant parameters in elite Olympic-distance triathlon. To ensure the transfer of these results to training, the identified anthropometric parameters should mainly be interpreted as prerequisites for performance, whereas the physiological variables are directly applicable to specific training programs.

6.2 Introduction

Triathlon is a classic endurance sport that consists of the sport disciplines of swimming, cycling and running, with individual events that vary greatly in distance (sprint-, short-, middle- and long-distance). Therefore, high training loads, which incorporate 9 to 12 training sessions for more than 20 hours per week (Friel & Vance, 2013) are indispensable for elite athletes, mainly independent of their preferred race distance. This high level of training should optimally be supported by a well-founded and structured training program. Consequently, it is important to identify performance relevant parameters, such as anthropometric and physiological parameters based on laboratory tests (Landers et al., 2000; Schabort et al., 2000), as a scientific basis for a well-structured training program.

Several studies have shown the importance of maximum oxygen uptake (VO_2max) and anaerobic thresholds (Millet et al., 2009, 2011) in endurance running or running in a triathlon. These parameters have significant correlations with race performance (Bassett, 2000; McLaughlin et al., 2010). Similar results have been found for swimming and cycling (Millet et al., 2009; Sleivert & Rowlands, 1996). However, these variables are only prerequisites for optimal performance; they are not performance differentiating in homogenous samples because of the small variation between athletes (Bassett, 2000; Sleivert & Rowlands, 1996; Stratton et al., 2009). Nonetheless, blood lactate concentrations from treadmill or cycle ergometer tests are useful parameters for predicting triathlon performance independent of an athlete's performance level (Hoffmann, Moeller, Seidel, & Stein, 2017; Schabort et al., 2000; Van Schuylenbergh et al., 2004). Moreover, physiological factors and anthropometric variables, such as percent body fat, body mass index (BMI), and the circumferences of several parts of the body could be important for performance in triathlon races (Knechtle et al., 2011).

In contrast to research that focuses on the individual disciplines of swimming, cycling and running, research on the indicators of overall triathlon performance has, to date, focused on predicting performance. Therefore, it has commonly used multiple linear regression analyses to develop equations to predict overall triathlon race time, using physiological and anthropometric parameters identified as predictive variables by various studies. Schabort et al. (2000), for example, identified body mass, blood lactate concentration during steady-state cycling, peak power output and VO_2peak during cycling, blood lactate while running at 15 $km \cdot h^{-1}$ and peak treadmill-running velocity as the best predictors of running time. Van

Schuylenbergh et al. (2004) found similar parameters for triathlon performance such as swimming and running velocity at maximal lactate steady state (MLSS), blood lactate concentration when running at MLSS, peak blood lactate concentration during a graded treadmill test and the oxygen uptake during a graded bicycle test at the 4-mmol·L^{-1} blood lactate threshold. Similar results supporting these findings (pelvis and shoulder width, as well as maximum respiratory rate, running pace at 3-mmol·L^{-1} blood lactate and maximum blood lactate during running as predicting parameters) were found by Hoffmann et al. (2017).

The results provided by these and similar studies can be seen as a first important step to uncover the performance structure of triathlon. The term performance structure, which was first characterized by Letzelter and Letzelter (1982) and Hohmann and Brack (1983), describes the situation where the factors influencing a sport are identified and prioritized using statistical methods. Accordingly, uncovering performance structure can provide a scientific basis for training programs and adjustments to them when necessary. Letzelter and Letzelter (1982) are of the opinion that structuring the performance of a sport is one of the main objectives of training science following three fundamental and irreversible steps of hierarchization based on theory, internal order and prioritization.

Structural equation modeling (SEM) is of special interest for uncovering the performance structure of triathlon, since this computational approach involves the steps of hierarchization, internal order and prioritization. Therefore, SEM delivers considerably more information about the performance structure than regression models and makes it possible to identify indicators that explain race performance through a more complex modeling process. SEM, which was introduced in the field of social and behavioral science (Hox & Bechger, 1998), merged three historical statistical traditions: path analysis, simultaneous-equation models and factor analysis (Rosseel, 2012). Today, the areas of application of SEM are diverse, including psychology, political science, education, business-related disciplines (Jais, 2007) and sport science (Felser et al., 2015; Ostrowski & Pfeiffer, 2007).

To the best of our knowledge, the performance structure of triathlon based on anthropometric and physiological parameters has not yet been investigated. Moreover, previous studies on triathlon performance mostly focused on recreational and not elite athletes (Kohrt et al., 1987; Millet et al., 2011; Miura et al., 1997; Sleivert & Wenger, 1993), probably due to the limited availability of elite athletes for experimental research. Based on the premise of using data of regular performance diagnostics to overcome this limitation, the aim of this study was

to analyze the performance structure of Olympic-distance triathlon of elite athletes using structural equation modeling (SEM). Several anthropometric and physiological variables measured during regular performance diagnostics over a period of four years of German male elite Olympic-distance triathletes provided the necessary data pool. After applying multiple linear regression models and artificial neural networks to predict individual race performance in a previous study (Hoffmann et al., 2017), the present study aimed to develop more complex models of triathlon performance taking anthropometric and physiological parameters collectively into account using SEM. In contrast to our earlier work (Hoffmann et al., 2017) this approach allowed us to compare the influence of latent variables in a theory-based model using consistent findings from the research literature.

6.3 Methods

Subjects

The same dataset used by Hoffmann et al. (2017) was used because of its high quality data on homogenous elite subjects. Eleven male German elite triathletes (age: 23.38 ± 2.79 years) competing in national or international championships were part of the dataset. Written informed consent, in the form of an athlete agreement and a cooperation agreement with the Institute for Applied Training Science (Leipzig, Germany) were mandatory. Participation in the performance diagnostics was voluntary and the triathletes could opt out at any time. After data acquisition, all the statistical analyses were conducted anonymously. Table 6.1 shows the descriptive characteristics (mean and standard deviation (SD) and coefficient of variation (CV = (SD/Mean)*100)) of the triathletes.

Table 6.1 Descriptive statistics for variables on German elite triathletes (N = 11).

	Mean ± SD	**CV (%)**
Anthropometric		
Age [yrs]	23.38 ± 2.79	11.93
Body height [cm]	187.0 ± 2.90	1.55
Body weight [kg]	74.46 ± 4.28	5.75
Seat height [cm]	96.38 ± 1.59	1.65
Shoulder width [cm]	40.16 ± 2.24	5.58
Pelvis width [cm]	28.65 ± 1.61	5.62
Thorax width [cm]	28.27 ± 1.29	4.56
Thorax depth [cm]	21.06 ± 1.41	6.70
Quetelet Index [$g \cdot cm^{-1}$]	398.15 ± 21.32	5.35
BMI [$kg \cdot m^{-2}$]	21.29 ± 1.17	5.50
Body fat [%]	10.70 ± 1.36	12.71
Body fat [kg]	8.00 ± 1.35	16.88
Lean body mass [kg]	66.46 ± 3.27	4.92
Physiological		
VO_2max [$mL \cdot min^{-1}$]	5457.67 ± 292.56	5.36
VO_2max [$mL \cdot min^{-1} \cdot kg^{-1}$]	72.02 ± 4.29	5.96
PL3 [$m \cdot s^{-1}$]	5.08 ± 0.23	4.53
Max running pace [$m \cdot s^{-1}$]	5.22 ± 0.27	5.17
Max running pace mobi [$m \cdot s^{-1}$]	6.92 ± 0.17	2.46
LA max mobi [$mmol \cdot L^{-1}$]	9.18 ± 1.30	14.16
VCO_2 max mobi [mL]	6472.75 ± 431.74	6.67
Max distance mobi [m]	1762.69 ± 136.70	7.76

RMV max mobi [mL·min^{-1}]	187.73 ± 12.40	6.61
RR max mobi [breaths·min^{-1}]	63.18 ± 10.10	15.99
BLC 3 min [mmol·L^{-1}]	8.08 ± 1.31	16.21
BLC 6 min [mmol·L^{-1}]	9.13 ± 1.29	14.13
BLC 10 min [mmol·L^{-1}]	8.62 ± 1.38	16.01
Normalized overall race time Olympic distance [min]	113.79 ± 3.21	2.82

Notes: PL3 = running pace at 3-mmol·L^{-1} blood lactate; mobi = mobilization test; LA max mobi = maximum blood lactate in mobilization test; RMV max mobi = maximum respiratory minute volume in mobilization test; RR max mobi = maximum respiratory rate in mobilization test; BLC 3, 6, 10 min = blood lactate concentration 3, 6, 10 min after load in mobilization test.

Experimental Procedure

The data used in this study were derived from laboratory tests performed between 2008 and 2012 at the Institute for Applied Training Science (Leipzig, Germany) within the framework of national squad investigations. Overall, twenty-five sets of variables from eleven triathletes were used. Anthropometric and physiological parameters used in the analyses (Table 6.1) were collected through routine and highly standardized laboratory tests. A more detailed description of the data acquisition and preprocessing steps, which were necessary to develop the structural equation models, were presented in the study of Hoffmann et al. (2017).

Data analysis

Normalization of race time

As described in Hoffmann et al. (2017), normalization was necessary to obtain comparable individual race times independent of the various triathlon races in which the subjects participated. The reference factor to normalize individual race times was calculated as the mean

value of the overall race times of the Top 10 athletes at the World Triathlon Series (WTS, International Triathlon Union) between 2009 and 2012.

reference factor = mean (overall race times of the Top 10 WTS athletes of all races within the WTS 2009, 2010, 2011, 2012)

Up to two races of the WTS are sprint-distance triathlons. To use these race times, the same approach was applied to determine a factor that transformed sprint-distance triathlon race time into an Olympic-distance equivalent.

The normalization yielded a mean ± SD of overall race time in an Olympic-distance triathlon of 6,827.57 ± 192.56 [s] (approximately 1:54 h) for male elite triathletes. The conversion factor for changing sprint-distance race times into an Olympic-distance equivalent was 2.08 ± 0.03. These results were comparable to those reported by Landers et al. (2000) and a closer look at the Top 10 ranked athletes in the WTS from 2009 until 2012 showed that the mean ± SD of overall race times were consistent with the values used, considering that German elite triathletes commonly have a Top 20 position in WTS races.

Preselection of variables

In general, the statistical analyses applied after race-time normalization built upon the results of three different preselection approaches described in the following paragraph. Given the available data, the general structure of the structural equation model consisted of two latent variables (anthropometric and physiological factor), which explained overall race performance. The latter is measured through the normalized race times. The latent constructs “anthropometric” and “physiological” were selected to compare their influence on race performance.

