EMINENCE IN ENDURANCE CYCLING SPORTS

JESSE KORF

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN

PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN KINESIOLOGY AND HEALTH SCIENCE

YORK UNIVERSITY

TORONTO, ONTARIO

OCTOBER 5, 2023

©Jesse Korf, 2023

Abstract

The objectives of this dissertation were to define and explore eminence in endurance cycling. To accomplish these objectives, both qualitative and quantitative approaches were used. Chapter 1 involved a narrative review of the literature and provides a historical account of the development of the body of knowledge around eminence in sport. In Chapter 2, a definition of eminence in a specific context of endurance cycling sports was explored. This was done by way of a Delphi method including two panels, the first involving athletes and the second involving technical experts such as coaches and performance directors. The next two chapters utilized quantitative designs where the research questions explored the developmental histories of athletes who met the criteria of eminence. For all quantitative analyses, the Union Cycliste International (UCI) database, which includes all UCI sanctioned race results from across the globe (~ 4.7 million race results), was interrogated. Several notable findings were revealed. In Chapter 2, criteria for the achievement of eminent status in endurance cycling sports were established, which included the need to win at least three high level races including the World Championships, Olympic Games and some sport specific races and general classifications such as the Tour de France and the Tour of Flanders in Road cycling and the World Cup general classification in Mountain Bike. Chapter 3 revealed that becoming an eminent cyclist was more strongly related to a diversified approach to racing across multiple cycling sports rather than fully specializing in a single cycling sport. Findings also revealed that those that achieved eminent status reached their first podium in elite races at a much younger age than those who did not. In Chapter 4, it was revealed that whilst junior to elite predictability varied between sports, junior performance in these cycling sports has a higher predictive value than suggested in other pieces of research on talent identification and development. In summary, this research program provides a framework for

researchers to explore eminence in sport specific settings and provides guidance for administrators and coaches to increase the likelihood of athletes becoming eminent through specific selection and development approaches.

Dedication

To my parents, Jan and Marlene. Your examples of hard work and love for sport have spurred the direction of my current career and instilled the values of hard work, dedication, fun and exploration. Your unwavering dedication to make the best out of every situation is something that has helped shape core values in me that help me through daily life, including work and this journey to a PHD. I can still remembers sitting in Physical Education classes multiple days a week before even going to elementary school. Playing with older kids, doing multiple sports and always being active. Through formative years of engagement and sport, foundations around teaching, coaching, exploration and a drive to excel formed. This was only because you demonstrated these virtues and supported me at every step. Even whilst disengaged with school through my teenage years and the frustration that must brought for you, you have been nothing but supportive and encouraging.

To Liva, the love of my life. Thank you for being a rock and supporting me through this ongoing journey. Thank you for being by my side and for allowing me to chase all opportunities to date. The sacrifices you made by doing so, moving to multiple countries and different continents over the years, sacrificing the roots you built in those countries and being side by side on these adventures is something that I will forever hold dear.

Acknowledgements

To Dr. Joe Baker. I vividly remember coming into your office whilst working for Cycling Canada to discuss talent issues. Somehow, this conversation evolved from exploring and challenging theorical constructs and how these played out in the Canadian Olympic and Paralympic landscape to finding a pathway to complete a PHD. Without your early encouragement and ongoing guidance, the realisation of this dissertation could have never happened. I would also like to acknowledge Dr. Kathryn Johnston. After a blistering start which saw the majority of course requirements and research completed in the first year, a prolonged period of limited activity followed. This due to changes in my professional and geographic environment. Your injection of energy and editing support sped things back up again and most certainly provided the momentum to push this work to completion.

To my friends Peter Keen and Kevin Field, your expertise and mentorship started me out on this journey. Your curiosity and vast experience led to the formation of the topic of this dissertation, and I'll be forever grateful for your kindness, expert guidance and for opening your networks to me. From a cycling research, strategy and leadership perspective, you taught me more than anyone. On a personal note, I am lucky to be able to call you both friends.

This section would not be complete without a special mention of Dr. Bob Heere. Whilst not a part of this thesis nor my time at York University, he was the one that started me off on my academic journey. Back in 2009, Bob somehow managed to get me accepted by the University of Texas in a doctoral program. Whilst this turned out to be a false start, the seeds you planted eventually sprouted albeit in a different field.

I also want to give a special thanks to Marne Fechner and Michael Smith. Marne as the CEO and Michael chair of the High-Performance committee of the board at AusCycling respectively, hired me into my current role and gave me an opportunity to translate my research findings into a real-world scenario with one of the greatest Olympic teams on the planet.

Table of contents

List of Tables

List of Figures

List of Appendices

Chapter One: General Introduction

NOTE: All references are provided after Chapter 5

Introduction

The traits and characteristics of those who advance the frontiers of human capabilities have fascinated scientists for centuries. Studying those who operate at, and indeed advance, the edges of human performance may hold lessons that could benefit athlete development for all ages and stages of competition. For example, studying the journeys that elite athletes take to reach Olympic level performance may inform the pathway for others who are training for international competition. Moreover, identifying mental skills that drive elite performance (e.g., flow states; Csikszentmihalyi, 1988) may have relevance for health, wellness, and performance in the general population.

Identifying appropriate groups and/or individuals to study, however, can be a challenge due to a lack of demarcation criteria when it comes to defining elite populations. With a long list of names including 'elite', 'expert', 'skilled', and 'Olympic medalist', along with publications listing a broad spectrum of performance as 'elite' (e.g., 'elite' 9-year-olds), inadequate study designs and imprecise terminology can compromise our understanding of which traits and characteristics have broader benefits and why (McAuley et al., 2022; Swann et al. 2016). This issue has been highlighted in several recent publications that call for a more robust taxonomy around the notion of indexing and categorizing the varying levels of performance (Baker et al., 2015; McKay et al., 2022; Swann et al., 2016). More specifically, these studies highlighted, amongst other things, a need for a subdivision within the elite category. This notion is central to the current dissertation.

Historical background

Early work

The word *genius* can be traced to ancient Rome (Leisman, 2012). Under the rule of Augustus, the term acquired its double meaning of "inspiration" and "talent" (*Oxford Latin Dictionary* [Oxford: Clarendon Press, 1982, 1985 reprinting], entries on *genius*, p. 759, and *gigno*, p. 764.). More scientific explorations of the term 'genius' occurred many years later (1860s) with the work of Francis Galton (*Hereditary Genius*; Galton, 1869). Using his cousin's work, Charles Darwin's *Origin of Species* (1859) as a model, Galton set out to study the basis of human exceptionality in Judges, Statesmen, Commanders, Literary Men, Men of Science, Poets, Painters, Musicians, senior Classics at Cambridge, Oarsmen and Wrestlers. Whilst findings of the study built the foundation for future work in the area, the use of the term "genius" in this setting received little follow up, and even Galton, upon reflection, did not feel that the term was appropriate:

> The fault in the volume that I chiefly regret is the choice of its title of Hereditary Genius, but it cannot be remedied now. There was not the slightest intention on my part to use the word genius in any technical sense, but merely as expressing an ability that was exceptionally high, and at the same time inborn. (p.viii, Galton, 1892).

Instead, Galton positioned the word "eminence" as the epitome for exceptionality studies: "After some provisional verification, I applied this same law to mental faculties, working it backwards to obtain a scale of ability, and to be enabled thereby to give precision to the epithets employed. Thus, the rank of first in 4,000 or thereabouts is expressed by the word "eminent". (p.xii)

In 1904, those building on the work of Galton found themselves in a position like current scientists exploring ‘talent’ and ‘expertise’. That is, scientific exploration was being constrained by a lack of definitions and/or clearly theory to drive these definitions. This was a challenge that Charles Spearman set out to address within the rapidly developing field of ‘intelligence’. Indeed, Spearman’s work on intelligence (Spearman et al., 1904) was driven by a desire to address the lack of clarity around many elements that would find their way into the field of skill acquisition, such as attention and hand-eye coordination, in addition to his proposed method of conceptualizing and measuring intelligence (i.e., Spearman’s general factor of intelligence, more widely known as Spearman’s *g*).

As evident in the early work of Galton, it was generally believed at the time that ability of any kind was either largely or exclusively attributable to inherited, biological factors. As such, much of the early work in the field focussed on the exploration of genetics in gifted individuals. One notable example is the Genetic Studies of Genius, begun by Lewis Terman in the 1920s (Terman et al., 1925). In this ambitious endeavor, which went on the become the oldest and longest running longitudinal investigation in the history of psychology, a sample of a thousand children who were selected based on standardized tests and ratings from teachers were tracked over time. Both mental and physical traits of these children were assessed as well as a range of cultural and environmental factors. Conclusions included, amongst other things, the earliest mentions of the effects of parent education and other familial influences on the development of gifted children. These studies are important to the field of talent research as they provided the first insights into how gifted individuals progress over time.

The 1930s gave rise to the notion of ‘motor educability’. First mentioned by McCloy (1934), this notion explored an individual’s ability to develop high skill quickly. A number of publications on the topic followed but it was Gire and Espenschade (1942) who proposed specific measurements that might predict the learning of motor skills. In effect, this work ignited a discussion around talent identification and talent selection that continues to this day. In their 1942 article, the authors proposed that "*a test of motor educability which would analyze accurately the ability to learn, or the aptitude of the individual for learning, would contribute to a better understanding of physical performance and would provide an effective tool for the administration of the physical education program*." (p.1). While this concept makes intuitive sense, support for it is weak, despite many attempts by practitioners and academics to determine the concept's feasibility, operationalization, and measurement.

The Rise of 'Expertise'

Early research on human exceptionality was dominated by studies focusing on heredity, which explored generally fixed qualities in individuals (such as intelligence, anthropomorphic qualities, familial characteristics, etc.). This body of work underpinned a general notion that exceptional performance was due to nature rather than nurture (Collins et al., 2000; Haldane, 1946; Hogben, 1933; Robinson, 2004; Ridley 2003; Wachs, 1992). Interestingly, despite the emergence of research on motor educability, it was research on cognitive perceptual skills that re-positioned the role and importance of learning in public and scientific debates. This work began with exploring differences between novice and master players in chess (first by De Groot, 1965 and later Simon and Chase, 1973) which showed that higher level players were superior at recognizing and recalling patterns of chess pieces on the board, but this capacity was limited to patterns that had meaning and structure to the game of chess (i.e., unstructured information was not recalled any better between skilled and less-skilled groups). It was concluded that the superior performance of stronger players (which does not appear in random positions), derives from the ability of those players to encode the position into larger perceptual chunks, each consisting of a familiar sub-configuration of pieces.

Similar findings on ‘chunking’ and pattern recognition were found in Go (Reitman, 1976) and Bridge (Charness, 1979). In one of the earliest studies of this phenomenon in sport, Allard and Starkes (1980) examined skilled and less-skilled volleyball players. Looking at both the speed and accuracy of performance in detecting the presence of a ball in rapidly changing slides, differences between experienced players and non-players showed markedly better performance for experienced players. Although not in accuracy, the speed in which experts could detect a specific target in various sport related situations was shown to be sensitive to improvement by means of experience or training.

In the 1990s, Ander Ericsson, considered by many as the ‘father of expertise’ began his explorations into this area. With several highly influential publications (e.g., Chase & Ericsson, 1980; 1981; 1982; Ericsson et al, 1980; Ericsson & Smith, 1991), two research achievements have stood out. The first is skilled memory theory which he pioneered together with Chase. In this seminal work, the researchers looked to increase the memory of digits and identify which cognitive processes underpinned the acquisition of this skill. It was found that two components in particular drove skilled memory. The first was to code digit groups into meaningful units and the second was to adopt retrieval structures to draw those units from long-term memory. These findings would have a profound impact on future research and training of cognitive skills which together with physical skill and other factors are key drivers to developing expertise, and ultimately eminence.

The second seminal stream of work from Ericsson is on ‘deliberate practice’ (Ericsson et al., 1993), a topic that would bring expertise research into popular culture. Ericsson and colleagues found it takes individuals in music and chess thousands of hours (roughly 10,000) of focussed training on areas of weakness. Furthermore, they noted this type of engagement was perceived as repetitive and unenjoyable but critical to the development of expertise. Although the work on deliberate practise was not originally conducted within sports, it has been widely accepted in the general population due to the popularization and transformation to the 10,000-hour rule via bestselling books such as *Outliers* (Gladwell, 2008) and *The Talent Code* (Coyle, 2009). The notion that it takes considerable practice/training to become an expert in any domain has since been extensively studied and found to be insufficient for explaining success in sports. Many differences in quantities of deliberate practice have been reported between sports, ranging from 3,000 hours to over 10,000 to reach expertise (Issurin, 2017). Regardless, the notion that deliberate practice is one of the key contributors to athletic greatness is largely uncontested.

While research on talent was relatively sparse throughout the 20th century, the 2000s saw an uptick of research publications on talent (Baker et al., 2020). It included several notable publications that made a large impact in popular culture. Nobel laureate Daniel Kahneman, who has advanced knowledge in several psychological domains amongst others in decision-making, together with Klein (Kahneman & Klein, 2009) published work on the conditions necessary for intuitive expertise. Although not directly derived from sport specific research, this work and the publication of his popular book on the matter *Thinking, Fast and Slow* (2011) had a profound impact on knowledge development around expertise. It also tied in with classical sports-related notions such as ‘flow’ (Csikszentmihalyi, 1988; Swann, 2016; Swann et al., 2012) and the traditional concept of ultra-instinct (‘mushin no shin’, or ‘mind without mind’) in martial arts that have a Zen tradition (such as Kendo, Judo, Karate, and Aikido) .

The 2000s and 2010s saw further growth in research on the psychological factors that contribute to the development of expertise, particularly as they relate to engagement in extensive training over time (e.g., Dweck's concept of Growth Mindset (Dweck, 2009) and Duckworth's notion of Grit (Duckworth et al., 2007) is another general quality in psychology that has been linked to expert development (Duckworth & Quinn, 2009). Sport scientists Macnamara and Collins (2011) proposed a set of psychological skills linked to expert performance which they aptly named *Psychological Characteristics for Developing Excellence*.

While the above factors speak to the development of particular psychological and mental skills (i.e., mental strategies that can be developed in any individual), other qualities of elite performers are largely fixed. For example, being tall benefits those who look to become elite basketball or volleyball players (Berri et al., 2011). Conversely, being 'shorter' helps gymnasts and divers rotate around their axis more efficiently and therefore could be a requirement to become an Olympian in those sports (Benardot, 2013). Similarly, having a higher percentage of fast-twitch fibers (e.g., muscle fibers that can facilitate fast and powerful contractions) can improve the likelihood of success in long jump or the 100-meter sprint but impede the ability to perform endurance and ultra-endurance tasks (e.g., marathons or Ironman-distance triathlons). These two qualities (stature and muscle fibre type) are generally not responsive to training (Wilson et al., 2012).

As such, the development of expertise, and the attainment of eminence in sport, cannot be fully understood without some discussion of the role of genetics. As Tucker and Collins (2012) note "Although deliberate training and other environmental factors are critical for elite performance, they cannot by themselves produce an elite athlete. Rather, individual performance thresholds are determined by our genetic make-up, and training can be defined as the process by which genetic potential is realised" (p.2).

To complicate matters further, there is growing evidence suggesting the environment influences how genes express themselves in aspects like behaviour (for examples see, Kendler & Baker 2007 for a review). With these fixed and non-fixed influences of genetics on the development of expertise, some have tried to find genetic markers that can predict future performance. For example, McAuley and colleagues (McAuley et al., 2021; 2023; 2024) tried to identify the influence and application of genetics in soccer through a targeted program of research. Whilst the role of genetics on performance is undeniable, our understanding of the relationships amongst key DNA sequence variations and the psychological and physiological determinants of performance is still limited (Joyner, 2019).

Exploring Eminence

Tracing back to Galton's seminal work (Galton, 1869), and more recent research (Baker et al., 2019), the apex of human performance (e.g., the top performers within expert groups) in this research program was aptly named 'eminence'. In a 2019 study, Baker et al proposed eminence as reflecting the highest level of sport attainment, beyond even expertise. Perhaps the largest impact from any investigation into the notion of eminence came from the Great British Medallists project (Hardy et al., 2017; Rees et al., 2016). Studies exploring the differences and histories of the most elite athletes (e.g., multi-Olympic medallists, generational super-stars, hall of famers and the like) juxtaposed with slightly-less-elite populations (i.e., who might have won a single major trophy in their careers) are rare but incredibly important for understanding the factors explaining skill differences (Baker et al., 2015). The Great British Medallist project presented a review of current literature on expertise development that explored three topics: (a) the performer; (b) the environment; and (c) practice and training and used a novel way to index different levels of expertise. This publication was followed by an extensive qualitative investigation of 32 Olympians from Great Britain (Hardy et al., 2017).

Prior to the Great British Medallists project, the 1990s saw several publications on the psychology of Olympians. Daniel Gould, for example, explored the mental skills (e.g., mental preparation, coping strategies, team cohesion, focus, commitment, and others) that affect Olympic performance in a series of investigations (e.g., Gould et al., 1991; 1993; 1999; Gould & Eklund, 1992) and whether interventions by sport psychologists were effective in developing these skills. This work was largely done with US Olympic athletes and coaches in wrestling and figure skating.

Defining eminence

Whilst introduced in the 19th century, the concept of 'eminence' as the ultimate stage of sport skill attainment (that is as a subcategory within the 'expert' classification/population) is relatively new (Baker et al., 2015). However, it is clearly related to other research on the process of skill acquisition and expertise development more generally. Unfortunately, while research on talent and expertise has been around since the mid 19th century and the total body of research on talent has increased nearly exponentially in the past two decades (Baker et al., 2020), there are still large gaps in our understanding. When looking at sport research done on expert populations between 1990 and 2015, for example, the cumulative number of inquiries is limited to only 307 publications (Rees et al., 2016). Within these publications, only 36 included female populations and even fewer used longitudinal designs. In addition, use of the term 'expert' in research studies has been very broad and inconsistent. For instance, studies on high school athletes or populations from university sports are often categorized as expert samples (Swann et al., 2015).

This lack of measurement and conceptual precision presents a challenge for investigators, administrators, coaches, and athletes looking to explore the frontiers of human performance. With differences in ability being evident at the highest levels of sport (i.e., all-stars vs non-all-stars in North-American professional sports, Olympic medal contenders vs "average" athletes who qualify for Olympic Games, team leaders versus helpers in World Tour cycling, Ballon d'Or nominees vs starters in Champions League soccer, Grand Slam winners versus tennis players ranked in the top 250 in ATP etcetera) it is reasonable to assume there are differences between those who compete at the university or even high school level and those super-elite athletes that are genuinely considered the 'world's-best'.

To allow for differentiation between these tiers of performers within the expert category, Baker and colleagues (2015) proposed an updated taxonomy that investigators could use to appropriately classify performers for research purposes. Akin to the classification by Rees et al (2016), Baker and colleagues made a case to index performers based on three classifications; (1) training level, (2) skill level and (3) competition level. To establish training levels, performers were categorized as either not engaging in training, training inconsistently or training regularly while the level of competition within the taxonomy ranged from competing at regional/local levels to the state/provincial level and from national to international. Skill levels started at 'naïveté', followed by 'novice' and progressing to 'basic', 'intermediate', 'advanced' and ultimately 'expert' and 'eminence' at the international competition level. To underpin their case for eminence and sub-dividing this group at the zenith of human performance from the expert class they argued that "It would be valuable to establish a set of clear criteria for the identification of an "eminent" athlete to overcome some of the limitations of previous work on exceptional performers" (Baker et al., 2019. p.5).

This highlights an on-going consideration for researchers studying the highest levels of skill. Performers who ultimately fall within the eminent category are difficult to access and the limited numbers of participants can lead to underpowered studies from a statistical standpoint. However, these statistical considerations only serve to emphasize the importance of clearly defining skill levels to avoid comparing 'apples and oranges' (i.e., the expert from the eminent). Furthermore, even given these statistical limitations, there is tremendous academic and social value in studying eminence in athletic, academic, and political domains (Ploutz-Snyder et al., 2014). Moreover, books by or about top athletes continue to place on international bestselling charts, highlighting the continued interest people have in eminent performers in many areas of society. Finally, understanding the factors affecting the development of eminent performers informs a broader understanding of human physiological, cognitive, and mental capabilities.

On the Need for Sport-Specific Investigations

As it concerns the development of such expertise, and ultimately eminence, early work identified there are primary and secondary influences to consider (Baker & Horton, 2004). Examining sports by indexing their specific primary influences (involving the performer namely, training requirements, psychological requirements, and genetic propensities) and secondary influences (those shaping the environment; namely socio-cultural and contextual) (Baker & Horton, 2004) reveals vast differences between sports. This complexity of sports is well captured by Pol et al. (2020), who argued there are two main characteristics of constraints relevant in sports: "(1) they act at different nested timescales, and (2) they are circularly interdependent (bottom-up and top-down)." (p9). An example of differences in nested timescales would be the timescale at which a human cell reacts at a change in environment compared to the time required for a whole team to react to an environmental change. Circular independence, on the other hand, reflects the process by which a gene influences changes in a cell, a cell influences organs, organs influence the overall athlete, the athlete influences a whole team (and vice versa).

From this perspective, organismic levels (e.g., genes, cells, tissues, organs, players, teams) are related through circular causality (i.e., individual training having effects on team performance, team performance on the entire competition, and, at the same time, the level of competition affecting team performance and team performance affecting how individuals train). Both the Baker and Horton (2004) and the Pol et al. (2020) models describe variables that differ based on the demands of the sport, the performer(s), the sports culture, and the general culture in which the athlete is embedded.

Collectively, this research and the theoretical models on which it is grounded in emphasizes high degrees of nuance that are not currently captured in the evidence base (e.g., Baker et al., 2020). For instance, the factors discussed by Baker and Horton (2004) and Pol et al. (2020) suggest sport and regional profiles need to be understood to effectively 1) evaluate talent and 2) predict the future development of the athlete and the sport. Obviously, these issues are conceptually too broad and cumbersome to be resolved in a single research project. However, focusing on understanding what 'eminence' means in a single sport is a strong place to start. Categorizing an eminent class could have benefits not only to those interested in talent and expertise development, but may have trickle-down effects for those examining human performance in sport genetics, exercise physiology, sport psychology, biomechanics, sport sociology, sport management and sport governance, among other disciplines. Setting definitions of eminence in a sport-specific setting (i.e., cycling) might help to advance not only studies of talent, expertise and skill acquisition, but might have second order effects on other scientific fields that operate with novice – intermediate – expert classifications. To this end, a logical starting place is trying to answer the question, "*how do we define/measure eminence*?"

Approaches to defining and exploring eminent populations

Over the past decade, there has been increasing interest in a sub-group of the population of expert performers, to understand what makes this group unique. For instance, Rees et al. (2016) categorized super-elites as those that have won multiple Olympic Gold medals or World Championships in their respective sports. However, while this was a reasonable criterion for their project exploring Olympic performers, it is too crude to be useful across all sports. For instance, not all sports participate in the Olympic Games, nor is it likely that all players on a World Championship team would be considered of equal quality.

The taxonomy proposed by Baker and colleagues (2015) allows for a more inclusive measure. They proposed that multiple indicators could be considered and started with an early suggestion of four potential criteria to explore eminence in sports: (1) MVP arguments, (2) Hall of Fame arguments, (3) career length arguments and (4) Lotka-Price arguments[1]. While this initially proposed set can be defended, it is overly centered around the North

---

[1] The Lotka-Price Law was initially found in studies of scientific productivity (Lotka, 1926; Price, 1963). In general, it proposes that in a given domain the number of performers making a specific contribution (e.g., publishing n papers or winning n events) is proportional to $1/n^2$. This inverse square relationship suggests that for every 100 performers winning a single event, 25 will win two events, 11 will win three, and so forth.

American cultural sport paradigm. Most sports that have their origins in Europe or Asia do not have MVPs or a Hall of Fame. Moreover, Lotka – Price arguments might not be universal since they are influenced by the depth and breadth of participation in a domain and these factors vary considerably across sports. Given cultural and organizational differences between sports and in both history and depth of competition, it stands to reason that criteria to demarcate eminence will vary from one sport to the next to accommodate the uniqueness of each sport.

In order to establish these criteria, we discuss three different approaches:

1) ***Field defined*** - Coaches and athletes performing at the elite, or potentially eminent, level have both a lived experience and intimate knowledge of varying levels of performers within top (e.g., international) competitions. Having experts within their respective field set the criteria for eminence would, therefore, not only yield a high level of sport-specific accuracy, it would also simultaneously generate 'buy in' from the community having potential positive net effects on the impact of publications. For this class of exploration, Delphi-type studies would be particularly valuable (Brady, 2015; Pawlowski, 2004). A limitation of this kind of an approach, however, is that true, internationally-recognized experts are limited in numbers and challenging to recruit for research.
2) ***Mathematically defined*** - As often the case, domain specific experts could set parameters to define eminence on a sport-by-sport basis. Multiple statistical methods for inquiry could be used as designs. Lotka-Price methods could fit this kind of approach as proposed by Baker et al. (2017). Equally viable approaches would include the utilization of Bayesian rating systems such as Glicko (Glickman, 1998), True skill (Doswell et al, 1990) and/or ELO (Elo, 1978) type ratings. These rating

systems allow for rankings based on head-to-head encounters which can drive scores of individuals up or down. They are usually adjusted to account for large differences in scores (i.e., if someone with a high score beats someone with a low score, there is no change) and some even include adjustments of the whole ranking list without direct head-to-head encounters (for example, if player A beats player B, the score of player A goes up. If subsequently, player B beats player C, the scores of both player B and A go up). This allows for very strong ranking systems in which not all players have to compete against each other, making it particularly practical for individual sports. As with any system, it has limitations when considering applicability to determine eminent populations. One such limitation would be that the cut-off point between expert and eminent would remain somewhat arbitrary and outcomes might not align with perceptions of the actual experts themselves or consumers (fans) of any particular sport. A second limitation is that whilst exceptionally potent in individual sports, these systems are less useful in team sport settings. As players on teams move around, tracking scores of teams over time becomes blurry and oftentimes even inconsistent.

3) ***As defined by the general population*** - Another method would be to utilize the wisdom of crowds (Surowiecki, 2005). Large scale, unstructured online inquiries amongst fan bases could be used to define eminence. This kind of design could be particularly strong as it has been reported to reduce systematic biases when compared to closed design reviews, benefitting scientific quality of qualitative research should sufficiently large populations be recruited (Iyer & Graham, 2012).This kind of methodology would be in line with work by Gigerenzer and Gaissmaier (2011) on the use of simple heuristics.

Ultimately, more robust ways to define skill groups in a taxonomy of expertise can be achieved through multiple methods. For the field to advance and explore the nuances between sports, or indeed expert performance of any kind, domain specific criteria need to be formulated in a more detailed fashion. Allowing for this domain specificity and demarcation of subgroups in the elite field holds a promise of increased understanding through more specific inquiries.

In cycling sports, there is no Hall of Fame, and best international rider awards are not given, and whilst career length might be an indicator of eminence, it might also be that one enjoys a long career in cycling without ever winning a race (by being a very good and trustworthy helper in a road team, for example). If one wishes to examine the characteristics of eminent cyclists to gain more knowledge on developing talent in the sport, psychological or physiological developmental histories or establish requirements to win the most significant cycling races, it would be helpful to have guidelines regarding how to determine the parameters of the population under investigation. As such, this research program focuses on defining and exploring eminence in endurance-based cycling sports (i.e., longer duration cycling sports).

Research Objectives

The purpose of this dissertation was to define and explore eminence in endurance cycling. This introductory chapter on the literature of eminence and expert performance in sport is followed by three investigations into eminence in endurance cycling sports. A more detailed description of each chapter is provided below.

Chapter 2 sought to establish inclusion criteria for eminence in endurance cycling through a Delphi study with two groups: an athlete group consisting of World Championship medallists and a group of technical leaders (e.g., coaches, technical directors and performance directors) who have guided athletes to a minimum of three Olympic or World Championship medals.

Chapter 3 explored the characteristics of those performers that have met the inclusion criteria established in Chapter 2 through the analysis of all elite race results globally between 2010 and 2020. This is done by building on methods that have been successfully used in similar investigations in ski sports (Barth et al., 2018). More specifically, this study investigated the prevalence and impact of taking a multi-sport approach to elite cycling compared to an approach that focused on participation in a single sport.

Chapter 4 investigated the relevance of junior performance to future elite performance in these cycling sports. Whilst general scoping reviews indicated that junior performance has low predictability for elite performance (Li et al., 2018), cycling specific research done on national samples in Germany (Schumacher et al., 2011) and Denmark (Moesch et al., 2011) provide conflicting findings which warrant a global investigation.

The dissertation concludes (**Chapter 5**) with a summary and general discussion of this program of research. This chapter also provides implications for various stakeholders such as administrators, athletes, and coaches.

Chapter Two: Defining Eminence in Endurance cycling sports: A Delphi study with some of the World's most successful cyclists, coaches and directors

Jesse Korf and Joseph Baker

NOTE: All references are provided after Chapter 5

Abstract

In order to define criteria for eminence in endurance cycling, a Delphi study with World Championship medal winning athletes and technical leaders in relevant cycling sports was conducted. The Delphi study involved two independent groups: athletes (n=9) and technical leaders (n=13). Participants in the athlete group had all won at least one World Championship or Olympic medal while the technical leaders have supported a minimum of three such performances. A survey using a balanced Likert-scale design with a randomized question sequence was conducted followed by a round of virtual discussions using an anonymous chatroom and video conference media. Findings from the survey were validated and discussed in the separate groups and consensus was set at 70%. Cross-group outcomes were only shared upon completion of the study. The definition for eminence in endurance cycling sports generated through this study indicated cyclists must win a minimum of three races in an exclusive subset of races containing World Championships, Olympic Games, Monuments and Grand Tours or World Cup overall classification over the course of their careers to be considered eminent.

Key Words: talent, eminence, cycling, Delphi study, expertise, coaching

Introduction

Professional teams in mountain bike and road cycling operate on multi-million-dollar annual budgets (Van et al., 2016). In addition, many nations invest millions of dollars per year with aims of Olympic medal performances in cycling sports (Groot, 2008; Hogan & Norton, 2000; Houlihan & Zheng, 2016). Through their performances, top cyclists (e.g., the most prolific winners in peak international races such as the Olympic Games and World Championships) have a major impact on both society and the cycling industry at large (e.g., through audiences of tens of millions of spectators for some races). In addition, these riders can help drive innovation and marketing of equipment such as bikes and clothing which eventually trickle down to consumer products. From a scientific perspective, however, we know very little about performers at the highest levels of achievement (i.e., the 'eminent', Baker et al., 2019) either in general or in specific sports such as cycling. Furthermore, there have been no retrospective or longitudinal studies on the development of expert or eminent cyclists to date (Baker et al., 2020). This study explores the concept of eminence in cycling disciplines, by exploring how key stakeholders conceptualize and define this term. Defining what eminence in cycling sports looks like is a first step to informing future research examining performance characteristics, team selection, pathways, and other aspects of athlete development.

Importantly, cycling is an umbrella phrase for a number of sports with different characteristics. It includes sports that are measured by time (often described as Centimeter, Grams, and Seconds [CGS] sports as a class) such as Road cycling, Track Endurance, Track Sprint, Mountain Bike, BMX, Cyclo Cross (cx), and E-racing. These disciplines can be further sub-divided into endurance (e.g., Track Endurance, Road, Mountain Bike, Cyclo Cross) and power sports (e.g., Track Sprint, BMX Race), each with vastly different race characteristics and performance demands. Cycling also includes artistic sports such as BMX Freestyle and Game sports such Cycle Ball.

In endurance cycling, there have been several research investigations of physiological characteristics, including examples of athletes who might be considered ‘eminent’ (Baker et al., 2019; Craig & Norton, 2001; Franciscus & Merkes, 2020; Gardner et al., 2005; Schumacher et al., 2006; van Erp, 2019). However, there has been little scientific exploration on the *development* of elite cyclists, with the exception of two studies that looked at the correlations between junior and elite performance (e.g., Mosteart et al., 2021; Henriksen et al., 2013) and a recent publication on relative age effects in road cycling (Mosteart et al., 2021).

Capturing and understanding the nuances between cycling sports, or indeed expert performance of any kind, requires the establishment of domain specific criteria to clearly determine athlete skill/achievement level. This domain specificity and demarcation of skill groups holds the promise to increased understanding through more specific inquiries (e.g., clearer designs and methods). To this end, this study attempts to define the upper-most echelon of elite performers in endurance cycling, to establish sport specific criteria for indexing aptitude, as proposed by Baker and colleagues (2017). This taxonomy could subsequently be used to explore topical issues in talent research in cycling in particular, or as a model for other sport specific contexts.

Methods

In order to establish what eminence means in endurance cycling (e.g., what do cyclists have to do to be considered eminent?), a Delphi study with World Championship medal winning athletes and technical leaders (i.e., coaches, sport directors and performance directors) in relevant cycling sports was conducted.

Study design

The study was comprised of three phases. Phase one involved participant recruitment. Using the network of contacts from the primary investigator, participants from multiple endurance-based cycling sports were contacted to participate in the study. Out of these participants, two panels were created. To meet the high standard for inclusion in this expert panel, the first group, athletes (A; n=9), had to have obtained a minimum of one Olympic or World Championship medal. Final members of the athlete group were from Canada, the USA and the Netherlands representing Road, Track and Mountain Bike events.

A second group, technical leaders (TL), was included due to unique elements of professional cycling; for instance, mountain-bike and road are structured differently. In addition, many National Sporting Organisations (NSO) have national teams that compete in the Olympics and/or World Championships in specific cycling sports. Across different environments, coaches, trainers, sport directors or technical directors have different roles relative to training design, team design and race deployment, and all would be considered topical experts. Members of the TL group had to have led one or more athletes to a minimum of three Olympic or World Championship medals. Members of this group were from Belgium, The Netherlands, Great Britain, Australia, Switzerland, Canada, Spain, and United States. The athlete and technical leader group combines a total of 148 World Championship and Olympic medals won, and can be viewed in Table 1.

Starting questions were formed through informal conversations with both athletes and coaches that have participated in Olympic cycling competitions to ensure that questions had relevancy. These starting questions helped develop the questions for the starting survey.

In the second phase, participants completed the aforementioned survey that used a balanced Likert-scale design with a randomized question sequence. This approach was chosen to ensure that both groups had similar starting points. Using a Likert-scale with a randomized question sequence also allowed the researcher to compare findings from both groups to examine initial convergence or divergence between views of independent groups. As a result, it allowed for items to be removed from subsequent discussion or highlighted during the discussion during the first of three interview rounds in phase two.