- “Theoretical preselection”: The core idea of modeling the performance structure in a sport is to build a theory-based model. Based on the research literature, including the studies mentioned in the introduction, the following variables comprise the measurement model of anthropometric and physiological (run) factors:
    - body weight, BMI and body fat (Hoffmann et al., 2017; Knechtle et al., 2011; Landers et al., 2000)
    - relative VO_2max, maximum running pace and running pace at 3-mmol·L^{-1} blood lactate (Hoffmann et al., 2017; Schabort et al., 2000; Suriano & Bishop, 2010)

- "Computational preselection": An exploratory factor analysis (EFA) was conducted to preselect relevant independent variables, using the principal-component method. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy of 0.726 (based on anthropometric variables) and 0.697 (based on physiological variables) showed a "middling" suitability (Kaiser & Rice, 1974) for both EFAs. Varimax rotation yielded a final solution of four anthropometric factors ("body composition", "body fat", "height" and "segment width") and three physiological factors ("lactate", "respiration and velocity" and "respiration and velocity in mobilization tests"). A suppression level of 0.5 was used to indicate decisive variables (Hair, 1995) and to exclude variables with poorer explanatory value for one general factor. Most of the variables showed a strong relationship to a single factor. A more detailed description of the EFA can be found in Hoffmann et al. (2017).
- "Expertise preselection": A dominance paired comparison was conducted to identify performance relevant parameters, based on the expertise of four professional German triathlon coaches (Hoffmann et al., 2017). The dominance paired comparison helped raters to prioritize influencing variables in a systematic and objective way. To ensure the ability to use SEM, given the sample size, two dominance paired comparisons were conducted separately that yielded the five most relevant anthropometric and physiological parameters (see Table 6.2). The anthropometric parameters mostly described the body composition of the athletes. The selected physiological parameters consisted of respiratory, lactate, and velocity-related variables.

Table 6.2 Results of dominance paired comparisons with national triathlon coaches for anthropometric and physiological variables (Hoffmann et al., 2017).

Five most important parameters	
Anthropometric	**Physiological**
Body weight [kg]	Absolute VO_2max [$mL \cdot min^{-1}$]
BMI [$kg \cdot m^{-2}$]	Relative VO_2max [$mL \cdot min^{-1} \cdot kg^{-1}$]
Body fat [%]	Running pace at 3-$mmol \cdot L^{-1}$ blood lactate [$m \cdot s^{-1}$]
Body fat [kg]	Max running pace [$m \cdot s^{-1}$]
Lean body mass [kg]	Max running pace in mobilization test [$m \cdot s^{-1}$]

These three different approaches of preselection provided the theoretical basis for SEM using performance relevant variables selected through theoretical considerations and through computational and expertise-based approaches.

Structural Equation Modeling

The sample size is considered small for using SEM; therefore, the application is exploratory in nature. As stated by MacCallum et al. (1999), a small sample size is not a basic obstacle to SEM, but it requires that the factors be well determined and the computations of the factor analysis or the structural equation model need to converge on an appropriate solution. To avoid further limiting the validity of the results, besides the small sample size, missing data (less than 10 %) were handled using full information maximum likelihood (FIML), which is a common method in SEM (Enders & Bandalos, 2001).

The conceptual framework of SEM consists of a measurement and a structural model. The measurement model consists of observed or measured variables, traditionally depicted as rectangles. The structural model consists of latent or unobserved variables, traditionally depicted as ovals. A line between two variables symbolizes the causal effect of a latent variable on an observed variable or a latent variable (Schreiber et al., 2006). R software (Version 3.4.3,

https://cran.r-project.org) with the Lavaan package was used to perform SEM (more information is provided by Rosseel, 2012).

Robust maximum likelihood (MLR) estimation was used with robust (Huber-White) standard errors and a scaled test statistic (asymptotically equal to the Yuan-Bentler test statistic). A "robust" estimator was selected due to the small sample size, which could otherwise result in biased standard errors and test statistics. Numerous goodness-of-fit indicators exist to assess model fit (Schreiber et al., 2006). Given the available data, the Comparative Fit Index (CFI), the Tucker-Lewis-Index (TLI), also known as the Non-Normed Fit Index (NNFI), the Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR) were used (for details, see: Hooper et al., 2008; Hox & Bechger, 1998; Hu & Bentler, 1999):

- CFI >= 0.95 for acceptance
- TLI >= 0.9 for acceptance and >= 0.95 for 'good' model fit
- RMSEA < 0.06 to 0.08 with confidence interval for acceptance
- SRMR < 0.08 for acceptance

6.4 Results

The preselection procedures provided the theoretical basis for SEM using performance relevant variables selected through theoretical considerations and computational and expertise-based approaches. The models are presented in the following sections.

SEM based on theoretical considerations

The structural equation model based on theoretical considerations and previous research results had a good fit to the data for describing the performance structure of triathlon, with TLI = 0.99, CFI = 0.99 and RMSEA = 0.03. The variable body fat [kg or %] was not included in the model because it reduced the model fit.

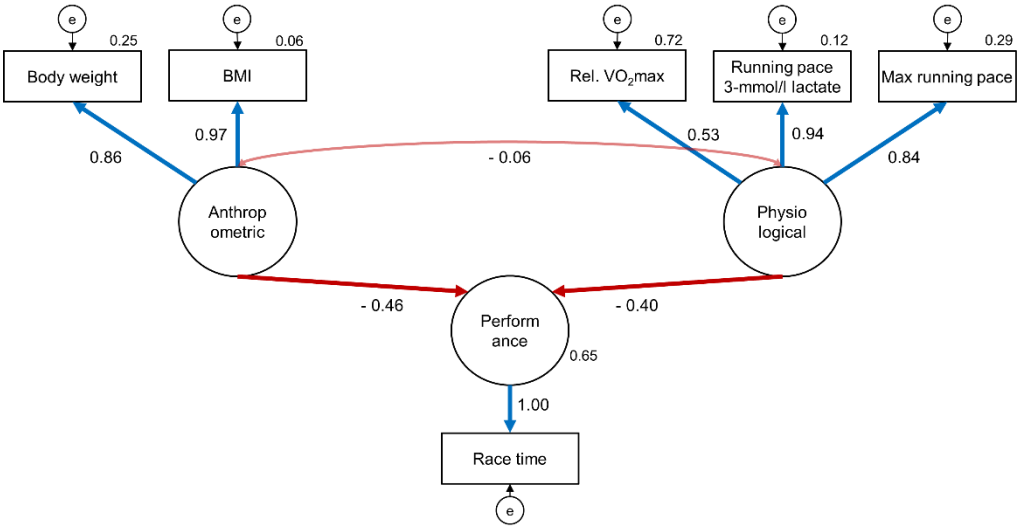


Figure 6.1 Structural equation model of anthropometric and physiological parameters chosen by theory-based preselection (completely standardized solution). TLI = 0.99; CFI = 0.99; RMSEA = 0.03; SRMR = 0.11; chi-square = 7.32; degrees of freedom = 7.

The structural part of the model consists of the latent constructs Anthropometric and Physiological (ovals) that influence performance. These latent constructs are causally linked to the performance parameters (rectangles), which are observed variables and therefore part of the measurement model (Figure 6.1). The standardized coefficients shown in Figure 6.1 indicate, for example, if the maximum running pace increased by 0.84 SD, with all other parameters held constant, that the factor Physiological would be expected to increase by 1 SD. Therefore, all the performance parameters in the SEM presented in Figure 6.1 show high and comparable

effects to their respective latent construct, despite of the relative VO_2max that is less important. The similar coefficients of Anthropometrics (standardized coefficient = -0.46) and Physiological (standardized coefficient = -0.40) for performance show that these two latent variables have similar negative effects on overall race time. This means that larger values of the latent constructs Anthropometric and Physiological provoke a lower overall race time by which the latent construct Performance is defined. Table 6.3 further shows the unstandardized coefficients of each parameter, which are expressed in their original units. This allows a direct interpretation: if, for example, body weight increased by 3.68 kg, while all other parameters were constant, the related Anthropometric factor would increase by 1 unit. The variance accounted for by a latent construct can be calculated as 1 minus the variance of a specific observed variable (the small number above each rectangle), e.g., Physiological accounts for 71 % of the variance in maximum running pace.

Table 6.3 Standardized and unstandardized coefficients of SEM derived from theory-based preselection.

Observed variable	**Latent construct**	**Unstandardized coefficient**	**Standardized coefficient**
BMI	Anthropometric	1.14	0.97
Body weight	Anthropometric	3.68	0.86
PL3	Physiological	0.25	0.94
Max running pace	Physiological	0.23	0.84
Rel. VO_2max	Physiological	2.27	0.53

Notes: PL3 = running pace at 3-mmol·L^{-1} blood lactate

SEM based on computational preselection

Based on EFA preselection, one structural equation model could be found that combined anthropometric and physiological parameters. The model had a good fit, with TLI = 0.99, CFI = 0.99 and RMSEA = 0.06, but it could not be accepted due to negative variances for some of the observed variables.

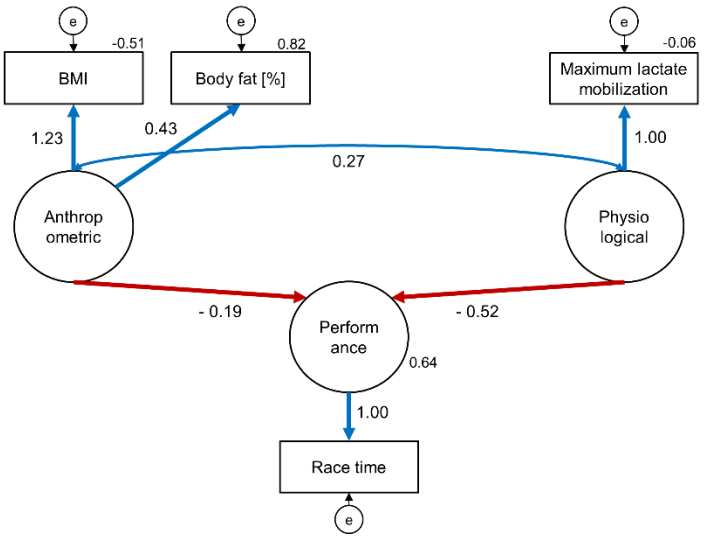


Figure 6.2 Structural equation model of anthropometric and physiological parameters chosen by EFA (completely standardized solution). TLI = 0.99; CFI = 0.99; RMSEA = 0.04; SRMR = 0.04; chi-square = 1.08; degrees of freedom = 1.

The structural part of the model is the same as that shown in Figure 6.1. Table 6.4 shows the standardized and the unstandardized coefficients of each parameter. Due to negative variances for the observed variables BMI and the maximum lactate in mobilization test, the structural equation model had to be rejected.

Table 6.4 Standardized and unstandardized coefficients of SEM derived from EFA preselection.

Observed variable	Latent construct	Unstandardized coefficient	Standardized coefficient
BMI	Anthropometric	1.44	1.23
Body fat [%]	Anthropometric	0.59	0.43
LA max mobi	Physiological	1.30	1

Notes: LA max mobi = maximum blood lactate in mobilization test

SEM based on preselection by professional expertise

Based on preselection by the dominance paired comparison, the following structural equation model was tested, which also had a good fit to the data, with TLI = 0.95, CFI = 0.97 and RMSEA = 0.11.

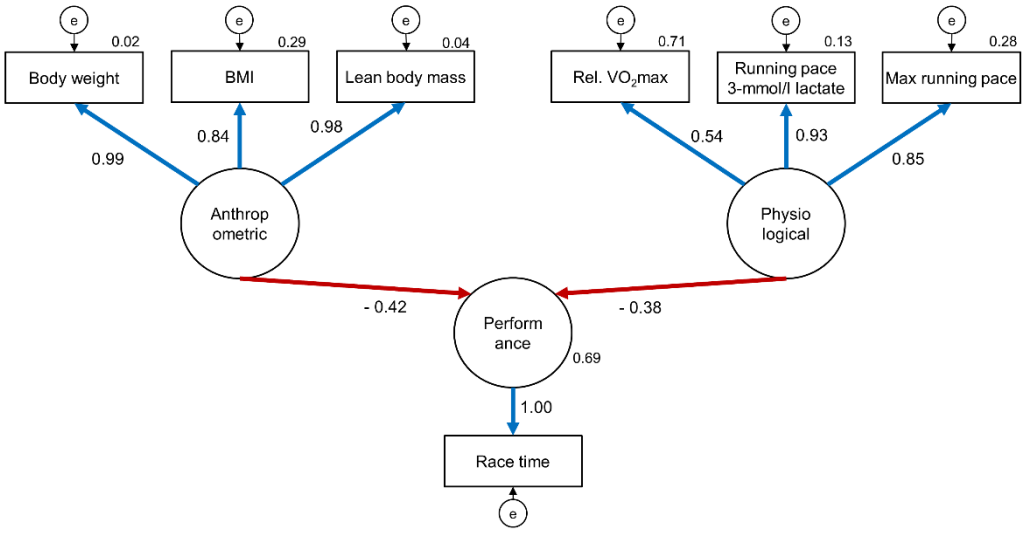


Figure 6.3 Structural equation model of anthropometric and physiological parameters chosen by dominance paired comparison (completely standardized solution). TLI = 0.95; CFI = 0.97; RMSEA = 0.11; SRMR = 0.1; chi-square = 16.60; degrees of freedom = 12.

The structural part of the model is again the same. The standardized coefficients in Figure 6.3 show, for example, if body weight increased by 0.99 SD while all other parameters were constant, the factor Anthropometric would be expected to increase by 1 SD. Again, all the performance parameters in the SEM presented in Figure 6.3 show high effects to their respective latent construct, despite of the relative VO_2max. The coefficients for Anthropometric (standardized coefficient = -0.42) and Physiological (standardized coefficient = -0.38) indicate they have nearly the same effect on performance. Larger values of these latent constructs provoke a lower overall race time by which the construct Performance is defined. Table 6.5 shows the unstandardized coefficients of each parameter: e.g. if body fat [%] increased by 4.23 while all other parameters were constant, the Anthropometric factor would increase by 1 unit. The variance accounted for by a latent construct can again be calculated as 1 minus the variance of a specific observed variable, e.g. Physiological accounts for 87% of the variance in running pace at 3-mmol·L^{-1} blood lactate.

Table 6.5 Standardized and unstandardized coefficients of SEM derived from dominance paired comparisons.

Observed variable	**Latent construct**	**Unstandardized coefficient**	**Standardized coefficient**
BMI	Anthropometric	0.98	0.84
Body weight	Anthropometric	4.23	0.99
Lean body mass	Anthropometric	3.21	0.98
PL3	Physiological	0.25	0.93
Max running pace	Physiological	0.23	0.85
Rel. VO_2max	Physiological	2.32	0.54

Notes: PL3 = running pace at 3-mmol·L^{-1} blood lactate

6.5 Discussion

The aim of the current study was to assess the performance structure of the Olympic-distance triathlon using anthropometric and physiological parameters of male elite triathletes. Anthropometric and physiological variables measured during routine laboratory tests provided the database without interfering with individual training programs or season calendars of the elite triathletes.

Three structural equation models based on different parameter preselection approaches were tested:

- A model based on theoretical considerations derived from the existing research literature, which corresponded to the usual approach to SEM;
- An alternative model based on preselection using EFA as a computational approach choosing the most relevant parameters; and
- A model based on the preselection of the most relevant parameters by the expertise of professional triathlon coaches.

The model derived from the existing literature and theoretical considerations, which consisted of two anthropometric variables and three physiological variables in the measurement model and two latent variables (Anthropometrics and Physiology) in the structural model, provided a good model fit to the data (CFI = 0.99; TLI = 0.99; RMSEA = 0.03) and yielded the best results of the three models (Figure 6.1, Table 6.3).

Structural equation models

The necessary reduction of the large number of variables within the dataset in relation to the small sample size was a problem, and this was combined with a major issue in SEM, namely the selection and identification of performance-relevant variables on a theoretical basis. The results suggest, whereas the computational approach, as an objective procedure to identify parameters highly related to performance should be rejected, SEM that uses an expertise-based approach relying on the knowledge and experience of national coaches seems to be an appropriate way to build theory-based SEM. Nevertheless, the model created based on theoretical considerations derived from the literature yielded the best fit to the data, which could have resulted from the fact that the cited studies had mainly used regression or correlation

analyses, and therefore, had already identified relationships between the parameters and performance. The three resulting models are discussed in the following sections.

SEM based on theoretical considerations

SEM in general, is a theory-based method which implies that consistent research findings in the literature should provide a valuable basis for analysis. Compared to the two other approaches, the model based on previous research results and theoretical considerations, which consisted of two anthropometric and three physiological variables, provided a good model fit and the best results found in the study (Figure 6.1, Table 6.3), even though some of the considered studies were not conducted with elite triathletes or triathletes competing at Olympic distances. On closer inspection, the anthropometric variables body weight and BMI reflect the common body type described by Knechtle et al. (2011). Knechtle et al. (2011) found medium to large effect sizes for these two variables for total race time and time in running split for recreational ironman triathletes and referenced other studies that investigated runners (Hoffman, 2008; Knechtle, Duff, Welzel, & Kohler, 2009). This relationship is plausible regarding the importance of the running split in elite Olympic-distance triathlons. The running speed-related physiological variables running pace at 3-mmol·L^{-1} blood lactate and maximum running pace contribute more to the overall race performance in an elite triathlon than relative VO_2max (Figure 6.1). Several studies have shown that relative VO_2max is comparable over elite triathletes (Suriano & Bishop, 2010) and only has a prerequisite function. Therefore, its contribution to clarifying performance structure in a comprehensible manner is smaller. Overall, the effects of anthropometric and physiological variables on overall race performance in the structural equation model were very similar, which corresponds to the explanations mentioned above.

SEM based on a computational preselection

The purely statistical preselection approach using EFA is devoid of subjective influences, and prioritizes variables based on their influence on a general factor. To construct a structural equation model using measurable parameters instead of the principal components of the EFA, variables with a small variance were identified and only the variables with the highest loadings on each general factor were used, which likely resulted in a loss of information, and therefore, a lack of explanatory value in the model. The model based EFA preselection consisted of two anthropometric variables and one physiological variable and provided an acceptable model fit (Figure 6.2, Table 6.4), but had to be rejected due to an error in the model. Consistent with the

results of Knechtle et al. (2011), BMI and body fat [%] were associated with overall race time. The reason for this could be again the importance of the running split in an Olympic-distance triathlon: the findings of Bale, Bradbury, and Colley (1986) demonstrated the importance of percent body fat to the performance of elite runners. The relevance of the physiological variable maximum blood lactate concentration in mobilization tests could be evidence for its influence on the ability to tolerate increased running paces or long and fast final spurts in the third discipline. The importance of maximum or peak blood lactate concentrations on isolated running performance or performance in a sprint triathlon has already been shown (Noakes, Myburgh, & Schall, 1990; Slattery, Wallace, Murphy, & Coutts, 2006). Nevertheless, the model based on computational preselection must be rejected due to negative variances for two measured variables, as this is not possible. This could possibly be attributed to poor parameter selection, which was not theory-based in this case, and selecting inappropriate parameters, even though they were identified separately in other studies.

SEM based on preselection by professional expertise

Dominance paired comparisons, as prioritization based on professional expertise, also seems to be a plausible preprocessing step because the knowledge and experience of national coaches can be used to build a theory-based model. However, this could be problematic due to the manual specification of the maximum number of variables, which is necessary because the dominance paired comparisons rank all the variables, and the number of variables has to be limited in SEM. The reason why the extended model, in comparison to the theory-based one, had slightly poorer model fit indices (Figure 6.3) may be due to the greater number of variables, which increased the number of degrees of freedom. A general drawback of subjective assessments could be that a variable, such as VO_2max, was selected and prioritized as a common parameter to characterize the endurance of heterogeneous groups (Butts et al., 1991; Miura et al., 1997). In contrast, VO_2max normally has only a prerequisite function instead of a predictive function in homogeneous groups (Sleivert & Rowlands, 1996). The model created through the preselection method of dominance paired comparisons included the same variables as the theory- based model, with the addition of lean body mass. This variable corresponds to the variable body fat in the model based on EFA because lean body mass [kg] is calculated via body weight [kg] minus body fat [kg]. It seems that the national coaches who participated in our study are familiar with the actual state of research on performance-relevant parameters,

which is to be welcomed. The effects of the anthropometric and physiological variables on overall race performance were very similar to those in the theory-based model.

These results show that parameter selection based on substantiated knowledge from scientific studies on performance relevant parameters (though not in their entirety) and preselection based on the expertise of national coaches both seem to be good working approaches for achieving the major assumption of SEM: i.e. the theory-based selection of parameters to determine the relationships and the underlying structure within a complex model, such as the performance structure of elite triathlon.