Broadly, the goals of this survey were to gain insights in three areas: 1.establishing tiers in race levels, 2. establishing qualifications required to be considered eminent, and 3. gaining early insights into pathways towards eminence. Establishing tiers in competition was important as there are many races throughout the year, yet not all races are considered of equal prestige. Riders choose to peak for races (i.e., World Championships or the Tour de France) and as such, pre-select a small subset of races they deem most important. Establishing the most important races, as reflected by those that the best riders focus on, was considered important to assess which races should be included in the demarcation criteria when investigating eminence. Furthermore, beyond race results, other qualifications might be relevant in determining whether a rider is eminent. Popularity or skill, for example, might be considered by some to be important markers.

Phase three involved discussions with the expert groups to develop a cohesive and comprehensive framework for assessing cycling eminence. Given the limited guidance around Delphi studies, we attempted to increase the analytical robustness of the study design in several ways. First, both athlete (Panel A) and technical leader (Panel TL) groups were interviewed independently from each other to increase the validity of findings (i.e., through convergence) and reduce or buffer any biases that might exist in either population. Second, both groups helped improve the study design as issued raised by one group helped tailor questions for the other group. This allowed for similar topics to be discussed independently by both groups yet ensure that no critical issues were missed. Overall, the goal was to establish objective demarcation criteria for eminence in cycling that could be used in further quantitative analyses. The ability to discuss and revisit previous positions was introduced via online methods. Participants were given the option to participate via a virtual, anonymous chatroom (Leapchat.com), or via video conference call (Zoom rooms version 5.0).

The discussion portion of the study involved a prespecified number of rounds (n=3). This design was chosen due to several factors. First, participants came from three different continents making it challenging to meet at the same time. Second, participants were in a pre-Olympic year posing considerable constraints on their availability. Third, the global COVID pandemic made it impossible to set of group discussion in-person even if participants were in relative proximity to each other.

A methods journal containing structure, decision rules, and justifications to changes was used during the onset of the study and notes were kept during the online discussions as well as throughout the data collection stages. Consensus for both the survey and discussion were set *a priori* ($\tilde{x}$ = 70% agreement on a criterion) as this was considered appropriate given the selected method of two independent groups and the diversity of participants. Furthermore, this is only slightly below the mean consensus used in medical studies using a single group design ($\tilde{x}$ =75%) (Diamond et al., 2014).

Following the survey round, the following criteria were used to include, discuss or reject items: Accepted: >70% in both groups, discussed: >70% in only one group, rejected: <70% in both groups. Both items from the survey that were over the threshold in both groups (accepted criteria) and those that were both under the threshold for both groups (dropped criteria) were presented back to the individual groups for final review.

Items with uncertainty were discussed amongst the panels individually (same group) and cross group findings were shared after the study was completed.

Ethics

All participants provided informed consent to participate in the study. The interview guide, informed consent forms and overall study design received ethics approval from the authors' institutional ethics review board.

Results

The first topic of inquiry for the panelists centered on whether there was a specific subset of races within endurance cycling events that were considered to be of the highest quality. To explore this possibility, panelists discussed elements of race 'prestige', depth in the field of competitors, and historical trends regarding proportion of top athletes who would peak for different events. Interestingly, after discussions with both panels in Round 1, Cyclo Cross was included in the category of events to be considered. Subsequently, there was discussion around which Cyclo Cross races or classifications should be included. Ultimately, based on panel feedback only the Cyclo Cross World Championships was included as the key race of establishing eminence in this sport. For example, A3 noted: "Cyclo Cross should also get consideration. But the world cup system is changing, so many different races and it is not very international. Women maybe more than men. The only race we should consider are the World Championships as I am not sure on how to place the other races." Similarly, A1 stated "most of the top performing Cyclo Cross riders are top Mountain Bikers or top Roadies that cross over and it is only a part of their season. They come in with good form in races that matter but are also picking and choosing based on the calendar so it doesn't interfere with their primary sport. We should only look at events later in the season and perhaps only Worlds as it is the only race where you have everybody participating." For the races and classifications for Road, Mountain-Bike and Track (endurance) categories, both panels found consensus on the same subset of races (displayed in Table 2).

After establishing this set of key events, five items were used to help guide the discussion around qualifications to reach eminence (Table 3). First, career length (A=74.2%, TL=76.6%) and winning the World Championships, Tour de France (men)/Giro d'Italia Femminile (Giro Rosa), a monument or an Olympic medal (A=71.4%, TL=76.6%) were accepted as important considerations for individuals to be qualified as eminent.

Whether or not cyclists can be considered eminent after a relatively short career (A=77.2%, TL=58.4%) was brought into the panels for further discussion. The possibility of being considered eminent after a single short period (A=51.4%, TL=46.6%), the years ranked in the UCI top 10 (A=60%, TL=63.4%) and the opportunity to become eminent without winning any of the agreed upon 'premier league' races (A=48.6%, TL=46.6%) were 'rejected' at this stage, and put forward to the panels for further review.

Panel discussions

In Phase 3 of the Delphi (the panel discussions), both groups considered whether a cyclist could be classified as eminent after a relatively short career. Both groups found that while an athlete might show traits of eminence, ultimately one must win multiple premier league events (e.g., World Championships, Olympic Games or Grand Tours) to be considered eminent. Both groups supported the need for further differentiation between pre-eminent stages and consensus was reached on a qualification of 'emerging eminence' to reflect cyclists who had won at least one of the listed premier events.

Beyond performance milestones, various developmental traits and circumstances were also discussed. For instance, the panel raised whether the ability or perhaps need to overcome adversity was an important element of eminence, not dissimilar from discussion points noted in the Great British Medalists study (Hardy et al., 2017). Furthermore, the ability to win in multiple cycling sports was also broadly discussed, similar to previous inquiries in ski sports. More specifically, Barth et al (2018) found that both serial winners (those who won multiple events in the same disciple) and multiple winners (those who won in varying disciplines) are common in the top of the elite categories.

Both points were captured well in one of the arguments in the Technical Leader (TL) group. For example, TL 11 noted:

> One must be able to win multiple top races, over multiple seasons, while demonstrating versatility and composure. Versatility = crossing disciplines or types of races. Composure = ability to handle high pressure, both on and off the bike. Including, when needed, coming back from setbacks. Top races are: world championships, monument classics (or like), world cups, Olympics, grand-tours (GC or stages). These are the best races, the most prestigious. In addition, overcoming adversity should be considered. For example: - Vos: winner in Cyclo Cross, Track and Road – now returning to winning ways from a 1-year career set back. Ferrand-Prevot: wins in Cyclo Cross, Road, Mountain Bike. Wiggins: went from Track - to - Road, wins in short 1-day events as well as grand tours. Theo Bos: from Track Sprint to Road and back There are more athlete examples. However, these few all demonstrate multiple wins, versatility, and composure.

Finally, milestones and transitional phases were discussed to gain understanding of the potential pathways to eminence. A7 shared:

> I think there is a ladder of achievement in the sport that is important, like winning at local level, regional level, provincial or state level, national level and then international level. Some skip a level or two, especially if they come from another sport that they have participated in at a high level. I believe that in climbing the ladder to eminence, it is important to have exposure to as many different types of races and disciplines within the sport of cycling as possible. It is also important to aim for success at one level before moving onto the next, as jumping into too high of a level that the athlete is not yet ready for.

Similarly, TL 3 believed there are three stages athletes move through during their careers and necessary to achieve the aptitude linked to eminence.

> Much like Maslow's hierarchy of needs, a cyclist and coach duo can evolve in the same way. Initially, riders are purely physiological specimens who need complete structure and guidance. I would call this the 'robot stage'. They are sponges of information and follow commands almost without question. Then, as they evolve, they develop more knowledge and autonomy and become more independent professionals which I call the 'learning' phase. Finally, as riders evolve further their autonomy allows the coach to play a much smaller role. In essence act as a sounding board and this is where athletes are the driving force behind the process and peak performance is achieved. I call this the 'expert' phase.

TL 6 offered the following commentary around the development of eminent cyclists:

> I think the evolution can be very different between cyclists. Some are so extremely talented that even without significant experience, they can immediately compete at the highest level (Mathieu van der Poel, Remco Evenepoel, Egan Bernal and Tadej Pogacar

as some current examples). However, others need to build up experience before gradually growing to the level of eminence as indicated above. To my experience, it is also very important that general athletic development via different sports, including different disciplines in cycling should be the primary focus until post-puberty. Great competitors at the age of 12-16 seldom become great adult cyclists, at least on the road. If young riders need a very professional approach and very high volumes of training to be able to win a race, they probably won't make it at a later age. It is crucial to maintain sufficient spare capacity to increase specific cycling training workload as long as possible during athletic development. This is also important to avoid mental fatigue even before 'the real game' starts. I have seen too many young cyclists fading out at an early age because of too professional an approach too early on.

While many points were discussed, consensus was reached in both groups through the discussion on three items to expand on the key 'prestige' or 'premier league' races identified in Phase 2. After the survey and discussions phases, the group showed consensus on the following criteria of eminence:

1. Cyclists must have won a minimum of three (3) 'premier league' races (as defined in Table 2).
2. This may occur in the same sport or via a combination of cycling sports.
3. Cyclists who have won at least 1 but less than three 'premier league' events reflect 'emerging eminence'.
4. The pace or rate at which one accumulates criteria 1 and 2 is not relevant.

Discussion

Although the fields of skill acquisition and expertise are increasing at rapid rates, studies of eminence are rare due to challenges around access and sparsity of subjects. Yet, the topic has been part of the academic debate since 1869 (i.e., Galton's initial work) and recent discussions highlight the value that rare, small sample sized studies might provide to scientific understanding (Ploutz-Snyder et al., 2014). In sport, some publications have begun exploring eminent groups such as comparisons between champions and super champions (Collins et al., 2016; Collins & Macnamara, 2017; Farrow, 2017) as well as medallists versus non-medallists or multiple medallists in Olympic sports (Güllich, 2017, Hardy et al., 2017). However, contextual differentiators (such as differences between individual sports and team sports, Olympic versus non-Olympic, depth of competition) and sport specific taxonomies have been lacking, creating potential risks of over-generalisation between the field of expertise as a whole and its relevance for individual sports that have unique cultures, traditions, and development systems.

The objective of this study was to define sport specific criteria for the highest level of achievement in endurance-based cycling sports, allowing for more robust inquiries about eminence in cycling and a potential framework for future inquiries. Panel members in this study highlighted several areas for future exploration such as the value of domain specialization (e.g., focusing on only Road or only Mountain Bike) versus multi-domain engagement (e.g., participating in Road and Mountain Bike) for the development of expertise and eminence in cycling sports, which have implications for programming, training, and athlete pathway designs.

A greater understanding of the costs and benefits of single versus multisport cycling engagement during key phases of development could contribute to selection and elite program design within the professional and Olympic cycling communities. A number of exceptional cyclists have managed to win races at the highest international levels in multiple sports. For example, several athletes have managed to be successful in Track and Road cycling, including Bradley Wiggins, Elia Viviani, Amy Pieters and Chloe Dygert. Similarly, multiple World Champions such as Marianne Vos, Wout van Aert and Mathieu van der Poel have been able to attain success in both Cyclo Cross and Road cycling. On the Track, one of the premier disciplines, the Team Pursuit, often sees medal-contending nations fielding multiple riders who also compete on the Road. Interestingly, the most recent Olympic Games in Tokyo saw Tom Pidcock of the United Kingdom, a top contender in some of the world's major road races and a member of the world's wealthiest road cycling team (Ineos Grenadiers) win the men's Mountain Bike race. This ability to compete in multiple cycling sports appears to be becoming more frequent. While this has not yet been confirmed in cycling, the possibility of crossing over between disciplines in endurance sports has been noted in cross-country skiing (Barth et al., 2018). All this to say, the Delphi approach raised several intriguing areas of future exploration.

Limitations

While this investigation had notable strengths, there were several limitations to our design. First, although it is possible that in-person conversations may have led to different and/or more fruitful discussions, the relative scarcity of participants at this level of skill, their geographical location around the globe, and limitations due to the SARS-COVID19 pandemic made this impossible for this study. In the current study, these concerns were somewhat mitigated by increasing the flexibility to participants in how they engaged with the discussions, such as offering the choice to partake in either video conference discussions or anonymized chat rooms.

In addition, the number of participants in some areas was very low. For example, the investigators only managed to recruit a single female Technical Leader, which likely reflects the low number of female technical leaders in the sport and the stringency of the inclusion criteria. Mountain Bike and CycloCross also had a relatively low number of representatives. While this does not necessarily constrain the design nor the value of Delphi-type studies, the possibility exists that this may have skewed potential findings. Further, this study only focused on endurance-based cycling sports, meaning sports such as BMX SX, BMX Freestyle, Track sprint, MTB Downhill and others were excluded from inquiry. Future work should consider the unique elements of performance defining eminence in these sports.

Conclusions

Using Delphi methods to define definitions of target groups can create a rich discussion and has proven to be effective in setting definitions for eminent populations. While further discussion is needed regarding the creation of a taxonomy with well-defined indicators of categorization at lower levels of skill, both World or Olympic medal winning athletes and technical leaders agreed that in order to be considered amongst those at the highest level of achievement, cyclists must win a minimum of three races in an exclusive subset of races containing World Championships (all cycling sports), Olympic Games (Track, Mountain bike and Road), Monuments (Road) or World Cup overall classification (Mountain bike) over the course of their careers to be considered eminent. Active cyclists who have won at least one of these events can be considered emerging eminent. These definitions provide a more defined taxonomy for researchers exploring this topic. The findings of this study might help future researchers considering the various sciences related to cycling sports, such as exercise physiology, psychology, sport management, biomechanics and kinesiology, to better define their sample groups. Conversely, practitioners working with professional or Olympic populations might have a more informed way to qualify research articles in their field when assessing validity of findings for their own groups. This work builds on the call for more detailed taxonomies in sports by Baker et al (2019) and is the first to provide definitions for eminence in a sport specific context.

Table 1: Demographics of expert panels

Athletes (A) and Technical Leaders (TL) might be/have been active in multiple cycling sports.

		Olympic & World Championship Medal Totals	Sample Size	Track	Road	Mountain Bike	Cyclo Cross
A		31	9	3	8	1	1
	Male		3	1	3	0	0
	Female		6	2	5	1	1
TL		117	13	7	13	4	2
	Male		12	6	12	4	2
	Female		1	1	1	0	0
Totals							
	Sample		22	10	21	5	3
	Medals	148		74	54	4	16

Table 2: 'Premier League' races and classifications in cycling

	Male	**Female**
Road	World Championships, Olympic Games, the Monuments (*Milan-San Remo, Paris – Roubaix, Tour of Flanders, Liege - Bastogne – Liege, Giro di Lombardia*), the three Grand Tours (*Giro d'Italia, Tour de France, Vuelta a Espana)*	World Championships, Olympic Games, the Monuments *(varies per year)* and the Giro Rosa
Mountain Bike	Olympics, World Championships, overall World Cup winner	Olympics, World Championships and the overall World Cup winner
Track	Olympics, World Championships	Olympics, World Championships
Cyclo Cross	World Championships	World Championships

Table 3: establishing qualifications of being considered eminent

Item	Athletes	Technical Leaders	Accepted(A)/Rejected (R)/ Discussed (D)
Career length is an important differentiator between non-eminent cyclists and eminent cyclists	3.71 (74.2%)	3.83 (76.6%)	A
In order to be labelled eminent, one has to have won World Championships, The Tour de France/Giro Rosa, a monument or Olympic Games	3.57 (71.4%)	3.83 (76.6%)	A
A cyclist can be considered Eminent after a single short and successful period	2.57 (51.4%)	2.33 (46.6%)	R
The number of years ranked in the top 10 of the UCI ranking is an important differentiator	3.00 (60%)	3.17 (63.4%)	R
A cyclist might be considered eminent without winning races in the "premier league" events you described	2.43 (48.6%)	2.33 (46.6%)	R
A cyclist can be considered eminent with a relatively short career	3.86 (77.2%)	2.92 (58.4%)	D

Chapter Three: Characteristics of Eminent Endurance Athletes in Road, track, Mountain Bike and Cyclo Cross. A Cohort Study

Jesse Korf and Joseph Baker

NOTE: All references are provided after Chapter 5

Abstract

Talent identification and development are recognized as critical components of building successful sport teams. The sport of cycling has received recent scientific attention, as the act of identifying talent early in the competitive pathway is believed to pay significant dividends in later success. For this reason, the present study explored the developmental histories of the world's 'best' (i.e., eminent) cyclists who competed between 2010 and 2020 to determine whether developmental histories influence race results, age, and progression rates. In addition, we explored whether the odds of becoming an 'eminent' cyclist differed for those who 'specialized' (e.g., focus on one specific cycling sport) compared to those who competed in multiple cycling sports whilst competing at the elite level. All race results listed on the Union Cyclisme International (UCI) database for athletes between the years 2010-2020 (N= 4,760,647) were considered. Basic descriptive statistics (e.g., age and time to different development milestones) and effect sizes were computed. Further, podium stability across the cycling sports and patterns of success (e.g., specialization versus non-specialization) were examined using Gini coefficients and odd ratios. Overall, findings revealed that multiple pathways to eminence in endurance cycling exist, but that cycling's most significant races are often won by the same subset of athletes. Results also showed that doing multiple cycling sports as an elite athlete in Track, Road, Mountain Bike and Cyclo Cross increased the odds of becoming eminent in any specific sport compared to those who specialized in a single cycling sport.

Key Words: talent, eminence, cycling, Mountain Bike, Road cycling, Cyclo Cross, Track cycling

Introduction

National teams and professional teams in cycling sports such as Mountain Biking and Road, often operate considerable annual budgets (Van Rees et al., 2022) to support athlete compensation, operations, logistics, management, sport scientists, coaches, and equipment. It is with these significant budgets that teams invest in cutting edge science in the hope of achieving podium performances at major competitions such as the Olympics or the Tour de France (Groot, 2008; Hogan & Norton, 2000; Houlihan & Zheng, 2016, Van Rees et al., 2022). To more effectively allocate the limited resources available in many sports, researchers and practitioners have focused significant effort in trying to predict the next 'superstars' through various testing and assessment practices (commonly termed talent identification; TID). These practices are common across most sports and continue to gain research interest (Baker et al., 2019).

Of the available evidence on TID in sport to date, there appears to be a variety of practices, frameworks, and theories that underpin the approaches used. For example, a recent review by Baker et al (2020) illuminated the various tests and assessments that are employed within many sport domains. Findings from the review revealed that most TID research has been unidimensional (i.e., focused on one aspect of sport performance such as physical or physiological components), but multidimensional approaches (e.g., examining multiple areas of sport performance such as physical, physiological, psychological, perceptual cognitive etc.; Baker et al., 2020 Johnston et al., 2018) are becoming more common. These tests are often designed to be sport specific (i.e., mirroring the sport demands) and/or examine aspects of an athlete's background, such as sport exposure history (e.g., what sports where played, in what capacity, and for how long), or demographic information (e.g., age, birthplace, family sport profiles etc.). Conversely, talent *development* research has generally focused on the historical practice and competition experience of athletes in relation to future performance, such as the relationships between hours of training (Baker & Young, 2014) and the number and nature of competition(s) attended early in development with measures of later success. A recent study (see Chapter 2) emphasized the importance of competition experience to attaining 'eminence' in cycling. More specifically, World- and Olympic medal-winning athletes and technical leaders agreed that to be considered amongst the 'best' (i.e., eminent), cyclists needed a minimum of three wins (first place finishes) in an exclusive subset of 'premier league' races (e.g., World Championships, Olympic Games or the five Monuments for Road).

Despite the importance of understanding these developmental trajectories in elite cycling, investigations focusing on how athletes reach eminence or the preceding phases before reaching eminence ('emerging eminence') are rare, and very few accounts of retrospective or longitudinal studies of elite athletes are available (Given, 2019; Johnston et al., 2018). Similarly, while some studies have provided insights into the physical characteristics of peak performing elite cyclists in Tour de France, World Tour and Olympic settings (Gardner et al., 2005; Granier et al., 2018; Martin et al., 2007; Padilla et al., 2000; Phillips & Hopkins, 2020; Van Erp, 2019; Van Erp et al., 2021), broader developmental histories and/or research in other performance-related areas are limited.

Despite the restricted work in cycling sports, developments in other sports have provided some promising findings (Janssens et al., 2022; Van Bulck et al., 2021). For instance, researchers exploring successful athletes in endurance sports such as cross-country skiing, have made distinctions between 'serial winners' (i.e., athletes winning in a single ski discipline but not others) and 'multiple winners' (i.e., athletes who win in multiple ski disciplines) (Barth et al. 2018). Moreover, prior research has illuminated relationships between the nature and type of sports that athletes participated and competed in during their pathways. However, recent attention has focused on clarifying the meaning and importance of key concepts. For instance, Mosher and colleagues (2020) questioned the consistency and reliability of how concepts such as 'specialization' have been defined and measured in previous work. They also questioned whether some widely held notions around the need to sample or specialize should be broadly applied across all sport contexts. The relevance of this finding to the development of elite cyclists may be important given the number of athletes who have had success in multiple cycling sports (e.g., Bradley Wiggins, Marianne Vos, Wout van Aert and Chloe Dygert).

In this examination we explore the value of specialization within a category of sports, where specialization relates to *competing in a single cycling sport at the elite level* and multisport reflects *competing in multiple cycling sports as an elite*. Understanding the value of specialized versus multisport patterns of engagement within a category of sports is important for developing evidence-informed approaches to athlete training and development. To this end, this study sought to answer two fundamental questions: 1. what did the sport engagements and performance profiles look like for the world's best cyclists who competed between the years of 2010 and 2020, and 2. in order to become an eminent cyclist, is it more beneficial to specialize or compete in multiple cycling sports as an elite?

Methods

Study design

This study examined the competition histories of eminent or 'emerging eminent' cyclists in endurance cycling sports (Chapter 2). Data for these analyses came from race results listed on the Union Cyclisme International (UCI) database for athletes between the years 2010-2020 (N= 4,760,647). Participants were considered ‘eminent’ if they met the criteria for ‘eminence’ in endurance cycling (i.e., winning three World Championships, Olympic Games, Grand Tours, Monuments or World Cup overall classifications) as defined in Chapter 2.This preliminary assessment identified 56 athletes (see Figure 1 for reference), across male and female disciplines who participated in endurance-based events (i.e., Road, Track Endurance, Mountain Bike and Cyclo Cross). This group consisted of 33 males and 23 females who reached ‘eminent’ status at an average age of 26.9 (±4.1) years (male 26.7 [±3.6], female 27.2 [±4.7]). Further, 132 athletes met the criteria for ‘emerging eminence’, which meant they won at least one, but fewer than three premier races (i.e., World Championships, Olympic Games, Grand Tours, Monuments or World Cup overall classifications; Chapter 2).

Following this, the performance development of these riders was indexed by comparing the time and rate at which the ‘eminent’ and ‘emerging eminent’ samples reached key developmental milestones, compared to the ‘non-eminent’ sample (i.e., the rest of the elite-aged cycling population in the UCI database). These milestones included a first podium at ‘premier’ races (e.g., World Championships, Olympics, Tour de France, Tour of Flanders etc.), a first podium finish in a UCI sanctioned elite competition and wins at ‘premier’ races. In addition, we considered whether serial winners (i.e., those that win exclusively in one cycling sport) differed from multiple winners (i.e., those that win in multiple cycling sports). This was done by using Barth et al’s (2018) methodological approach from Alpine Ski racing. Finally, we explored whether specializing or combining cycling endurance sports at the elite levels differed amongst the ‘eminent’ and ‘emerging eminent’ groups.

Data Analysis

All data analyses were performed using RStudio (version 1.4.1106), and statistical significance was set at $p=<0.05$. Using independent sample T-test comparisons of the age at which athletes achieved different milestones between the ‘elite’, ‘eminent’ and ‘emerging eminent’ groups examined whether the ‘speed’ (as a product of age) in which milestones were achieved was a predictor of achieving ‘eminent’ status. More specifically, data were explored by comparing the overall ‘elite’ population (i.e., any athlete that has participated in an elite, UCI sanctioned race in this period) to the ‘eminent’ group and between the ‘emerging eminent’ and ‘eminent’ groups by effect size using Cohen’s d per cycling sport and gender where effect sizes have been indicated as very small ($d = < 0.2$), small ($d = 0.2$ to 0.5), medium ($d= 0.5$ to 0.8), and large ($d = >0.8$) (Cohen, 1988).

Similar to Barth et al’s (2018) approach, ‘podium consistency’ (i.e., the same athletes holding podium spots with little variability between podium winners) was explored. Using the same methodologies, Lorenz curves and Gini coefficients were used to calculate the consistency of podium winners in premier league competitions for the four endurance sports by sex. This approach and comparison of studies is particularly relevant given the similarities between skiing and cycling as both govern multiple sports (e.g., Cross-Country and Downhill in skiing and Track and Mountain Bike in cycling) Finally, we examined whether specializing in a single sport or combining cycling sports increased to odds of becoming an eminent or emerging eminent cyclist.

Results

Podium stability

To examine the stability of podium finishes (e.g., how often did the same athletes have multiple podium finishes) Gini-coefficients and Lorenz-curves were calculated (shown in Figures 2 through 5).

These analyses showed relative stability amongst podium performers across the cycling sports. Road measured slightly higher podium stability for males (G=0.351) than for females (G=0.331). In Mountain Bike, podium stability was higher than in Road for both males (G=0.361) and females (G=0.353). In Track cycling, podium stability for males (G=0.326) was slightly lower than females (G=0.32) whilst Cyclo Cross measured the lowest podium stability (G=0.298 for males and G=0.298 for females) of all endurance cycling sports.

Performance Milestones

Four performance milestones were explored:

1. The first podium finish in any elite UCI race;
2. The first podium in a premier race (see Chapter 2 for races that are included in the 'premier' category);
3. The first win in a premier race (i.e., achieving emerging eminent status); and
4. The third premier win (i.e., achieving eminent status).

To determine any differences in the age at which the first milestone was achieved, the 'eminent' group (n=60) was compared with to overall elite population (n=92602). For the subsequent milestones, the 'eminent' group was compared to the 'emerging eminent' group (n=132; i.e., since the majority of the overall 'elite' group did not achieve this milestone).

Elite versus Eminent

Between the 'elite' and 'eminent' populations, large effects for the rate at which athletes achieved their first podium result in an elite race were found (t=13.15, df=92660; d=2.14, p=0.001). Moreover, these large effects were observed in all cycling sports examined and for both sexes. Male Road boasted the highest effect (t=6.85, df=29926; d=2.61, p=0.001) and female Road, whilst still observing a large effect, saw the lowest (t=2.6, df=7905; d=1.06, p=0.001). The average age at which 'eminent' athletes achieved their first elite podium was 20.7 (±3.2) whilst the average age for the overall 'elite' population was 31.4 (±36.3). Significant differences between the 'elite' and 'eminent' population for this first milestone were found for all sports and sexes, with the exception of Mountain Bike.

Emerging Eminent versus Eminent

Results for the comparisons between 'emerging eminent' and 'eminent' cyclists are more advanced milestones are listed in Table 4. Between the 'emerging eminent' (n=132) and 'eminent' groups (n=60), small effects (i.e., Cohen's d= < 0.5) between the first 'premier podium' to 'first win' (t=2.34, df=190; d=0.36, p=0.02), and from their 'first win' to 'third win' (t=2.34, df=190; d=0.38, p=0.02) were found. At a per-sport level, Road showed a large effect (t=2.75, df, 49; d=0.89, p=0.008) in the male group for the amount of time it took between the first elite podium and premier race podium finish. A large effect (t=3.31, df=49; d=1.06, p=0.002) was also found for the speed at which athletes managed to go from their first win of a premier race to their third. For females on the Road, however, the speed at which athletes went from their first elite UCI race podium to their first premier podium was small and not statistically significant (t=-0.83, df=17; d=0.40, p=0.41). The speed at which those athletes managed to get from one 'premier' win to three 'premier' wins also showed only a very small effect which was not statistically significant (t=-0.31, df=17; d=0.14, p=0.78).

Mountain Bike (n=6) saw an equal spilt of eminent athletes between sexes (n=3 for each) and yielded little insight between the effects of reaching milestones by age between both groups. Similarly, in Track cycling 29 athletes (male n=14, female n=15) reached 'eminent' status but no statistically significant effects were found in the rate at which milestones were reached. However, in Cyclo Cross (n=5), 3 males and 2 females reached 'eminent' status and the effect size for the time it took for males between their 'first premier podium' to their 'first win' was large (t=4.87, df=3; d=4.42, p=0.02).

Overall, results of speed at which eminent athletes met milestones versus non-eminent athletes where faster (i.e., at an earlier age for the first milestone (getting a podium finish in any UCI elite category race). Beyond that first milestone however, results were mixed. Eminent athletes in specific sports and genders, showed much faster progression (e.g., male Road, female Track and male Cyclo Cross) whilst eminent athletes in other groups showed no such signs (e.g., male Track, female Road and Mountain Bike in general).

Serial Winners vs Multiple Winners

The proportion of serial and multiple winners can be found in Table 5. Race results for athletes were compared across the 'elite', 'emerging eminent', and 'eminent' athletes (N=92,794). Amongst the 'elite' group (n= 92,602), 9% were serial winners and 2% had won events in multiple sports. Within the 'emerging eminent' athlete group (n=132), 65% were serial winners and 14% multiple winners whilst 'eminent athletes' (n=60) recorded 72% serial winners and 28% multiple winners. On a per sport and per sex basis, 'eminent' road females (n=7) recorded the highest percentage of multiple winners with 57%, followed by 'eminent' female track athletes (n=15) with 40%. Male Mountain Bike athletes had the highest percent of specialized athletes with 100%, whilst also having the smallest sample (n=3). Amongst the more elite cyclists, serial winners were more prevalent in the 'eminent' and 'emerging eminent' groups than multiple winners ('eminent': 72% serial, 28% multiple, 'emerging': 65% serial, 14% multiple).

Specialization versus Multisport

'Eminent' cyclists in general engaged in more cycling sports (83% multisport and 17% specialized) at the elite level (i.e., doing both Road and Track, or Mountain Bike and Cyclo Cross etc.) compared to the overall 'elite' population (52% multisport and 48% specialized) and the emerging eminent group (72% multisport, 28% specialized).

As well, odds ratios (Table 6) exploring whether the proportion of becoming 'eminent' in this data set was higher with different combinations of sports versus specialization for each sport broadly showed that combining specific sports was favorable. The outcomes of these analyses are discussed below for each sport.

Road

For Road, the proportion of becoming 'eminent' were found to improve significantly when adding elite competition in Mountain bike (males OR=25.6, CI =19.9-32.8, p=0.00 and females OR=13.4, CI =9.45-19.2, P=0.00), Cyclo-cross (males OR =38.9, CI +30.2-50.1, p=0.00 and females OR=15.2, CI=10.8-21.3, p=0.00) or Track (males OR=55.4, CI=44.5-69.0, p=0.00 and females OR=27.9, CI=20.9-37.2, p=0.00).

Track

For Track, adding Road was the only discipline that was found to increase the proportion of becoming 'eminent' (males OR=1.91, CI=1.69-2.15, p=0.00 and females OR=2.21, CI=1.89- 2.59, p=0.00).

Mountain Bike

For Mountain Bike all matches were found to be favorable over full specialization with Cyclo Cross being the strongest (male: OR=5.50, CI=4.88-6.21, p=0.00 and female: OR=4.21, CI=3.47-5.11, p=0.00). Track boasted (male: OR=2.94, CI=1.81-4.79, p=0.00 and female: OR=2.70, CI=1.51-4.82, p=0.008) and Road (male: OR=1.90, CI=1.63-2.21, p=0.00 and female: OR=1.69, CI=1.32-2.18, p=0.00).

Cyclo Cross

For Cyclo Cross, only two sports were favorable matches compared to specialization for men (Road: OR=4.58, CI=3.64-5.75, p=0.00 and Mountain Bike: OR=8.72, CI=7.10-10.7, p=0.00), whilst for women, all three matches improved the odds of becoming eminent in Cyclo Cross over specialization (Road: OR=6.75, CI=4.76-9.58, p=0.00, Mountain Bike: OR=14.8, CI=10.6-20.6, p=0.00 and Track: OR=4.56, CI=1.85-11.2, p=0.009).

Discussion

This investigation examined the developmental histories of the world's 'best' cyclists to explore differences between cyclists who achieved the highest levels of success (i.e., eminence) compared to cyclists who did not achieve the same performance standard (i.e., the emerging eminent and elite cyclists). Our analyses revealed several interesting results, which are summarized and discussed below.

Podium Stability

The podium stability analysis showed that 33% to 36% of podium spots in the most prestigious Road and Mountain Bike races were won by the same riders. When considering that professional teams are arguably driven by two general outcomes, namely value for their sponsors (e.g., through media exposure and association to inspirational performances or role models) or performance results (often measured in UCI points, race wins, or podiums), when making rider acquisitions administrators undoubtedly consider return on investment (for example, UCI points per Dollar invested). With relatively high podium stability, it would be beneficial for teams to prioritise the identification and acquisition of individuals who frequent podiums more regularly. These podium stability numbers further indicate that predicting future performance, although differing slightly from one cycling sport to the next, may be possible, at least to a point.

Performance Milestones

The difference in average age of achieving a first elite podium between eminent riders and the overall elite population suggests that early podium performances have some predictability for future performance. Large effects for this milestone were observed across both sexes whilst the speeds at which other milestones were reached varied. Interestingly, the average age for the second milestone (i.e., achieving a first podium result at a 'premier' race) saw a medium effect for males yet a very small effect for females. Even though this effect was not statistically significant for females, it might warrant further interrogation as to why the differences in effect are so large between these groups.

The rate at which athletes met milestones was strongest in men's Road and Cyclo Cross and female Track, suggesting that predictability of future performance is highest in those categories. However, it might also be that specific biases and contextual circumstances influence these effects. For example, in Road racing, athletes operate in a team environment with specific roles (e.g., compared to Mountain Bike where they operate largely independently). As such, it might be that those that win early get more opportunities to race for their individual result with the support of the team (e.g., as opposed to playing a support role for someone else), which ultimately ends up influencing their likelihood of success.

Serial Winners vs Multiple Winners

When examining the percentages of serial versus multiple winners at each population (i.e., 'elite', 'emerging eminent', and 'eminent') two general observations can be made. The first is that whilst the absolute percentage of both serial and multiple winners rose at every step up in performance/skill, the proportion of multiple winners grew at a greater rate than serial winners. The second observation is that whilst the relative percentage of multiple winners increased, the vast majority in all three populations (of those that have won at least two races in the UCI elite category) were serial winners.