Performance structure versus performance prediction

Whereas common performance prediction models of triathlon (Hoffmann et al., 2017; Schabort et al., 2000; Van Schuylenbergh et al., 2004) have identified relevant performance parameters through multiple linear regression models or artificial neural networks, no complex models combining different types of parameters, such as anthropometric and physiological parameters, have so far been published about triathlon performance to the best of our knowledge. The additional benefit of the current results, in comparison to the preceding study (Hoffmann et al., 2017), is that they show the influence of each single parameter (e.g. body weight or maximum running pace) on overall race performance as well as the influence of the primary factors anthropometrics and physiology on overall race performance, which can be enlightening.

Both approaches, performance prediction and performance structure, require a preselection method to reduce the number of performance parameters due to the small sample size when working with elite athletes. The preselection methods considered for use can profoundly affect model structure depending on the purpose of the models (Stachowiak, 1973). The parameters used in each computational approach, therefore, differ depending on the preselection method and whether the aim is performance prediction or performance structure. For example, the best fitting prediction model in Hoffmann et al. (2017) used the variables of the dominance paired comparisons, whereas the best fitting structural model was found using theoretical considerations derived from the research literature.

Limitations

The sample in this study was elite, small, and homogenous, which limits generalizability to other triathlete cohorts. However, the generalizability of results to other triathlete cohorts was

not the aim of this study, since we explicitly focused on elite athletes. National squads in triathlon are generally small compared to other sports and elite athletes are often reluctant to participate in experiments. Additionally, individual training schedules and differences in season calendars complicate experimental laboratory studies. Therefore, one of our aims was to assess the performance structure of male elite Olympic-distance triathlon using data from regular performance diagnostics. To overcome the drawback of having a small number of available athletes, we used the algorithm used by Hoffmann et al. (2017) to increase the number of datasets used in the statistical analyses, by collecting performance diagnostics over a period of four years. We only included datasets if two requirements were fulfilled: (1) the availability of a complete set of variables from laboratory tests, and (2) a finished Olympic-distance triathlon race within 8 weeks after each performance diagnostic.

Furthermore, we did not take the results of laboratory tests for swimming and cycling into account, as described in Hoffmann et al. (2017). Nevertheless, the general tactical behavior in elite Olympic-distance triathlon races allows the use of running diagnostics alone to generate meaningful results. The swimming and cycling disciplines in elite Olympic-distance triathlon only have a prerequisite function, whereas the running discipline is normally the critical factor for success (Fröhlich et al., 2008; Vleck et al., 2006). Therefore, the results of the present structural equation models, with only running specific physiological parameters, can be considered appropriate even if there is potential for improvement in future studies.

Conclusion, practical recommendations and outlook

Two structural equation models to determine the performance structure of the elite Olympic-distance triathlon were developed without interfering with individual training programs, through triathlete participation in a standardized experimental study. The advantage of SEM in the field of training science is its ability to model complex patterns, such as the performance structure of a sport, if sufficiently good datasets are available. As we showed, it is also possible to obtain data for elite athletes, who are generally reluctant to participate in regular scientific studies. The structural equation model that was derived from the existing research literature and theoretical considerations consisted of two anthropometric variables and three physiological variables provided a good model fit (CFI = 0.99; TLI = 0.99; RMSEA = 0.03) and the best results of the three models in the study. Obviously, SEM is a powerful analytical procedure that is able to identify performance relevant variables in the elite Olympic-distance triathlon.

Collecting a dataset from a different cohort of elite triathletes in the future would be necessary to confirm the results of the structural equations models in the present study.

To ensure the transfer of the results into training, the identified anthropometric parameters BMI and body weight, in the case of the theory-based model, can serve a prerequisite function and it may be possible to transfer this information into the field of recreational athletes. Absolute and relative VO_2max needs to be viewed in the same way. Both the running pace at 3-mmol·L^{-1} blood lactate and the maximum running pace, as identified physiological variables, can be influential in specific training programs.

Finally, the results of the present study indicate that future research should focus on collecting larger samples to use SEM better. Even though larger samples could lead to other preselection methods that may be preferable, based on our results, it would be beneficial to focus on the scientifically based identification of performance relevant variables to improve training programs and the performance of athletes, in general, especially in sports with high amounts of training, such as triathlon.

7 General Discussion and Outlook

The purpose of the present thesis was to investigate the field of performance prediction and performance structure of Olympic-distance triathlon. For this purpose, the thesis focused on (1) the identification of performance-relevant parameters in recreational triathlon to predict individual race performance, (2) the identification of performance-relevant parameters in elite Olympic-distance triathlon to predict individual race performance and (3) the modeling the performance structure of elite Olympic-distance triathlon. To clarify the research questions involved, different computational approaches were used within the three presented studies to account for the special circumstances of triathlon.

This chapter summarizes and discusses the main findings of the studies described in Chapters 4 to 6, considers the implications and recommendations for future research and finally closes with a general conclusion.

7.1 Requirements to develop performance structure and prediction models

This section summarizes and discusses the requirements surrounding the performance parameters necessary to develop performance structure and performance prediction models.

7.1.1 Assessment of performance parameters

The selection of parameters collected through performance diagnoses in the field or laboratory has an important effect on the performance structure and prediction models as well as their results. The current literature shows very different test protocols, hindering both comparison and reconciliation with current knowledge.

Anthropometric measurements, especially of segment lengths, are often not sufficiently specified regarding standardization within studies or when different concepts are used (examples are Knechtle et al. (2011) and Landers et al. (2000)). Therefore, only a classification - instead of a comparison - of the results from the present study can be done. The influence of different test procedures becomes even clearer if looking at the acquisition of physiological parameters: besides the fact that simulated as well as real triathlon performances were used

within studies (Millet & Bentley, 2004; Van Schuylenbergh et al., 2004), the step tests differ in test design, step length, increments between steps, etc. The running-cycling-running trial used by Millet and Bentley (2004) as well as the graded tests on bicycle and treadmill with specific increments and lengths used by Van Schuylenbergh et al. (2004) can serve as examples. As a result, performance tests are more or less race-specific and account for different physiological load conditions. Comparisons between recreational and elite triathletes or even generalizable performance models are therefore not possible, especially because of differentiation in test protocols due to the great variation in performance. In general, it is expected that there are bigger differences in performance parameters in a heterogeneous group of recreational triathletes. Elite triathletes are more often homogenous because some performance parameters have a prerequisite rather than a differentiating function (Sleivert & Rowlands, 1996; Stratton et al., 2009). It can be stated that studies with elite triathletes more often have small sample sizes, because national squads are normally small, which leads to less variation. Therefore, the whole cohort of elite athletes could be tested, which makes the modeling process statistically more relevant. In contrast, recreational triathletes can be tested in larger quantities, which will lead to more variation within the data, which can be beneficial for a better generalization of the results.

With regard to the samples within this thesis, the elite triathletes were relatively homogeneous. Anthropometric characteristics had only small variations and the physiological variables showed a partially similar distribution: VO_2max showed a small variation because of its prerequisite nature in samples consisting of elite triathletes. Maximum lactate values as well as the lactate values after load showed higher CVs. Different individual strengths in the three disciplines likely affected the results of the running-specific step tests conducted within the national squad.

The recreational sample within this thesis showed a comparable performance level (small CV in overall race time), although the sample showed a relatively large spread concerning age. Anthropometric parameters are comparable to other studies using recreational triathletes (Knechtle et al., 2011; Kohrt et al., 1987; Sleivert & Wenger, 1993). Physiological parameters are more difficult to compare due to the specific step test conducted. The athletes' own recorded training volumes were high, given their recreational background, and in comparison to other studies (Hue et al., 1998; Kohrt et al., 1987; Kohrt et al., 1989). All collected parameters have a high CV, which underlines the statement above.

7.1.2 Normalization of race time

To obtain comparable overall race times independent of the individually-selected triathlon races triathletes took part in, race time normalization is necessary for the modeling process. A standardized laboratory setting is necessary to determine the individual performance metrics of an athlete. This step is not necessary if all participants within a study take part in the same triathlon race

For studies working with an elite sample, it is very difficult for all athletes to participate in the same triathlon race due to different race calendars and training schedules. Even recreational triathletes in cohorts larger than 10 are unlikely to participate in the same event. To create performance prediction models or performance structure models, it is indispensable to have a performance metric and this metric should preferably be competition-like.

The normalization procedures used within this thesis are not to be found in other scientific studies, but are acceptable given the results. The mean and SD of normalized overall race time for both recreational triathletes over a sprint distance (4,326.29 ± 239.19 s; approximately 1:12 h) and elite athletes over an Olympic distance (6,827.57 ± 192.56 s; approximately 1:54 h) are comparable to other studies (Landers et al., 2000; Taylor & Smith, 2014). Influencing factors can be, for example, the league membership due to different regulations or the homogeneity of the sample.

7.1.3 Selection of performance variables

Before creating performance structure or performance prediction models, the abundance of measured parameters has to be brought in line with the available sample. Within this thesis, different statistical methods have been used to reduce the number of collected parameters without losing too much information before creating the performance prediction or performance structure models.

The purely statistical approach using an exploratory factor analysis is in general devoid of subjective influences, and prioritizes variables based on their influence to a general latent factor. Therefore, variables with a small variance will typically be removed, which could result in a lack of explanation of variance in later models.

The dominance paired comparison, as an objective method for prioritization based on professional expertise, also seems to be a plausible preprocessing step. A reduction of variables similar to an exploratory factor analysis cannot be achieved because the maximum number of variables used in the computational models must be specified manually; also, the method does not have high demands regarding the number of coaches consulted.

A salient example for this is provided by the parameter VO_2max, which is a common parameter characterizing the endurance of heterogeneous groups (Butts et al., 1991; Miura et al., 1997). Perhaps this is one of the reasons why national triathlon coaches selected absolute as well as relative VO_2max as important parameters. Concerning homogenous groups, it becomes apparent that VO_2max normally has a premising instead of a differentiating character, because of the small variation in such groups of athletes (Sleivert & Rowlands, 1996). This could possibly be a drawback of subjective assessments compared to a purely statistical approach like exploratory factor analysis, which completely removed VO_2max.

A major difference between the exploratory factor analysis and the dominance paired comparison concerns the conditions for the application: whereas the dominance paired comparison as a preselection method has no conditions, the exploratory factor analysis has some requirements the data set must fulfill (Ferguson & Cox, 1993). On the other hand, the exploratory factor analysis is devoid of subjective influences and is able to deal with large sets of variables. Depending on the computational approach followed, the appropriate preselection method should therefore be selected (e.g. if linear or non-linear relationships should be taken into account).

7.2 Prediction of recreational and elite triathlon performance

The general aim of the studies in Chapters 4 and 5 was to reveal relationships between different performance prerequisites and triathlon race performance, quantified through overall race time in recreational and elite triathlon. Thereby, standardized laboratory tests of recreational triathletes competing over sprint distance and regular performance diagnoses of elite triathletes competing over Olympic distance were used. The identified anthropometric and physiological parameters were used to predict individual overall race time using different computational approaches. A further step towards analyzing the performance structure of triathlon was the development of appropriate models, which will also be discussed.