Specialization versus Multisport

Overall, findings revealed that athletes who compete in multiple cycling disciplines as an elite athlete have increased odds of becoming 'eminent' (e.g., winning three or more important international competitions). Further to this, the percentage of those who engaged in multiple cycling sports increased from the 'elite' to the 'emerging eminent' groups, and from the 'emerging eminent' to the 'eminent' groups. These findings may have important implications for athlete development, as there continues to be a debate regarding the benefits of sport specialization (i.e., focussing on one sport competitively from an early age) and diversifying (i.e., multiple sport participation in a competitive manner). This corroborates other research in the field indicating that sport specialization is not necessarily beneficial for certain sports (examples include Alpine Skiing; Barth et al. [2018] and Speed Skating; Record of the Olympiad Games [2021]).

In the context of the present study, the notion of 'sport mixing' presents an interesting case for reaching elite levels. This could be for many reasons such as complementary development (i.e., building skills and abilities in one sport helps to supplement or complement another sport's needs). A finding such as this could also better support the health and longevity of athletes in sport as early sport specialization has been reported to relate to health issues such as physical and psychological burnout, stress-related musculo-skeletal injuries such as stress factors etc.). What is left to be determined is whether participating in sports in a single broad domain (all under one sport umbrella such as cycling) provides sufficient ‘diversification’ to circumvent these negative ramifications.

These findings may have important implications for researchers and practitioners working in competitive cycling, particularly those working in athlete identification, selection, and development. For instance, those looking to maximize performance of athletes on their professional road cycling team, may wish to examine the potential value of strategically supporting their athletes into Track or Mountain Bike races. As well, those working in athlete development could consider these results in the context of program development and training periodization. The specific combinations of sports to optimize such mixing at pre-elite stages, however, is unknown. Preliminary results from our present paper, however, illuminate that not restricting cycling athletes to a single sport. might improve their capabilities.

Limitations

While this study presents novel findings for the field of elite Road, Track, Cyclo Cross and Mountain Bike, there were multiple limitations that are important to acknowledge. First, this cohort includes those who competed between 13 and 3 years ago and, as a result, proportions might have changed as advancements in resourcing and training prescription progressed and it would not include past or future eminent performers. Second, some individuals, most notably those who competed in the last few years in the cohort, may still be selected in future elite world championships. Therefore, our results may not truly reflect the full effects of the sample’s competitive performance. Finally, the introduction of a minimum salary for athletes on women’s world tour (Road) teams in 2020 could change some of the relationships and outcomes described in this article, emphasizing the need for replication of these analyses in future samples.

Future Directions

These analyses highlight several areas for future work. For example, examining the next cohort of 10 years of sample data (e.g., from 2020-2030) may help to identify changes and trends (if any) in the areas of investigation included in this study (gender differences, sport differences, etc.). It may also be beneficial for researchers and practitioners to consider how commercial influences affect performance and development trends within these populations. For instance, athletes competing in multiple sports may appeal to varying populations of fans, which may increase the marketability of multiple winners. As world tour teams professionalize, commercial influences may accentuate these findings in future cohorts.

Conclusions

Elite sport is on a constant search for competitive advantages. Overall, findings from this analysis reveal that early performance at the elite level could be an indicator of future performance. Further, having a mixture of cycling sports as an ‘elite’ athlete increased the odds of becoming an ‘eminent’ athlete in a specific sport. That is to say, a cyclist might only win in one sport, but the results from this study showed that supplementing a primary sport with a secondary cycling sport is beneficial. These findings have the potential to catalyze broad changes in the approach coaches or administrators at national or professional cycling teams structure race schedules for athletes. This information, along with the work of others in the field of athlete development (constraint-based approaches to learning, for example; Davids, 2010), support the notion that striking a balance of mixing sports and mixing competition is beneficial for the athlete in multiple ways.

Figure 1: Pathways to eminence. The ages by which eminent endurance cyclists achieved their milestones

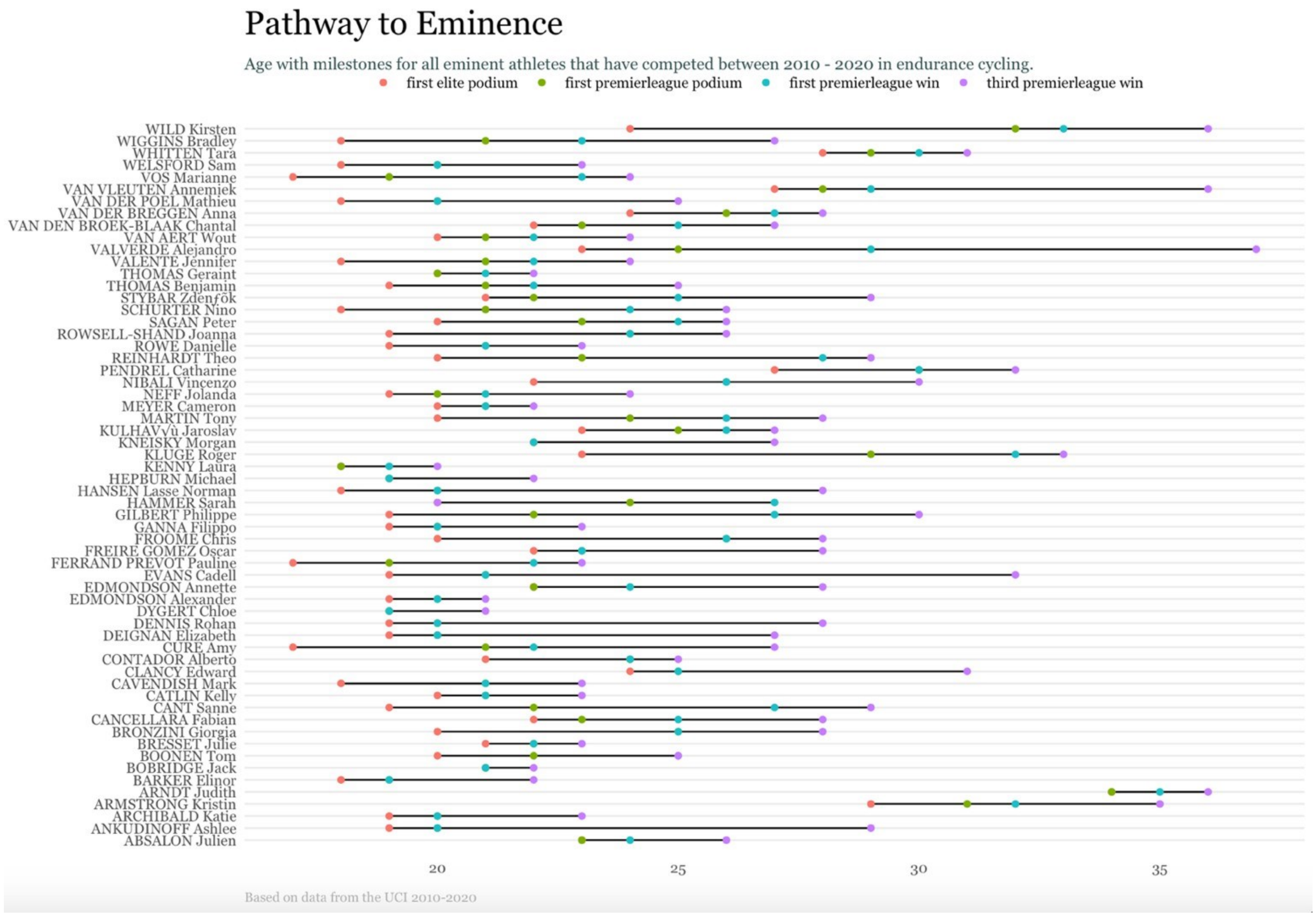


Table 4: Career milestones by age comparison between 'elite', 'emerging eminent' cyclists and 'eminent' cyclists in endurance sports

		First Elite Podium					First Premier Podium					First Premier Win					Third Premier Win
	Stage	Age (SD)	t	d	Effect Size	Sig	Age (SD)	t	d	Effect Size	Sig	Age (SD)	t	d	Effect Size	Sig	Age (SD)
All (n=92,794)	Elite (n=92602)	31.4 (±6.3)															
	Emerging (n=132)	20.2 (±2.5)					24.1 (±4.0)					25.5 (±4.6)					
	Eminent (n=60)	20.7 (±3.2)	13.15	2.14	large	0.001	22.7 (±3.4)	2.34	0.36	small	0.02	23.9 (±3.9)	2.34	0.38	small	0.02	26.9 (±3.5)
Sex																	
Male (n=74713)	Elite (n=74588)	31.5 (±6.2)															
	Emerging (n=92)	20.0 (±2.3)					24.2 (±4.3)					24.5 (±4.7)					
	Eminent (n=33)	20.2 (±1.8)	10.47	2.47	large	0.001	22.2 (±2.3)	2.54	0.50	med.	0.01	23.4 (±3.1)	1.25	0.48	small	0.02	26.7 (±3.6)
Female (n=18332)	Elite (n=18265)	31.1 (±6.9)															
	Emerging (n=40)	20.7 (±2.9)					24.0 (±3.4)					25.6 (±4.5)					
	Eminent (n=27)	21.3 (±4.3)	7.38	1.70	large	0.001	23.3 (±4.5)	0.72	0.16	very small	0.52	24.4 (±4.6)	1.06	0.26	small	0.31	27.2 (±4.7)
Road																	
Male (n=29966)	Elite (n=29915)	31.8 (±6.0)															
	Emerging (n=38)	20.9 (±2.3)					26.8 (±4.7)					28.7 (±4.3)					
	Eminent (n=13)	20.4 (±1.5)	6.85	2.61	large	0.001	23.1 (±1.9)	2.75	0.89	large	0.008	24.5 (±2.6)	3.31	1.06	large	0.002	28.3 (±3.5)
Female (n=7919)	Elite (n=7900)	31.3 (±6.4)															
	Emerging (n=12)	21.7 (±2.4)					25.1 (±3.6)					26.8 (±3.6)					
	Eminent (n=7)	25.0 (±5.4)	2.6	1.06	large	0.02	26.7 (±4.8)	-0.83	0.40	small	0.41	27.6 (±5.0)	-0.31	0.14	very small	0.78	31.0 (±4.4)
Mountain Bike																	
Male (n=37752)	Elite (n=37748)	31.9 (±7.2)															

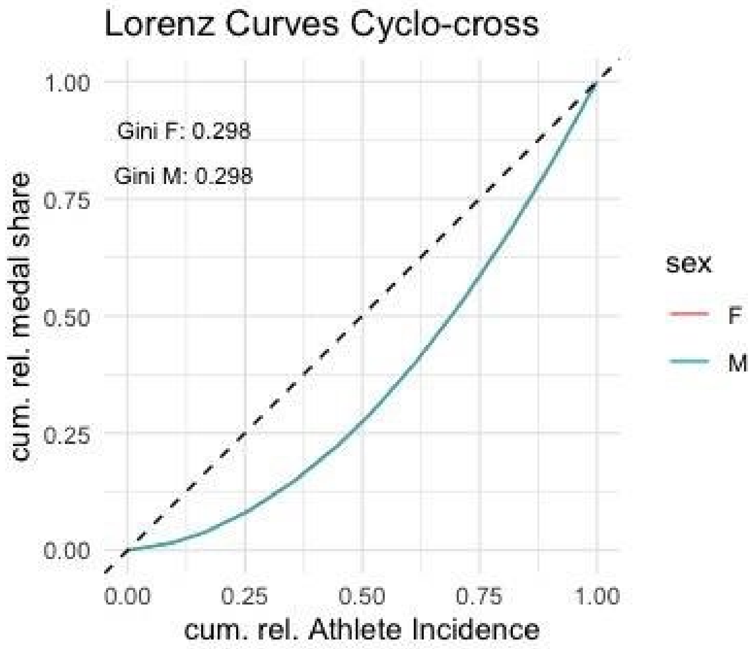
		Emerging (n=1)	22.0					26.0										
		Eminent (n=3)	21.3 (±2.9)	2.55	2.13	large	0.86	23.0 (±2.0)	NA	NA	NA	0.32	24.7 (±1.2)	NA	NA	NA	0.03	26.3 (±0.6)
	Female (n=7136)	Elite (n=7128)	32.6 (±6.4)															
		Emerging (n=5)	19.8 (±3.3)					22.8 (±2.2)					25.8 (±4.0)					
		Eminent (n=3)	22.3 (±4.1)	2.79	1.91	large	0.38	24.0 (±5.3)	-0.46	0.34	small	0.66	24.3 (±4.9)	0.48	0.34	small	0.66	26.3 (±4.9)
Track																		
	Male (n=10773)	Elite (n=10708)	29.4 (±6.4)															
		Emerging (n=51)	19.2 (±2.2)					22.1 (±2.6)					22.7 (±2.6)					
		Eminent (n=14)	20.0 (±1.9)	5.50	1.99	large	0.001	21.6 (±2.6)	0.64	0.20	Very small	0.52	22.4 (±2.6)	0.38	0.11	Very small	0.72	25.4 (±3.9)
	Female (n=4891)	Elite (n=4855)	29.2 (±6.4)															
		Emerging (n=21)	20.5 (±3.1)					23.9 (±3.4)					25.1 (±4.1)					
		Eminent (n=15)	19.8 (±2.9)	5.67	1.89	large	0.001	22.0 (±3.9)	1.36	0.51	med.	0.14	22.9 (±2.4)	1.86	0.53	med.	0.13	25.8 (±4.3)
Cyclo Cross																		
	Male (n=8369)	Elite (n=8364)	33.6 (±7.5)															
		Emerging (n=2)	22.0 (±0.0)					27.5 (±2.1)					31.5 (±7.8)					
		Eminent (n=3)	19.7 (±1.5)	3.21	2.57	large	0.004	21.0 (±1.0)	4.87	4.42	large	0.02	22.3 (±2.5)	4.86	1.86	large	0.14	26.0 (±2.6)
	Female (n=3533)	Elite (n=3529)	33.3 (±9.2)															
		Emerging (n=2)	18.0 (±0.0)					21.0 (±1.4)					22.5 (±0.7)					
		Eminent (n=2)	18.0 (±1.4)					20.5 (±2.1)					24.5 (±3.5)					26.0 (±4.2)

d Refers to Cohen’s d. Effect sizes of < 0.20 = Very Small, <0.50 = Small, <0.80 = Medium, all other scores = Large
Significant codes: 0***, 0.001**, 0.01*
med= medium effect size

Table 5: Serial winners and multiple winners per sport, sex and stage

		Stage	Career multisport / Specialization	Serial Winners / Multiple winners
	All (n=92,794)			
		Elite (n=92602)	52% / 48%	9% / 5%
		Emerging (n=132)	64% / 36%	65% / 14%
		Eminent (n=60)	83% / 17%	72% / 28%
Sex				
	Male	Elite (n=74588)	47% / 53%	8% / 4%
		Emerging (n=92)	63% / 37%	67% / 13%
		Eminent (n=33)	73% / 27%	76% / 24%
	Female	Elite (n=18265)	75% / 25%	12% / 5%
		Emerging (n=40)	68% / 32%	60% / 18%
		Eminent (n=27)	96% / 4%	59% / 41%
Road				
	Male	Elite (n=29915)	38% / 62%	13% / 1%
		Emerging (n=38)	24% / 76%	66% / 16%
		Eminent (n=13)	31% / 69%	77% / 23%
	Female	Elite (n=7900)	62% / 38%	8% / 2%
		Emerging (n=12)	42% / 58%	75% / 25%
		Eminent (n=7)	86% / 14%	43% / 57%
Mountain Bike				
	Male	Elite (n=37748)	30% / 70%	2% / 5%
		Emerging (n=1)	100% / 0%	100% / 0%
		Eminent (n=3)	100% / 0%	100% / 0%
	Female	Elite (n=7128)	46% / 54%	9% / 5%
		Emerging (n=5)	40% / 60%	60% / 20%
		Eminent (n=3)	100% / 0%	67% / 33%
Track				
	Male	Elite (n=10708)	57% / 43%	11%/ 7%
		Emerging (n=51)	92% / 8%	67% / 12%
		Eminent (n=14)	100% / 0%	79% / 21%
	Female	Elite (n=4855)	60% / 40%	15% / 6%
		Emerging (n=21)	86% / 14%	52% / 14%
		Eminent (n=15)	100% / 0%	60% / 40%
Cyclo Cross				
	Male	Elite (n=8364)	71% / 29%	2% / 5%
		Emerging (n=2)	100% / 0%	100% / 0%
		Eminent (n=3)	100% / 0%	33% / 67%
	Female	Elite (n=3529)	74% / 26%	5% / 5%
		Emerging (n=2)	100% / 0%	100% / 0%
		Eminent (n=2)	100% / 0%	50% / 50%

Figure 2: Lorenz curves medal concentrations in Cyclo Cross



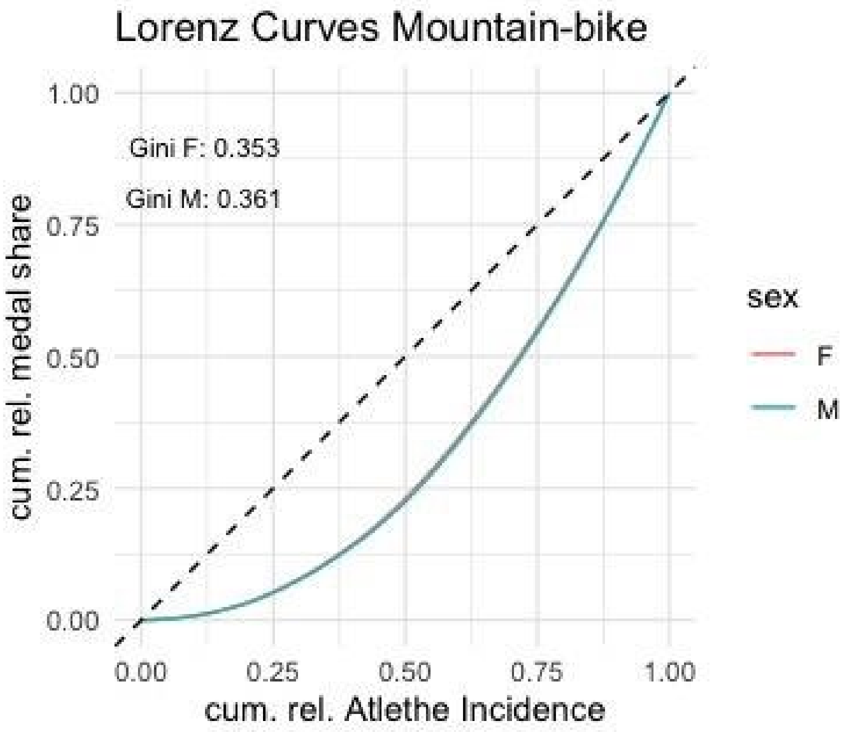
F= female

M= male

Cum.rel. = cumulative relative

Given the same Gini-coeficient, the male and female line in the Lorentz curve overlap. As a result, in modelling, only a single line has been displayed.

Figure 3: Lorenz curves medal concentrations in Mountain Bike

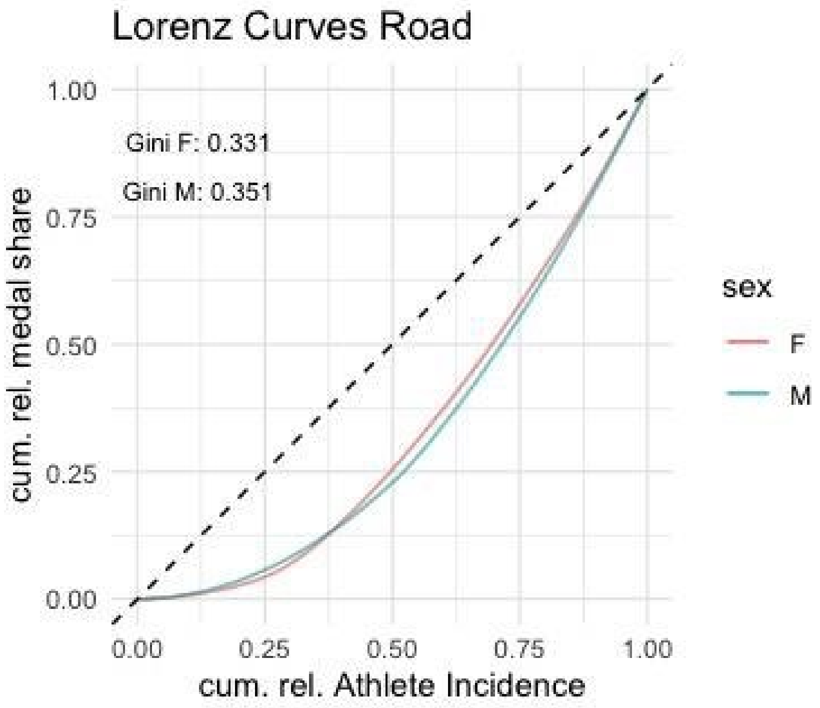


F= female

M= male

Cum.rel. = cumulative relative

Figure 4: Lorenz curves medal concentrations in Road

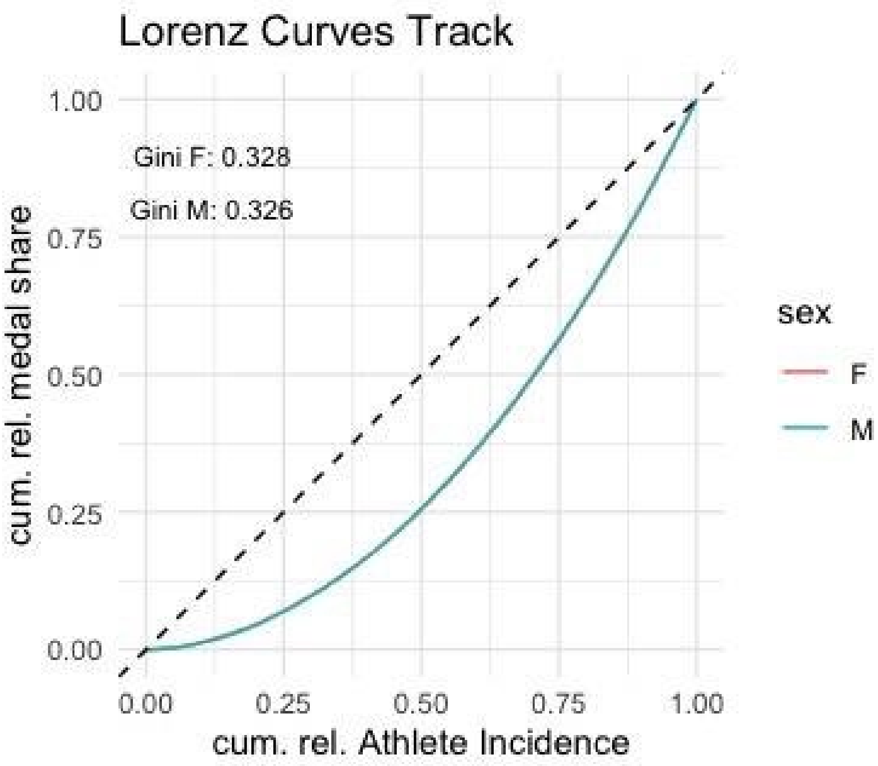


F= female

M= male

Cum.rel. = cumulative relative

Figure 5: Lorenz curves medal concentrations in Road



F= female

M= male

Cum.rel. = cumulative relative

Table 6: Odds ratios of achieving eminent cycling, comparing specific matches of endurance cycling sports to specialization in a single endurance cycling sport

		Odds of becoming 'eminent'		
		OR	**95% CI**	**P-value**
Male				
	Track vs Track + Road	1.91	(1.69-2.15)	0.0000
	Track vs Track + Mountain Bike	1.36	(0.83-2.23)	0.2133
	Track vs Track + Cyclo Cross	0.53	(0.19-1.46)	0.2207
	Road vs Road + Mountain Bike	25.6	(19.9-32.8)	0.0000
	Road vs Road + Cyclo Cross	38.9	(30.2-50.1)	0.0000
	Road vs Road + Track	55.4	(44.5-69.0)	0.0000
	Mountain Bike vs Mountain Bike + Road	1.90	(1.63-2.21)	0.0000
	Mountain Bike vs Mountain Bike + Track	2.94	(1.81-4.79)	0.0000
	Mountain Bike vs Mountain Bike + Cyclo Cross	5.50	(4.88-6.21)	0.0000
	Cyclo Cross vs Cyclo Cross + Road	4.58	(3.64-5.75)	0.0000
	Cyclo Cross vs Cyclo Cross + Mountain Bike	8.72	(7.10-10.7)	0.0000
	Cyclo Cross vs Cyclo Cross + Track	1.81	(0.64-5.04)	0.2563
Female				
	Track vs Track + Road	2.21	(1.89-2.59)	0.0000
	Track vs Track + Mountain Bike	1.70	(0.94-3.04)	0.0746
	Track vs Track + Cyclo Cross	0.81	(0.34-1.92)	0.6437
	Road vs Road + Mountain Bike	13.4	(9.45-19.2)	0.0000
	Road vs Road + Cyclo Cross	15.2	(10.8-21.3)	0.0000
	Road vs Road + Track	27.9	(20.9-37.2)	0.0000
	Mountain Bike vs Mountain Bike + Road	1.69	(1.32-2.18)	0.0000
	Mountain Bike vs Mountain Bike + Track	2.70	(1.51-4.82)	0.0008
	Mountain Bike vs Mountain Bike + Cyclo Cross	4.21	(3.47-5.11)	0.0000
	Cyclo Cross vs Cyclo Cross + Road	6.75	(4.76-9.58)	0.0000
	Cyclo Cross vs Cyclo Cross + Mountain Bike	14.8	(10.6-20.6)	0.0000
	Cyclo Cross vs Cyclo Cross + Track	4.56	(1.85-11.2)	0.0009

Chapter Four: Junior Performance as Predictors of Elite Performance

Jesse Korf, Kathryn Johnston, and Joseph Baker

NOTE: All references are provided after Chapter 5

Abstract

Several research studies have examined the relationship between early sport experiences and later sport performances. While the research interest has grown in many sport domains, much less has been done in the cycling disciplines of Road, Track, Cyclo Cross, and Mountain Bike. The following study investigated whether junior performance predicted elite World Championship performance in these disciplines using a retrospective examination of 10 years of data (from 2010 – 2020) held on the Union Cyclisme International (UCI) database. This included 4,760,647 athletes varying in age, country of representation, and competition level. Odds ratios were calculated for a) cycling sport and level distinguishing between top 3 (podium) performances at junior World Championships, top 8, top 25 or general participation without considerations for rankings, b) the number of World Championship medal winning athletes who did not compete at junior World Championships and c) progression rates on an inter- (in the same sport) and intra- (between different cycling sports) sport level for athletes at the U19 to U23 and elite levels. Findings revealed that performance at the junior World Championships has varying predictability on future performance.

Key Words: talent, cycling, junior performance, talent identification

Introduction

The term 'cycling' is unique in referring to a category of different sports, each with distinct characteristics and performance demands. Many of the sports under this categorization are further sub-divided by how performance is measured, such as timed events (e.g., Time Trials), bunch races (e.g., Road Races), game sports (e.g., Cycle Ball), and artistic events (e.g., BMX Freestyle). Cycling sports can also be categorized by their primary musculo-skeletal and energy system requirements, into such broad grouping as 'endurance' (e.g., Road cycling, Cyclo-Cross, and Track Endurance) or 'power' sports (e.g., Track Sprint, and BMX Race).

There is growing in interest in these cycling sports, as reflected in the body of research on physiological characteristics of elite athletes in cycling sports such as Road, Mountain Bike, Track, and BMX Race (e.g., Abbiss et al., 2013; Bejder et al., 2019; Côté & Vierimaa, 2014; Craig & Norton, 2001; Debraux & Bertucci, 2011; Gardner et al., 2005; Granier et al., 2018; Kabush & Orlick, 2001; Kordi et al., 2020; Louis et al., 2013; Martin et al., 2007; Mostaert et al., 2020; Mujika & Padilla, 2001; Padilla et al., 2000; Sánchez-Muñoz et al., 2018; van Erp, 2019). Evidence thus far has helped to create physiological profiles of athletes in their respective sports showcasing the types of 'requirements' for successful performance at elite levels of competition (van Erp, 2019). For example, Rylands and Roberts (2019) provided support for athletes in the sport of BMX Race by identifying specific physiological levels of torque, power, and cadence as well as key technical abilities, such as gate starting skill. While this has helped to lay an important foundation of evidence, the literature examining successful trajectories (i.e., following an athlete over multiple years from amateur to the professional or Olympic level) based on longitudinal or retrospective (i.e., over multiple years or decades) designs are slim. This is a notable issue as understanding the sequences of events, rates of progression and value of critical races/events in an athlete's development are foundational for training prescription and the development of evidence-based sport policies.

This relationship also directly impacts the nature of decisions regarding athlete selection (sometimes termed talent selection; TS), as professional teams in Road cycling, Cyclo Cross and Mountain Bike, as well as National Team programs in Track, Road and Mountain Bike, have increasing pressure to make accurate selections at younger and younger ages (for relevant work in sport generally, see Cooke et al., 2010; De Bosscher & De Rycke, 2017; Green, 2005). Limited resources and competition with other teams for talent means decisions based on talent evaluations, development, and training prescription need to be made as accurately as possible. These decisions often involve assessment of athletes in the Under 19 (junior), Under 23 and early Elite (i.e., 19 years and over for those sports that do not have a U23 category or 23 and over for those sports that do) categories. From this perspective, the relationship between performance as a junior competitor and success as an adult, elite cyclist becomes increasingly important.

Research suggests that one-third of those competing at the World Championships in the elite categories of multiple sports have been 'successful' junior athletes (Schumacher et al., 2006). In this sense, 'successful' is defined as reaching the elite level. Schumacher and colleagues concluded that "successful junior athletes have, in several cycling disciplines[2], significantly better results in the adult category" (p.1155). Shortly after, contradicting evidence was published by Moesch and colleagues (2011). In their work, Moesch et al. (2011) investigated Danish CGS sports (including cycling, but without specifying which specific cycling domains were being reporting), showcasing that elite athletes in their sample

[2] Findings were not specified for which cycling domains were under investigation.

intensified their training regimes at an older age than their near-elite peers. Moreover, the authors found that elite groups recorded fewer years at the junior national team stage and more years in the senior national team stage. They concluded that career planning with less training at early ages and specialization later seems more beneficial for young athletes in CGS sports (Moesch et al., 2011).

These contrasting findings are important as this field of inquiry has been generally underexplored. While these studies provide a starting point to examine the correlation between junior and elite performance, a more in-depth investigation of the different cycling sports and potential differences between sexes would add more specific data to this area. To this end, this study retrospectively investigated the Junior World Championships cycling performance for successful athletes in the elite Road, Track, Cyclo Cross and Mountain Bike disciplines. The authors hypothesized that many cyclists competing in endurance sports at the elite level have competed in junior world championships. Second, that it is common for athletes to compete in different, or multiple, cycling sports at the U19 level compared to when they reach the elites.

Methods

Participants

Race results listed on the Union Cyclisme International (UCI) database between 2010-2020 (N= 4,760,647) were explored using RStudio (version 1.4.1106). This allowed analysis of the entire population of elite, international cyclists who performed in this particular decade. Athletes who competed at U19 World Championships during this period were included (N=12,404) in various disciplines including Road (n=3,474), Track (n=6,381), Mountain Bike (n=1,830) and Cyclo Cross (n=719). Geographic regions included in this sample included Africa (n=54), Asia (n=45), Europe (n=51), The Americas (n=43), and

Oceania (n=8). Age was provided in the UCI data base, but only by year (i.e., 17 or 18 years old), which precluded deriving accurate age means and standard deviations (e.g., the data set indicated that 4,992 athletes competed in the U19 World Championships at age 17 and 7,412 at age 18).

Study Design

To test our hypothesis that cyclists competing in at the elite world championship level were more likely to have competed in junior world championships, we followed a multi-step approach. The first step involved identifying the athletes in the 10-year sample who participated in the U19 World Championships, and within that sample, determining whether placement provided any predictive value for future participation. More specifically, the sample was analysed using simple percentages to determine the proportion of the population that competed at junior World Championships at some point in the decade between 2010 and 2020 and ended up participating at the Elite World Championships. This same approach was repeated for each sport and sex. Next, these steps were repeated for those who finished in the top 3 at the Junior World Championships, the top 8 and the top 25 to examine whether level of performance as a junior held any predictive value beyond general participation. Finally, the Elite World Championship fields were examined to calculate the percentage of participants who never participated in the junior categories.

To test the second hypothesis (i.e., that athletes would be more likely to participate in different, or multiple, cycling sports at the U19 level compared to the elite level), percentages of 1. cross-participation (e.g., an athlete competing in the road cycling World Championships and also competed in the Mountain Bike World Championships in the same year), 2. linear progression (i.e., from junior to elite, or from junior through U23 to elite in the same sport), and 3. cross-progression (for example; an athlete competing in the junior Track World

Championships and elite Road World Championships) were calculated and displayed at both a total population overview and by sex.

In this sense, ‘cross-participation’ refers to athletes who participated in a particular cycling sport at the U19 World Championships, (e.g., Track), and a second cycling sport (e.g., Road) and the U19 World Championship level. Conversely, ‘cross-progression’ describes those who competed in one U19 World Championships, (e.g., example Cyclo Cross), and progressed to participate in an Elite World Championship in a different cycling sport (e.g., Mountain Bike). Finally, ‘linear progression’ describes those who competed exclusively in one cycling sport (for example, an athlete who competed in U19 Track and also competed in Elite Track).

Data Analyses

Data were extracted from the UCI website into R and basic descriptive statistics were used to determine the size of the cohort (N = all individual entries at the U19 UCI World Championships in Mountain bike, Track, Road or Cyclo-Cross between 2010 and 2020). Subsequently, the data set was filtered by sport and country levels to establish the sport- and region-specific demographics.

For each U19 entry, we examined the broader UCI data set to establish whether the cyclist participated in an Elite World Championships at some point after their U19 performance. The proportion of participation was calculated as P(A|B) where A equals participation in Elite World Championships and B, participation in U19 World Championships (U19). These steps were subsequently repeated with adjusted measures for B based on their placement (i.e., being top 3 in U19, top 8 and top 25). Finally, A and B were reversed to calculate the probability of an Elite World Championships participant competing at the U19s at some point in their career.

Finally, we examined cross-participation and cross-progression. These examinations were done using an upper triangular matrix listing all cycling sports and World Championships age categories (Cyclo Cross: U19, Elite; Mountain Bike: U19, U23 and Elite; Road: U19, U23 and Elite; Track: U19, Elite) with the frequency of occurrence for each match in the data set expressed in percentages.

Results

Proportiony of U19 World Championships athletes competing at the Elite World Championships

The proportion between U19 and elite participation (Table 7) was highest for those who finished in the top 3 at the U19 level (P:0.26 for Females [F] and P:0.22 for Men [M]). In the between sport comparisons, Mountain Bike had the highest probabilities to U19 success leading to elite participation as an adult (P:0.32 [F] and P: 0.40 [M]). As a trend, probabilities dropped in each sport as placements went down. Of those who won medals at the Elite World Championships, the probability that they did not compete in U19 World Championships was 0.34 for Females and 0.20 for Men.

Differences between males and females

Significant differences between males and females were found for all cycling sports under investigation (Table 7). Links between U19 performance and elite performances was highest for those who finished in the top 3 at the U19 level (26.1% for Females [F] and 21.7 for Men [M]). The proportion of top 3, top 8, and top 25 junior World Championship performers participating at future elite world championships across categories was 4 to 8 percentage points higher for females than it was for males.

Differences between sport disciplines

Differences between cycling sports was also observed (Tables 8 and 9). Cyclo Cross had the lowest proportions (0.05 for F and 0.06 for M) while Track cycling was highest (i.e., 0.22 for F and 0.13 for M). These results indicate those who competed in Track as a junior were two to four times more likely to participate at the elite World Championship than their Cyclo Cross peers. Similar to the trend noted above for the overall sample, the probability of junior performance leading to future elite participation dropped in each sport as placements went down.

Cross participation, linear and cross progression in cycling disciplines

Cross-participation (Tables 8 and 9) was highest in the U19 categories, particularly between Track and Road with 20% competing in both events (35% for F and 17% for M). The linear progression rate from junior to elite on a per sport level was highest in Track with 35% (38% for F and 34% for M) and lowest in Mountain Bike with 11% (16% for F and 10% for M). In terms of cross-progression, Mountain Bike (as a junior) to Cyclo-Cross (as an elite) was the most frequent with 11% of riders reporting this pattern of participation (14% for F and 9% for M).

Discussion

World junior level competition is not a mandatory step for elite success

Perhaps the most important finding from this research is that a linear pathway from junior levels to elite World Championship performance, whilst not the only pathway, is common in Road, Track, Cyclo Cross or Mountain Bike sports. More specifically, based on the current sample of participants from 2010-2020, of those that competed at the Elite World Championships, the probability that they competed in U19 World Championships was 0.66 for Females and 0.80 for Men.

This finding, that certain early performance milestones are indicative for later elite performance, runs somewhat counter to work done by others in the field (e.g., Barreiros et al., 2014). The work by Barreiros and colleagues (2014), for example, explored the competitive sport pathways of 395 athletes spanning the sports of soccer, volleyball, swimming, and judo, noting that only a third of international pre-junior athletes reappeared as senior athletes. In the present study's sample, only one third of females and a quarter of males who competed at elite World Championships did not compete at the U19 junior level in these respective sports.

This relationship likely reflects the dynamic nature of athlete development. Specifically, there are many dynamic variables at both the athlete and system levels, which ebb and flow throughout the developmental pathway. For example, athletes at the U19 level are not fully developed performers, but will continue to develop from physiological, social, cognitive, and emotional perspectives. Moreover, the system surrounding the athlete will change, as reflected in alterations to rules and equipment, as well as changes in the complicated environmental interactions that shape an athlete's development (e.g., coaches, training environment, support etc.). Finally, the influence of luck and random chance is rarely integrated into models of athlete development, but their impact cannot be easily dismissed (see Bailey, 2007; Pluchino et al., 2018).

Interestingly, the proportion that athletes who reached elite World Championship levels had competed in U19 World Championships was lower for females than it was for males. This finding does corroborate with other research in the field examining sex/gender differences in milestones and developmental histories (see Peters et al., 2022 for more detail). This result may be connected to the disproportionate resources and research that exists for males compared to females in sport (Baker et al., 2020). In many of the cycling sports examined in the current analysis, there has been a longer history of male participation, which has led to the faster accumulation of resources and funding available for recruiting, selecting, and developing male athletes. This, in turn, provides a stronger foundation for developmental pathways for male athletes, without similar funding and resource opportunities available to female athletes.

Similarly, socio-cultural differences, as reflected, for example in the available positions for males and females on professional teams in these sports may differentially affect the trajectory for women in cycling sports. For example, a male athlete who competes at an early age in the sport of Mountain Bike may recognize there are significant and substantial earning opportunities for becoming a professional athlete and may choose to stay in the sport. Conversely, a female athlete may recognize that opportunities for satisfactory earning potential are more limited, leading to different choices for their sport and career, which ultimately affects the rates of progression from junior to elite success.

As seen with the growing trend towards equitable earnings in competitions in sports such as golf and tennis, hopefully this gap in opportunities will close and that effective and inclusive opportunities will continue to grow to support female cyclists and those who identify as women in sport. It will be important for future work to explore these questions, perhaps in the form of mixed-methods or qualitative explorations, as these types of research questions are important for creating more equitable, inclusive, and supportive sport environments for all participants.

Sport-specific differences

Another notable finding was that the pathway for cyclists who reached the World Championships varied by sport/discipline – with the largest effects found in Mountain Bike and the smallest effects in Cyclo Cross. We suspect this finding appeared in the data for various reasons, one of which is the peak age of the athletes in each of the sports (i.e., the average age that athletes tend to reach their highest level of performance), which varies significantly between different types of sports and is typically connected to the performance requirements for the sport. For example, the peak age for female gymnasts is 16 whereas the peak age for female distance runners is 37 (Allan et al., 2015). The performance demands needed in gymnastics requires high degrees of flexibility, power and strength, all of which require exposure to deliberate training from a young age. Conversely, the metabolic system for endurance sports continue to improve later in life, and thus peak age is later. At this point in time, research on the peak age for Road, Track, Cyclo Cross and Mountain Bike is limited, but preliminary searches for peak performance age for male elite road cyclists indicate some sport specificity (Kholkine et al. 2023). Ultimately, however, in sports with an older peak age there is more time between U19 competition and later World Championship performance, which makes the pathway longer and more prone to dynamic influences (Den Hartigh et al., 2018).

In addition, there are likely important contextual and environmental factors between cycling sports that influence these differences. Cyclo Cross, for example, has lower participation numbers, and very few 'full-time professionals'. In addition, it sees very little investment from National Sporting Federations, as their investment is often linked to Olympic outcomes (De Bosscher et al. 2019), which are not relevant to Cyclo Cross, at least currently.

Closely connected to this is the degree of 'cross sport' mixing in this data. Top performing riders in Cyclo Cross might get picked up by professional Road or Mountain Bike teams (examples include some prolific winners on the road such as Wout van Aert, Mathieu van der Poel and Marianne Vos; www.procyclingstats.com). When looking at the overall data rather than these few exemplars, we note that 9% of those who participated at the U19 Track World Championships and 4% of Cyclo Cross U19s, participated at the Elite Road World Championships.

Limitations

While this study presents several novel findings for the fields of elite Road, Track, Cyclo Cross and Mountain Bike, there are multiple limitations that are important to acknowledge. The first is that this cohort includes those who competed between 13 and 3 years ago and hence, probabilities might have changed as advancements in resourcing and training prescription have occurred. In a similar sense, some who competed in the last few years and are captured in this cohort may still be selected in future elite world championships, and thus their full trajectory may not be captured in this study. From an environmental perspective, the introduction of a minimum salary for athletes on the women's' world tour (Road) teams (which occurred in 2020) may influence the decisions and pathways. We expect this change to improve both drop-out and retention rates of females in the professional Road space.

Future Directions

With these study findings and limitations in mind, it would be both helpful and interesting for future work to build on the current study's design and capture the next cohort of 10 years of sample data (e.g., from 2020-2030). Investigating this sample may help to identify the changes and trends (if any) in the aforementioned areas of investigation (sex differences, sport differences, etc.). As this work helps to understand a portion of the pathway (as we only investigated U19, U23 and elite results) in these cycling sports, it could be valuable to repeat this study in another decade. A valuable addition might also be to look at U15, U17 and other non-elite categories in order to more strongly identify the various pathways that might exist for talented cyclists. This work could mimic that of McCue et al., (2019), which examined the various entry points and pathways for elite baseball players to reach the professional level within Major League baseball, ultimately noting 17 distinct pathways.

Conclusions

Overall, this retrospective study reveals that performance at junior World Championships has varying predictability on future performance. For those who competed at the junior age, ranking does seem to have relevance for understanding future success. However, a large proportion of participants at the elite World Championships did not compete at that level as a junior. These findings emphasize that a single, linear pathway to elite cycling does not exist. This emphasizes the difficulties in predicting later success based on early identification, and raising caution when selecting athletes into sport specific pathways too early. This information could be important for both professional and national team programs when evaluating talent as it may help inform selection policies and talent identification models.

Table 7: Probabilities for Junior-to-elite world championship performances. An overview of probabilities of elite world championship participation by sport, gender and placement.

		All	Cyclo Cross	Mountain Bike	Road	Track
Top 3 at Junior Worlds						
	Female	0.261	0.132	0.319	0.297	0.335
	Male	0.217	0.285	0.401	0.201	0.234
Top 8 at Junior Worlds						
	Female	0.188	0.084	0.248	0.220	0.238
	Male	0.139	0.167	0.274	0.130	0.158
Top 25 at Junior Worlds						
	Female	0.161	0.047	0.224	0.171	0.218
	Male	0.098	0.090	0.178	0.088	0.135
Any spot but competed at Junior Worlds						
	Female	0.149	0.046	0.213	0.147	0.217
	Male	0.070	0.063	0.146	0.054	0.133
Not competing in Junior Worlds						
	Female	0.344	0.079	0.485	0.319	0.419
	Male	0.199	0.182	0.399	0.158	0.281

Table 8: Progression rates and mixing of athletes in U19, U23 and elite spaces in endurance-based cycling sports for both sexes combined.

	Cyclo Cross-Elite	Cyclo Cross-Junior	MountainBike-Elite	Mountai Bike-U23	Mountain Bike-Junior	Road-Elite	Road-U23	Road-Junior	Track-Elite	Track-Junior
Cyclo Cross-Elite		19%	9%	9%	11%	10%	10%	8%	4%	3%
Cyclo Cross-Junior			2%	7%	13%	4%	9%	11%	2%	3%
Mountain Bike-Elite				9%	11%	10%	2%	3%	2%	1%
Mountain Bike-U23					18%	2%	3%	1%	1%	1%
Mountain Bike-Junior						3%	4%	6%	2%	4%
Road-Elite							23%	22%	17%	9%
Road-U23								24%	17%	18%
Road-Junior									14%	20%
Track-Elite										35%

Table 9: Progression rates and mixing of athletes in U19, U23 and elite spaces in endurance-based cycling sports for females.

	Cyclo Cross-Elite	Cyclo Cross-Junior	Mountain Bike-Elite	Mountai Bike-u23	Mountain Bike-Junior	Road-Elite	Road-u23	Road-Junior	Track-Elite	Track-Junior
Cyclo Cross-Elite		6%	15%	7%	14%	15%	4%	15%	5%	4%
Cyclo Cross-Junior			0%	1%	3%	0%	1%	2%	0%	1%
MountainBike-Elite				12%	16%	13%	1%	4%	4%	2%
MountainBike-U23					21%	3%	4%	3%	1%	1%
Mountain-bike-Junior						4%	2%	12%	3%	4%
Road-Elite							8%	19%	27%	13%
Road-U23								11%	7%	8%
Road-Junior									21%	35%
Track-Elite										38%

Table 10: Progression rates and mixing of athletes in U19, U23 and elite spaces in endurance-based cycling sports for males.

	Cyclo Cross-Elite	Cyclo Cross-Junior	Mountain Bike-Elite	Mountain Bike-U23	Mountain Bike-Junior	Road-Elite	Road-U23	Road-Junior	Track-Elite	Track-Junior
Cyclo Cross-Elite		24%	8%	10%	9%	8%	11%	6%	3%	2%
Cyclo Cross-Junior			3%	7%	15%	4%	10%	12%	2%	4%
Mountain Bike-Elite				9%	10%	9%	2%	3%	1%	1%
Mountain Bike-U23					18%	2%	3%	1%	1%	1%
Mountain Bike-Junior						2%	4%	5%	2%	3%
Road-Elite							27%	22%	15%	8%
Road-U23								26%	20%	16%
Road-Junior									11%	17%
Track-Elite										34%

Chapter Five: General Discussion

NOTE: All references are provided after Chapter 5

General Discussion

There has been a longstanding appreciation for exceptional performers. Ancient Olympians were celebrated in the Greek city states (Swattling, 1999). The Romans held games and engaged in theatre to the joy of its citizens (Schädler & Bagnal, 2013). Jousting games drew crowds during medieval times (Bennett et al., 2018). In more modern times, a large population of the world is captivated by the Olympic Games, whilst professional sport leagues and athletes are watched by hundreds of millions of people, generating billions of dollars (Gratton et al., 2012). For good or bad, many sport stars become idols and role models to children and their performances generate memories that can last a lifetime. Examples such as these demonstrate the socio-cultural value of 'eminent' athletes across the globe, and learning more about eminent performers may not only demystify the pathway to becoming one, it may also help to translate findings into practice for both sport practitioners and the broader public (e.g., when academic findings on expert performers become the basis for best-selling books such as Thinking Fast and Slow (Kahneman, 2011), The Talent Code (Coyle, 2009), and Outliers (Gladwell, 2008)).

In an attempt to add to our knowledge of eminence and demystify eminent performance in a sport specific setting, this dissertation explored two main objectives. The first was to examine whether a consensus exists within the research landscape regarding the criteria for 'eminence' in endurance cycling sports and the second was to explore the developmental histories of cyclists who met established criteria for 'eminence' with available race result data. In Chapter 1, a historical investigation was conducted on 'eminence', including definitional, conceptual and practical components. Chapter 2 subsequently built on this foundation by establishing evidence-informed criteria for 'eminent' endurance cyclists using the knowledge and experiences of a panel experts including coaches, technical directors, performance directors and athletes from varying competitive levels. Findings informed the research described in Chapter 3, which explored the characteristics of endurance cyclists who met the criteria for eminence from Chapter 2. Finally, in Chapter 4, junior-level race results were used to predict elite, adult performance, and by extension, the potential of reaching 'eminent' status.

To the author's knowledge, this research project is the first cycling-specific inquiry into eminence, and one of very few studies looking at the development of elite cyclists more generally. Using the present research as a basis for defining eminence in endurance cycling, the hope is that future research will use this as a stepping-stone to build the research area from multiple perspectives (competitive levels, countries, ages, etc.). The following sections expand on some of the key findings of the preceding chapters, position them in the context of the evolving field, and discuss implications for athletes, coaches, administrators, and researchers.

Blurry terms and their impact on our knowledge of the development of elite performers

Over the past decade, it has become evident that the terminology in talent research as it pertains to the 'top performers' is limited, and that this has important implications for how this research is applied (Chapter 1). For example, many words have been used to describe 'eminent' performers. Terms such as 'super champions', 'super elites', 'eminent athletes', 'multiple medallists' are just some of the terms that appear in sport science literature. Yet, none of these entities are the result of agreed upon sport-specific criteria for defining group membership (Chapter 2).

The implications of the use of blurry terms are further highlighted in Chapter 3 where the criteria for specialization in cycling was not clear. More specifically, depending on whether engagement in multiple different forms of cycling (e.g., Road and Mountain Bike) is seen as specialization (i.e., as being within the same 'sport') or as multi-sport (e.g., both are listed as different sports by relevant sport governing bodies (e.g., IOC, and UCI) has important consequences for how research results (including those in Chapter 4) are interpreted. This fuzzy terminology in the context of elite sport (as noted by Johnston et al., 2023) is problematic as specificity and rigor in measurement are critical for improvement.

Part of the reason why terms such as eminence may be so blurry in sport is because target populations for Olympic medalists, “super-champions” or “super elites” are, by definition, small. This challenge is further exacerbated when those groups are sub-divided into individual sports, making it hard to acquire enough data for most inferential statistical analyses. That said, there is a vested interest from professional and national teams alike to select and develop athletes in order to achieve the performance goals that are set by the teams or critical stakeholders (Gowthorp et al., 2017, Van Reeth et al., 2022).

Baker et al (2020) noted that certain groups, like females, and athletes in sports like cycling, sailing, shooting or from countries like Africa, Asia and South America are underrepresented in much of the talent research. What this means is that there are significant gaps in our knowledge that directly impact how athlete development models are delivered. More traditional or widely investigated models for male athletes in sports such as soccer or rugby in England and Australia may have limited utility in settings where they are not representative. For practitioners, this could result in approaches that are less effective, or even adverse, for a group of athletes or a sports program. For researchers, it might mean that hypotheses generated from research in one context (e.g., English male soccer) may not be appropriate in other contexts.

To counter these potential effects, more sport-specific and female-specific studies are needed, as are analyses of countries that are underrepresented in the literature. To this end, Chapters 3 and 4 explored differences between males and females in the pathways to eminence, with particular focus on rates of achievement and progression from junior to elite World Championship participation. These findings underscore the need for caution of generalizing findings in different talent development contexts.

On the odds of becoming an eminent endurance cyclist

Findings from the present dissertation revealed that the proportion, and therefore potentially future odds, of becoming an eminent endurance cyclist were higher when combining cycling sports (e.g., Road *and* Mountain Bike) than specialising in a single cycling sport (i.e., road or mountain bike alone). Moesch et al (2011) concluded that late specialization (e.g., specializing in one single sport) is required for elite success in centimeter, seconds, and grams sports. In cycling, most endurance sports would fall under this definition, yet the evidence in our analyses differed from other work in this area. We speculate that this is the case because either the definition of specialization does not capture cycling sports well or because opportunities to develop or transfer specific skills between cycling sports are higher than in other sports.

In Chapter 3, the research helped to understand longitudinal race day development and career milestones by sport and sex. Importantly, analysis of training and event histories revealed that the volume of events in Road Cycling and Track Cycling were higher than in Mountain Bike or Cyclo Cross. This could impact the relationship between specializing versus diversifying in obtaining eminence (Chapter 2). This potential effect, however, cannot fully explain the differences in odds ratios for specializing versus diversifying cycling sports in this population.

Other potential explanations as to why multiple cycling sports may be more beneficial could relate to skill acquisition and motor learning. For example, the constraints-based approach to motor learning (Renshaw et al., 2010) proposes that motor skills can be developed through constraining aspects from three groups of factors. The theory speaks to opportunities of constraining the *performer* (i.e., limiting the athlete in executing a motor task. For example, by asking them to do something with one arm instead of two), the *task* (i.e., making the drill, assignment, or game itself more or less challenging) or the *environment* (for example, changing the size of the field of play or the type of surface). When combining endurance cycling sports, one could argue that having athletes alternate between bikes with different gearing (e.g., a road bike with multiple gears and a track bike with a fixed gear) and execute similar bike handling skills on varying surfaces (such as tarmac, dirt, a smooth track) may manipulate key task and performer constraints that consequently lead to improved skill. Future work is needed to test the validity of this speculation.

Overuse injuries (Aicale et al., 2018) and burn-out (Goodger et al., 2007) are other reasons that could explain the benefits from engaging in multiple cycling sports. More specifically, the ability to sustain high volumes of training has been identified as critical in the development of elite endurance athletes. For endurance cyclists, it is not uncommon to accumulate over a thousand hours of training in a year (Seiler & Tønnessen, 2009). It has also been well documented that high volumes of work in a monotonous fashion or environment increase the risks of both overuse injuries and burn-out (Clarson et al., 2013; Gustafsson et al., 2008). As a result, strategically combining cycling sports throughout the year for an elite rider may decrease monotony and reduce the associated risks, whilst still promoting the required physiological and psychological adaptations. That said, these hypotheses are largely speculative but present interesting avenues for future exploration.

Junior performance as a predictor of elite performance

During the design of this research program, the original intent for the exploration of junior performance (Chapter 4) was to explore any indicators of 'eminence' at this age. Whilst the dataset allowed for the identification of athletes who met the criteria for eminence, and an exploration of their developmental histories in the elite categories, the dataset was limited in scope. It did not extend back far enough to explore the developmental histories of eminent riders to allow for a complete analyses of developmental factors on adult success. Whilst this was disappointing, an alternative research question explored whether performance at the Junior World Championships predicted participation and performance at Elite World Championships (i.e., the only event that was part of the nomination criteria for eminence in all examined sports).

As the number of professional and development teams (and their budgets) increased, there became an increasingly competitive battle for 'talent' (Van Reeth et al., 2022). When looking at the probabilities of U19 to elite World Championship representation, two observations stand out – the first was that those finishing higher in the rankings had a higher probably of future elite representation (perhaps a more intuitive finding). The second was that in Track and Mountain Bike, relatively high percentages (28% to 49%) of the field had never competed at junior World Championships (this is especially the case for females with an average of 34%). This finding was particularly interesting (and perhaps less intuitive) for Track as national teams run Track programming as opposed to professional teams (in many contexts). These organizations invest hundreds of thousands of dollars per year in resourcing junior and academy-like programs (e.g., the Academy in Great Britain, NextGen in Canada, Podium Potential categorisation in Australia and so forth), which focus on progressing junior athletes to senior levels (as per published financial reports of National Sporting Organisations and their national/Olympic funding agencies). Potential selection biases in these Track programs might be an area for future inquiry and consideration for administrators. Another intriguing area of future work concerns why the percentages of those who compete at elite World Championships without a junior World Championship experience were higher for females than for their male peers. As noted above, male-specific talent development research vastly outnumbers female-specific work, suggesting we know considerably less about the development of female athletes. The differences between these groups in our analyses may also reflect differences in income potential (which has been lower for females than males in professional team settings) amongst other social-cultural expectations (i.e., expectations of seeking employment, getting a university degree or starting a family). All this to say, the findings of the current dissertation echo recent calls for more research on female-specific samples.

Practical Implications

What does this work mean for athletes and coaches?

Reaching the required level of aptitude to compete in the elite categories in UCI sanctioned races is impressive, a feat that only a small proportion of competitive cyclists ever achieve. When considering the notion of 'eminence', the percentage of people achieving this level of accomplishment is obviously considerably lower. Between the different cycling sports that have been explored, only the top 0.04% to 0.4% (Chapter 3) of cyclists who competed in the Elite categories between 2010 and 2020 met the inclusion criteria.

Whilst only 52 athletes in our data sample met these criteria, there are things that coaches and athletes aspiring to reach this level can take away from their developmental histories. For example, athletes and coaches should consider supplementing their primary cycling sport with a second one. For instance, Mountain Bike athletes might consider supplementing their training and racing regime with Cyclo Cross. For Road, the strongest recommendation appears to be Track (at least based on the data from Chapter 3). The opposite recommendations (i.e., supplementing Cyclo Cross with Mountain Bike and Track with Road) also hold up. These effects increase the odds of achieving eminent status (Chapter

3), suggesting they would be beneficial for any elite cyclist. Whilst not investigated, it would be reasonable to assume that these effects also apply at lower (i.e., younger age) categories.

Junior performance seems to be a relatively strong predictor of elite performance in most of the cycling sports examined. That is to say, good results at the junior World Championships seem to predict future representation at elite World Championships. In that sense, coaches and athletes can track performance progression and consider how viable their pathways might be towards, or at, the under 19 categories. That said, however, a meaningful number of elite athletes competing at the world championships never represented their country as a junior. This might be because they were not in the sport yet, or their performance at that stage did not merit selection. This phenomenon was particularly strong in females (Chapter 4). When looking at Road, Track and Mountain Bike, the percentage of athletes at Elite World Championships who did not compete at that level as a junior was 32%, 42% and 49% respectively. This suggests that junior performance for females at this level was not a requirement for future elite performance. This result could be used to encourage later developers (i.e., those that mature at a relatively later age than their peer group) and late entrants (i.e., that come into the sport at a later age) to continue training and competing since future success remains a possibility. The same could be said for males in these sports; however, the correlation between junior performance and national representation with future World Championship representation (in Mountain Bike 40% of Elite males never competed at the U19 World Championships whilst in Track and Road, these percentages were found to be 28% and 16% respectively) was higher than found in their female peers. Conversely, coaches and athletes should be aware that even amongst those who medalled at junior World Championships, most never ended up representing their country at the elite stage. Put more simply, the significance of junior performance should not be overestimated by either coaches or athletes in these endurance-based cycling sports.

What does it mean for sport administrators?

When considering that professional teams are arguably driven by two general outcomes, namely value for their sponsors (e.g., through media exposure and association to inspirational performances or role models) or performance results (often measured in UCI points, race wins, or podiums), administrators consistently consider return on investment when making rider acquisitions (for example, UCI points per Dollar invested). From this perspective, it would be reasonable to prioritise 'stars' of the sport in your acquisition strategy, potentially even against high salaries. This was demonstrated in Chapter 3 which showed that those events that fetch the highest amount of UCI points and attention (e.g., the Tour de France, World Championships etcetera) are often won by the same small sub-set of athletes. A potential downside of this strategy, however, might be that it could deplete the financial capacity required to fill other spots on the team.

Administrators at professional teams (e.g., technical directors and talent scouts) might also consider other findings of this research program when making rider acquisition decisions. For example, when looking at a rider acquisition strategy in women's Road, one might want to consider balancing selection of younger talent with keeping an eye on the emergence of older talent. Analyses of pathways to eminence (Chapter 3) and progression rates in national representation from junior to elite World Championships (Chapter 4) suggest both pathways are strongly represented at the top of the sport and, therefore, should be considered in an optimised rider acquisition strategy.

The progression rates from Track and Mountain Bike junior performance to elite Road performance as well as the odds of becoming eminent when combining certain sports (Chapter 3) may also be relevant for administrators. These findings suggest it would be beneficial to a) deliberately structure opportunities for athletes to cross-over between certain sports (rather than just focussing on racing in one sport), and b) for Road teams and national team Track programs to collaborate for mutual gain. The latter proposal reflects the reality that national teams often oversee Track programs and the odds of achieving eminent status on the Road seem to be highest when deliberately combining Road racing with Track racing (followed by Mountain Bike and Cyclo Cross). Conversely, national teams on the Track might benefit by this as the burden to develop athletes is shared and could enhance performance of their best athletes at World Championship and Olympic races.

What does it mean for sport researchers?

To establish the criteria for eminence, we used two independent groups in a Delphi-based study. Whilst there is some foundation in the literature pertaining to the use of Delphi approaches, guidelines around this process are not clear. In Chapter 2, the criteria for eminence were established by an athlete group and a technical expert group (e.g., coaches, technical directors, and performance directors). When considering the establishment of criteria using a Delphi approach in other sports/contexts, future research might consider different stakeholders to be a part of this type of design. For instance, arguments could be made for either the inclusion of athletes exclusively, coaches/technical experts exclusively or perhaps even other stakeholder groups, such as researchers or fans. When considering these stakeholder groups, one should reflect on the amount of expert knowledge each participant or group could integrate into the conversation. After all, the foundational idea around a Delphi study, is to engage a diversity of leading experts on a particular subject (Keeney et al., 2006). As such, less expert voices (such as fans for example) might be chosen to be deliberately left out – even though those lines are blurry for what constitutes an 'expert' in the first place.

It was felt on reflection that the design of the Delphi was robust. In particular, the inclusion of an athlete group and a technical director group in a blinded fashion was felt to be a strength of the study. That said, there were ideas to improve the design even further that future researchers might wish to consider. For example, to reduce the potential for bias even further and to maximize inclusion, researchers may choose a “mixed method” approach that includes both a Delphi study (e.g., introducing a blinding effect into the Delphi study design) and a quantitative analysis to help validate suggestions from the panel. Such examples include leveraging the wisdom of a crowd (Welinder et al., 2010) and/or a quantitative method using a class of Bayesian statistical analysis (Bayes & Hume, 1763).

In terms of defining participation criteria, Baker et al (2017) called for an updated taxonomy for researchers in skill acquisition and expertise. Part of the reason for this call was that the ‘expert’ or ‘elite’ categories were considered to be too broad. An example that has been referenced by Baker and colleagues can be illustrated in the sport of swimming. The criteria for subjects to be considered ‘elite’ might see a 16-year-old swimmer who competes for state titles all the up to all-time greats such as Michael Phelps. To assume that physical capabilities, skill levels and psychological aptitudes of those two are similar is problematic. As a result, the capacity to generalize research findings is compromised because of the broad spectrum of performance within these categories. It was for this reason that the researchers argued a further differentiation was required.

One of the challenges in investigating eminent athletes is the lack of clear inclusion criteria (Chapter 1). To establish criteria, one could argue that most valuable player awards, hall of fame induction or Olympic medals are good criteria (as Baker et al., 2017) argued in their original paper). This, however, may not be applicable to many sports and cultures beyond North America. For example, many European sport leagues do not award MVP’s, many sports do not have halls of fame and not all sports are Olympic. Even amongst Olympic sports, there are those for which the Games are the highest level of competition (e.g., athletics) and those for which it is not (e.g., tennis and soccer). Hence, it might be more appropriate to consider inclusion criteria on a sport-by-sport level.

Interestingly, our results indicated that inclusion criteria might not be the same even within sports that have multiple disciplines under their umbrella. Both Delphi groups found that it was be inappropriate to consider the same races across cycling sports (Chapter 2). For Track for example, the overall world cup standings were excluded from the criteria whilst it was included for Mountain Bike. This suggests a need for future research to establish appropriate, sport specific criteria before starting their investigations.

The need for clear criteria and definitions was also highlighted in the current work when examining ‘specialization’. The definition of sports specialization reads as: “intense, year-round engagement in a single sport with the exclusion of other sports” (Ericsson, 2020. p.14). Endurance cycling technically consists of four different sports (Road, Track, Mountain Bike and Cyclo Cross). In popular terms, however, they are all done on a bike and hence could be considered one sport. From this perspective, any sport on a bike is ‘cycling’ and therefore those who engage in both Mountain Bike and Road cycling are still specializing. Conversely, amongst cycling, artistic cycling (doing acrobatics on a bike) and BMX Freestyle (doing 1-minute runs on a BMX course with high skill arial tricks) require not only completely different skills, they are judged sports where other cycling sports are timed sports. All require different skills and abilities, vary in equipment, and have different physiological demands. To complicate matters further, cycling consists of individual events (e.g., Mountain Bike XCO, Road Individual Time Trial, BMX Race etc.) and team events (e.g., Team Pursuit, Team Time Trial, Indoor cycling etc.). This ‘diversity’ amongst cycling sports undermines the conclusion that cycling is a single sport and athletes who engage in multiple ‘sub-sports’ under the cycling umbrella are at the same risk of other specialized athletes. Given the recent attention to exploring the mechanisms of risk associated with specialization (Baker et al., 2021), these results from cycling (and perhaps similar sports like triathlon that have a degree of diversity built in) highlight several areas for future research.

This example highlights the need to further clarify the definitions of specialization and diversification. In the case of these 'umbrella sports' (e.g., ski sports, cycling, athletics etcetera) the current definitions do not fit in the opinion of the writer. What about different positions in a team sport? A goal keeper does fundamentally different things to a forward or winger in hockey and soccer. As such, a class or multiple classes that sit between specialization or multi-sport to fit these sports is required and the following repositioning of definitions, including two new ones, are proposed below:

(Early, simplified thoughts. Needs iterative improvement)

Specialization

Adding further refinement to Ericsson (2020) his definition, bringing it to: engaging in intense year-round engagement in a single sport, and within that sport being exclusively active in a single discipline or position within said sport with the exclusion of engagement in other sports or disciplines.

Similar-discipline

Those engaging in a single sport, yet in multiple disciplines of a similar nature (e.g., using similar energy systems for similar durations, similar motor skill and/or similar cognitive perceptual skills). Examples include an athlete engaging in the 100m sprint, 200m sprint and hurdles in athletics. An athlete engaging in both the 100m freestyle and butterfly in Swimming or a cyclists competing in both the 4-km individual and team pursuits on the Track or a Basketball player being active in both 5 on 5 and 3 on 3 Basketball.

Multi-discipline

Those engaging in multiple disciplines of a differing nature (e.g., different energy systems, higher skill variety between disciplines and/or different environments and equipment) in a single sport. Examples include an athlete competing in Road and Mountain Bike cycling or an athlete engaging in both Downhill and Cross Country skiing.

Multi-sport

Describing those that engage in completely different sports altogether, requiring vastly different skill. Such as Football and Basketball or Ice dancing and Baseball for example.

Using the above continuum, the term 'diversification' effectively becomes scalable (i.e., one can engage in varying degrees of diversification) rather than a binary term (i.e., one either does or does not engage in diversification). Based on the above rationale, one can engage in some level of diversification in similar-discipline and increase levels of diversification through multi-discipline and ultimately multi-sport.

Future work

As alluded to in the previous paragraph, there are a number of other cycling sports administered by the UCI that were not included in this research (e.g., Track Sprint, BMX Race, BMX Freestyle, Mountain Bike Gravity, Indoor cycling, Trials and Artistic cycling). Researchers may wish to explore these sports using similar or improved methods to see if the current findings can be generalized amongst other cycling sports. By extension, it might also be of interest to investigate whether similar effects can be found in other sports that operate in similar contexts (i.e., having multiple sports or disciplines under their umbrella) such as athletics, swimming, skiing and speedskating.

It is noteworthy that the cohort that was studied in this thesis includes those who competed between 13 and 3 years ago at the time of writing. In this or future decades, effects may change as advancements in resourcing and training prescription occur. To determine whether the findings of this study hold up over time, it may be worthwhile to conduct a similar analysis on future cohorts.

Given the findings on the odds of achieving eminence in endurance cycling by matching certain cycling sports and the relatively high number of multiple winners amongst eminent and emerging eminent cyclists, an obvious area for future inquiry is determining why this was the case. While hypotheses have been offered earlier in this general discussion, they remain largely speculative and require exploration.

A final suggestion for future work relates to exploring potential selection biases in the transitions from U19 to elite. For example, when looking at the progression rates from U19 athletes per cycling sport to Elite domains, Track was nearly double that of other sports. It is also the only cycling sport that was investigated where a national team controls programming at both the junior and elite levels. This raises the question of whether track is a sport that requires earlier specialization, due to talent scarcity beyond the junior age, for example, or whether there might be selection biases at play. Future work can benefit from exploring these questions, which may ultimately help to better support athletes and more effectively use resources or sporting organisations.