7.2.1 Multiple linear regression analysis to predict performance

Besides correlations between single performance indicators and a measure of performance (Bentley, Wilson, Davie, & Zhou, 1998; Sleivert & Wenger, 1993; Zhou et al., 1997), linear regressions are widely used to generate knowledge about performance-relevant parameters, and even to develop performance prediction models (Landers et al., 2000; Landers et al., 2008; Schabort et al., 2000; Van Schuylenbergh et al., 2004). Within this thesis, multiple linear regression models were successfully developed to predict overall triathlon race time as a measure of performance in recreational and elite triathlon. The boundary conditions will be discussed in the following section.

The anthropometric parameters that predict performance depend on tactical behavior, which depends on factors such as race distance (and therefore the priority of the three disciplines), if drafting is allowed etc. and are therefore difficult to generalize. Elite triathletes have a homogenous body shape and composition, which makes it interesting that statistically-relevant parameters for linear regression models could be found. Each of the identified parameters is already known as performance-relevant to one of the three disciplines of triathlon (Anderson, 1996; Barnes & Kilding, 2015; Williams et al., 1987), but only partially to triathlon itself. Although the interpretation of the parameters found for elite triathletes (pelvis width and shoulder width) was in line with common literature, the findings for recreational triathletes (leg length and arm span) were more difficult to classify. General conditions influencing linear regression models, and therefore performance prediction models, include race distance - especially the proportion of each discipline, which differs over the four common race distances in triathlon (see section 2.1) - and the general sport-specific background of an athlete. The three disciplines show different performance-relevant parameters compared to each discipline alone (Barnes & Kilding, 2015; Lätt et al., 2010). Therefore, the percentage contribution of each of the three disciplines compared to overall race distance influences how individual strengths and prerequisites or even training years of an athlete come out. This should be more relevant for recreational triathletes than for elite triathletes, because their sport-specific backgrounds are much more diverse.

The general importance of physiological parameters, especially for performance prediction, is indisputable (Schabort et al., 2000; Suriano & Bishop, 2010). Blood lactate concentrations are common in research, and have previously been used to predict overall triathlon performance (Schabort et al., 2000; Van Schuylenbergh et al., 2004). Hue (2003)

demonstrated that lactate concentration measured at the end of the cycling phase of a simulated cycle-run test appears to be a performance predictor in triathlon, which confirms the present findings. With regard to different tactics in races with or without drafting, Hue (2003) - in reference to Hausswirth et al. (1999) - highlighted the importance of specific test protocols and their influence on the results of performance prediction models. Based on the findings within this thesis, the workload within test protocols should therefore be defined by a relative parameter such as % VO_2max or % lactate threshold and not by absolute parameter values. This should lead to a better linear regression fit and a more individual performance prediction model.

In general, the studies within this thesis showed that performance prediction in triathlon were possible using anthropometric and physiological parameters measured through laboratory tests as well as performance diagnoses in recreational and elite athletes. Regardless of whether recreational or elite triathletes are brought into focus, the lactate concentration at particular performance thresholds has an important impact on prediction models. Since triathlon is an endurance sport, this is not surprising since these variables describe the ability of an athlete to realize a higher (running) pace with the same use of metabolic processes, and could therefore directly lead to a faster race time. An important finding within this thesis is the fact that these parameters are statistically linked to overall race performance and should therefore be of special interest within the creation of training schedules because they can be decisively controlled. The specific parameters depend on the variety of possible performance diagnoses and step tests, and should be investigated further with regard to their prediction capability.

For classification of the results of linear regression models within this thesis, the standard error of the estimate (SEE) was used, as it is a common indicator in linear regression (Montgomery, Peck, & Vining, 2001; Weisberg, 2005). Up to now, studies in this field used different model fit parameters – if they are stated at all – so SEE seems to be a good parameter to classify significant and valid regression models for interpretation considering the regarded sample.

7.2.2 Artificial neural networks to predict performance

The second performance prediction approach, using artificial neural networks (ANNs), provided two performance prediction models which explained a medium and large degree of variance. In general, the present thesis confirms that ANNs could be an alternative and valuable computational approach for performance prediction without the restrictions of distribution and

independence of variables, as stated by Edelmann-Nusser et al. (2002) and Silva et al. (2007). Furthermore, the combination of ANNs with a preselection based on the expertise of professional triathlon coaches seems to be beneficial, since this kind of preselection uses knowledge and practical experience without consideration of technical or statistical conditions of ANNs, which are minimal.

The ANN based on the preselected anthropometric parameters body weight, BMI, lean body mass and absolute as well as relative body fat in elite Olympic-distance triathlon explained a medium percentage of variance. The variations in parameters such as body weight or BMI are normally too small to obtain significant results in small and homogenous samples (Table 5.1). In the present thesis, the professional triathlon coaches were partially responsible for young athletes in national squads, where the mentioned variables have a higher influence and a greater variance than in elite triathletes. However, absolute as well as relative body fat show larger coefficients of variation, which could be an indication that these are useful predictive variables in general. Knechtle et al. (2011) already stated this for recreational male triathletes competing over long distance races.

The ANN based on the preselected physiological parameters maximum running pace, running pace at 3-mmol·L^{-1} blood lactate, maximum running pace in the mobilization test and absolute as well as relative VO_2max in elite Olympic-distance triathlon explained a large percentage of variance and led to the best prediction model within this thesis. The parameters maximum running pace, running pace at 3-mmol·L^{-1} blood lactate and maximum running pace in the mobilization test allow a valid assessment of the anaerobic capacities of an athlete. This could again underline the importance of the running discipline: nearly all races of the WTS were actually won during the running discipline, especially in the final spurt. High lactate values and high running paces could therefore be important factors for overall triathlon race time. Again, absolute as well as relative VO_2max were preselected through the professional triathlon coaches and therefore included in the prediction model; even if VO_2max normally has only a premising instead of a predicting character in homogenous groups (Sleivert & Rowlands, 1996).

The accuracy of the prediction results of the ANNs could be attributed to the fact that performance prediction is quite a complex, non-linear problem (Edelmann-Nusser et al., 2002) which can be well modeled by an ANN. Nevertheless, ANNs have rarely been used to predict race performance, possibly because the network design of an ANN requires substantial input

concerning the number of neurons, layers and training algorithm (Zhang, Eddy Patuwo, & Y. Hu, 1998), but with the benefit of fewer restrictions concerning the data set.

A general drawback of an ANN concerns the model architecture: while minimizing the overall error of the network by adapting the weights, there is no meaningful way to extract the priority of each performance parameter, which must be seen as a limitation when computing performance prediction models as done within this thesis. On the other hand, the prediction models using ANNs outperformed those from a purely statistical approach comprising factor analysis and multiple regressions. Furthermore, a closer look at the SEE, as an overall error metric used for the prediction models within this thesis, revealed that these are smaller than the performance variation of individual athletes and of the Top 10 athletes in the WTS 2014. This highlights the benefits of ANNs and the benefit this non-linear approach could have in the context of performance prediction.

7.2.3 Practical implications

The reported performance prediction models within this thesis showed that specific influencing parameters generally exist. The creation of such models succeeded, although there are confounding variables like the environment, the conditions of the race, the terrain, etc. which add variance to the modeled equations. However, the identified parameters and prediction models could allow a more objective talent selection procedure by defining minimum physical requirements (e.g. for specific age groups). Talent identification programs could also use information on advantageous anthropometric requirements to direct young athletes to the sport of triathlon. The design of training schedules could be influenced by focusing on optimal training levels (e.g. to improve specific lactate levels; more detailed in section 7.3). These aspects concerning performance prediction models were also underlined by Landers et al. (2000). They illustrated the importance of low levels of adiposity for overall race time and most of the subdisciplines in elite triathlon and of proportionally longer segmental lengths for successful swimming outcome.

Especially for recreational triathletes, assessment of their split and overall pace might be a major problem when determining an individual race tactic. Selection of an optimum individual pace in each discipline can prevent a decline in performance or even a dropout. Therefore, prediction of their individual race performance could be a useful tool.

Performance prediction models could be enhanced by using information about the number of previous races and personal best times (Knechtle et al., 2015; Rüst et al., 2012), and also by information about previous races such as overall or split times (Gilinsky et al., 2014), in particular for samples of recreational triathletes because of their heterogeneous characteristics compared to elite triathletes.

7.3 Structure of triathlon performance in elite triathletes

The general aim of the study in Chapter 6 was to reveal relationships between different performance prerequisites and triathlon race performance, quantified through overall race time in elite Olympic-distance triathlon, to investigate performance structure. Thereby, standardized regular performance diagnoses of elite triathletes competing over Olympic distances were used. Identified anthropometric and physiological parameters were applied to develop performance structure models for this specific setting.

The necessary reduction of a large number of variables collected through common performance diagnoses in relation to small national squads could successfully be combined with a major use of SEM, namely the identification of performance-relevant variables on a theoretical basis. SEM implies that consistent research findings in the literature should provide a valuable basis for analysis. The results presented within this thesis indicate that structural equation models based on theoretical considerations derived from research literature yielded the best fit to the data, which could have resulted from the fact that the cited studies mainly used regression or correlation analyses, and therefore had already identified relationships between the parameters and performance. Even the structural equation model, using an expertise-based approach that relies on the knowledge and experience of national triathlon coaches, seems to be an appropriate way to build theory-based SEMs.

The structural model based on previous research results and theoretical considerations consists of two anthropometric and three physiological variables and provides a good model fit. The identified anthropometric variables body weight and BMI reflect the common body type described by Knechtle et al. (2011). In line with Knechtle et al. (2011), Knechtle et al. (2009) and Hoffman (2008), who all found connections between these two variables and running time, the SEM within this thesis underlines the importance of the running split in elite Olympic-distance triathlon for overall race time; or, more specifically, the fundamental relevance of BMI

and body weight as basic prerequisites. The importance of the running discipline can be further illustrated by the fact that the running speed-related physiological variables of running pace at 3-mmol·L^{-1} blood lactate and maximum running pace contribute more to the overall race performance in an elite triathlon than relative VO_2max. Several studies have shown that relative VO_2max is comparable over elite triathletes (Suriano & Bishop, 2010) and has a prerequisite function. Therefore, its contribution to clarifying performance structure is understandably smaller. The mentioned physiological variables could be further evidence for the necessary ability to tolerate increased running paces or long and fast final spurts during the run phase of elite Olympic-distance triathlon. Overall, the effects of anthropometric and physiological variables on overall race performance in the structural equation model were very similar.