Concluding Remarks

Results from this dissertation highlight the reality that there are more questions about eminent performers and their development (both in general and in endurance cycling sports) than answers. These findings emphasize how little we know about eminent performance in general, and in specific sport settings in particular. Whilst scientific inquiry into the matter started nearly a millennium and a half ago, investigations into these individuals have been limited. As such our understanding is still very much developing. That said, there are some important takeaways from the current work for both researchers and practitioners regarding athlete selection and development in cyclists. It is my hope that with this work, athletes, coaches, researchers, and administrators can make more informed decisions and will continue to build on our collective understanding of cycling performance.

References

Abbiss, C. R., Ross, M. L. R., Garvican, L. A., Ross, N., Pottgiesser, T., Gregory, J., & Martin, D. T. (2013). The distribution of pace adopted by cyclists during a cross-country mountain bike World Championships. *Journal of Sports Sciences*, *31*(7), 787–794.

Aicale, R., Tarantino, D., & Maffulli, N. (2018). Overuse injuries in sport: a comprehensive overview. *Journal of Orthopaedic Surgery and Research*, *13*(1), 1-11.

Allen, S. V., & Hopkins, W. G. (2015). Age of peak competitive performance of elite athletes: A systematic review. *Sports Medicine*, *45*(10), 1431-1441

Bahník, Š., & Vranka, M. A. (2017). Growth mindset is not associated with scholastic aptitude in a large sample of university applicants. *Personality and Individual Differences*, *117*, 139-143.

Bailey, R. (2007). Talent development and the luck problem. *Sport, Ethics and Philosophy*, 1(3), 367-377.

Bejder, J., Bonne, T. C., Nyberg, M., Sjøberg, K. A., & Nordsborg, N. B. (2019). Physiological determinants of elite mountain bike cross-country Olympic performance. *Journal of Sports Sciences*, *37*(10), 1154–1161.

Baker, J., Cobley, S., & Fraser-Thomas, J. (2009). What do we know about early sport specialization? Not much! *High Ability Studies*, *20*(1), 77–89.

Baker, J., Mosher, A., & Fraser-Thomas, J. (2021). Is it too early to condemn early sport specialisation? *British Journal of Sports Medicine*, 55(3), 179-180.

Baker, J., Wilson, S., Johnston, K., Dehghansai, N., Koenigsberg, A., de Vegt, S., & Wattie, N. (2020). Talent Research in Sport 1990–2018: A Scoping Review. *Frontiers in Psychology*, *11*.

Baker, J., Longo, A. F., Siffredi, C. R., Cardey, M. L., Aquilino, G. D., & Lentini, N. A. (2016). Age of peak performance in Olympic sports: A comparative research among disciplines. *Journal of Human Sport and Exercise*, *11*(1), 31-41.

Baker, J., Schorer, J., Lemez, S., & Wattie, N. (2019). Understanding high achievement: The case for eminence. *Frontiers in Psychology*, *10*.

Baker, J., Wattie, N., & Schorer, J. (2015). Defining expertise: A taxonomy for researchers in skill acquisition and expertise. In J. Baker and D. Farrow (Eds.) *Routledge handbook of sport expertise* (pp. 145-155). Routledge

Barth, M., Güllich, A., & Emrich, E. (2018). The rich get richer and the poor get poorer - the Matthew mechanism as an approach to explain selection effects and the occurrence of multiple medalists in the "production" of international success in alpine ski racing. *Current Issues in Sport Science (CISS)*. *3*, 008-028.

Bayes, T., & Hume, D. (1763). Bayes's Theorem. In *Proceedings of the British Academy* (Vol. 113, pp. 91-109).

Benardot, D. (2013). Gymnastics. *The Encyclopaedia of Sports Medicine: An IOC Medical Commission Publication*, *19*, 596-606.

Bennett, P. E., Carpenter, S., & Gardiner, L. (2018). Chivalric games at the course of Edward III. *Medium Ævum*, *87*(2), 304-342.

Berri, D. J., Brook, S. L., & Fenn, A. J. (2011). From college to the pros: Predicting the NBA amateur player draft. *Journal of Productivity Analysis*, *35*, 25-35.

Bentham, J. (1890). *Utilitarianism*. Progressive Publishing Company

Brady, S. R. (2015). Utilizing and adapting the Delphi Method for use in qualitative research. *International Journal of Qualitative Methods*, *14*(5), 160940691562138.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. *Cognitive skills and their acquisition*, 141-189.

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. In *Psychology of learning and motivation* (Vol. 16, pp. 1-58). Academic Press.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1).

Clarsen, B., Myklebust, G., & Bahr, R. (2013). Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) overuse injury questionnaire. *British journal of sports medicine*, *47*(8), 495-502.

Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American psychologist*, *55*(2), 218.

Collins, D. J., & Macnamara, A. (2017). Making champs and super-champs-current views, contradictions, and future directions. *Frontiers in Psychology*, *8*(MAY), 1–8.

Collins, D.J., MacNamara, Á., & McCarthy, N. (2016). Super champions, champions, and almosts: Important differences and commonalities on the rocky road. *Frontiers in Psychology*, *6*(JAN), 1–11.

Craig, N. P., & Norton, K. I. (2001). Characteristics of track cycling. *Sports Medicine, 31,* 457–468.

Darwin, C. & Kebler, L. (1859) On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life . London: J.

Davids, K. (2010). The constraints-based approach to motor learning: Implications for a non-linear pedagogy in sport and physical education. *Motor learning in practice* (pp. 23-36). Routledge.

De Bosscher, V., Shibli, S., & Weber, A. C. (2019). Is prioritisation of funding in elite sport effective? An analysis of the investment strategies in 16 countries. *European Sport Management Quarterly*, *19*(2), 221-243.

Debraux, P., & Bertucci, W. (2011). Determining factors of the sprint performance in high level BMX riders. *Computer Methods in Biomechanics and Biomedical Engineering*, *14*(SUPPL.1), 53–55.

Den Hartigh, R.J., Hill, Y., & Van Geert, P.L.C. (2018). The development of talent in sports: A dynamic network approach. *Complexity*. doi: 10.1155/2018/9280154.

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology*, *67*(4), 401–409.

Dimaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Dweck, C. S. (2009). Mindsets: Developing talent through a growth mindset. *Olympic Coach*, *21*(1), 4-7.

Ericsson, K. A., Chase, W. G., & Falcon, S. (1980). Acquisition of a memory skill. *Science*, *208*(4448), 1181-1182.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, *100*(3), 363.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. *The Cambridge handbook of expertise and expert performance*, *38*(685-705), 2-2.

Ericsson, K. A. (2020). Towards a science of the acquisition of expert performance in sports: Clarifying the differences between deliberate practice and other types of practice. *Journal of sports sciences*, *38*(2), 159-176.

Erp, T. Van, Kittel, M., & Lamberts, R. P. (2021). Demands of the Tour de France: A Case Study of a World-Class Sprinter (Part I). *International Journal of Sports Physiology and Performance*, *16*(9), 1363–1370.

Erp, T.Van. (2019). Load, Intensity and Performance in Professional Road Cycling. *PhD Thesis, Vrije Universiteit van Amsterdam*.

Farrow, D. (2017). Super-elite athletes: Some complimentary observations from Australia and some lessons for sports expertise research: Comment on Hardy et al. In *Progress in Brain Research* (1st ed., Vol. 232). Elsevier B.V.

Ford, P., de Ste Croix, M., Lloyd, R., Meyers, R., Moosavi, M., Oliver, J., Till, K., & Williams, C. (2011). The Long-Term Athlete Development model: Physiological evidence and application. *Journal of Sports Sciences*, *29*(4), 389–402.

Franciscus, P., & Merkes, J. (2020). *Improving Sprint Performance in Road Cycling*.

Frumkin, P., & Galaskiewicz, J. (2004). Institutional Isomorphism and Public Sector Organizations. *Journal of Public Administration Theory*.

Gagné, F. (2000). A differentiated model of giftedness and talent (DMGT). *Systems and models for developing programs for the gifted and talented*, *2*.

Galton, F (1869). *Hereditary Genius*. http://galton.org.

Gardner, S. A., Martin, T. D., Barras, M., Jenkins, G. D., & Hahn, G. A. (2005). Power output demands of elite track sprint cycling. *International Journal of Performance Analysis in Sport*, *5*(3), 149–154.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual review of psychology*, *62*, 451-482.

Given, J .(2019). Case Study: A Finnish junior's eight-year road to cycling's World Tour.

Goodger, K., Gorely, T., Lavallee, D., & Harwood, C. (2007). Burnout in sport: A systematic review. *The sport psychologist*, *21*(2), 127-151.

Gould, D., Hodge, K., Petlichkoff, L., & Simons, J. (1990). Evaluating the effectiveness of a psychological skills educational workshop. *The Sport Psychologist*, *4*(3), 249-260.

Gould, D., Tammen, V., Murphy, S., & May, J. (1991). An evaluation of US Olympic sport psychology consultant effectiveness. *The Sport Psychologist*, *5*(2), 111-127.

Gould, D., Eklund, R. C., & Jackson, S. A. (1992). 1988 US Olympic wrestling excellence: I. Mental preparation, precompetitive cognition, and affect. *The sport psychologist*, *6*(4), 358-382.

Gould, D., Eklund, R. C., & Jackson, S. A. (1992). 1988 US Olympic wrestling excellence: II. Thoughts and affect occurring during competition. *The sport psychologist*, *6*(4), 383-402.

Gould, D., Eklund, R. C., & Jackson, S. A. (1993). Coping strategies used by US Olympic wrestlers. *Research quarterly for Exercise and Sport*, *64*(1), 83-93

Gould, D., Guinan, D., Greenleaf, C., Medbery, R., & Peterson, K. (1999). Factors affecting Olympic performance: Perceptions of athletes and coaches from more and less successful teams. *The sport psychologist*, *13*(4), 371-394.

Gowthorp, L., Toohey, K., & Skinner, J. (2017). Government involvement in high performance sport: An Australian national sporting organisation perspective. *International Journal of Sport Policy and Politics*, *9*(1), 153-171.

Granier, C., Abbiss, C. R., Aubry, A., Vauchez, Y., Dorel, S., Hausswirth, C., & Meur, Y. Le. (2018). Power output and pacing during international cross-country mountain bike cycling. *International Journal of Sports Physiology and Performance*, *13*(9), 1243–1249.

Gratton, C., Liu, D., Ramchandani, G., & Wilson, D. (2012). *The global economics of sport*. Routledge.

Grix, J., & Carmichael, F. (2012). Why do governments invest in elite sport? A polemic. *International Journal of Sport Policy*, *4*(1), 73–90.

Groot, L. (2008). *The Contest for Olympic Succes as a Public Good*. *151*(4), 1–46.

Gulbin J, Croser M, Morley E. (2013). An integrated framework for the optimisation of sport and athlete development: A practitioner approach. *Journal of Sports Sciences, (2013), 1319-1331, 31(12)*

Güllich, A. (2017). International medallists' and non-medallists' developmental sport activities–a matched-pairs analysis. *Journal of Sports Sciences*, *35*(23).

Gustafsson, H., Hassmén, P., Kenttä, G., & Johansson, M. (2008). A qualitative analysis of burnout in elite Swedish athletes. *Psychology of sport and exercise*, *9*(6), 800-816.

Haldane, J. B. S. (1946). The interaction of nature and nurture. *Annals of eugenics*, *13*(1), 197-205.

Hardy, L., Barlow, M., Evans, L., Rees, T., Woodman, T., & Warr, C. (2017). Great British medalists: Psychosocial biographies of Super-Elite and Elite athletes from Olympic sports. In *Progress in Brain Research* (1st ed., Vol. 232). Elsevier B.V.

Henriksen, K., & Christensen, M. K. (2013). Athletes' careers in Denmark. *Athletes' careers across cultures*, 77-89.

Hogan, K., & Norton, K. (2000). The "price" of olympic gold. *Journal of Science and Medicine in Sport*, *3*(2), 203–218.

Hogben, L. (1933). Nature and nurture.

Hopwood, M. (2013). *The Developmental History of Athletes Questionnaire: Towards a comprehensive understanding of the development of sport expertise* (Doctoral dissertation, Victoria University).

Houlihan, B., & Zheng, J. (2016). The Olympics and elite sport policy: Where will it all end? *Delivering Olympic and Elite Sport in a Cross Cultural Context: From Beijing to London*, *3367*, 1–18.

Janssens, B., Bogaert, M., & Maton, M. (2022). Predicting the next Pogačar: a data analytical approach to detect young professional cycling talents. *Annals of Operations Research*.

Johnston, K., & Baker, J. (2019). Waste reduction strategies: Factors affecting talent wastage and the efficacy of talent selection in sport. *Frontiers in Psychology*, *10*.

Johnston, K., McAuley, A. B., Kelly, A. L., & Baker, J. (2023). Language games and blurry terminology: Can clarity enhance athlete development?. *Frontiers in Sports and Active Living*, *5*, 1150047.

Johnston, K., Wattie, N., Schorer, J., & Baker, J. (2018). Talent Identification in Sport: A Systematic Review. In *Sports Medicine* (Vol. 48, Issue 1, pp. 97–109). Springer International Publishing.

Kabush, D., & Orlick, T. (2001). Focusing For Excellence: Lessons From Elite Mountain Bike Racers. In *Journal of Excellence-Issue No. 5 Kabush and Orlick* (Vol. 39).

Keeney, S., Hasson, F., & McKenna, H. (2006). Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *Journal of advanced nursing*, *53*(2), 205-212.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: a systematic review. *Psychological medicine*, *37*(5), 615-626.

Kholkine, L., Latré, S., Verdonck, T., & de Leeuw. (2023). Age of peak performance in professional road cycling, *Journal of Sports Sciences*.

Kliethermes, S. A., Marshall, S. W., LaBella, C. R., Watson, A. M., Brenner, J. S., Nagle, K. B., ... & Beutler, A. I. (2021). Defining a research agenda for youth sport specialisation in the USA: the AMSSM Youth Early Sport Specialization Summit. *British Journal of Sports Medicine*, *55*(3), 135-143.

Kordi, M., Folland, J., Goodall, S., Haralabidis, N., Maden-Wilkinson, T., Sarika Patel, T., Leeder, J., Barratt, P., & Howatson, G. (2020). Mechanical and morphological determinants of peak power output in elite cyclists. *Scandinavian Journal of Medicine and Science in Sports*, *30*(2), 227–237.

Leisman, G. (2012). Genius. *Functional Neurology, Rehabilitation, and Ergonomics*, *2*(2), 93.

Lloyd, R. S., Oliver, J. L., Faigenbaum, A. D., Howard, R., de Ste Croix, M. B. A., Williams, C. A., Best, T. M., Alvar, B. A., Micheli, L. J., Thomas, D. P., Hatfield, D. L., Cronin, J. B., & Myer, G. D. (2015). Long-term athletic development- Part 1: A pathway for all youth. In *Journal of Strength and Conditioning Research* (Vol. 29, Issue 5, pp. 1439–1450). NSCA National Strength and Conditioning Association.

Louis, J., Billaut, F., Bernad, T., Vettoretti, F., Hausswirth, C., & Brisswalter, J. (2013). Physiological demands of a simulated BMX competition. *International Journal of Sports Medicine*, *34*(6), 491–496.

Martin, J. C., Davidson, C. J., & Pardyjak, E. R. (2007). Understanding sprint-cycling performance: The integration of muscle power, resistance, and modeling. *BRIEF REVIEW International Journal of Sports Physiology and Performance* (Vol. 2).

McAuley, A. B., Hughes, D. C., Tsaprouni, L. G., Varley, I., Suraci, B., Roos, T. R., ... & Kelly, A. L. (2021). Genetic association research in football: A systematic review. *European Journal of Sport Science*, *21*(5), 714-752.

McAuley, A.B.T., Baker, J. & Kelly, A.L. Defining “elite” status in sport: from chaos to clarity. *Ger J Exerc Sport Res* 52, 193–197 (2022).

McAuley, A. B., Baker, J., Johnston, K., Varley, I., Herbert, A. J., Suraci, B., ... & Kelly, A. L. (2023). Talent inclusion and genetic testing in sport: A practitioner’s guide. *Current Issues in Sport Science (CISS)*, *8*(1), 008-008.

McAuley, A. B., Baker, J., Suraci, B., & Kelly, A. L. (2024). Genetics: Understanding the Influence and Application of Genetics in Soccer. In *Talent Identification and Development in Youth Soccer* (pp. 197-211). Routledge.

McCue, M., Baker, J., Lemez, S., & Wattie, N. (2019). Getting to first base: Developmental trajectories of major league baseball players. *Frontiers in Psychology,* 10, 2563.

McKay, A. K., Stellingwerff, T., Smith, E. S., Martin, D. T., Mujika, I., Goosey-Tolfrey, V. L., ... & Burke, L. M. (2022). Defining Training and Performance Caliber: A Participant Classification Framework. *International journal of sports physiology and performance*, *1*(aop), 1-15.

Millar, P., Clutterbuck, R., & Doherty, A. (2020). Understanding the adoption of long-term athlete development in one community sport club. *Managing Sport and Leisure*, *25*(4), 259–274.

Moesch, K., Elbe, A. M., Hauge, M. L. T., & Wikman, J. M. (2011). Late specialization: The key to success in centimeters, grams, or seconds (cgs) sports. *Scandinavian Journal of Medicine and Science in Sports*, *21*(6).

Mosher, A., Baker, J., & Fraser-Thomas, J. (2020). Comparing psychological constructs in early specializing and non-specializing youth boys hockey players. *Current Issues in Sport Science (CISS)*, *5*, 004-004.

Mostaert, M., Laureys, F., Vansteenkiste, P., Pion, J., Deconinck, F. J., & Lenoir, M. (2020). Discriminating performance profiles of cycling disciplines. *International Journal of Sports Science & Coaching*.

Mostaert,M., Vansteenkiste,P ., Pion., Deconinck,F., & Lenoir, M (2021) The importance of performance in youth competitions as an indicator of future success in cycling, European Journal of Sport Science.

Mujika, I., & Padilla, S. (2001). *Physiological and performance characteristics of male professional road cyclists*. Sports Med (7).

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: an example, design considerations and applications. *Information & management*, *42*(1), 15-29.

Padilla, S., Nigo Mujika, I. , Orbananos, J., Orbananos, O., & Angulo, F. (2000). Exercise intensity during competition time trials in professional road cycling. *Medicine and Science in Sports and Exercise, 32,* 850-856.

Pawlowski, S. D. (2004). The Delphi Method as a Research Tool : An Example , Design Considerations and Applications 1 Introduction 2 Overview of the Delphi method. *Information & Management*, *42*(1), 15–29.

Peters, C. M., Hendry, D. T., & Hodges, N. J. (2022). A scoping review on developmental activities of girls' and women's sports. *Frontiers in Sports and Active Living*, *4*, 903886.

Phillips, K. E., & Hopkins, W. G. (2020). Determinants of cycling performance: A review of the dimensions and features regulating performance in elite cycling competitions. *Sports Medicine - Open*, *6*(1), 23.

Ploutz-Snyder, R. J., Fiedler, J., & Feiveson, A. H. (2014). Justifying small-n research in scientifically amazing settings: Challenging the notion that only “big-n” studies are worthwhile. In *Journal of Applied Physiology* (Vol. 116, Issue 9, pp. 1251–1252). American Physiological Society.

Pluchino, A., Biondo, A. E., & Rapisarda, A. (2018). Talent versus luck: The role of randomness in success and failure. Advances in Complex systems, 21(03n04), 1850014

Pol, R., Balagué, N., Ric, A., Torrents, C., Kiely, J., & Hristovski, R. (2020). Training or synergizing? Complex systems principles change the understanding of sport processes. *Sports Medicine-Open*, *6*(1), 1-13.

Record of Medals at Games of the Olympiad (2021). Retrieved via https://olympics.com/ioc/documents/olympic-games/records-and-medals

Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., ... & Warr, C. (2016). The great British medalists project: a review of current knowledge on the development of the world's best sporting talent. *Sports medicine*, *46*(8), 1041-1058.

Renshaw, I., Chow, J. Y., Davids, K., & Hammond, J. (2010). A constraints-led perspective to understanding skill acquisition and game play: a basis for integration of motor learning theory and physical education praxis?. *Physical Education and Sport Pedagogy*, *15*(2), 117-137.

Ridley, M. (2003). *Nature via nurture: Genes, experience, and what makes us human*. HarperCollins Publishers.

Robinson, G. E. (2004). Beyond nature and nurture. *Science*, *304*(5669), 397-399.

Rylands, L. P., & Roberts, S. (2019). Performance Characteristics in BMX Racing: A Scoping Review. *Journal of Science and Cycling*, *8*(1), 3-10.

Sánchez-Muñoz, C., Muros, J. J., & Zabala, M. (2018). World and Olympic mountain bike champions' anthropometry, body composition and somatotype. *Journal of Sports Medicine and Physical Fitness*, *58*(6), 843–851

Schädler, U., & Bagnall, R. S. (2013). Games, Greek and Roman. *The encyclopedia of ancient history*, *5*, 2841-2844.

Schumacher, Y. O., Mroz, R., Mueller, P., Schmid, A., & Ruecker, G. (2006). Success in elite cycling: A prospective and retrospective analysis of race results. *Journal of Sports Sciences*, *24*(11), 1149–1156.

Seiler, S. (2010). What is best practice for training intensity and duration distribution in endurance athletes?. *International journal of sports physiology and performance*, *5*(3), 276-291.

Seiler, S., & Tønnessen, E. (2009). Intervals, thresholds, and long slow distance: the role of intensity and duration in endurance training. *Sportscience*, *13*.

Skulmoski, G. J., Hartman, F. T., & Krahn, J. (2007). The Delphi method for graduate research. *Journal of Information Technology Education: Research*, *6*(1), 1-21.

Spearman, C. (1961). " General Intelligence" Objectively Determined and Measured.

Stöggl, T. L., & Sperlich, B. (2015). The training intensity distribution among well-trained and elite endurance athletes. *Frontiers in physiology*, *6*, 295.

Swaddling, J. (1999). *The ancient Olympic games*. University of Texas Press.

Swann, C., Keegan, R. J., Piggott, D., & Crust, L. (2012). A systematic review of the experience, occurrence, and controllability of flow states in elite sport. *Psychology of sport and exercise*, *13*(6), 807-819.

Swann, C., Moran, A., & Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, *16*, 3-14

Swann, C., Piggott, D., Schweickle, M., & Vella, S. A. (2018). A review of scientific progress in flow in sport and exercise: normal science, crisis, and a progressive shift. *Journal of Applied Sport Psychology*, *30*(3), 249-271.

Terman, L. M., & Oden, M. H. (1925). *Genetic studies of genius: The gifted child grows up*. Stanford University Press.

Van, D., Daniel, R., & Larson, J. (2016). *The Economics of Professional Road Cycling*.

Wachs, T. D. (1992). *The nature of nurture*. Sage publications.

Wilson, Jacob M.; Loenneke, Jeremy P.; Jo, Edward; Wilson, Gabriel J.; Zourdos, Michael C.; Kim, Jeong-Su. (2012). The Effects of Endurance, Strength, and Power Training on Muscle Fiber Type Shifting. Journal of Strength and Conditioning Research 26(6):p 1724-1729.

Van Bulck, D., Vande Weghe, A., & Goossens, D. (2021). Result-based talent identification in road cycling: discovering the next Eddy Merckx. *Annals of Operations Research*.

Van, D., Daniel, R., & Larson, J. (2016). *The Economics of Professional Road Cycling*.

Van Hooren, B., & Croix, M. D. S. (2020). Sensitive periods to train general motor abilities in children and adolescents: do they exist? A critical appraisal. *Strength & Conditioning Journal*, *42*(6), 7-14.

Van Reeth, D.(2022). The Economics of Professional Road Cycling, second edition.

Welinder, P., Branson, S., Perona, P., & Belongie, S. (2010). The multidimensional wisdom of crowds. *Advances in neural information processing systems*, *23*.

Appendices

Appendix A. Description of different cycling sports mentioned in this thesis.

Artistic Cycling

A judged sport where cyclists perform a set of tricks (called exercises) routine on a flat surface, not unlike a dancing using a specialized, fixed-gear bike.

BMX Race

A sport done on a specific BMX Race course with several hills and corners, starting on either a 5 meter or 8 meter hill with a gated start. Each race will feature up to 8 athletes. Athletes use BMX bikes and elite riders typically complete a BMX Race course between 30-60 seconds.

BMX Freestyle

A judged sport where athletes look to impress through acrobatics on a BMX bike during 60 second individual runs on a park with several ramps and 'features'. Athletes get scores based on smoothness and impressiveness.

Cycle Ball

A team sport on bicycles that is similar in association to indoor football (soccer). It is played indoors with team sizes of 5 or 6 athletes each side. Athletes use the wheels of their bikes to maneuver, pass or shoot the ball.

Cyclo Cross

A sport that is a race that starts with a bunch of athletes featuring a mass start and specific Cyclo Cross bikes. Athletes navigate a number of laps on an off-road course which typically futures mud, sand, grass, hurdles and steep climbs. Races for the elite categories typically last between 40 and 60 minutes.

Mountain Bike

Similar to Cyclo Cross, Mountain Bike races are "off road" races. In this study, only the Olympic version, Cross Country Olympic (MTB XCO) is explored. MTB XCO done on a fixed circuit with specific Mountain Bike bicycles and courses that typically feature climbs, flowing downhill sections, rock gardens and other features to navigate. Races typically vary from 40 minutes to an hour and a half.

Road

Consisting of Individual Time Trials, Team Time Trials and Road Races, These events are held on any kind of road surface and can feature different terrains (i.e., flat, hilly, mountainous). Individual time trials are individual races against the clock on specific time trial bikes and typically vary between 10 and 60 minutes in duration. Team time trails are similar, yet executed by a professional or national team made up of several members. Road races are subdivided between one day races (i.e., Tour of Flanders, Paris Roubaix, World Championships etc.) which typically last between 2 and 6 hours and grand tours (i.e., Tour de France, Giro d'Italia, Paris – Nice, Vuelta d'Espagna etc.) which are comprised of individual stages and might take anywhere between 4 and 21 days.

Track Endurance

Track endurance events are held on 250-meter-long oval-shaped tracks and are comprised of both bunch (for example, an omnium which features up to 24 individual riders at a time) and timed events (such as individual and team pursuits) and may last anywhere between 3 and 60 minutes.

Track Sprint

Held on the same tracks as track endurance events, track sprint events are much shorter in duration and typically last between 30 and 60 seconds.

Appendix B. Eminence Interview Guide

Establishing Eminence in Endurance Based Cycling Sports

Objective

Develop a better understanding of what eminence looks like in Endurance Based Cycling sports, to explore sport specific stages in elite Cycling and to explore which factors differentiate "super-champions" from the rest.

1. On the protocol form, please address/revise/clarify the following:

· Part B, Q3 - Recruitment – please provide more detail regarding the recruitment process.

· The protocol states that participants will be anonymous, but they are not anonymous to the researchers since they are being recruited through their own networks/prior knowledge:

o The participant data will be confidential, but not anonymous to the researchers Please address.

· The protocol states that there are no risks to participants however, the participation in a focus group or an interview with individuals known to them, may result in participants feeling uncomfortable or concerned about loss of reputation, etc. Please address/revise.

To submit additional/ revised documentation, please reply and attach the files to this email. The files will be automatically uploaded into the online ethics review system and date-stamped as the most recently submitted for review.

Interview Guide

The Delphi study will consist of two interview sets and each set will be made up of two rounds.

The two interview sets will be geared to exploring:

1. *Eminence – what does it look like in Cycling?*
2. *What are underpinning characteristics that differentiate eminent cyclists from those that are not?*

Within those two sets, two rounds of interaction will take place:

1. *Individual answers of panelists*
2. *Distribution of answers in an anonymous fashion followed by a group-wide discussion to arrive at jointly agreed upon positions*

As interview set #2 will be dependent on outcomes of set#1, interview questions for set#2 will be developed post completion of round 2 of set#1.

Panelists may choose to submit their interview answers via SurveyMonkey. If preferred, interviews can also be conducted via Skype/Zoom or in person and will be recorded, with permission.

Now that we have had the chance to go through the information letter, would it be okay if we began the interview? During this session, we will explore what Eminence in cycling looks like. During this process, I may ask you questions that are relevant to the information you provide me. These questions will mainly be related to your personal opinions and experiences. That being said, you may add any information that you think is pertinent at any point in time. Do you give me permission to audio record this interview?

Please keep in mind that after our interview, I will collect answers from all panelists and redistribute them in an anonymous fashion for discussion.

The interview commences

Explanation of the word "Eminence" in the context of this study. The word 'eminence' is often used in this document and will be used a lot throughout the entirety of the study. In this setting, eminence refers to the highest achievable level within the sport. Synonyms to eminence in the sport context include: "super-champions" and "super-elite". In popular terms, athletes that are eminent in North American sports are often described as: "(future) hall of famers", "MVP's" and "superstars".

Set #1 , Round 1

Section A: Establishing tiers in race levels:

1. If you should name a set of races as the "premier league" of cycling, which races would be in it?
2. If you were to pool all races from all endurance disciplines, which would be in the "premier league" which would be "second division", "third division" for males?
3. If you were to pool all races from all endurance disciplines, which would be in the "premier league" which would be "second division", "third division" tiers for females?

Section B: Establishing qualifications of being considered eminent:

4. If the absolute best cyclists in the world would be categorized as "eminent", what would one have to do to be considered in that class?

Section C: Establishing pathway to become eminent:

5. If eminence would be considered to be the level or category that describes those that have reached the highest possible level within elite cycling, how many levels would you distinguish between entry to eminence within elite cycling?
6. Would there be any additional thoughts based on the topic at hand that should be elaborated on or that you would like to add?

Thank you. The final items are some statements for which I'd like to ask you if you strongly agree, agree, are neutral, somewhat disagree or strongly disagree.

1. (Balanced Likert design, sequence to be randomized)
    a. Career length is an important differentiator between non-eminent cyclists and eminent cyclists (career length +1)
    b. A cyclist can be considered eminent with a relatively short career (career length -1)
    c. The amount of years ranked in the top 10 of the UCI ranking is an important differentiator between non-eminent cyclists and eminent athletes (duration of peak +1)
    d. A cyclist can be considered eminent after a single short and successful period (duration of peak -1)
    e. In order to be labelled eminent, one has to have won World Championships, the Tour de France, a Monument or Olympic Games (assumed premier league +1)
    f. A cyclist might be considered eminent without winning races in the "premier league" events as described above (assumed premier league -1)
    g. Cyclists that are eminent have the ability to be dominant in multiple cycling disciplines (domain specificity +1)
    h. Most cyclists that are considered eminent only have the ability to be dominant in a singly cycling sport even if they would apply themselves in a different discipline (domain specificity -1)

All findings will be shared amongst panelists followed by an anonymous group discussion which will be the next stage. Timelines for this discussion will be very broad to allow all to participate without impeding with competition or training scheduling. These timelines and other general information can be found in the study guideline.

After the discussion period is concluded, the first phase will be finished.

The second phase of the study will be designed to explore the characteristics, or the traits and behaviors that separates eminent athletes from those that are not. Questions and areas of exploration can include such things as psychological traits and skills, physical aptitude, training volumes, amount of race days, whether or not these athletes are strictly active in one sport or if they cross over to different domains or other differentiators that are found to be important. This phase is targeted to commence in October 2020.

To conclude, I would like to thank you for your time and patience today. Please remember that in no shape or form will your identity be exposed in presentations and publications and that you will remain anonymous throughout the entire length of this project. As a reminder, you have the right to withdraw until the publication of the findings.

Appendix C. Informed Consent

1. **Invitation to Participate**
   You are invited to participate in this research study which looks to integrate innovative scientific research and sport-driven data to improve models of athlete development in endurance cycling sports.

2. **Purpose of the Letter**
   The purpose of this letter is to provide you with information required for you to make an informed decision regarding participation in this research.

3. **Purpose of this Study**
   The purpose of this research project is to enhance our understanding of key development-related factors that affect expert athlete development and performance. The sport-driven data can help improve athlete development models, training approaches and talent identification systems in cycling.

4. **Inclusion Criteria**
   Individuals who currently or previously have coached or competed in an endurance cycling sport and have at least one podium performance at elite World Championships or Olympic Games are eligible to participate in this study.

5. **Study Procedures**
   If you agree to participate, you will be asked to participate in interviews with the principle researcher, Jesse Korf, which will contain questions pertaining to your experiences in Cycling. The study will consist of two phases. Participants will be asked to fill out one questionnaire per phase consisting of open questions and five-point Likert-scale (strongly agree to strongly disagree) items and participate in a discussion. Both the questionnaires and the discussions can be done either online or via conversation. Content will be made anonymous and shared with participants with the aim of coming to (or as close as possible) agreed upon answers as a group.
   The first contact point (questionnaire) of each phase will be to accrue information on an individual level. The second (discussion) to attempt to come to shared definitions on the group level.

6. **Possible Risks and Harms**
   There are no known or anticipated risks or discomforts associated with participating in this study. However, the participation in a focus group or an interview with individuals known to them, may result in participants feeling uncomfortable or concerned about loss of reputation.

7. **Expected Benefits**
   There are no direct benefits to you but the results from this study will extend our understanding of pathways to expertise and assist with developing and modifying programs to enhance the conditions optimal for development of expertise in cycling sports.

8. **Compensation**
   No compensation is available to participants

9. **Voluntary Participation**
   Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time.

10. **Confidentiality**
    In the interest of privacy and confidentiality, all of your responses in this interview will be transcribed, returned to you for validation and identified using a participant code. As such, only the research team will have access to your personal information and **you will remain anonymous** throughout the course of the research. Further, confidentiality will be provided to the fullest extent possible by law. That being stated, participant data will not be anonymous to the researchers.

11. **Data collection and Storage**
    Data will be collected during interviews via an audio recording device. These recordings will be transferred to a secure computer with a two-phases password protection and the recordings on the original device will be deleted immediately. The data will only be stored in the two-phase password protected system and accessible by the research team for up to 7 years, after which all recordings will be deleted permanently.

- [ ] I consent to the interview being held with the lead researcher to be recorded with the audio recording device.

12. **Contacts for Further Information**
    Any queries about your participation in this project may be directed to the Principal investigator(s):

    Mr. Jesse Korf

    jesse_jcu@hotmail.com

+1 905-691-3829

OR

Dr. Joseph Baker

bakerj@yorku.ca

+1 416-736-2100 ext. 22361

If you have any queries or complaints about the project, or the way you have been treated, you are of course, free to contact the Manager of the York University Research Ethics Department (Ms. Alison Collins-Mrakas, 309 York Lanes, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, +1 416 736 5914, acollins@yorku.ca

13. Publication

You will remain anonymous throughout the course of the research and reporting of the findings in presentations in various conferences and coaching workshops and publications in various scientific journals and coaching magazines. If you would like to receive a copy of any potential study results, please contact one of the investigators.