The structural model based on professional expertise also seems to be an alternative approach in which theory can be substituted by the knowledge and experience of national coaches, as appropriate. Of course, national coaches are using theoretical knowledge besides their wide variety of practical experiences. A possible drawback of the subjective assessment, through a dominance paired comparison as preselection before creating the SEM, could be that a variable, such as VO_2max, was selected and prioritized as a common parameter to characterize the endurance of heterogeneous groups (Butts et al., 1991; Miura et al., 1997). In homogeneous groups, VO_2max normally has a prerequisite instead of a predictive function (Sleivert & Rowlands, 1996). The structural model created after the preselection method of dominance paired comparisons included nearly the same variables as the theory-based model, which demonstrates that the national coaches who participated in the study are familiar with the current state of research on performance-relevant parameters. It would be interesting to see how this depends on the level of knowledge of triathlon coaches through different proficiency levels. Overall, the results of the structural model based on expertise show that parameter selection based on substantiated knowledge from scientific studies on performance-relevant parameters (though not in their entirety) and preselection based on the expertise of national triathlon coaches both seem to be good working approaches for achieving the major assumption of SEM: the theory-based selection of parameters to determine the relationships and the underlying structure within a complex model, such as the performance structure of elite triathlon.

The structural model based on computational preselection had to be rejected, which could possibly be attributed to poor parameter selection. The lack of explanatory value in the

model could result from the purely statistical parameter selection using EFA, which was not theory-based and likely resulted in a loss of information as described in section 6.5.

Whereas common performance prediction models of triathlon (Hoffmann et al., 2017; Schabort et al., 2000; Van Schuylenbergh et al., 2004) have identified relevant performance parameters through multiple linear regression models or ANNs, no complex models combining different types of parameters, such as anthropometric and physiological parameters, have so far been published about triathlon performance. The additional benefit of structural equation models, in comparison to the preceding prediction models (Hoffmann et al., 2017), is that they show the influence of each single parameter (e.g. body weight or maximum running pace) on overall race performance as well as the influence of the primary anthropometric and physiological factors on overall race performance, which can be enlightening.

Practical implications

The advantage of SEM in the field of training science is its ability to model complex patterns, such as the performance structure of a sport, if sufficiently good datasets are available. The structural equation models used to determine the performance structure of elite triathlon within this thesis show that meaningful knowledge could be generated without interfering with individual training programs through triathlete participation in a standardized experimental study. It is clear that SEM is a powerful analytical procedure that is able to identify performance-relevant variables in elite Olympic-distance triathletes.

To ensure the transfer of results into training, the identified anthropometric parameters BMI and body weight, in the case of the theory-based model, can serve a prerequisite function. It may be possible to transfer this information into the field of recreational athletes, in the sense of threshold values that need to be reached to achieve a top position. Absolute and relative VO_2max should also be viewed in the same way. Both running pace at 3-mmol·L^{-1} blood lactate and maximum running pace, as identified physiological variables, can be influential in designing or optimizing training schedules by focusing on optimal training levels (e.g. to improve specific lactate levels). One main focus should therefore be on aerobic threshold training including tempo runs and extensive interval training (Pöhlitz & Valentin, 2015). Thus, the aerobic threshold continues to develop, independent of the athlete's performance level. This results in a faster race pace with regard to the importance of the parameter running pace at 3-mmol·L^{-1} blood lactate in the running split of Olympic-distance triathlon. Mainly for

recreational athletes, the improvement of the aerobic capacity (VO_2max) and running economy can be achieved with speed variations and interval training with high (running) paces. Repeated heats of 3-5-7 minutes with heart rates between 90 and 100 % and comparable long or shorter breaks will optimize individual aerobic capacity (Pöhlitz & Valentin, 2015).

7.4 Limitations and implications for future research

Both general approaches within this thesis, performance prediction and performance structure, require a preselection method to reduce the number of parameters to obtain due to the small sample sizes when working with elite or in some cases even recreational triathletes. The preselection methods considered for use can profoundly affect model structure depending on the purpose of the models (Stachowiak, 1973). The parameters used in each computational approach therefore depend on the preselection method and whether the aim is performance prediction or performance structure. For example, the best fitting prediction model in Hoffmann et al. (2017) used the variables after preselection by dominance paired comparisons, whereas the best fitting structural model was found using theoretical considerations derived from the research literature, which should be kept in mind.

The samples within this thesis have in common that they were small regarding the computational methods used, which must be mentioned as a potential limitation. The cohorts of both recreational and elite triathletes were both small and homogenous, focusing on a specific race distance, which limits the generalizability of the findings to other triathlete cohorts. However, generalizability was not the main aim of this thesis since we explicitly focused on detecting methods to create performance prediction and performance structure models and to generate ideas about performance-relevant parameters for specific settings in the field of triathlon, especially for elite athletes over the Olympic distance. National squads in triathlon are generally small compared to other sports and elite athletes are often reluctant to participate in experiments. The first point was the reason only male (elite) triathletes could be considered within this thesis because the female national squad was unfortunately too small for reliable computations. Additionally, individual training schedules and differences in season calendars complicate experimental laboratory studies. In the case of elite triathletes, we used the internally-developed algorithm to increase the number of datasets used in statistical analyses, by collecting performance diagnoses over a period of four years to overcome the drawback of having a small number of available athletes.

Furthermore, the results of laboratory tests for swimming and cycling were not taken into account. This opens a vast potential for future research to generate more comprehensive knowledge about performance-relevant parameters. Nevertheless, the general tactical behavior in modern triathlon, especially in elite Olympic-distance triathlon, allows the use of running diagnoses to generate meaningful results. The swimming and cycling disciplines in elite Olympic-distance triathlon more often have a prerequisite function, whereas the running discipline is normally the critical factor for success (Fröhlich et al., 2008; Vleck et al., 2006). Therefore, the results of the present thesis, with only running-specific physiological parameters, can be considered appropriate even if there is potential for improvement in future studies.

Overall, the results of the present thesis indicate that future research should focus on collecting larger samples for better application of the applied computational methods. Even though larger samples could lead to other preselection methods that may be preferable, based on our results it would be beneficial to focus on the scientifically-based identification of performance-relevant variables to improve training programs and the performance of athletes in general, especially in sports with high amounts of training, such as triathlon. Besides collecting larger samples, future research should focus on the application of more specific and more comparable laboratory tests, preferably combining the three single disciplines, to determine more extensive and specific performance prediction and performance structure models.

8 Conclusion

The present thesis investigated performance prediction and performance structure models and identified performance-relevant parameters in the field of triathlon, with special consideration to different computational approaches. Therewith, this thesis provides valuable research combining a traditional theoretical field in training science with the practical application to the growing sport of triathlon. Special attention was paid to the application of different computational approaches to reveal the potential of different methods in the field of performance prediction and performance structure, because literature regarding this aspect is relatively sparse.

The present thesis aimed to overcome research gaps and to gain more detailed insights about triathlon and performance-relevant parameters. For this purpose, recreational as well as elite triathletes were investigated and the results of performance diagnoses were used to develop multiple regression models and ANNs, as well as structural equation models. Essentially, the research presented in this thesis revealed the following findings:

- The prediction of triathlon performance using linear or non-linear approaches based on actual or routine performance diagnoses is possible for recreational and elite triathletes. This confirms common findings in literature and expands the knowledge about the use of multiple linear regression models and non-linear ANNs in this specific field. The application of non-linear computations could be an especially promising approach with regard to the complex construct of sport performance.
- To deal with small sample sizes, which is unavoidable while working with elite triathletes, it seems to be possible to gather larger datasets through collecting data of performance diagnoses over several years and to combine them with appropriate performance measurements. Another finding concerns the selection and reduction of the large number of variables collected through performance diagnoses: both theory-based and expertise-based approaches seem to work well when preselecting performance-relevant variables before creating performance prediction or structural models.

- The structural equation models lead to a better understanding of the performance structure in triathlon, even if they are exploratory in nature. They focus on physiological variables such as specific lactate values, which are trainable to a certain degree. Relevant anthropometric parameters on the other side can be useful prerequisites in the composition of national squads, for example. Moreover, even an adaptation of the selection process of talented trainees could be possible: a focus on single performance-relevant parameters derived from performance structure models can be a good addition to classic qualifying heats.

Taken together, the present thesis adds some valuable work to the literature, reinforcing prior findings and expanding on them by delivering new insights on the process of clarifying the performance structure in triathlon. The relationships between performance parameters and race performance, irrespective of whether one considers performance prediction or performance structure models, can help to develop training processes and talent diagnostics. Moreover, the computational methods have the potential to be used in other settings and sport disciplines as well, which should encourage further research studies.

References

Ackland, T. R., Blanksby, B. A., Landers, G., & Smith, D. (1998). Anthropometric profiles of elite triathletes. *Journal of Science and Medicine in Sport*, *1*(1), 52–56. https://doi.org/10.1016/S1440-2440(98)80008-X

Aggarwal, C. C. (2018). *Neural Networks and Deep Learning: A Textbook*. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-94463-0

Anderson, T. (1996). Biomechanics and running economy. *Sports Med*, *22*(2), 76–89.

Arbuckle, J. L. (2014). Amos (Version (Version 23.0)) [Computer software]. Chicago: IBM SPSS.

Atkinson, G., & Nevill, A. M. (2001). Selected issues in the design and analysis of sport performance research. *Journal of Sports Sciences*, *19*(10), 811–827. https://doi.org/10.1080/026404101317015447

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2018). *Multivariate Analysemethoden [Multivariate Methods of Analysis]: Eine anwendungsorientierte Einführung [An application-oriented introduction]* (15., vollst. überarb. Auflage 2018). Berlin: Springer Berlin; Springer Gabler.

Bale, P., Bradbury, D., & Colley, E. (1986). Anthropometric and training variables related to 10km running performance. *British Journal of Sports Medicine*, *20*(4), 170–173. https://doi.org/10.1136/bjsm.20.4.170

Barnes, K. R., & Kilding, A. E. (2015). Running economy: measurement, norms, and determining factors. *Sports Med - Open*, *1*(1), 357. https://doi.org/10.1186/s40798-015-0007-y

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). *Latent variable models and factor analysis: A unified approach* (3rd ed.). *Wiley series in probability and statistics*. Hoboken, N.J: Wiley.

Basset, F. A., & Boulay, M. R. (2000). Specificity of treadmill and cycle ergometer tests in triathletes, runners and cyclists. *Eur J Appl Physiol*, *81*(3), 214–221. https://doi.org/10.1007/s004210050033

Bassett, D. R. (2000). Limiting factors for maximum oxygen uptake and determinants of endurance performance. *Med Sci Sports Exerc*, *32*(1), 70–84. https://doi.org/10.1097/00005768-200001000-00012

Bentley, D. J., Millet, G. P., Vleck, V. E., & McNaughton, L. R. (2002). Specific aspects of contemporary triathlon: Implications for physiological analysis and performance. *Sports Medicine*, *32*(6), 345–359. https://doi.org/10.2165/00007256-200232060-00001

Bentley, D. J., Wilson, G. J., Davie, A. J., & Zhou, S. (1998). Correlations between peak power output, muscular strength and cycle time trial performance in triathletes. *Journal of Sports Medicine and Physical Fitness*, *38*(3), 201–207.

Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation [Research methods and evaluation]: Für Human- und Sozialwissenschaftler [For human and social scientists]* (4., überarb. Aufl., [Nachdr.]). *Springer-Lehrbuch Bachelor, Master*. Heidelberg: Springer-Medizin-Verl.

Bottoni, A., Gianfelici, A., Tamburri, R., & Faina, M. (2011). Talent selection criteria for olympic distance triathlon. *Journal of Human Sport and Exercise*, *6*(2 (Suppl.)), 293–304. https://doi.org/10.4100/jhse.2011.62.09

Bunker, R. P., & Thabtah, F. (2019). A machine learning framework for sport result prediction. *Applied Computing and Informatics*, *15*(1), 27–33. https://doi.org/10.1016/j.aci.2017.09.005

Butts, N. K., Henry, B. A., & Mclean, D. (1991). Correlations between VO2max and performance times of recreational triathletes. *J Sport Med Phys Fit*, *31*(3), 339–344.

Chatard, J. C., & Wilson, B. (2003). Drafting distance in swimming. *Medicine & Science in Sports & Exercise*, *35*(7), 1176–1181. https://doi.org/10.1249/01.MSS.0000074564.06106.1F

Deutsche Triathlon Union e.V. (2018). *Triathlon in Deutschland - Zahlen, Fakten & Hintergründe [Triathlon in Germany - Figures, facts & background]*. Frankfurt. Retrieved from www.dtu-info.com

Edelmann-Nusser, J. (2005). *Sport und Technik [Sports and technology]: Anwendungen moderner Technologien in der Sportwissenschaft [Application of modern technology in sport science]*. *Berichte aus der Sportwissenschaft*. Aachen: Shaker.

Edelmann-Nusser, J., Hohmann, A., & Henneberg, B. (2002). Modeling and prediction of competitive performance in swimming upon neural networks. *European Journal of Sport Science*, *2*(2), 1–10. https://doi.org/10.1080/17461390200072201

Enders, C., & Bandalos, D. (2001). The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(3), 430–457. https://doi.org/10.1207/S15328007SEM0803_5

Farrell, J. E. (2001). Efficient method for paired comparison. *Journal of Electronic Imaging*, *10*(2), 394. https://doi.org/10.1117/1.1344187

Felser, S., Behrens, M., Bäumler, M., & Bruhn, S. (2015). Ein Modellansatz zur Aufklärung der Leistungsstruktur beim Short Track : eine Studie anhand empirischer Daten deutscher Short-Track-Athleten [A modeling approach to investigate the performance structure in short track using empirical data of German short track athletes]. *Leistungssport*, *45*(2), 17–23.

Ferguson, E., & Cox, T. (1993). Exploratory Factor Analysis: A Users Guide. *International Journal of Selection and Assessment*, *1*(2), 84–94. https://doi.org/10.1111/j.1468-2389.1993.tb00092.x

Friel, J., & Vance, J. (2013). *Triathlon science*. Champaign, Illinois: Human Kinetics.

Fröhlich, M., Klein, M., Pieter, A., Emrich, E., & Gießling, J. (2008). Consequences of the Three Disciplines on the Overall Result in Olympic-distance Triathlon. *International Journal of Sports Science and Engineering*, *2*(4), 204–210.

Gilinsky, N., Hawkins, K. R., Tokar, T. N., & Cooper, J. A. (2014). Predictive variables for half-Ironman triathlon performance. *J Sci Med Sport*, *17*(3), 300–305. https://doi.org/10.1016/j.jsams.2013.04.014

Hair, J. F. (1995). *Multivariate data analysis with readings* (4th ed). Englewood Cliffs, N.J.: Prentice Hall.

Hausswirth, C., Le Meur, Y., Bieuzen, F., Brisswalter, J., & Bernard, T. (2010). Pacing strategy during the initial phase of the run in triathlon: Influence on overall performance. *Eur J Appl Physiol*, *108*(6), 1115–1123. https://doi.org/10.1007/s00421-009-1322-0

Hausswirth, C., Lehénaff, D., Dréano, P., & Savonen, K. (1999). Effects of cycling alone or in a sheltered position on subsequent running performance during a triathlon. *Med Sci Sports Exerc*, *31*(4), 599–604. https://doi.org/10.1097/00005768-199904000-00018

Haykin, S. S. (2009). *Neural networks and learning machines* (3rd ed.). Upper Saddle River: Pearson Education.

Heiberger, R. M., & Holland, B. (2004). *Statistical analysis and data display: An intermediate course with examples in S-plus, R, and SAS*. New York: Springer.

Hoffman, M. D. (2008). Anthropometric characteristics of ultramarathoners. *International Journal of Sports Medicine*, *29*(10), 808–811. https://doi.org/10.1055/s-2008-1038434

Hoffmann, M., Moeller, T., Seidel, I., & Stein, T. (2015). Prediction of elite triathlon performance by multiple linear regression models. *Book of Abstracts of the 20th Annual Congress of the European College of Sport Science*, 314.

Hoffmann, M., Moeller, T., Seidel, I., & Stein, T. (2017). Predicting Elite Triathlon Performance: A Comparison of Multiple Regressions and Artificial Neural Networks. *International Journal of Computer Science in Sport*, *16*(2), 101–116. https://doi.org/10.1515/ijcss-2017-0009

Hohmann, A., & Brack, R. (1983). Theoretische Aspekte der Leistungsdiagnostik im Sportspiel [Theoretical aspects of performance diagnoses in sport games]. *Leistungssport*, *13*(2), 5–10.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Hottenrott, K., & Seidel, I. (2017). *Handbuch Trainingswissenschaft - Trainingslehre [Handbook training science - training]*. *Beiträge zur Lehre und Forschung im Sport: Band 200*. Schorndorf: Hofmann.

Hox, J. J., & Bechger, T. M. (1998). An Introduction to Structural Equation Modeling. *Family Science Review*, *11*, 354–373.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hue, O. (2003). Prediction of Drafted-Triathlon Race Time From Submaximal Laboratory Testing in Elite Triathletes. *Can J Appl Physiol*, *28*(4), 547–560. https://doi.org/10.1139/h03-042

Hue, O., Le Gallais, D., Boussana, A., Chollet, D., & Prefaut, C. (2000). Performance level and cardiopulmonary responses during a cycle-run trial. *International Journal of Sports Medicine*, *21*(4), 250–255. https://doi.org/10.1055/s-2000-8883

Hue, O., Le Gallais, D., Chollet, D., Boussana, A., & Préfaut, C. (1998). The influence of prior cycling on biomechanical and cardiorespiratory response profiles during running in triathletes. *Eur J Appl Physiol Occup Physiol*, *77*(1-2), 98–105. https://doi.org/10.1007/s004210050306

Hue, O., Le Gallais, D., Chollet, D., & Préfaut, C. (2000). Ventilatory threshold and maximal oxygen uptake in present triathletes. *Can J Appl Physiol*, *25*(2), 102–113. https://doi.org/10.1139/h00-007

Jais, S.-D. (2007). *The successful use of information in multinational companies: An exploratory study of individual outcomes and the influence of national culture*. *Research in management accounting & [and] control*. Wiesbaden: Deutscher Universitäts-Verlag.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modelling with SIMPLIS command language*. Chicago, Ill., Hillsdale, N.J.: SSI Scientific Software International; Lawrence Erlbaum Associates.

Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood, IL: SSSI. Scientific Software International.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark Iv. *Educational and Psychological Measurement*, *34*(1), 111–117. https://doi.org/10.1177/001316447403400115

Knechtle, B., Duff, B., Welzel, U., & Kohler, G. (2009). Body mass and circumference of upper arm are associated with race performance in ultraendurance runners in a multistage race--the Isarrun 2006. *Research Quarterly for Exercise and Sport*, *80*(2), 262–268. https://doi.org/10.1080/02701367.2009.10599561

Knechtle, B., Wirth, A., Rüst, C. A., & Rosemann, T. (2011). The Relationship between Anthropometry and Split Performance in Recreational Male Ironman Triathletes. *Asian J Sports Med*, *2*(1), 23–30.

Knechtle, B., Zingg, M. A., Rosemann, T., & Rüst, C. A. (2015). The aspect of experience in ultra-triathlon races. *SpringerPlus*, *4*, 278. https://doi.org/10.1186/s40064-015-1050-3

Knussmann, R., & Barlett, H. L. (1988). *Anthropologie : Handbuch der vergleichenden Biologie des Menschen [Anthropology : Handbook of comparative biology of humans]* (2nd ed). Stuttgart: Fischer.

Kohrt, W. M., Morgan, D. W., Bates, B., & Skinner, J. S. (1987). Physiological responses of triathletes to maximal swimming, cycling, and running. *Med Sci Sports Exerc*, *19*(1), 51–55.

Kohrt, W. M., O'Connor, J. S., & Skinner, J. S. (1989). Longitudinal assessment of responses by triathletes to swimming, cycling, and running. *Med Sci Sports Exerc*, *21*(5), 569–575.

Landers, G., Blanksby, B. A., Ackland, T. R., & Monson, R. (2008). Swim Positioning and its Influence on Triathlon Outcome. *International Journal of Exercise Science*, *1*(3), 96–105.

Landers, G., Blanksby, B. A., Ackland, T. R., & Smith, D. (2000). Morphology and performance of world championship triathletes. *Ann Hum Biol*, *27*(4), 387–400.

Lätt, E., Jürimäe, J., Mäestu, J., Purge, P., Rämson, R., Haljaste, K., . . . Jürimäe, T. (2010). Physiological, biomechanical and anthropometrical predictors of sprint swimming performance in adolescent swimmers. *J Sport Sci Med*, *9*(3), 398–404.

Lembeck, M., Starringer, G., & Schönfelder, M. (2009). Trainingsverhalten von Freizeit- und Breitensportlern im Triathlonsport: Analyse und Empfehlungen [Training behaviour of recreational athletes in triathlon: Analysis and recommendations]. In M. Engelhardt, B. Franz, G. Neumann, & A. Pfützner (Eds.), *23. Internationales Triathlon-Symposium, Erding 2008* (pp. 73–114). Hamburg: Czwalina.