Appendix D. Survey items

Section A: Establishing tiers in race levels:

1. If you should name a set of races as the "premier league" of cycling, which races would be in it?
2. If you were to pool all races from all endurance disciplines, which would be in the "premier league" which would be "second division", "third division" for males?
3. If you were to pool all races from all endurance disciplines, which would be in the "premier league" which would be "second division", "third division" tiers for females?

Section B: Establishing qualifications of being considered eminent:

4. If the absolute best cyclists in the world would be categorized as "eminent", what would one have to do to be considered in that class?

Section C: Establishing pathway to become eminent:

5. If eminence would be considered to be the level or category that describes those that have reached the highest possible level within elite cycling, how many levels would you distinguish between entry to eminence within elite cycling?
6. Would there be any additional thoughts based on the topic at hand that should be elaborated on or that you would like to add?

Appendix E. Example of data capture in the Delphi studies

Transcript of Cycling Eminence study round 1
Technical Leaders
2
Q1 If only the absolute best cyclists in international racing would be
named “eminent”, describe in your own words what an athlete has
to do to earn this title?
Answered: 12 Skipped: 0
RESPONSES DATE
1 A cyclist that over that several years (4yrs+) are consistently are winning the Top (World Tour) races, in different parts of the season and in different kind or races.
4/27/2020
10:13 PM

Attain podium finishes in senior global individual competitions on at least 5 occasions over a minimumof 3 seasons
4/26/2020
9:18 PM
3
For me it is less about what an athlete has to do than what an athlete has to be: An athlete has to be the best complete combination of all of the characteristics that lead to success. Performance capacity stemming from physical talent is only one piece of this equation. Equal value goes to: attitudinal and mental plasticity and openness, physical and immune system robustness (illness/injury avoidance), discipline and self-determination, ability to handle set-backs and failures and come back from them with greater maturity and experience that contributes to future growth, genetic response rate to stimuli. These are but a few.
4/25/2020
4:49 PM
4 They need to show that they are the best in world at least 2 times when everyone is able to compete. (with no luck) Just skill, fitness and power, endurance and mental strength.
4/18/2020
10:59 AM
5
Display an ability to consistently place in the top 3 in races on the level of monuments, grand tour GC, world championships and Olympics from year to year. Athletes that have a consistent list of success on this level can be thought of as eminent and a level above other athletes that have been occasionally successful.
4/14/2020
12:48 PM
6
Dominate a generation of cyclists by performance as well as by opinion leading personality. Unfortunately, in this survey it is entirely unclear whether the questions address ROAD CYCLING,
TRACK CYCLING, MTB, CX? My answers focus on ROAD CYCLING.
4/14/2020
12:30 PM
7 To win the biggest races in his or her discipline. More in a year and year after year.
4/14/2020
7:58 AM
8 Repeated podium success spanning a career, this could occur in one day or stage race results. 4/14/2020

1:00 AM
9 Win the Tour de France, The Giro d'Italia, The Vuelta a España, and/or World Champinships or some of the "monuments", not once but multiple times
4/13/2020
6:35 PM
10 Consistently win the highest level events, along their athletic career (Juniors to Elite to World class) 4/13/2020
6:08 PM
11
Able to win multiple top races, over multiple seasons, while demonstrating versatility and composure. Versatility = crossing disciplines or types of races. Composure = ability to handle high pressure, both on and off the bike. Including, when needed, coming back from set-backs. Top races = world championships, monument classics (or like), world cups, Olympics, grand-tours (GC or stages). The best races, most prestigious. For example: - Vos: CX, track and road - returning from a 1 year career set back - Ferrand-Prevot: CX, road, MTB - Wiggins: track - to - road, short 1-day events to grand tours - Theo Bos: from track sprint to road and back There are more athlete examples. However, these few all demonstrate multiple wins, versatility, and composure.
4/13/2020
4:42 PM
12 He/she must achieve several times top performances, such as podiums, in the Monuments of cycling and/or at world championships.
4/13/2020
4:17 PM
1/1
3

Appendix F: R Code of the quantitative analyses

```
#### Install packages

install.packages("tidyverse")
install.packages("dplyr")
install.packages("FactoMineR")
install.packages("reshape2")
install.packages("mvnormtest")
install.packages("PlayerRatings")
install.packages("hrbrthemes")
install.packages("pastecs")
install.packages("effsize")
install.pckages("broom")
install.packages("ggpubr")
install.packages("rstatix")
install.packages("cluster")
install.packages("vcd")
install.packages("pivottabler")
install.packages("stats")
install.packages("ISLR")
install.packages("tidyr")
install.packages("elo")
install.packages("readxl")
install.packages("readr")
install.packages("data.table)
install.packages("gglorenz")
install.packages("Hmisc")
install.packages("ineq")
install.packages("ggplot2")
```

```
library(ineq)
library(ggplot2)
library(pivottabler)
library(stats)
library(ISLR)
library(tidyr)
library(elo)
library(readxl)
library(readr)
library(data.table)
library(gglorenz)

library(dplyr)
library(tidyverse)
library(pastecs)
library(effsize)
library(mvnormtest)
library(PlayerRatings)
library(hrbrthemes)
library(ggpubr)
library(rstatix)
library(broom)
```

```
library(cluster)

library(vcd)

 ####General Notices###


# Placement out of order

# fix example: dataset$placement <- as.numeric(as.character(dataset$placement))


#### Clean up header and row 1


# remove row 1


result <- result[-c(1), ]


#rename vectors


names(result) <- c("sport", "discipline", "competition", "location", "level", "year", "cat",
"placement", "bib", "name", "lastname", "firstname", "country", "team", "sex", "age",
"phase", "heat", "time", "norace", "points")


##### create smaller dataframes


essentials <- result [c("sport","discipline","competition", "placement", "name", "sex", "age",
"year" )]



##eminent dataframe


## TT Worlds

place <-subset(result, placement=="1")

tt <- subset(place, competition=="Championnats du Monde UCI CLM/ UCI TT World
Championships")

tt2 <- subset(place, competition=="Championnats du Monde UCI CLM/ UCI TT World
Championships (Men Elite)")
```

```
tt3 <- subset(place, competition=="Championnats du Monde UCI CLM/ UCI TT World Championships (Women Elite)")

tt4 <- subset(place, competition=="Championnats du Monde UCI / UCI World Championships")

tt5<- subset(place, competition=="Championnats du Monde Route UCI / UCI Road World Championships")

tt <-rbind(tt,tt2,tt3,tt4,tt5)

tt <- subset(tt, sport=="Road")

tt <- subset(tt, discipline=="Individual Time Trial")

tt <- subset(tt, level=="Elite")



## Road race worlds

rr <-subset(place, competition=="Championnats du Monde Route UCI / UCI Road World Championships")

rr2 <-subset(place, competition=="Championnats du Monde UCI / UCI World Championships")

rr3 <-subset(place, competition=="Championnats du Monde UCI CL / UCI RR World Championships")

rr4 <-subset(place, competition=="Championnats du Monde UCI CL / UCI RR World Championships (Men Elite)")

rr5 <-subset(place, competition=="Championnats du Monde UCI CL / UCI RR World Championships (Women Elite)")

rr<-rbind(rr,rr2,rr3,rr4,rr5)

rr <- subset(rr, sport=="Road")

rr <- subset(rr, level=="Elite")

rr <- subset(rr, discipline=="Individual Road Race")


##Track Worlds

tw <-subset(place, sport=="Track")

tw <-subset(tw, cat=="CM")

tw <-subset(tw, level=="Elite")

##filter for endurance

ip <-subset(tw, discipline=="Individual Pursuit")

md <-subset(tw, discipline=="Madison")
```

```
pr <-subset(tw, discipline=="Points Race")
omn<-subset(tw, discipline=="Omnium")
tp <-subset(tw, discipline=="Team Pursuit")
scr <-subset(tw, discipline=="Scratch")
remove(tw)
twt <- rbind(ip,tp,md,pr,omn,scr)

###remove md duplicates manually

####Cyclocross
cx <-subset(place, sport=="Cyclo-cross")
cx <-subset(cx, cat=="CM")
cx <-subset(cx, level=="Elite")

### MTB Worlds
mtb <-subset(place, sport=="Mountain-bike")
mtb <-subset(mtb, cat=="CM")
mtb <-subset(mtb, level=="Elite")
mtb <-subset(mtb, discipline=="country Olympic")

### Monuments
rm <-subset(place, sport=="Road")
rm <-subset(rm, level=="Elite")
#Flanders
flanders <-subset(rm, competition=="Ronde van Vlaanderen / Tour des Flandres")
#Roubaix
roubaix1 <-subset(rm, competition=="Paris - Roubaix")
roubaix2 <-subset(rm, competition=="Paris-Roubaix")
roubaix <-rbind(roubaix1,roubaix2)
###Liege
liege1 <-subset(rm, competition=="Liège-Bastogne-Liège")
liege2 <-subset(rm, competition=="Liège - Bastogne - Liège")
```

```
liege3 <-subset(rm,competition=="Liège-Bastogne-Liège Femmes")
liege <-rbind(liege1, liege2, liege3)
###Milan - San Remo
msr1 <- subset(rm, competition=="Milano-Sanremo")
msr2 <-subset(rm, competition=="Milano - Sanremo")
msr <-rbind(msr1,msr2)
###Lombardia
lomb1 <- subset(rm, competition=="Giro di Lombardia")
lomb2 <- subset(rm, competition=="Il Lombardia")
lomb <-rbind(lomb1,lomb2)

##Add womens road others manually

##Olympic Games
#road
rdoly <- subset(rm, cat=="JO")
#track
oomn <- subset(troly, discipline=="Omnium")
otp <- subset(troly, discipline=="Team Pursuit")
troly <-rbind(otp,oomn)
#MTB
omtb <-subset(place, sport=="Mountain-bike")
olymtb <-subset(omtb, cat=="JO")



##### Premier League dataframe
premierleague <-rbind(olymtb,troly,rdoly,lomb,msr,liege,roubaix,flanders,mtb,cx,twt,rr,tt)

write.csv(premierleague, "premierleagueraw.csv")

##clean csv in Excel, add birtdate quadrant for RAE analysis
```

```
### Check for RAE's

##whole sample
group_by(raeall, control) %>%
summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raeall)
effect

### by sex
##Female
group_by(raefemale, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raefemale)
effect
##MAle
group_by(raemen, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )
```

```
effect <- cohen.d(rae ~ control,data=raemen)
effect

###by sport
group_by(raeroadmen, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raeroadmen)
effect

group_by(raeroadfemale, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raeroadfemale)
effect

group_by(raemtbfemale, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raemtbfemale)
```

```
effect

group_by(raemtbmen, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raemtbmen)
effect

group_by(raetrackmen, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raetrackmen)
effect

group_by(raetrackfemale, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raetrackfemale)
effect
```

```
group_by(raecxfemale, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raecxfemale)
effect

group_by(raecxmen, control) %>%
  summarise(
    count = n(),
    mean = mean(rae, na.rm = TRUE),
    sd = sd(rae, na.rm = TRUE)
  )

effect <- cohen.d(rae ~ control,data=raecxmen)
effect

### clean up and subset dataset eminent

###imported t2eminent includes age of first elite podium, first premier league podium, first premier league win and age of eminent status

##Visualize pathways for eminent group

eminentpathway <- ggplot(t2eminent, aes( age,name )) + geom_line(aes(group = name)) + geom_point(aes(color = stage)) + scale_y_discrete(expand = c(.02, 0)) + labs(title = "Pathway to Eminence",

subtitle = "Age with milestones for all eminent athletes that have competed between 2010 - 2020 in endurance cycling.", caption = "Based on data from the UCI 2010-2020 ") + xlab("Age") + theme_minimal() + theme(axis.title = element_blank(),
```

```
panel.grid.major.x = element_blank(),
panel.grid.minor = element_blank(),
legend.title = element_blank(),
legend.justification = c(0, 1),
legend.position = c(.1, 1.075),
legend.background = element_blank(),
legend.direction="horizontal",
text = element_text(family = "Georgia"),
plot.title = element_text(size = 20, margin = margin(b = 10)),
plot.subtitle = element_text(size = 10, color = "darkslategrey", margin = margin(b = 25)),
plot.caption = element_text(size = 8, margin = margin(t = 10), color = "grey70", hjust = 0))

### visualize age and age ranges per sport and sex eminent group
vars<-t2eminent[c("sport", "sex","elp","plp","emerging","eminent")]
names(vars)[3] <- "first elite podium"
names(vars)[4] <- "first premier league podium"
names(vars)[5] <- "first premier league win"

###summary tables
stat.desc(vars)
eminentmale <-subset(vars, sex=="M")
eminentfemale <- subset(vars, sex=="F")
stat.desc(eminentmale)
stat.desc(eminentfemale)

eminentmtb <-subset(vars, sport=="Mountain-bike")
eminentmtbmale <-subset(eminentmtb, sex=="M")
eminentmtbfemale <-subset(eminentmtb, sex=="F")
stat.desc(eminentmtb)
stat.desc(eminentmtbfemale)
stat.desc(eminentmtbmale)
```

```
eminentrd <-subset(vars, sport=="Road")
eminentrdmale <-subset(eminentrd, sex=="M")
eminentrdfemale <-subset(eminentrd, sex=="F")
stat.desc(eminentrd)
stat.desc(eminentrdfemale)
stat.desc(eminentrdmale)


eminenttr <-subset(vars, sport=="Track")
eminenttrmale <-subset(eminenttr, sex=="M")
eminenttrfemale <-subset(eminenttr, sex=="F")
stat.desc(eminenttr)
stat.desc(eminenttrfemale)
stat.desc(eminenttrmale)

eminentcx <-subset(vars, sport=="Cyclo-cross")
eminentcxmale <-subset(eminentcx, sex=="M")
eminentcxfemale <-subset(eminentcx, sex=="F")
stat.desc(eminentcx)
stat.desc(eminentcxfemale)
stat.desc(eminentcxmale)


tars<-emerandemi[c("sport", "sex","elp","plp","emerging","eminent")]
names(tars)[3] <- "first elite podium"
names(tars)[4] <- "first premier league podium"
names(tars)[5] <- "first premier league win"
stat.desc(tars)

allmale <-subset(tars, sex=="M")
allfemale <- subset(tars, sex=="F")
```

```
stat.desc(allmale)
stat.desc(allfemale)

allmtb <-subset(tars, sport=="Mountain-bike")
allmtbmale <-subset(allmtb, sex=="M")
allmtbfemale <-subset(allmtb, sex=="F")
stat.desc(allmtb)
stat.desc(allmtbfemale)
stat.desc(allmtbmale)


allrd <-subset(tars, sport=="Road")
allrdmale <-subset(allrd, sex=="M")
allrdfemale <-subset(allrd, sex=="F")
stat.desc(allrd)
stat.desc(allrdfemale)
stat.desc(allrdmale)


alltr <-subset(tars, sport=="Track")
alltrmale <-subset(alltr, sex=="M")
alltrfemale <-subset(alltr, sex=="F")
stat.desc(alltr)
stat.desc(alltrfemale)
stat.desc(alltrmale)

allcx <-subset(tars, sport=="Cyclo-cross")
allcxmale <-subset(allcx, sex=="M")
allcxfemale <-subset(allcx, sex=="F")
stat.desc(allcx)
stat.desc(allcxfemale)
stat.desc(allcxmale)
```

```
ttars<-emergingclean[c("sport", "sex","elp","plp","emerging")]
names(ttars)[3] <- "first elite podium"
names(ttars)[4] <- "first premier league podium"
names(ttars)[5] <- "first premier league win"
stat.desc(ttars)

tallmale <-subset(ttars, sex=="M")
tallfemale <- subset(ttars, sex=="F")
stat.desc(tallmale)
stat.desc(tallfemale)

tallmtb <-subset(ttars, sport=="Mountain-bike")
tallmtbm <-subset(tallmtb, sex=="M")
tallmtbfemale <-subset(tallmtb, sex=="F")
stat.desc(tallmtb)
stat.desc(tallmtbfemale)
stat.desc(tallmtbm)



tallrd <-subset(ttars, sport=="Road")
tallrdmale <-subset(tallrd, sex=="M")
tallrdfemale <-subset(tallrd, sex=="F")
stat.desc(tallrd)
stat.desc(tallrdfemale)
stat.desc(tallrdmale)



talltr  <-subset(ttars,  sport=="Track")
talltrmale  <-subset(talltr,  sex=="M")
talltrfemale <-subset(talltr, sex=="F")
```

```
stat.desc(talltr)
stat.desc(talltrfemale)
stat.desc(talltrmale)

tallcx <-subset(ttars, sport=="Cyclo-cross")
tallcxmale <-subset(tallcx, sex=="M")
tallcxfemale <-subset(tallcx, sex=="F")
stat.desc(tallcx)
stat.desc(tallcxfemale)
stat.desc(tallcxmale)

### analysis of variance for milestones between emerging eminence and eminence

##cleaned up dataset aovmilestones

aovfemale<-subset(aov_milestones, sex=="F")
#general
group_by(aov_milestones, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aov_milestones)
effect

group_by(aov_milestones, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
```

```
effect <- cohen.d(plp ~ stage,data=aov_milestones)
effect

group_by(aov_milestones, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aov_milestones)
effect

##BySex
#Male
aovmale<-subset(aov_milestones, sex=="M")

group_by(aovmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovmale)
effect

group_by(aovmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovmale)
```

```
effect

group_by(aovmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovmale)
effect

#Female
aovfemale<-subset(aov_milestones, sex=="F")

group_by(aovfemale, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovfemale)
effect

group_by(aovfemale, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovfemale)
effect
```

```
group_by(aovfemale, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovfemale)
effect

####ROAD

#Male
aovrd<-subset(aov_milestones, sport=="Road")
aovrdmale<-subset(aovrd, sex=="M")

group_by(aovrdmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovrdmale)
effect

group_by(aovrdmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovrdmale)
effect
```

```
group_by(aovrdmale, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovrdmale)
effect

#Female
aovfemale<-subset(aovfemale, sport=="Road")

group_by(aovfemale, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovfemale)
effect

group_by(aovfemale, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovfemale)
effect

group_by(aovfemale, stage) %>%
```

```
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovfemale)
effect

####CX

#Male
aovcx<-subset(aov_milestones, sport=="Cyclo-cross")
aovcxm <-subset(aovcx, sex=="M")

group_by(aovcxm, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovcxm)
effect

group_by(aovcxm, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovcxm)
effect
```

```
group_by(aovcxm, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovcxm)
effect

#Female
aovcxf<-subset(aovtr, sex=="F")

group_by(aovcxf, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovcxf)
effect

group_by(aovcxf, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovcxf)
effect

group_by(aovcxf, stage) %>%
  summarise(
```

```
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovcxf)
effect


### MTB

#Male
aovmtb<-subset(aov_milestones, sport=="Mountain-bike")
aovmtbm <-subset(aovmtb, sex=="M")

group_by(aovmtbm, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovmtbm)
effect

group_by(aovmtbm, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
plpmilestones.aov <- aov(plp ~ stage, data = aovmtbm)
summary(plpmilestones.aov)
aov_residuals <- residuals(object = plpmilestones.aov )
```

```
shapiro.test(x = aov_residuals )

group_by(aovmtbm, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
plwmilestones.aov <- aov(emerging ~ stage, data = aovmtbm)
summary(plwmilestones.aov)
aov_residuals2 <- residuals(object = plwmilestones.aov )
shapiro.test(x = aov_residuals2 )

#Female
aovmtbf<-subset(aovmtb, sex=="F")

group_by(aovmtbf, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovmtbf)
effect

group_by(aovmtbf, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovmtbf)
```

```
effect

group_by(aovmtbf, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovmtbf)
effect

####Track

#Male
aovtr<-subset(aov_milestones, sport=="Track")
aovtrm <-subset(aovtr, sex=="M")

group_by(aovtrm, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovtrm)
effect

group_by(aovtrm, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
```

```
effect <- cohen.d(plp ~ stage,data=aovtrm)
effect

group_by(aovtrm, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovtrm)
effect

#Female
aovtrf<-subset(aovtr, sex=="F")

group_by(aovtrf, stage) %>%
  summarise(
    count = n(),
    mean = mean(elp, na.rm = TRUE),
    sd = sd(elp, na.rm = TRUE)
  )
effect <- cohen.d(elp ~ stage,data=aovtrf)
effect

group_by(aovtrf, stage) %>%
  summarise(
    count = n(),
    mean = mean(plp, na.rm = TRUE),
    sd = sd(plp, na.rm = TRUE)
  )
effect <- cohen.d(plp ~ stage,data=aovtrf)
effect
```

```
group_by(aovtrf, stage) %>%
  summarise(
    count = n(),
    mean = mean(emerging, na.rm = TRUE),
    sd = sd(emerging, na.rm = TRUE)
  )
effect <- cohen.d(emerging ~ stage,data=aovtrf)
effect

###full result dataframe for eminent category

eminent <- rbind(absalon,ankudinof, anrchibald, armstrong, arndt, barker, blaak,
bobridge,boonen, bresset, bronzini, cancellara, cant, catlin, cavendish, clancy, contador,
deignan, dennis, dygert, edmondson, edmondsonannette, evans, freire, froome, ganna, gilbert,
hammer, hansen, hepburn, kenny, kluge, kneisky, kulhavy, martin, meyer, neff, nibali,
pendrel, reinhardt, rowe, rowsell, sagan, schurter, stybar, thomas, valverde,
vanaert,vanvleuten, vdbreggen, vdpoel, vos, welsford, whitten, wiggins, wild)

###population percentiles
rd <-subset(result,sport=="Road")
rd <-subset(rd,level=="Elite")
rdw <-subset(rd,sex=="F")
rdm <-subset(rd,sex=="M")
rdw <-rdw[c("name")]
rdm <-rdm[c("name")]
n_distinct(rdw, na.rm = TRUE)
n_distinct(rdm, na.rm = TRUE)

mtb <-subset(result,sport=="Mountain-bike")
mtb <-subset(mtb,discipline=="country Olympic")
mtb <-subset(mtb,level=="Elite")
mtbw <-subset(mtb,sex=="F")
mtbm <-subset(mtb,sex=="M")
```

```
mtbw <-mtbw[c("name")]
mtbm <-mtbm[c("name")]
n_distinct(mtbm, na.rm = TRUE)
n_distinct(mtbw, na.rm = TRUE)

tr <-subset(result,sport=="Track")
tro <-subset(tr,discipline=="Omnium")
trtp <-subset(tr,discipline=="Team Pursuit")
trip <-subset(tr,discipline=="Individual Pursuit")
trmad <-subset(tr,discipline=="Madison")
trs <-subset(tr,discipline=="Scratch")
tr <-rbind(tro,trtp,trip,trmad,trs)
tr <-subset(tr,level=="Elite")
trw <-subset(tr,sex=="F")
trm <-subset(tr,sex=="M")
trw <-trw[c("name")]
trm <-trm[c("name")]
n_distinct(trm, na.rm = TRUE)
n_distinct(trw, na.rm = TRUE)

cx <-subset(result,sport=="Cyclo-cross")
cx <-subset(cx,level=="Elite")
cxw <-subset(cx,sex=="F")
cxm <-subset(cx,sex=="M")
cxw <-cxw[c("name")]
cxm <-cxm[c("name")]
n_distinct(cxm, na.rm = TRUE)
n_distinct(cxw, na.rm = TRUE)


#####
library(dplyr)
```

```
library(cluster)
library(vcd)
library(pivottabler)
library(stats)
library(ISLR)
library(tidyr)
library(elo)
library(readxl)
library(readr)
library(data.table)
library(Hmisc)
library(purrr)

############# 1.ELO #############

#############################################################################
###################### execute result##################################
#data result"
        rm()
        gc()

        ############# library #############
        library(dplyr)
        library(cluster)
        library(vcd)
        library(pivottabler)
        library(ineq)
        library(ggplot2)
        library(gglorenz)
        library(stats)
        library(ISLR)
        library(tidyr)
```

```
library(elo)
library(readxl)
library(data.table)
library(Hmisc)
library(purrr)



############# 1.ELO #############

#data result
colnames(result)<-
c("X1","sport","discipline","competition","location","level","year","cat",

"placement","bib","name","lastname","firstname","country","team","sex",
                "age","phase","heat","time","norace","points")

data<-data.table(result)

data1<-data[,.(year,sport,discipline,competition,placement,name,sex,X1)]


data2<-data1[,new_discipline:=fifelse(sport=="Road" & (discipline == "Stage 8b : Individual Time Trial" | discipline == "Team Time Trial" | discipline == "Individual Time Trial" | discipline == "individual time trial"),"Time Trial",
                          fifelse(sport=="Mountain-bike" & discipline=="country Olympic","Country Olympic",
                            fifelse(sport=="Track" & discipline=="Madison","Madison",
                              fifelse(sport=="Track" &


discipline=="Individual Pursuit","Individual Pursuit",
                                fifelse(sport=="Track" & discipline=="Team Pursuit","Team Pursuit",
                                  fifelse(sport=="Track" & discipline=="Omnium","Omnium",
```

```
                                                  fifelse(sport=="Track" &
discipline=="Points Race","Points Race",
                                                     fifelse(sport=="Track" &
discipline=="Scratch","Scratch",
                                                        fifelse(sport=="Cyclo-
cross","",

fifelse(sport=="Bmx-racing",""

,"Other"))))))))))]



        data2<-data2[order(year,sport,discipline,competition,new_discipline,sex)]



        #create new variables


        data2<-data2[,points.Home :=1/placement]
        data2<-data2[,points.Visitor :=rep(0,nrow(data2))]
        data2<-data2[,away.team :=rep('No Aplica',nrow(data2))]

        data3<-data2[, group_c := paste0(year, '_', sport,'_',new_discipline,'-',sex)]

        group<-data.table(table(data3$group_c))
        group<-group[,order:=c(seq(1,nrow(group)))]

        data4 <- merge(data3, group,
                by.x = "group_c", by.y = "V1",
                all.x = TRUE, all.y = FALSE)


        table5<-
data.table(X1=numeric(0),group_c=character(0),order=numeric(0),final.elos.e1.=numeric(0))


        n=max(data4$order)
        for (i in 1:n){
```

```
  table1<-data4[order==i]
  names(table1)
  e1<-elo.run(score(points.Home, points.Visitor) ~ name + away.team, data =
table1, k = 1)
  teams<-data.frame(e1$teams)
  rating<-data.frame(final.elos(e1))
  table2<-cbind(teams,rating)
  table3<-merge(table1, table2,
        by.x = "name", by.y = "e1.teams",
        all.x = TRUE, all.y = FALSE)
  table4<-table3[,.(X1,group_c,order,final.elos.e1.)]
  table5<-rbind(table5,table4)

}


result_final<-merge(data, table5,
          by.x = "X1", by.y = "X1",
          all.x = TRUE, all.y = FALSE)



result_final

rm(data,data1,data2,data3,data4,e1,group,table1,table2,table3,table4,table5,rating,teams,i,n)
names(result_final)

############# 2.count #############

eminent_count_by_year<-eminent%>%
  group_by(name,sport,discipline,competition,year)%>%
  arrange(name,year)%>%
```

```
  group_by(name,year)%>%
  mutate( racedays = row_number())

eminent_count_by_age<-eminent%>%
  group_by(name,sport,discipline,competition,age)%>%
  arrange(name,year)%>%
  group_by(name,year)%>%
  mutate( racedays = row_number())



emerging_count_by_year<-emerging%>%
  group_by(name,sport,discipline,competition,year)%>%
  arrange(name,year)%>%
  group_by(name,year)%>%
  mutate( racedays = row_number())

emerging_count_by_age<-emerging%>%
  group_by(name,sport,discipline,competition,age)%>%
  arrange(name,year)%>%
  group_by(name,year)%>%
  mutate( racedays = row_number())

result_count_by_year<-result%>%
  group_by(name,sport,discipline,competition,year)%>%
  arrange(name,year)%>%
  group_by(name,year)%>%
  mutate( racedays = row_number())

result_count_by_age<-result%>%
  group_by(name,sport,discipline,competition,age)%>%
  arrange(name,year)%>%
  group_by(name,year)%>%
```

```
mutate( racedays = row_number())

############# 3.correlation #############
############# 3.1.correlation eminent#############

eminent$discipline<-as.character(eminent$discipline)
eminent$sport<-as.character(eminent$sport)
eminent$new_discipline<-ifelse((eminent$sport=="Road" &
eminent$discipline=="Individual Time Trial"),"Individual Time trial",
                              ifelse((eminent$sport=="Road" &
eminent$discipline!="Individual Time trial"),"Others",
                                 ifelse((eminent$sport=="Mountain-bike" &
eminent$discipline=="country Olympic"),"country Olympic",
                                    ifelse((eminent$sport=="Mountain-bike" &
eminent$discipline %in% c("country marathon","country Marathon","country marathon
Series")),"marathon",
                                       ifelse(eminent$sport=="Cyclo-cross"
,eminent$discipline,
                                          ifelse((eminent$sport=="Track"  &
eminent$discipline %in% c("Omnium","Individual Pursuit","Team
Pursuit","Scratch","Madison","Points Race")),eminent$discipline,NA))))))

eminent2<-
unique(eminent[which(!is.na(eminent$new_discipline)),c("sport","new_discipline","name")])

eminent2$cat<-ifelse((eminent2$sport=="Road" &
eminent2$new_discipline=="Individual Time trial"),1,
                    ifelse((eminent2$sport=="Road" &
eminent2$new_discipline=="Others"),2,
                       ifelse((eminent2$sport=="Mountain-bike" &
eminent2$new_discipline=="country Olympic"),3,
                          ifelse((eminent2$sport=="Mountain-bike" &
eminent2$new_discipline=="marathon"),4,
                             ifelse((eminent2$sport=="Track" &
eminent2$new_discipline=="Omnium"),5,
                                ifelse((eminent2$sport=="Track" &
eminent2$new_discipline=="Individual Pursuit"),6,
```

```
                                        ifelse((eminent2$sport=="Track" &
eminent2$new_discipline=="Team Pursuit"),7,
                                           ifelse((eminent2$sport=="Track" &
eminent2$new_discipline=="Scratch"),8,
                                              ifelse((eminent2$sport=="Track" &
eminent2$new_discipline=="Madison"),9,
                                                 ifelse((eminent2$sport=="Track"
& eminent2$new_discipline=="Points Race"),10,

ifelse((eminent2$sport=="Cyclo-cross" & eminent2$new_discipline=="Men Elite"),11,

ifelse((eminent2$sport=="Cyclo-cross" & eminent2$new_discipline=="Women
Elite"),12))))))))))))

        eminent2<-eminent2[,c("name","cat")]
        cat<-as.data.frame.vector(table(eminent2$cat))
        colnames(cat)<-c("count_cat")
        cat$cat<-paste0("cat",rownames(cat))
        eminent2<-as.data.frame.matrix(table(eminent2$name,eminent2$cat))
        colnames(eminent2)<-
c("cat1","cat2","cat3","cat4","cat5","cat6","cat7","cat8","cat9","cat10","cat11","cat12")

       eminent3<-combn(colnames(eminent2), 2) %>%
         t() %>%
         as_tibble() %>%
         mutate(Count = map2_int(V1, V2, ~sum(eminent2[[.x]] & eminent2[[.y]])))

        eminent3<-merge(eminent3,cat,by.x="V1",by.y="cat")
        colnames(eminent3)<-c("V1","V2","Count","count_cat1")
        eminent3<-merge(eminent3,cat,by.x="V2",by.y="cat")
```

```
colnames(eminent3)<-c("V1","V2","Count","count_cat1","count_cat2")
eminent3$max<-apply(eminent3[,c("count_cat1","count_cat2")], 1, max)
eminent3$perc<-round(eminent3$Count/eminent3$max,2)

eminent3$V1<-ifelse(eminent3$V1=="cat1","Road/Individual Time trial",
        ifelse(eminent3$V1=="cat2","Road/Others",
          ifelse(eminent3$V1=="cat3","Mountain-bike/country Olympic",
            ifelse(eminent3$V1=="cat4","Mountain-bike/marathon",
              ifelse(eminent3$V1=="cat5","Track/Omnium",
                ifelse(eminent3$V1=="cat6","Track/Individual
Pursuit",
                  ifelse(eminent3$V1=="cat7","Track/Team
Pursuit",

ifelse(eminent3$V1=="cat8","Track/Scratch",

ifelse(eminent3$V1=="cat9","Track/Madison",

ifelse(eminent3$V1=="cat10","Track/Points Race",

ifelse(eminent3$V1=="cat11","Cyclo-cross/Men Elite",

ifelse(eminent3$V1=="cat12","Cyclo-cross/Women Elite","NA")))))))))))


eminent3$V2<-ifelse(eminent3$V2=="cat1","Road/Individual Time trial",
        ifelse(eminent3$V2=="cat2","Road/Others",
          ifelse(eminent3$V2=="cat3","Mountain-bike/country Olympic",
            ifelse(eminent3$V2=="cat4","Mountain-bike/marathon",
              ifelse(eminent3$V2=="cat5","Track/Omnium",
                ifelse(eminent3$V2=="cat6","Track/Individual
Pursuit",
                  ifelse(eminent3$V2=="cat7","Track/Team
Pursuit",

ifelse(eminent3$V2=="cat8","Track/Scratch",
```

```
ifelse(eminent3$V2=="cat9","Track/Madison",

ifelse(eminent3$V2=="cat10","Track/Points Race",

ifelse(eminent3$V2=="cat11","Cyclo-cross/Men Elite",

ifelse(eminent3$V2=="cat12","Cyclo-cross/Women Elite","NA"))))))))))))))

        pt_eminent <- PivotTable$new()
        pt_eminent$addData(eminent3)
        pt_eminent$addColumnDataGroups("V1")
        pt_eminent$addRowDataGroups("V2")
        pt_eminent$defineCalculation(calculationName="perc",
summariseExpression="max(perc)")
        pt_eminent$renderPivot()
        rm(eminent2,eminent3)
        pt_eminent


        ############# 3.2.correlation emerging#############

        emerging$discipline<-as.character(emerging$discipline)
        emerging$sport<-as.character(emerging$sport)
        emerging$new_discipline<-ifelse((emerging$sport=="Road" &
emerging$discipline=="Individual Time Trial"),"Individual Time trial",
                        ifelse((emerging$sport=="Road" &
emerging$discipline!="Individual Time trial"),"Others",
                            ifelse((emerging$sport=="Mountain-bike" &
emerging$discipline=="country Olympic"),"country Olympic",
                                ifelse((emerging$sport=="Mountain-bike" &
emerging$discipline %in% c("country marathon","country Marathon","country marathon
Series")),"marathon",
                                    ifelse(emerging$sport=="Cyclo-cross"
,emerging$discipline,
```

```
                                        ifelse((emerging$sport=="Track" &
emerging$discipline %in% c("Omnium","Individual Pursuit","Team
Pursuit","Scratch","Madison","Points Race")),emerging$discipline,NA))))))


       emerging2<-
unique(emerging[which(!is.na(emerging$new_discipline)),c("sport","new_discipline","name
")])


       emerging2$cat<-ifelse((emerging2$sport=="Road" &
emerging2$new_discipline=="Individual Time trial"),1,
                ifelse((emerging2$sport=="Road" &
emerging2$new_discipline=="Others"),2,
                   ifelse((emerging2$sport=="Mountain-bike" &
emerging2$new_discipline=="country Olympic"),3,
                      ifelse((emerging2$sport=="Mountain-bike" &
emerging2$new_discipline=="marathon"),4,
                         ifelse((emerging2$sport=="Track" &
emerging2$new_discipline=="Omnium"),5,
                            ifelse((emerging2$sport=="Track" &
emerging2$new_discipline=="Individual Pursuit"),6,
                               ifelse((emerging2$sport=="Track" &
emerging2$new_discipline=="Team Pursuit"),7,
                                  ifelse((emerging2$sport=="Track" &
emerging2$new_discipline=="Scratch"),8,
                                     ifelse((emerging2$sport=="Track" &
emerging2$new_discipline=="Madison"),9,
                                        ifelse((emerging2$sport=="Track"


& emerging2$new_discipline=="Points Race"),10,


ifelse((emerging2$sport=="Cyclo-cross" & emerging2$new_discipline=="Men Elite"),11,


ifelse((emerging2$sport=="Cyclo-cross" & emerging2$new_discipline=="Men Junior"),12,


ifelse((emerging2$sport=="Cyclo-cross" & emerging2$new_discipline=="Men Under
23"),13,


ifelse((emerging2$sport=="Cyclo-cross" & emerging2$new_discipline=="Women Under
23"),14,
```

```
ifelse((emerging2$sport=="Cyclo-cross" & emerging2$new_discipline=="Women
Elite"),15,16)))))))))))))))

emerging2<-emerging2[which(emerging2$cat<16),c("name","cat")]
cat<-as.data.frame.vector(table(emerging2$cat))
colnames(cat)<-c("count_cat")
cat$cat<-paste0("cat",rownames(cat))
emerging2<-as.data.frame.matrix(table(emerging2$name,emerging2$cat))
colnames(emerging2)<-
c("cat1","cat2","cat3","cat4","cat5","cat6","cat7","cat8","cat9","cat10","cat11","cat12","cat13
","cat14","cat15")

emerging3<-combn(colnames(emerging2), 2) %>%
  t() %>%
  as_tibble() %>%
  mutate(Count = map2_int(V1, V2, ~sum(emerging2[[.x]] & emerging2[[.y]])))

emerging3<-merge(emerging3,cat,by.x="V1",by.y="cat")
colnames(emerging3)<-c("V1","V2","Count","count_cat1")
emerging3<-merge(emerging3,cat,by.x="V2",by.y="cat")
colnames(emerging3)<-c("V1","V2","Count","count_cat1","count_cat2")
emerging3$max<-apply(emerging3[,c("count_cat1","count_cat2")], 1, max)
emerging3$perc<-round(emerging3$Count/emerging3$max,2)

emerging3$V1<-ifelse(emerging3$V1=="cat1","Road/Individual Time trial",
           ifelse(emerging3$V1=="cat2","Road/Others",
              ifelse(emerging3$V1=="cat3","Mountain-bike/country
Olympic",
```

```
                                ifelse(emerging3$V1=="cat4","Mountain-bike/marathon",
                                    ifelse(emerging3$V1=="cat5","Track/Omnium",
                                        ifelse(emerging3$V1=="cat6","Track/Individual
Pursuit",
                                            ifelse(emerging3$V1=="cat7","Track/Team
Pursuit",


ifelse(emerging3$V1=="cat8","Track/Scratch",

ifelse(emerging3$V1=="cat9","Track/Madison",

ifelse(emerging3$V1=="cat10","Track/Points Race",

ifelse(emerging3$V1=="cat11","Cyclo-cross/Men Elite",

ifelse(emerging3$V1=="cat12","Cyclo-cross/Men Junior",

ifelse(emerging3$V1=="cat13","Cyclo-cross/Men Under 23",

ifelse(emerging3$V1=="cat14","Cyclo-cross/Women Under 23",

ifelse(emerging3$V1=="cat15","Cyclo-cross/Women Elite","NA")))))))))))))))


          emerging3$V2<-ifelse(emerging3$V2=="cat1","Road/Individual Time trial",
                        ifelse(emerging3$V2=="cat2","Road/Others",
                            ifelse(emerging3$V2=="cat3","Mountain-bike/country
Olympic",
                                ifelse(emerging3$V2=="cat4","Mountain-bike/marathon",
                                    ifelse(emerging3$V2=="cat5","Track/Omnium",
                                        ifelse(emerging3$V2=="cat6","Track/Individual
Pursuit",
                                            ifelse(emerging3$V2=="cat7","Track/Team
Pursuit",


ifelse(emerging3$V2=="cat8","Track/Scratch",
```

```
ifelse(emerging3$V2=="cat9","Track/Madison",

ifelse(emerging3$V2=="cat10","Track/Points Race",

ifelse(emerging3$V2=="cat11","Cyclo-cross/Men Elite",

ifelse(emerging3$V2=="cat12","Cyclo-cross/Men Junior",

ifelse(emerging3$V2=="cat13","Cyclo-cross/Men Under 23",

ifelse(emerging3$V2=="cat14","Cyclo-cross/Women Under 23",

ifelse(emerging3$V2=="cat15","Cyclo-cross/Women Elite","NA")))))))))))))))

        pt_emerging <- PivotTable$new()
        pt_emerging$addData(emerging3)
        pt_emerging$addColumnDataGroups("V1")
        pt_emerging$addRowDataGroups("V2")
        pt_emerging$defineCalculation(calculationName="perc",
summariseExpression="max(perc)")
        pt_emerging$renderPivot()
        rm(emerging2,emerging3)
        pt_emerging

        ############# 3.2.correlation total#############

        result$discipline<-as.character(result$discipline)
        result$sport<-as.character(result$sport)
```

```
result$new_discipline<-ifelse((result$sport=="Road" & result$discipline=="Individual Time Trial"),"Individual Time trial",
ifelse((result$sport=="Road" & result$discipline!="Individual Time trial"),"Others",
ifelse((result$sport=="Mountain-bike" & result$discipline=="country Olympic"),"country Olympic",
ifelse((result$sport=="Mountain-bike" & result$discipline %in% c("country marathon","country Marathon","country marathon Series")),"marathon",
ifelse(result$sport=="Cyclo-cross" ,result$discipline,
ifelse((result$sport=="Track"  & result$discipline %in% c("Omnium","Individual Pursuit","Team Pursuit","Scratch","Madison","Points Race")),result$discipline,NA))))))


result2<-unique(result[which(!is.na(result$new_discipline)),c("sport","new_discipline","name")])


result2$cat<-ifelse((result2$sport=="Road" & result2$new_discipline=="Individual Time trial"),1,
ifelse((result2$sport=="Road" & result2$new_discipline=="Others"),2,
ifelse((result2$sport=="Mountain-bike" & result2$new_discipline=="country Olympic"),3,
ifelse((result2$sport=="Mountain-bike" & result2$new_discipline=="marathon"),4,
ifelse((result2$sport=="Track" & result2$new_discipline=="Omnium"),5,
ifelse((result2$sport=="Track" & result2$new_discipline=="Individual Pursuit"),6,
ifelse((result2$sport=="Track"  & result2$new_discipline=="Team Pursuit"),7,
ifelse((result2$sport=="Track" & result2$new_discipline=="Scratch"),8,
ifelse((result2$sport=="Track" & result2$new_discipline=="Madison"),9,
ifelse((result2$sport=="Track"  & result2$new_discipline=="Points Race"),10,
```

```
ifelse((result2$sport=="Cyclo-cross" & result2$new_discipline=="Men Elite"),11,

ifelse((result2$sport=="Cyclo-cross" & result2$new_discipline=="Men Junior"),12,

ifelse((result2$sport=="Cyclo-cross" & result2$new_discipline=="Men Under 23"),13,

ifelse((result2$sport=="Cyclo-cross" & result2$new_discipline=="Women Under 23"),14,

ifelse((result2$sport=="Cyclo-cross" & result2$new_discipline=="Women Elite"),15,16)))))))))))))))
```

```
result2<-result2[which(result2$cat<16),c("name","cat")]
cat<-as.data.frame.vector(table(result2$cat))
colnames(cat)<-c("count_cat")
cat$cat<-paste0("cat",rownames(cat))
result2<-as.data.frame.matrix(table(result2$name,result2$cat))
colnames(result2)<-c("cat1","cat2","cat3","cat4","cat5","cat6","cat7","cat8","cat9","cat10","cat11","cat12","cat13","cat14","cat15")

result3<-combn(colnames(result2), 2) %>%
  t() %>%
  as_tibble() %>%
  mutate(Count = map2_int(V1, V2, ~sum(result2[[.x]] & result2[[.y]])))

result3<-merge(result3,cat,by.x="V1",by.y="cat")
colnames(result3)<-c("V1","V2","Count","count_cat1")
result3<-merge(result3,cat,by.x="V2",by.y="cat")
colnames(result3)<-c("V1","V2","Count","count_cat1","count_cat2")
```

```
result3$max<-apply(result3[,c("count_cat1","count_cat2")], 1, max)
result3$perc<-round(result3$Count/result3$max,2)

result3$V1<-ifelse(result3$V1=="cat1","Road/Individual Time trial",
        ifelse(result3$V1=="cat2","Road/Others",
            ifelse(result3$V1=="cat3","Mountain-bike/country Olympic",
                ifelse(result3$V1=="cat4","Mountain-bike/marathon",
                    ifelse(result3$V1=="cat5","Track/Omnium",
                        ifelse(result3$V1=="cat6","Track/Individual
Pursuit",
                            ifelse(result3$V1=="cat7","Track/Team Pursuit",
                                ifelse(result3$V1=="cat8","Track/Scratch",

ifelse(result3$V1=="cat9","Track/Madison",

ifelse(result3$V1=="cat10","Track/Points Race",

ifelse(result3$V1=="cat11","Cyclo-cross/Men Elite",

ifelse(result3$V1=="cat12","Cyclo-cross/Men Junior",

ifelse(result3$V1=="cat13","Cyclo-cross/Men Under 23",

ifelse(result3$V1=="cat14","Cyclo-cross/Women Under 23",

ifelse(result3$V1=="cat15","Cyclo-cross/Women Elite","NA")))))))))))))))


result3$V2<-ifelse(result3$V2=="cat1","Road/Individual Time trial",
        ifelse(result3$V2=="cat2","Road/Others",
            ifelse(result3$V2=="cat3","Mountain-bike/country Olympic",
                ifelse(result3$V2=="cat4","Mountain-bike/marathon",
                    ifelse(result3$V2=="cat5","Track/Omnium",
                        ifelse(result3$V2=="cat6","Track/Individual
Pursuit",
```

```
                                                    ifelse(result3$V2=="cat7","Track/Team Pursuit",
                                                        ifelse(result3$V2=="cat8","Track/Scratch",

ifelse(result3$V2=="cat9","Track/Madison",

ifelse(result3$V2=="cat10","Track/Points Race",

ifelse(result3$V2=="cat11","Cyclo-cross/Men Elite",

ifelse(result3$V2=="cat12","Cyclo-cross/Men Junior",

ifelse(result3$V2=="cat13","Cyclo-cross/Men Under 23",

ifelse(result3$V2=="cat14","Cyclo-cross/Women Under 23",

ifelse(result3$V2=="cat15","Cyclo-cross/Women Elite","NA")))))))))))))))

          pt_result <- PivotTable$new()
          pt_result$addData(result3)
          pt_result$addColumnDataGroups("V1")
          pt_result$addRowDataGroups("V2")
          pt_result$defineCalculation(calculationName="perc",
summariseExpression="max(perc)")
          pt_result$renderPivot()
          rm(result2,result3)
          pt_result



          ############# 4.lorenz curves #############
          ############# 4.1.lorenz Curves Road#############


          result$flag<-ifelse(
            (result$sport=="Road" & result$sex=="M" & (result$competition %in% c(
              "Vuelta a España","World Championships",
              "Giro d'Italia","Giro Ciclistico d'Italia",
```

```
"Tour de France",

"Milano-Sanremo","Milano - Sanremo",

"Tour des Flandres Espoirs","Ronde van Vlaanderen - Tour des Flandres","Ronde van Vlaanderen / Tour des Flandres",

"Paris - Roubaix",

"Il Lombardia", "Championnats du Monde UCI CLM/ UCI TT World Championships","Championnats du Monde UCI CL / UCI RR World Championships",

"Liège-Bastogne-Liège","Liège - Bastogne - Liège","Championnats du Monde Route UCI / UCI Road World Championships","Championnats du Monde UCI / UCI World Championships",

"Championnats du Monde UCI CL / UCI RR World Championships (Men Elite)","Championnats du Monde UCI CLM / UCI TT World Championships (Men Elite)",

"Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR ME","Jeux Olympiques de la Jeunesse / Youth Olympic Games"))

 | (result$cat %in% c("1.UWT","CM","JO","2.UWT"))  ),

1,

ifelse((result$sport=="Road" & result$sex=="F" &

        (result$competition %in% c("Giro d'Italia Internazionale Femminile","Ronde van Vlaanderen - Tour des Flandres","Ronde van Vlaanderen / Tour des Flandres","World Championships","Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR ME","Championnats du Monde UCI CLM/ UCI TT World Championships","Championnats du Monde UCI CL / UCI RR World Championships","Championnats du Monde UCI CL / UCI RR World Championships (Men Elite)","Championnats du Monde UCI CLM / UCI TT World Championships (Women Elite)"))

    | (result$cat %in% c("1.WWT","CM","JO","2.WWT","CDM"))),

    1,0

))


sex_f<-result[which(result$flag==1 & result$sex=="F"),]

sex_f<-sex_f[which( is.na(sex_f$placement)==F),]

sex_f$placement2<-ifelse(sex_f$placement==1,1,ifelse(is.na(sex_f$placement),0,0))

sex_f$cum_placement<-cumsum(sex_f$placement2)
```

```
sex_m<-result[which(result$flag==1 & result$sex=="M"),]
sex_m<-sex_m[which(is.na(sex_m$placement)==F),]
sex_m$placement2<-
ifelse(sex_m$placement==1,1,ifelse(is.na(sex_m$placement),0,0))
sex_m$cum_placement<-cumsum(sex_m$placement2)

data<-
rbind(sex_f[,c("sex","cum_placement")],sex_m[,c("sex","cum_placement")])

gini_m<-round(ineq(sex_m$cum_placement,type="Gini"),3)
gini_f<-round(ineq(sex_f$cum_placement,type="Gini"),3)

LC_road<-data %>%
  ggplot(aes(x = cum_placement, colour = sex)) +
  stat_lorenz() +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "cum. rel. Athlete Incidence",
       y = "cum. rel. medal share",
       title = "Lorenz Curves Road") +
  annotate('text', x = 0.12, y = 0.8, label = paste("Gini M:",gini_m), size = 3)+
  annotate('text', x = 0.12, y = 0.9, label = paste("Gini F:",gini_f), size = 3)

############# 4.2.lorenz Curves Track #############
result$flag<-ifelse(
  (result$sport=="Track" & result$sex=="M" & result$cat %in% c("CM","JO") &
result$level %in% c("Elite") & result$discipline %in% c("Individual Pursuit","Team
Pursuit", "Madison", "Omnium", "Scratch")),
  1,
  ifelse((result$sport=="Track" & result$sex=="F" & result$cat %in%
c("CM","JO") & result$level %in% c("Elite") & result$discipline %in% c("Individual
Pursuit","Team Pursuit", "Madison", "Omnium", "Scratch")),
       1,0
```

```
))

sex_f<-result[which(result$flag==1 & result$sex=="F"),]
sex_f<-sex_f[which( is.na(sex_f$placement)==F),]
sex_f$placement2<-
ifelse(sex_f$placement==1,1,ifelse(is.na(sex_f$placement),0,0))
sex_f$cum_placement<-cumsum(sex_f$placement2)

sex_m<-result[which(result$flag==1 & result$sex=="M"),]
sex_m<-sex_m[which(is.na(sex_m$placement)==F),]
sex_m$placement2<-
ifelse(sex_m$placement==1,1,ifelse(is.na(sex_m$placement),0,0))
sex_m$cum_placement<-cumsum(sex_m$placement2)

data<-
rbind(sex_f[,c("sex","cum_placement")],sex_m[,c("sex","cum_placement")])

gini_m<-round(ineq(sex_m$cum_placement,type="Gini"),3)
gini_f<-round(ineq(sex_f$cum_placement,type="Gini"),3)

LC_track<-data %>%
  ggplot(aes(x = cum_placement, colour = sex)) +
  stat_lorenz() +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "cum. rel. Athlete Incidence",
       y = "cum. rel. medal share",
       title = "Lorenz Curves Track")+
  annotate('text', x = 0.12, y = 0.8, label = paste("Gini M:",gini_m), size = 3)+
  annotate('text', x = 0.12, y = 0.9, label = paste("Gini F:",gini_f), size = 3)
```

```
############# 4.3.lorenz Curves Mountain-bike #############

result$flag<-ifelse(
  (result$sport=="Mountain-bike" & result$sex=="M" & result$cat %in% c("CM","CDM","JO") & result$discipline=="country Olympic" & result$level %in% c("Elite")),
  1,
  ifelse((result$sport=="Mountain-bike" & result$sex=="F" & result$cat %in% c("CM","CDM","JO") & result$discipline=="country Olympic" & result$level %in% c("Elite")),
         1,0
  ))

sex_f<-result[which(result$flag==1 & result$sex=="F"),]
sex_f<-sex_f[which( is.na(sex_f$placement)==F),]
sex_f$placement2<-ifelse(sex_f$placement==1,1,ifelse(is.na(sex_f$placement),0,0))
sex_f$cum_placement<-cumsum(sex_f$placement2)

sex_m<-result[which(result$flag==1 & result$sex=="M"),]
sex_m<-sex_m[which(is.na(sex_m$placement)==F),]
sex_m$placement2<-ifelse(sex_m$placement==1,1,ifelse(is.na(sex_m$placement),0,0))
sex_m$cum_placement<-cumsum(sex_m$placement2)

data<-rbind(sex_f[,c("sex","cum_placement")],sex_m[,c("sex","cum_placement")])

gini_m<-round(ineq(sex_m$cum_placement,type="Gini"),3)
gini_f<-round(ineq(sex_f$cum_placement,type="Gini"),3)

LC_Mountain_bike<-data %>%
  ggplot(aes(x = cum_placement, colour = sex)) +
```

```
stat_lorenz() +
coord_fixed() +
geom_abline(linetype = "dashed") +
theme_minimal() +
labs(x = "cum. rel. Atlethe Incidence",
     y = "cum. rel. medal share",
     title = "Lorenz Curves Mountain-bike")+
annotate('text', x = 0.12, y = 0.8, label = paste("Gini M:",gini_m), size = 3)+
annotate('text', x = 0.12, y = 0.9, label = paste("Gini F:",gini_f), size = 3)


############# 4.4.lorenz Curves Cyclo-cross #############
result$flag<-ifelse(
  (result$sport=="Cyclo-cross" & result$sex=="M" & result$cat %in% c("CM") &
result$level %in% c("Elite")),
  1,
  ifelse((result$sport=="Cyclo-cross" & result$sex=="F" & result$cat %in%
c("CM") & result$level %in% c("Elite")),
         1,0
  ))


sex_f<-result[which(result$flag==1 & result$sex=="F"),]
sex_f<-sex_f[which( is.na(sex_f$placement)==F),]
sex_f$placement2<-
ifelse(sex_f$placement==1,1,ifelse(is.na(sex_f$placement),0,0))
sex_f$cum_placement<-cumsum(sex_f$placement2)


sex_m<-result[which(result$flag==1 & result$sex=="M"),]
sex_m<-sex_m[which(is.na(sex_m$placement)==F),]
sex_m$placement2<-
ifelse(sex_m$placement==1,1,ifelse(is.na(sex_m$placement),0,0))
sex_m$cum_placement<-cumsum(sex_m$placement2)


data<-
rbind(sex_f[,c("sex","cum_placement")],sex_m[,c("sex","cum_placement")])
```

```
gini_m<-round(ineq(sex_m$cum_placement,type="Gini"),3)
gini_f<-round(ineq(sex_f$cum_placement,type="Gini"),3)


LC_Cyclo_cross<-data %>%
  ggplot(aes(x = cum_placement, colour = sex)) +
  stat_lorenz() +
  coord_fixed() +
  geom_abline(linetype = "dashed") +
  theme_minimal() +
  labs(x = "cum. rel. Athlete Incidence",
       y = "cum. rel. medal share",
       title = "Lorenz Curves Cyclo-cross")+
  annotate('text', x = 0.12, y = 0.8, label = paste("Gini M:",gini_m), size = 3)+
  annotate('text', x = 0.12, y = 0.9, label = paste("Gini F:",gini_f), size = 3)



#####ELO rating career trend lines

####data
result_final$name<-as.character(result_final$name)
result_final$sex<-as.character(result_final$sex)

eminent_elo<-result_final[which(result_final$name %in% c(
    "ABSALON Julien","ANKUDINOFF Ashlee",
    "ARCHIBALD Katie","ARMSTRONG Kirstin",
    "ARNDT Judith",
    "BOONEN Tom","BOBRIDGE Jack",
    "BARKER Elinor","BRESSET Julie","BRONZINI Giorgia",
    "CANCELLARA Fabian",
```

```
"CANT Sanne", "CATLIN Kelly","CAVENDISH Mark",
"CLANCY Edward","CURE Amy","DEIGNAN Elizabeth","DENNIS Rohan",
"DYGERT Chloe","EDMONDSON Alexander",
"EDMONDSON Annette","FREIRE GOMEZ Oscar","FROOME Chris","GANNA Filippo","GILBERT Phillipe","HAMMER Sarah",
"HANSEN Lasse Norman", "HEPBURN Michael", "KENNY Laura", "KLUGE Roger", "KNEISKY Morgan", "KULHAVÝ Jaroslav",
"MARTIN Tony", "MEYER Cameron", "NEFF Jolanda","NIBALI Vincenzo", "PENDREL Catharine","REINHARDT Theo","ROWE Danielle",
"ROWSELL-SHAND Joanna","SAGAN Peter","SCHURTER Nino","ŠTYBAR Zdeněk","THOMAS Benjamin","VALENTE Jennifer","VALVERDE Alejandro",
"VAN AERT Wout","VAN DEN BROEK-BLAAK Chantal","VAN DER BREGGEN Anna","VAN DER POEL Mathieu","VAN VLEUTEN Annemiek",
"VOS Marianne","WELSFORD Sam","WHITTEN Tara","WIGGINS Bradley", "WILD Kirsten")),]


###ELO plots
eminent_elomale <-subset(eminent_elo,sex=="M")
eminent_elofemale <-subset(eminent_elo,sex=="F")



eloplotmale <-ggplot(eminent_elomale, aes(x = age, y = final.elos.e1.)) +geom_point() + geom_smooth()+facet_wrap(~sport)
labs(ylab="ELO rating", xlab="Age", title="ELO points by age", subtitle = "Elo ranking development by sport and age for eminent male athletes")
eloplotfemale <-ggplot(eminent_elofemale, aes(x = age, y = final.elos.e1.)) +geom_point() + geom_smooth()+facet_wrap(~sport)+labs(ylab="ELO rating", xlab="Age", title="ELO points by age", subtitle = "Elo ranking development by sport and age for eminent female athletes")


### UCI Race volume development


eminent_racedays<- eminent_count_by_age[which(eminent_count_by_age$name %in% c(
"ABSALON Julien","ANKUDINOFF Ashlee",
"ARCHIBALD Katie","ARMSTRONG Kirstin",
```

```
"ARNDT Judith",
"BOONEN Tom","BOBRIDGE Jack",
"BARKER Elinor","BRESSET Julie","BRONZINI Giorgia",
"CANCELLARA Fabian",
"CANT Sanne", "CATLIN Kelly","CAVENDISH Mark",
"CLANCY Edward","CURE Amy","DEIGNAN Elizabeth","DENNIS Rohan",
"DYGERT Chloe","EDMONDSON Alexander",
"EDMONDSON Annette","FREIRE GOMEZ Oscar","FROOME Chris","GANNA Filippo","GILBERT Phillipe","HAMMER Sarah",
"HANSEN Lasse Norman", "HEPBURN Michael", "KENNY Laura", "KLUGE Roger", "KNEISKY Morgan", "KULHAVÝ Jaroslav",
"MARTIN Tony", "MEYER Cameron", "NEFF Jolanda","NIBALI Vincenzo", "PENDREL Catharine","REINHARDT Theo","ROWE Danielle",
"ROWSELL-SHAND Joanna","SAGAN Peter","SCHURTER Nino","ŠTYBAR Zdeněk","THOMAS Benjamin","VALENTE Jennifer","VALVERDE Alejandro",
"VAN AERT Wout","VAN DEN BROEK-BLAAK Chantal","VAN DER BREGGEN Anna","VAN DER POEL Mathieu","VAN VLEUTEN Annemiek",
"VOS Marianne","WELSFORD Sam","WHITTEN Tara","WIGGINS Bradley", "WILD Kirsten")),]


eminent_racedays<-eminent_racedays[c("sport","discipline","competition", "name", "sex", "age", "year" )]
eminent_racedays<-unique(eminent_racedays)


###Emerging group
emerging_racedays<-emerging_count_by_age[which(emerging_count_by_age$name %in% c(
"ABBOTT Mara","ALAPHILIPPE Julien",
"ALBERT Niels","ALVARADO Ceylin del Carmen",
"ARU Fabio",
"AVILA VANEGAS Edwin Alcibiades","BAKER Georgia",
"BARBIERI Rachele",
"BASTIANELLI Marta",
"BELOMOINA Yana",
"BERNAL GOMEZ Egan Arley",
```

"BETTIOL Alberto",

"BOUDAT Thomas",

"BRENNAUER Lisa",

"BULLING Pieter",

"BURKE Steven",

"CARAPAZ Richard",

"CHAVES RUBIO Jhoan Esteban",

"CIOLEK Gerald",

"COQUARD Bryan",

"COSTA Rui",

"COURTNEY Kate",

"DAVISON Luke",

"DEGENKOLB John",

"DE JONG Thalita",

"DEMARE Arnaud",

"DHOORE Jolien",

"DIBBEN Jonathan",

"DIDERIKSEN Amalie",

"DOULL Owain",

"DRUYTS Kelly",

"DUMOULIN Tom",

"DURBRIDGE Luke",

"EMADI Kian",

"ERSHOV Artur",

"FRAME Alex",

"FREIBERG Michael",

"FUGLSANG Jakob",

"GATE Aaron",

"GAUMNITZ Stephanie",

"GAVIRIA RENDON Fernando",

"GEIST Kimberly",

"GERRANS Simon",

"GOSS Matthew",

"GOUGH Regan",

"GUARNIER Megan",

"HAYMAN Mathew",

"HAYTER Ethan",

"HEPBURN Michael",

"HERMIDA RAMOS Jose Antonio",

"HESJEDAL Ryder",

"HORNER Christopher",

"HOSKINS Melissa",

"HOUVENAGHEL Wendy",

"HUSHOVD Thor",

"IGLINSKIY Maxim",

"IRVINE Martyn",

"JEULAND Pascale",

"JOHANSEN Julius",

"JUNGELS Bob",

"KAJIHARA Yumi",

"KARALIOK Yauheni",

"KIRYIENKA Vasil",

"KOPECKY Lotte",

"KOVALEV Ivan",

"KRISTOFF Alexander",

"KÜNG Stefan",

"KWIATKOWSKI Michal",

"KWOK Ho Ting",

"LANGVAD Annika",

"LISS Lucas",

"LONGO BORGHINI Elisa",

"MACHAČOVÁ Jarmila",

"MADSEN Frederik Rodenberg",

"MANLY Alexandra",

"MARTIN Daniel",

"MOLLEMA Bauke",

"MORA VEDRI Sebastian",

"MORGAN Alexander",

"MØRKØV Michael",

"MULHERN Mitchell",

"MUNTANER JUANEDA David",

"NEBEN Amber Leone",

"NELSON Emily",

"NUYENS Nick",

"NYS Sven",

"O'BRIEN Kelland",

"O'SHEA Glenn",

"PAWLOWSKA Katarzyna",

"PEDERSEN Mads",

"PEDERSEN Rasmus",

"PHINNEY Taylor",

"PIETERS Amy",

"PINOT Thibaut",

"POELS Wouter",

"POOLEY Emma",

"PORTER Alexander",

"QUINTANA Nairo",

"RASMUSSEN Alex",

"RISSVEDS Jenny",

"RIVERA Coryn",

"RODRIGUEZ OLIVER Joaquim",

"ROGLIČ Primož",

"SAJNOK Szymon",

"SCARPONI Michele",

"SCOTSON Callum",

"SCOTSON Miles",

```
"STEWART Campbell",
"STRONG Corbin",
"TANFIELD Charlie",
"TEKLINSKI Adrian",
"TERPSTRA Niki",
"TOMIC Josephine",
"TORRES BARCELO Albert",
"VAN AVERMAET Greg",
"VAN DIJK Ellen",
"VAN SCHIP Jan Willem",
"VANSUMMEREN Johan",
"VERBEKE Grace",
"VILLUMSEN Linda",
"VIVIANI Elia",
"WHITE Emma",
"WIASAK Rebecca",
"WILLIAMS Lily",
"WLOSZCZOWSKA Maja",
"YALLOURIS Nicholas",
"ZAUGG Oliver")),]

emerging_racedays<-emerging_racedays[c("sport","discipline","competition",
"name", "sex", "age", "year" )]
emerging_racedays<-unique(emerging_racedays)

###Merge
racedays<-rbind(emerging_racedays,eminent_racedays)

###Track
```

```
track_racedays<- racedays[which(racedays$name %in% c("ANKUDINOFF
Ashlee",
"ARCHIBALD Katie","BOBRIDGE
Jack",
"BARKER Elinor","BRONZINI
Giorgia","CATLIN Kelly","CAVENDISH Mark",
"CLANCY Edward","CURE
Amy","DYGERT Chloe","EDMONDSON Alexander",
"EDMONDSON Annette","GANNA
Filippo","HANSEN Lasse Norman", "HEPBURN Michael", "KENNY Laura", "KLUGE
Roger",
"MEYER Cameron","REINHARDT
Theo","ROWE Danielle",
"THOMAS Benjamin","VALENTE
Jennifer","VAN VLEUTEN Annemiek",

"VOS Marianne","WELSFORD Sam","WHITTEN Tara","WIGGINS Bradley",
"WILD Kirsten","ABBOTT Mara","AVILA VANEGAS Edwin Alcibiades","BAKER
Georgia",
"BARBIERI Rachele",
"BASTIANELLI Marta","BRENNAUER Lisa",
"BULLING Pieter",
"BURKE Steven","COQUARD Bryan","DAVISON Luke","DHOORE
Jolien","DIBBEN Jonathan","DIDERIKSEN Amalie",
"DOULL Owain","DURBRIDGE Luke",
"EMADI Kian",
"ERSHOV Artur",
"FRAME Alex",
"FREIBERG Michael","GATE Aaron","GEIST Kimberly","GOUGH
Regan","HAYTER Ethan",
"HEPBURN Michael",
"HERMIDA RAMOS Jose Antonio","IRVINE Martyn","KAJIHARA
Yumi","KOPECKY Lotte","KWOK Ho Ting",
"LANGVAD Annika",
"LISS Lucas",
"LONGO BORGHINI Elisa","MADSEN Frederik Rodenberg",
"MANLY Alexandra",
"MARTIN Daniel","MORA VEDRI Sebastian",
```

"MORGAN Alexander",

"MØRKØV Michael",

"MULHERN Mitchell",

"MUNTANER JUANEDA David","PEDERSEN Mads",

"PEDERSEN Rasmus",

"PHINNEY Taylor",

"PIETERS Amy","SCOTSON Callum",

"SCOTSON Miles",

"STEWART Campbell",

"STRONG Corbin",

"TANFIELD Charlie","TORRES BARCELO Albert","VAN SCHIP Jan Willem","VIVIANI Elia",

"WHITE Emma", "WILLIAMS Lily","YALLOURIS Nicholas",

"ZAUGG Oliver")),]

###Road

road_racedays<- racedays[which(racedays$name %in% c("ANKUDINOFF Ashlee",

"ARMSTRONG Kirstin",

"ARNDT Judith",

"BOONEN Tom","BRONZINI Giorgia",

"CANCELLARA Fabian","CAVENDISH Mark","CURE Amy","DENNIS Rohan","DYGERT Chloe","FREIRE GOMEZ Oscar","FROOME Chris","GANNA Filippo","GILBERT Phillipe","MARTIN Tony","NIBALI Vincenzo","SAGAN Peter","ŠTYBAR Zdeněk","VALVERDE Alejandro","VAN AERT Wout","VAN DEN BROEK-BLAAK Chantal","VAN DER BREGGEN Anna","VAN DER POEL Mathieu","VAN VLEUTEN Annemiek",

"VOS Marianne","WIGGINS Bradley", "WILD Kirsten","ALAPHILIPPE Julien","ARU Fabio","BASTIANELLI Marta","BERNAL GOMEZ Egan Arley",