Letzelter, H., & Letzelter, M. (1982). Die Struktur sportlicher Leistungen als Gegenstand der Leistungsdiagnostik in der Trainingswissenschaft [Performance Structure as an item of performance diagnostic in training science]. *Leistungssport*, *12*(5), 351–361.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Marongiu, E., Crisafulli, A., Pinna, M., Ghiani, G., Degortes, N., Concu, A., & Tocco, F. (2013). Evaluation of reliability of field tests to predict performance during Ironman Triathlon. *Sport Sciences for Health*, *9*(2), 65–71. https://doi.org/10.1007/s11332-013-0147-8

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

Maszczyk, A., Gołaś, A., Pietraszewski, P., Roczniok, R., Zając, A., & Stanula, A. (2014). Application of Neural and Regression Models in Sports Results Prediction. *Procedia - Social and Behavioral Sciences*, *117*, 482–487. https://doi.org/10.1016/j.sbspro.2014.02.249

McLaughlin, J. E., Howley, E. T., Bassett, D. R., Thompson, D. L., & Fitzhugh, E. C. (2010). Test of the classic model for predicting endurance running performance. *Med Sci Sports Exerc*, *42*(5), 991–997. https://doi.org/10.1249/MSS.0b013e3181c0669d

Millet, G. P., & Bentley, D. J. (2004). The physiological responses to running after cycling in elite junior and senior triathletes. *International Journal of Sports Medicine*, *25*(3), 191–197. https://doi.org/10.1055/s-2003-45259

Millet, G. P., Bentley, D. J., & Vleck, V. E. (2007). The Relationships Between Science and Sport: Application in Triathlon. *International Journal of Sports Physiology and Performance*, *2*(3), 315–322. https://doi.org/10.1123/ijspp.2.3.315

Millet, G. P., & Vleck, V. E. (2000). Physiological and biomechanical adaptations to the cycle to run transition in Olympic triathlon: Review and practical recommendations for training. *British Journal of Sports Medicine*, *34*(5), 384–390. https://doi.org/10.1136/bjsm.34.5.384

Millet, G. P., Vleck, V. E., & Bentley, D. J. (2009). Physiological differences between cycling and running: lessons from triathletes. *Sports Med*, *39*(3), 179–206. https://doi.org/10.2165/00007256-200939030-00002

Millet, G. P., Vleck, V. E., & Bentley, D. J. (2011). Physiological requirements in triathlon. *J Hum Sport Exerc*, *6*(2 Suppl.), 184–204. https://doi.org/10.4100/jhse.2011.62.01

Miura, H., Kitagawa, K., & Ishiko, T. (1997). Economy during a simulated laboratory test triathlon is highly related to Olympic distance triathlon. *Int J Sports Med*, *18*(4), 276–280. https://doi.org/10.1055/s-2007-972633

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2001). *Introduction to linear regression analysis* (3. ed.). *Wiley series in probability and statistics Texts, references, and pocketbooks section*. New York, NY: Wiley. Retrieved from http://www.loc.gov/catdir/bios/wiley042/00051312.html

Muthén, L. K., & Muthén, B. O. (1998-2011). *Mplus User's Guide. Sixth Edition*. Los Angeles, CA: Muthén & Muthén.

Noakes, T. D., Myburgh, K. H., & Schall, R. (1990). Peak treadmill running velocity during the VO2 max test predicts running performance. *Journal of Sports Sciences*, *8*(1), 35–45. https://doi.org/10.1080/02640419008732129

O'Donoghue, P. (2010). *Research methods for sports performance analysis*. London: Routledge.

Olkin, I., & Sampson, A. R. (2001). Multivariate Analysis: Overview. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences* (pp. 10240–10247). New York: Elsevier Science.

Osborne J., & Waters E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, *8*(2). Retrieved from https://pareonline.net/getvn.asp?v=8&n=2

Ostrowski, C., & Pfeiffer, M. (2007). Modellansatz zur Aufklärung der Leistungsstruktur im Skilanglauf [Modeling approach for clarification of performance structure in cross-country skiing]. *Leistungssport*, *37*(2), 37-39.

Perl, J., & Pfeiffer, M. (2011). PerPot DoMo: antagonistic meta-model processing two concurrent load flows. *International Journal of Computer Science in Sport*, *10*(2), 85–92.

Pfützner, A. (1997). Koppeltraining - Hauptinhalt einer triathlonspezifischen Fähigkeitsentwicklung [Koppeltraining - Main content of a triathlon-specific skill improvement]. *Zeitschrift Für Angewandte Trainingswissenschaft*, *4*(2), 22–33.

Pöhlitz, L., & Valentin, J. (2015). *Trainingspraxis Laufen [Training practice running]: Beiträge zum Leistungstraining [Contributions to performance training]*. Aachen: Meyer & Meyer Verlag.

Pyrka, P., Wimmer, V., Fenske, N., Fahrmeir, L., & Schwirtz, A. (2011). Factor Analysis in Performance Diagnostic Data of Competitive Ski Jumpers and Nordic Combined Athletes. *Journal of Quantitative Analysis in Sports*, *7*(3). https://doi.org/10.2202/1559-0410.1300

R Development Core Team (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Rüst, C. A., Knechtle, B., Knechtle, P., Rosemann, T., & Lepers, R. (2011). Personal best times in an Olympic distance triathlon and in a marathon predict Ironman race time in recreational male triathletes. *Open Access Journal of Sports Medicine*, *2*, 121–129. https://doi.org/10.2147/OAJSM.S23229

Rüst, C. A., Knechtle, B., Wirth, A., Knechtle, P., Ellenrieder, B., Rosemann, T., & Lepers, R. (2012). Personal best times in an olympic distance triathlon and a marathon predict an ironman race time for recreational female triathletes. *Chinese J Physiol*, *55*(3), 156–162. https://doi.org/10.4077/CJP.2012.BAA014

Schabort, E. J., Killian, S. C., St Clair Gibson, A., Hawley, J. A., & Noakes, T. D. (2000). Prediction of triathlon race time from laboratory testing in national triathletes. *Med Sci Sports Exerc*, *32*(4), 844–849.

Schneider, D. A., & Pollack, J. (1991). Ventilatory threshold and maximal oxygen uptake during cycling and running in female triathletes. *International Journal of Sports Medicine*, *12*(4), 379–383. https://doi.org/10.1055/s-2007-1024698

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *The Journal of Educational Research*, *99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Silva, A. J., Costa, A. M., Oliveira, P. M., Reis, V. M., Saavedra, J., Perl, J., & Marinho, D. A. (2007). The Use of Neural Network Technology to Model Swimming Performance. *Journal of Sports Science & Medicine*, *6(1)*, 117–125.

Slattery, K. M., Wallace, L. K., Murphy, A. J., & Coutts, A. J. (2006). Physiological determinants of three-kilometer running performance in experienced triathletes. *Journal of Strength and Conditioning Research*, *20*(1), 47–52. https://doi.org/10.1519/R-16724.1

Sleivert, G. G., & Rowlands, D. S. (1996). Physical and physiological factors associated with success in the triathlon. *Sports Med*, *22*(1), 8–18.

Sleivert, G. G., & Wenger, H. A. (1993). Physiological predictors of short-course triathlon performance. *Med Sci Sports Exerc*, *25*(7), 871–876.

Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Wien [etc.]: Springer.

Stratton, E., O'Brien, B. J., Harvey, J., Blitvich, J., McNicol, A. J., Janissen, D., . . . Knez, W. (2009). Treadmill Velocity Best Predicts 5000-m Run Performance. *Int J Sports Med*, *30*(1), 40–45.

Suriano, R., & Bishop, D. (2010). Physiological attributes of triathletes. *J Sci Med Sport, 13*(3), 340–347. https://doi.org/10.1016/j.jsams.2009.03.008

Tarrow, S. (2010). The Strategy of Paired Comparison: Toward a Theory of Practice. *Comparative Political Studies*, *43*(2), 230–259. https://doi.org/10.1177/0010414009350044

Taylor, D., & Smith, M. F. (2014). Effects of deceptive running speed on physiology, perceptual responses, and performance during sprint-distance triathlon. *Physiol Behav*, *133*, 45–52. https://doi.org/10.1016/j.physbeh.2014.05.002

Tittel, K., & Wutscherk, H. (1972). *Sportanthropometrie : Aufgaben, Bedeutung, Methodik und Ergebnisse biotypologischer Erhebungen [Sports Anthropology : tasks, meanings, methodology and results of biotypological surveys]*. Leipzig: Barth.

Van Schuylenbergh, R., Eynde, B. V., & Hespel, P. (2004). Prediction of sprint triathlon performance from laboratory tests. *Eur J Appl Physiol*, *91*(1), 94–99. https://doi.org/10.1007/s00421-003-0911-6

Vleck, V. E., Burgi, A., & Bentley, D. J. (2006). The consequences of swim, cycle, and run performance on overall result in elite olympic distance triathlon. *International Journal of Sports Medicine*, *27*(1), 43–48. https://doi.org/10.1055/s-2005-837502

Weisberg, S. (2005). *Applied linear regression* (3. ed.). *Wiley series in probability and statistics*. Hoboken, NJ: Wiley-Interscience. Retrieved from http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10299721 https://doi.org/10.1002/0471704091

Williams, K. R., Cavanagh, P. R., & Ziff, J. L. (1987). Biomechanical studies of elite female distance runners. *Int J Sports Med, 8 Suppl 2*, 107–118.

Yong, A. G., & Pearce, S. (2013). A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis. *Tutorials in Quantitative Methods for Psychology*, *9*(2), 79–94. https://doi.org/10.20982/tqmp.09.2.p079

Zhang, G., Eddy Patuwo, B., & Y. Hu, M. (1998). Forecasting with artificial neural networks. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Zhou, S., Robson, S. J., King, M. J., & Davie, A. J. (1997). Correlations between short-course triathlon performance and physiological variables determined in laboratory cycle and treadmill tests. *Journal of Sports Medicine and Physical Fitness*, *37*(2), 122–130.

Appendix

Statutory Declaration

Hiermit erkläre ich, dass ich die vorliegende Dissertation mit dem Titel

> „Prediction and Structure of Triathlon Performance in Recreational and Elite Triathletes“

selbständig angefertigt und keine weiteren als die angegebenen Hilfsmittel benutzt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie (KIT) zur Sicherung guter wissenschaftlicher Praxis beachtet habe. Diese Arbeit wurde nicht bereits anderweitig als Prüfungsarbeit verwendet.

Bruchsal, den 04. November 2020 _____