"BETTIOL Alberto","BRENNAUER Lisa","CIOLEK Gerald",

"COQUARD Bryan",

"COSTA Rui","DEGENKOLB John","DEMARE Arnaud",

```
                                           "DHOORE Jolien","DIDERIKSEN
Amalie","FUGLSANG Jakob","THOMAS Geraint","DUMOULIN Tom","GERRANS
Simon",
                                           "GOSS Matthew","HAYMAN
Mathew","HAYMAN Mathew", "HORNER Christopher", "HUSHOVD Thor","IGLINSKIY
Maxim","JUNGELS Bob","KIRYIENKA Vasil",
                                           "KOPECKY Lotte","KRISTOFF
Alexander",
                                           "KÜNG Stefan",
                                           "KWIATKOWSKI Michal","LONGO
BORGHINI Elisa","MARTIN Daniel",
                                           "MOLLEMA Bauke","PHINNEY
Taylor",
                                           "PIETERS Amy","RIVERA Coryn",
          "RODRIGUEZ OLIVER Joaquim",

          "ROGLIČ Primož","SCARPONI Michele","TERPSTRA Niki","VAN
AVERMAET Greg",
          "VAN DIJK Ellen","VANSUMMEREN Johan","VIVIANI Elia")),]



         ###MTB
          mtbnames1<-subset(emerging_count_by_age, sport=="Mountain-bike")
          mtbnames2<-subset(eminent_count_by_age, sport=="Mountain-bike")
          mtbnames<-rbind(mtbnames1,mtbnames2)
          mtbnames<-mtbnames [c("name")]
          mtbnames<-unique(mtbnames)


           mtb_racedays<- racedays[which(racedays$name %in% c("ALVARADO Ceylin
del Carmen",
                                                      "BELOMOINA Yana",
                                                      "BERNAL GOMEZ Egan
Arley",
                                                      "COURTNEY Kate",
                                                      "HERMIDA RAMOS Jose
Antonio",
                                                      "KÜNG Stefan",
```

```
"LANGVAD Annika",
"NYS Sven",
"PEDERSEN Mads",
"RISSVEDS Jenny",
"VAN DIJK Ellen",
"VERBEKE Grace",
"WLOSZCZOWSKA
Maja",
"ABSALON Julien",
"BRESSET Julie",
"CANT Sanne",
"KNEISKY Morgan",
"KULHAVÝ Jaroslav",
"NEFF Jolanda",
"PENDREL Catharine",
"SCHURTER Nino",
"ŠTYBAR Zdeněk",
"VAN DER BREGGEN
Anna",
"VAN DER POEL
Mathieu",
"VAN VLEUTEN
Annemiek",
"VOS Marianne")),]

###CX
emerging_racedays<-
emerging_count_by_age[which(emerging_count_by_age$name %in% c(

###RACE day plots
track_racemale <-subset(track_racedays,sex=="M")
track_racefemale <-subset(track_racedays,sex=="F")
road_racemale <-subset(road_racedays,sex=="M")
```

```
road_racefemale <-subset(road_racedays,sex=="F")
mtb_racemale<-subset(mtb_racedays,sex=="M")
mtb_racefemale <-subset(mtb_racedays,sex=="F")



eminent_mtbmale <-subset(eminent_racemale,sport=="Mountain-bike")
eminent_mtbfemale <-subset(eminent_racefemale,sport=="Mountain-bike")

eminent_cxmale <-subset(eminent_racemale,sport=="Cyclo-cross")
eminent_cxfemale <-subset(eminent_racefemale,sport=="Cyclo-cross")


mrd<-rename(count(road_racemale, name, age), Freq = n)
frd<-rename(count(road_racefemale, name, age), Freq = n)
mtr<-rename(count(track_racemale, name, age), Freq = n)
ftr<-rename(count(track_racefemale, name, age), Freq = n)
mmtb<-rename(count(mtb_racemale, name, age), Freq = n)
fmtb<-rename(count(mtb_racefemale, name, age), Freq = n)
mcx<-rename(count(eminent_cxmale, name, age), Freq = n)
fcx<-rename(count(eminent_cxfemale, name, age), Freq = n)

roadmale <-ggplot(mrd, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",

color="darkred")+
theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 38)+ ylim(1,
90)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by age for male road athletes")

roadfemale <-ggplot(frd, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",
```

```
color="darkred")+
            theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 38)+ ylim(1,
85)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by age for female road athletes")


        trackmale <-ggplot(mtr, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",

color="darkred")+
            theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 35)+ ylim(1,
75)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by age for male track athletes")
        trackfemale <-ggplot(ftr, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",
                                                  color="darkred")+
            theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 35)+ ylim(1,
75)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by sport and age for female track athletes")
        mtbmale <-ggplot(mmtb, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",

color="darkred")+
            theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 38)+ ylim(1,
40)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by age for male mountainbike athletes")


        mtbfemale <-ggplot(fmtb, aes(x = age, y=Freq)) +geom_point(shape=18,
color="blue", alpha=0.2) + geom_smooth(linetype="dashed",

color="darkred")+
            theme_classic()+
geom_density_2d(alpha=0.4)+scale_color_brewer(palette="Dark2") +xlim(17, 38)+ ylim(1,
40)+labs(y="Racedays", title="UCI race volume by age", subtitle = "Race volume
development by age for female mountainbike athletes")
```

```
cxmale <-ggplot(mcx, aes(x = age, y=Freq)) +geom_point() +
geom_smooth()+labs( title="UCI race volume by age", subtitle = "Race volume development by sport and age for eminent female athletes")

cxfemale <-ggplot(fcx, aes(x = age, y=Freq)) +geom_point() +
geom_smooth()+labs( title="UCI race volume by age", subtitle = "Race volume development by sport and age for eminent female athletes")

############# 6. odds table 1 #############

library(dplyr)
library(cluster)
library(vcd)
library(pivottabler)
library(ineq)
library(ggplot2)
library(gglorenz)
library(stats)
library(ISLR)
library(tidyr)
library(elo)
library(readxl)
library(data.table)
library(Hmisc)
library(purrr)
library(kamila)
library(fmsb)
library(gtools)

colnames(result)<-
c("sport","discipline","competition","location","level","year","cat",

"placement","bib","name","lastname","firstname","country","team","sex",
                    "age","phase","heat","time","norace","points")
```

```
data<-data.table(result)


all<-data.table(data)

all<-all[,age:=fifelse(is.na(age),0,age)]

all<-all[age>0]



data<-all[,classif:=fifelse(sport=="Mountain-bike" & (cat=="CDM"|
cat=="CM"|cat=="JO"),1,
                    fifelse(sport=="Track" & ( cat=="CM"|cat=="JO"),1,
                        fifelse(sport=="Cyclo-cross" & (cat=="CM"),1,
                            fifelse(sport=="Road" & sex=="M" & (competition
%in% c(
                             "Vuelta a España","World Championships", "Giro
d'Italia","Giro Ciclistico d'Italia",
                             "Tour de France","Milano-Sanremo","Milano -
Sanremo",
                             "Ronde van Vlaanderen - Tour des Flandres","Ronde
van Vlaanderen / Tour des Flandres",
                             "Paris - Roubaix",
                             "Il Lombardia",
                             "Liège-Bastogne-Liège","Liège - Bastogne - Liège",
                             "Jeux Olympiques - c.l.m. ind. / Olympic Games - ind.
TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course
en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR
ME")),1,
                             fifelse(sport=="Road" & sex=="F" & (competition
%in% c("Giro d'Italia Internazionale Femminile","Ronde van Vlaanderen - Tour des
Flandres","Ronde van Vlaanderen / Tour des Flandres",
                                                   "World
Championships","Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux
Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne /
Olympic Games - IRR",
                                                   "Jeux Olympiques -
en ligne / Olympic Games - ind. RR ME")),1,0)))))]


#marcar los que son eminent#
```

```
data<-data %>% group_by(name) %>% summarise (premier=sum(classif))
data<-data.table(data)
data<-data[,eminent:=fifelse(premier>=3,1,0)]
data1 <- merge(all, data,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)

data1<-data1[,.(name,eminent,level)]
data1<-data1[!duplicated(data1[,1:3]),]
data1<-data1[level=="Elite"]

## First elite Podium  SEX ##



data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex)]
data2<-data2[podium==1 & level=="Elite"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ] # Orden directo
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)
#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
```

```
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-oddsratio(M, p.calc.by.independence=FALSE)
  odds$p.value
  odds$estimate
  odds$conf.int

  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev
})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)

t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
```

```
  M
  odds<-oddsratio(M, p.calc.by.independence=FALSE)
  odds$p.value
  odds$estimate
  odds$conf.int

  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev
})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)

t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4))

##ROAD##

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport)]
data2<-data2[podium==1 & level=="Elite" & sport=="Road"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ] # Orden directo
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
         by.x ="name", by.y ="name",
         all.x = TRUE, all.y = FALSE)

#SEX ROAD M#
```

```
sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)

  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
```

```
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#	Mountain-bike

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport)]
data2<-data2[podium==1 & level=="Elite" & sport=="Mountain-bike"]
```

```
data2 <- data2[with(data2, order(data2$name,data2$age)), ] # Orden directo
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
 cd<-nrow(data3[eminent==0 & sex=="M"])
 c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
 d<-(cd-c)

 ab<-nrow(data3[eminent==1 & sex=="M"])
 a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
 b<-(ab-a)
 M <- matrix(c(a, c, b, d), ncol = 2)
 M
 odds<-(oddsratio(M, p.calc.by.independence=FALSE))
 odds$p.value
 odds$estimate
 odds$conf.int
 rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
 rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
```

```
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
```

```
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))
#       Track


data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport)]
data2<-data2[podium==1 & level=="Elite" & sport=="Track"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ] # Orden directo
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev
```

```
})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})
```

```
t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))


#	Cyclo-cross

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport)]
data2<-data2[podium==1 & level=="Elite" & sport=="Cyclo-cross"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ] # Orden directo
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
```

```
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))
#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
```

```
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))


## First premier Podium ##

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport,classif)]
data2<-data2[podium==1 & level=="Elite" & classif==1]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
            by.x ="name", by.y ="name",
            all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
```

```
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)
```

```
  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))


#ROAD#

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport,classif)]
data2<-data2[podium==1 & level=="Elite" & classif==1 & sport=="Road"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
```

```
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)


#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
```

```
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})

t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))
```

```
#Mountain-bike #
data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport,classif)]
data2<-data2[podium==1 & level=="Elite" & classif==1 & sport=="Mountain-
bike"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)


#SEX M#

sex<-defmacro(x,y,expr={
 cd<-nrow(data3[eminent==0 & sex=="M"])
 c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
 d<-(cd-c)

 ab<-nrow(data3[eminent==1 & sex=="M"])
 a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
 b<-(ab-a)
 M <- matrix(c(a, c, b, d), ncol = 2)
 M
 odds<-(oddsratio(M, p.calc.by.independence=FALSE))
 odds$p.value
 odds$estimate
 odds$conf.int
 rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
```

```
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
```

```
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))


#Track#

data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport,classif)]
data2<-data2[podium==1 & level=="Elite" & classif==1 & sport=="Track"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)


#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)
```

```
  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)
```

```
  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#Cyclo-cross#


data2<-all[,podium:=fifelse((placement==1 | placement==2 |placement==3),1,0)]
data2<-data2[,.(podium,level,name,age,sex,sport,classif)]
data2<-data2[podium==1 & level=="Elite" & classif==1 & sport=="Cyclo-cross"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
            by.x ="name", by.y ="name",
```

```
        all.x = TRUE, all.y = FALSE)


#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))
```

```
#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

## First premier Win ##
```

```
data2<-all[,win:=fifelse((placement==1),1,0)]
data2<-data2[,.(win,level,name,age,sex,sport,classif)]
data2<-data2[win==1 & level=="Elite" & classif==1]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
```

```
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
```

```
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#Road

data2<-all[,win:=fifelse((placement==1),1,0)]
data2<-data2[,.(win,level,name,age,sex,sport,classif)]
data2<-data2[win==1 & level=="Elite" & classif==1 & sport=="Road"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
        by.x ="name", by.y ="name",
        all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
 cd<-nrow(data3[eminent==0 & sex=="M"])
 c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
 d<-(cd-c)

 ab<-nrow(data3[eminent==1 & sex=="M"])
 a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
 b<-(ab-a)
 M <- matrix(c(a, c, b, d), ncol = 2)
 M
 odds<-(oddsratio(M, p.calc.by.independence=FALSE))
 odds$p.value
 odds$estimate
 odds$conf.int
```

```
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
```

```
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#Mountain-bike

data2<-all[,win:=fifelse((placement==1),1,0)]
data2<-data2[,.(win,level,name,age,sex,sport,classif)]
data2<-data2[win==1 & level=="Elite" & classif==1 & sport=="Mountain-bike"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)
```

```
  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
```

```
 a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
 b<-(ab-a)
 M <- matrix(c(a, c, b, d), ncol = 2)
 M
 odds<-(oddsratio(M, p.calc.by.independence=FALSE))
 odds$p.value
 odds$estimate
 odds$conf.int
 rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
 rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#Track


data2<-all[,win:=fifelse((placement==1),1,0)]
data2<-data2[,.(win,level,name,age,sex,sport,classif)]
data2<-data2[win==1 & level=="Elite" & classif==1 & sport=="Track"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
```

```
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
```

```
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))


#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))
```

```
#Cyclo-cross

data2<-all[,win:=fifelse((placement==1),1,0)]
data2<-data2[,.(win,level,name,age,sex,sport,classif)]
data2<-data2[win==1 & level=="Elite" & classif==1 & sport=="Cyclo-cross"]
data2 <- data2[with(data2, order(data2$name,data2$age)), ]
data2<-data2[!duplicated(data2[,1:3]),]
data3 <- merge(data2, data1,
               by.x ="name", by.y ="name",
               all.x = TRUE, all.y = FALSE)

#SEX M#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="M"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="M"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="M"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="M"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev
```

```
})


t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))



#SEX F#

sex<-defmacro(x,y,expr={
  cd<-nrow(data3[eminent==0 & sex=="F"])
  c<-nrow(data3[eminent==0 & (age>=x & age<=y) & sex=="F"])
  d<-(cd-c)

  ab<-nrow(data3[eminent==1 & sex=="F"])
  a<-nrow(data3[eminent==1 & (age>=x & age<=y) & sex=="F"])
  b<-(ab-a)
  M <- matrix(c(a, c, b, d), ncol = 2)
  M
  odds<-(oddsratio(M, p.calc.by.independence=FALSE))
  odds$p.value
  odds$estimate
  odds$conf.int
  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev
```

```
})

t1<-sex(18,19)
t2<-sex(20,21)
t3<-sex(22,23)
t4<-sex(24,25)
t5<-sex(26,72)
cd+ab
(t(matrix(c(t1,t2,t3,t4,t5),ncol=5,nrow=4)))

##############################7##################################

##############single#########

colnames(result)<-
c("sport","discipline","competition","location","level","year","cat",

"placement","bib","name","lastname","firstname","country","team","sex",
                    "age","phase","heat","time","norace","points")
data<-data.table(result)

all<-data.table(data)
all<-all[,age:=fifelse(is.na(age),0,age)]
all<-all[age>0]

data<-all[,classif:=fifelse(sport=="Mountain-bike" & (cat=="CDM"|
cat=="CM"|cat=="JO"),1,
```

```
                    fifelse(sport=="Track" & ( cat=="CM"|cat=="JO"),1,
                        fifelse(sport=="Cyclo-cross" & (cat=="CM"),1,
                            fifelse(sport=="Road" & sex=="M" & (competition
%in% c(
                            "Vuelta a España","World Championships", "Giro
d'Italia","Giro Ciclistico d'Italia",
                            "Tour de France","Milano-Sanremo","Milano -
Sanremo",
                            "Ronde van Vlaanderen - Tour des Flandres","Ronde
van Vlaanderen / Tour des Flandres",
                            "Paris - Roubaix",
                            "Il Lombardia",
                            "Liège-Bastogne-Liège","Liège - Bastogne - Liège",
                            "Jeux Olympiques - c.l.m. ind. / Olympic Games - ind.
TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course
en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR
ME")),1,
                            fifelse(sport=="Road" & sex=="F" & (competition
%in% c("Giro d'Italia Internazionale Femminile","Ronde van Vlaanderen - Tour des
Flandres","Ronde van Vlaanderen / Tour des Flandres",
                                                "World
Championships","Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux
Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne /
Olympic Games - IRR",
                                                "Jeux Olympiques -
en ligne / Olympic Games - ind. RR ME")),1,0)))))]
    #marcar los que son eminent#


    data<-data %>% group_by(name) %>% summarise (premier=sum(classif))
    data<-data.table(data)
    data<-data[,eminent:=fifelse(premier>=3,1,0)]
    data1 <- merge(all, data,
            by.x ="name", by.y ="name",
            all.x = TRUE, all.y = FALSE)


    data1<-data1[,.(name,eminent,level,sport,sex)]
    data1<-data1[!duplicated(data1[,1:5]),]
```

```
data1<-data1[level=="Elite"]



single<-defmacro(x,y,z,w,t,expr={
  cd<-nrow(data1[eminent==0 & sex==t & (sport==x| sport==y | sport==z | sport==w)])
  c<-nrow(data1[eminent==0 & sex==t & sport==x])
  d<-(cd-c)
  cd
  c
  d
  ab<-nrow(data1[eminent==1 & sex==t & (sport==x| sport==y | sport==z | sport==w)])
  a<-nrow(data1[eminent==1 & sex==t & sport==x])
  b<-(ab-a)
  ab
  a
  b
  M <- matrix(c(a ,c, b, d), ncol = 2)
  M
  odds<-oddsratio(M, p.calc.by.independence=FALSE)
  odds$p.value
  odds$estimate
  odds$conf.int

  rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
  rev

})



t1<-single("Mountain-bike","Cyclo-cross","Road","Track","M")
t2<-single("Cyclo-cross","Road","Track","Mountain-bike","M")
```

```
t3<-single("Mountain-bike","Cyclo-cross","Road","Track","F")
t4<-single("Cyclo-cross","Road","Track","Mountain-bike","F")

(t(matrix(c(t1,t2,t3,t4),ncol=4,nrow=4)))

##
colnames(result)<-
c("sport","discipline","competition","location","level","year","cat",

"placement","bib","name","lastname","firstname","country","team","sex",
"age","phase","heat","time","norace","points")
data<-data.table(result)

all<-data.table(data)
all<-all[,age:=fifelse(is.na(age),0,age)]
all<-all[age>0]


data<-all[,classif:=fifelse(sport=="Mountain-bike" & (cat=="CDM"|
cat=="CM"|cat=="JO"),1,
fifelse(sport=="Track" & ( cat=="CM"|cat=="JO"),1,
fifelse(sport=="Cyclo-cross" & (cat=="CM"),1,
fifelse(sport=="Road" & sex=="M" & (competition
%in% c(
"Vuelta a España","World Championships", "Giro
d'Italia","Giro Ciclistico d'Italia",
"Tour de France","Milano-Sanremo","Milano -
Sanremo",
"Ronde van Vlaanderen - Tour des Flandres","Ronde
van Vlaanderen / Tour des Flandres",
"Paris - Roubaix",
"Il Lombardia",
"Liège-Bastogne-Liège","Liège - Bastogne - Liège",
"Jeux Olympiques - c.l.m. ind. / Olympic Games - ind.
TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course
```

```
en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR ME")),1,
                                        fifelse(sport=="Road" & sex=="F" & (competition %in% c("Giro d'Italia Internazionale Femminile","Ronde van Vlaanderen - Tour des Flandres","Ronde van Vlaanderen / Tour des Flandres",
                                                                                    "World Championships","Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne / Olympic Games - IRR",
                                                                                    "Jeux Olympiques - en ligne / Olympic Games - ind. RR ME")),1,0)))))]
        #marcar los que son eminent#


        data<-data %>% group_by(name) %>% summarise (premier=sum(classif))
        data<-data.table(data)
        data<-data[,eminent:=fifelse(premier>=3,1,0)]
        data1 <- merge(all, data,
                by.x ="name", by.y ="name",
                all.x = TRUE, all.y = FALSE)


        data1<-data1[,new_discipline:=fifelse(sport=="Road" & (discipline == "Stage 8b : Individual Time Trial" | discipline == "Team Time Trial" | discipline == "Individual Time Trial" | discipline == "individual time trial"),"Time Trial",
                                        fifelse(sport=="Mountain-bike" & discipline=="country Olympic","Country Olympic",
                                            fifelse(sport=="Track" & discipline=="Madison","Madison",
                                                fifelse(sport=="Track" & discipline=="Individual Pursuit","Individual Pursuit",
                                                    fifelse(sport=="Track" & discipline=="Team Pursuit","Team Pursuit",
                                                        fifelse(sport=="Track"  & discipline=="Omnium","Omnium",
                                                            fifelse(sport=="Track"  & 

discipline=="Points Race","Points Race",
                                                                fifelse(sport=="Track"  & discipline=="Scratch","Scratch",
```

```
                                                   fifelse(sport=="Cyclo-
cross","",

fifelse(sport=="Bmx-racing",""

,"Other"))))))))))]

        data1<-data1[,.(name,eminent,level,sport,sex,new_discipline)]
        data1<-data1[!duplicated(data1[,1:6]),]
        data1<-data1[level=="Elite"]

        single<-defmacro(x,y,z,w,v,t,expr={
         cd<-nrow(data1[eminent==0 & sex==t & (sport==x| sport==y | sport==z |
sport==w)])
         c<-nrow(data1[eminent==0 & sex==t & sport==x & new_discipline==v])
         d<-(cd-c)
         cd
         c
         d
         ab<-nrow(data1[eminent==1 & sex==t & (sport==x| sport==y | sport==z |
sport==w)])
         a<-nrow(data1[eminent==1 & sex==t & sport==x & new_discipline==v])
         b<-(ab-a)
         ab
         a
         b
         M <- matrix(c(a ,c, b, d), ncol = 2)
         M
         odds<-oddsratio(M, p.calc.by.independence=FALSE)
         odds$p.value
         odds$estimate
         odds$conf.int

         rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
```

```
rev

})


t1<-single("Road","Mountain-bike","Cyclo-cross","Track","Time  Trial","M")
t2<-single("Road","Mountain-bike","Cyclo-cross","Track","Other","M")

t3<-single("Track","Mountain-bike","Cyclo-cross","Road","Team Pursuit","M")
t4<-single("Track","Mountain-bike","Cyclo-cross","Road","Omnium","M")
t5<-single("Track","Mountain-bike","Cyclo-cross","Road","Madison","M")
t6<-single("Track","Mountain-bike","Cyclo-cross","Road","Individual
Pursuit","M")

(t(matrix(c(t1,t2,t3,t4,t5,t6),ncol=6,nrow=4)))

t1<-single("Road","Mountain-bike","Cyclo-cross","Track","Time  Trial","F")
t2<-single("Road","Mountain-bike","Cyclo-cross","Track","Other","F")

t3<-single("Track","Mountain-bike","Cyclo-cross","Road","Team Pursuit","F")
t4<-single("Track","Mountain-bike","Cyclo-cross","Road","Omnium","F")
t5<-single("Track","Mountain-bike","Cyclo-cross","Road","Madison","F")
t6<-single("Track","Mountain-bike","Cyclo-cross","Road","Individual  Pursuit","F")

(t(matrix(c(t1,t2,t3,t4,t5,t6),ncol=6,nrow=4)))


###multi################

colnames(result)<-
c("sport","discipline","competition","location","level","year","cat",

"placement","bib","name","lastname","firstname","country","team","sex",
```

```
                    "age","phase","heat","time","norace","points")
        data<-data.table(result)


        all<-data.table(data)
        all<-all[,age:=fifelse(is.na(age),0,age)]
        all<-all[age>0]


        data<-all[,classif:=fifelse(sport=="Mountain-bike" & (cat=="CDM"|
cat=="CM"|cat=="JO"),1,
                                fifelse(sport=="Track" & ( cat=="CM"|cat=="JO"),1,
                                    fifelse(sport=="Cyclo-cross" & (cat=="CM"),1,
                                        fifelse(sport=="Road" & sex=="M" & (competition
%in% c(
                                         "Vuelta a España","World Championships", "Giro
d'Italia","Giro Ciclistico d'Italia",
                                         "Tour de France","Milano-Sanremo","Milano -
Sanremo",
                                         "Ronde van Vlaanderen - Tour des Flandres","Ronde
van Vlaanderen / Tour des Flandres",
                                         "Paris - Roubaix",
                                         "Il Lombardia",
                                         "Liège-Bastogne-Liège","Liège - Bastogne - Liège",
                                         "Jeux Olympiques - c.l.m. ind. / Olympic Games - ind.
TT ME","Jeux Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course
en ligne / Olympic Games - IRR","Jeux Olympiques - en ligne / Olympic Games - ind. RR
ME")),1,
                                         fifelse(sport=="Road" & sex=="F" & (competition
%in% c("Giro d'Italia Internazionale Femminile","Ronde van Vlaanderen - Tour des
Flandres","Ronde van Vlaanderen / Tour des Flandres",
                                                                    "World
Championships","Jeux Olympiques - c.l.m. ind. / Olympic Games - ind. TT ME","Jeux
Olympiques - CLM Ind. / Olympic Games - ITT","Jeux Olympiques - Course en ligne /
Olympic Games - IRR",
                                                                    "Jeux Olympiques -
en ligne / Olympic Games - ind. RR ME")),1,0)))))]


        #marcar los que son eminent#
```

```
data<-data %>% group_by(name) %>% summarise (premier=sum(classif))
data<-data.table(data)
data<-data[,eminent:=fifelse(premier>=3,1,0)]
data1 <- merge(all, data,
          by.x ="name", by.y ="name",
          all.x = TRUE, all.y = FALSE)

data1<-data1[,.(name,eminent,level,sex)]
data1<-data1[!duplicated(data1[,1:3]),]
data1<-data1[level=="Elite"]



road<-all[(sport=="Road" ) & level=="Elite"]
road<-road[,name,sport]
road<-road[!duplicated(road[,1:2]),]
nrow(road)

Mountain<-all[(sport=="Mountain-bike" ) & level=="Elite"]
Mountain<-Mountain[,name,sport]
Mountain<-Mountain[!duplicated(Mountain[,1:2]),]
nrow(Mountain)

track<-all[(sport=="Track" ) & level=="Elite"]
track<-track[,name,sport]
track<-track[!duplicated(track[,1:2]),]
nrow(track)

cyclo<-all[(sport=="Cyclo-cross" ) & level=="Elite"]
cyclo<-cyclo[,name,sport]
cyclo<-cyclo[!duplicated(cyclo[,1:2]),]
nrow(cyclo)
```

```
data2 <- merge(data1, road,
          by.x = "name", by.y = "name",
          all.x = TRUE, all.y = FALSE)

colnames(data2)<-c("name","eminent","level","sex","road")

data2 <- merge(data2, track,
          by.x = "name", by.y = "name",
          all.x = TRUE, all.y = FALSE)

colnames(data2)<-c("name","eminent","level","sex","road","track")

data2 <- merge(data2, Mountain,
          by.x = "name", by.y = "name",
          all.x = TRUE, all.y = FALSE)

colnames(data2)<-c("name","eminent","level","sex","road","track","Mountain")

data2 <- merge(data2, cyclo,
          by.x = "name", by.y = "name",
          all.x = TRUE, all.y = FALSE)

colnames(data2)<-
c("name","eminent","level","sex","road","track","Mountain","cyclo")

data3<-data2[, group:= paste0(road, '_', track,'_',Mountain,'-',cyclo)]
data3<-data3[,.(name,eminent,group,sex)]
data3<-data3[!duplicated(data3[,1:3]),]
```

```
library(fmsb)
library(gtools) #MACROS


multi<-defmacro(x,y,z,expr={
 cd<-nrow(data3[eminent==0 & sex==z & (group==x| group==y)])
 c<-nrow(data3[eminent==0 & sex==z & group==y])
 d<-(cd-c)
 cd
 d
 c

 ab<-nrow(data3[eminent==1 & sex==z & (group==x| group==y)])
 a<-nrow(data3[eminent==1 & sex==z & group==y])
 b<-(ab-a)
 ab
 a
 b

 M <- matrix(c(a ,c, b, d), ncol = 2)
 M

 ab+cd
 odds<-oddsratio(M, p.calc.by.independence=FALSE)
 odds$p.value
 odds$estimate
 odds$conf.int

 rev<-c(odds$estimate,odds$conf.int[1],odds$conf.int[2],odds$p.value)
 rev

})
```

```
t1<-multi("NA_Track_NA-NA","Road_Track_NA-NA","M")
t2<-multi("NA_Track_NA-NA","NA_Track_Mountain-bike-NA","M")
t3<-multi("NA_Track_NA-NA","NA_Track_NA-Cyclo-cross","M")

t4<-multi("Road_NA_NA-NA","Road_NA_Mountain-bike-NA","M")
t5<-multi("Road_NA_NA-NA","Road_NA_NA-Cyclo-cross","M")
t6<-multi("Road_NA_NA-NA","Road_Track_NA-NA","M")

t7<-multi("NA_NA_Mountain-bike-NA","Road_NA_Mountain-bike-NA","M")
t8<-multi("NA_NA_Mountain-bike-NA","NA_Track_Mountain-bike-NA","M")
t9<-multi("NA_NA_Mountain-bike-NA","NA_NA_Mountain-bike-Cyclo-
cross","M")

t10<-multi("NA_NA_NA-Cyclo-cross","Road_NA_NA-Cyclo-cross","M")
t11<-multi("NA_NA_NA-Cyclo-cross","NA_NA_Mountain-bike-Cyclo-
cross","M")
t12<-multi("NA_NA_NA-Cyclo-cross","NA_Track_NA-Cyclo-cross","M")

t(matrix(c(t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,t11,t12),ncol=12,nrow=4))

##

t1<-multi("NA_Track_NA-NA","Road_Track_NA-NA","F")
t2<-multi("NA_Track_NA-NA","NA_Track_Mountain-bike-NA","F")
t3<-multi("NA_Track_NA-NA","NA_Track_NA-Cyclo-cross","F")

t4<-multi("Road_NA_NA-NA","Road_NA_Mountain-bike-NA","F")
t5<-multi("Road_NA_NA-NA","Road_NA_NA-Cyclo-cross","F")
t6<-multi("Road_NA_NA-NA","Road_Track_NA-NA","F")

t7<-multi("NA_NA_Mountain-bike-NA","Road_NA_Mountain-bike-NA","F")
t8<-multi("NA_NA_Mountain-bike-NA","NA_Track_Mountain-bike-NA","F")
```

```
t9<-multi("NA_NA_Mountain-bike-NA","NA_NA_Mountain-bike-Cyclo-cross","F")

t10<-multi("NA_NA_NA-Cyclo-cross","Road_NA_NA-Cyclo-cross","F")
t11<-multi("NA_NA_NA-Cyclo-cross","NA_NA_Mountain-bike-Cyclo-cross","F")
t12<-multi("NA_NA_NA-Cyclo-cross","NA_Track_NA-Cyclo-cross","F")

t(matrix(c(t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,t11,t12),ncol=12,nrow=4))

## Additional analyses Chapter 3

# Load the dplyr and knitr packages
library(dplyr)
library(knitr)

#load a data in R
result <- read.csv2("~/results.csv")



colnames(results_2021_10_08) <- results_2021_10_08[1, ]

# Create a subset of the data frame
df <- results_2021_10_08 %>% filter(sport %in% c("road", "track", "mountain-bike", "cyclo-cross") & jr_u23_elite_masters == "Elite")

##### TABLE 1 #####

# Calculate the average age, standard deviation, and average age of first podium group by gender and sport
answer <- df %>%
  group_by(sport, gender)%>%
  summarise(N = length(age),
    Average_age = mean(age, na.rm = TRUE),
```

```
          Standard_deviation = sd(age, na.rm = TRUE),
          Average_age_of_first_podium = mean(age[rank %in% c(1, 2, 3)], na.rm = TRUE))

# Add a new row for the overall results
answer <- rbind(answer, data.frame(sport = "All", gender = "All",
                             N = length(df$age),
                             Average_age = mean(df$age, na.rm = TRUE),
                             Standard_deviation = sd(df$age, na.rm = TRUE),
                             Average_age_of_first_podium = mean(df$age[df$rank %in% c(1, 2,
3)], na.rm = TRUE)))

# Print the answer data frame
kable(answer)

##compute the average age, standard deviation by sports and first podium
group_sport <- df %>%
  group_by(sport) %>%
  summarise(N=length(age),
    Average_age = mean(age, na.rm = TRUE),
          Standard_deviation = sd(age, na.rm = TRUE),
          Average_age_of_first_podium = mean(age[rank %in% c(1, 2, 3)], na.rm = TRUE))
group_sport

##comput the average age, standard deviation by gender and first podium
group_gender <- df %>%
  group_by(gender) %>%
  summarise(N=length(age),
    Average_age = mean(age, na.rm = TRUE),
          Standard_deviation = sd(age, na.rm = TRUE),
          Average_age_of_first_podium = mean(age[rank %in% c(1, 2, 3)], na.rm = TRUE))
group_gender
```

```
#### TABLE 2 ####

# Calculate the percentage of participants who achieved a podium and a first place for each group
answer2 <- df %>%
  group_by(sport, gender) %>%
  summarise(N=length(fullname),
    Percentage_of_podium = round(100 * n_distinct(fullname[rank %in% c(1, 2, 3)]) / n_distinct(fullname), 2),
        Percentage_of_first_place = round(100 * n_distinct(fullname[rank == 1]) / n_distinct(fullname), 2))

# Print the answer2 data frame
kable(answer2)
str(df)

## Find average age, standard deviation by both gender group by sport

overall_sport <- df %>%
  group_by(sport) %>%
  summarise(N=length(fullname),
    Percentage_of_podium = round(100 * n_distinct(fullname[rank %in% c(1, 2, 3)]) / n_distinct(fullname), 2),
        Percentage_of_first_place = round(100 * n_distinct(fullname[rank == 1]) / n_distinct(fullname), 2))

overall_sport

## Find average age, standard deviation by all sports group by gender
overall_gender <- df %>%
  group_by(gender) %>%
  summarise(N=length(fullname),
    Percentage_of_podium = round(100 * n_distinct(fullname[rank %in% c(1, 2, 3)]) / n_distinct(fullname), 2),
```

Percentage_of_first_place = round(100 * n_distinct(fullname[rank == 1]) / n_distinct(fullname), 2))

overall_gender

Filter the data by sport

sport <- "mountain-bike"

df_sport <- df[df$sport == sport,]

Group the data by full name and count wins in the sport

df_sport_wins <- df_sport %>%

group_by(fullname) %>%

summarise(wins = sum(rank == 1), # Count how many times they have won a race in the sport

other_sport = any(sport != first(sport))) # Check if they have won a race in any other sport

Calculate the percentages of interest

total <- nrow(df_sport_wins) # Total number of participants in the sport

perc_1 <- mean(df_sport_wins$wins > 0 & !df_sport_wins$other_sport) * 100 # Percentage of participants who have won a race in just the sport

perc_2 <- mean(df_sport_wins$wins > 0 & df_sport_wins$other_sport) * 100 # Percentage of participants who have won a race in that sport and at least one in another sport

Print the results

cat(paste0("The percentage of participants who have won a race in just ", sport, " is ", round(perc_1, 2), "%.\n"))

cat(paste0("The percentage of participants who have won a race in ", sport, " and at least one in another sport is ", round(perc_2, 2), "%.\n"))