



Predicting the Unpredictable

Why Algorithmic Talent Identification Struggles With Non-Linear Athlete Development

About this paper

This paper examines what happens when predictive analytics and machine learning are applied to talent identification in youth sport. It is written for performance directors, academy and pathway managers, coach developers and the governing-body staff who write selection policy, and for anyone who has been asked to buy, build or approve a tool that scores young athletes. The question it answers is a practical one. If a model is trained on the records of athletes who were selected and survived, what does it say about a 12-year-old who has barely started?

It is the first paper in the Athleet.AI strand, whose single argument is that AI in coaching and athlete development must respect the non-linear, human and developmental nature of athletic progress. Talent identification is where that argument is tested hardest, because it is the place in sport where a number most readily turns into a decision about a child, and where the consequences of that decision are least often followed up.

Editorial balance. Roughly half of the paper reports published evidence on athlete development, relative age, maturation and machine learning. About a third is my own analysis of how those bodies of evidence interact, which no study I have found tests directly, and the remaining part sets out a design stance and recommendations that are my proposals and are tagged as such.

HOW TO READ THE EVIDENCE TAGS

Every substantive claim carries a tag stating what kind of evidence supports it. Two tags are combined where a claim rests on both.

PEER-REVIEWED A primary study in a peer-reviewed journal.

REVIEW A systematic review, meta-analysis or narrative review.

CONSENSUS A governing body position, consensus statement or official guidance.

PREPRINT A working paper or preprint not yet peer reviewed.

INDUSTRY A company announcement, press report or industry source.

SCHOLARLY A scholarly book, edited volume or law journal article.

PRACTICE Practitioner opinion, including the author's own.

EVIDENCE GAP The available evidence does not answer the question.

ATHLEET.AI The author's own framework or analysis, proposed rather than proven.

CONTENTS

Contents

FRONT MATTER

A note from the author	5
Executive summary	6

PART ONE · HOW ATHLETES DEVELOP

1 The 14-year-old who would have been cut	9
2 What the development evidence shows	11

PART TWO · WHAT PREDICTION BRINGS TO THE PATHWAY

3 What predictive AI assumes	14
4 Survivorship in the training set	16
5 The prediction that makes itself true	18
6 Maturation as a confounder	20

PART THREE · A DIFFERENT JOB FOR AI

7 Describing the athlete in front of you	23
--	----

END MATTER

References	25
About the author	26

ILLUSTRATIVE MATERIAL

Every organisation, person and scenario described as illustrative in this paper is invented. Published research and official documents are cited and numbered, and invented material is never presented as evidence.

Research closed 23 January 2023.
Sources and events are as they stood at that date.

CITATION

Forrest, P. (2023) *Predicting the Unpredictable: Why Algorithmic Talent Identification Struggles With Non-Linear Athlete Development*. Athleet.AI Working Papers, AI in Coaching and Athlete Development, Paper 1. Athleet.AI, 25 January 2023. Available at: athleet.ai.

Comments and corrections are welcome and will be acknowledged in later versions.

PAPERS IN THIS STRAND

- **Paper one. *Predicting the Unpredictable. Why Algorithmic Talent Identification Struggles With Non-Linear Athlete Development*. 25 January 2023.**
- Paper two. *The Coach in the Loop*. Tacit Knowledge, Professional Judgement and the Risk of Deskilled Coaching. 27 February 2023.
- Paper three. *The Measured Child*. Datafication, Consent and Autonomy Across the Athlete Pathway. 19 April 2023.
- Paper four. *Maturation, Bias and the Training Set*. Developmental Fairness in AI-Supported Athlete Development. 24 May 2023.
- Paper five. *Load, Readiness and False Precision*. The Limits of Algorithmic Training Prescription in Long-Term Endurance Development. 21 June 2023.
- Paper six. *Democratised or Stratified?* AI, Competitive Fairness and the

Grassroots Coaching Gap. 26 July
2023.

A NOTE FROM THE AUTHOR

The list nobody keeps

Anyone who has coached young athletes for long enough carries a private list of the ones who were not supposed to make it. Mine is short and proves nothing statistically, yet it has shaped how I read every selection sheet I have seen since. The names on it share a pattern. They were small, or late to grow, or new to the sport, or simply having a poor season at the moment someone decided to count, and several of them were still training and improving years after the early stars had drifted away.

My advisory work puts me in rooms where the same question arrives from the other direction. Organisations want to know whether data can make selection fairer and cheaper, and whether the new generation of AI tools can help. Since ChatGPT appeared at the end of November, those conversations have gained a new energy, and I have heard coaches describe trying it for session ideas within weeks of its release. The appetite is real, and much of it is healthy.

I wrote this paper because talent identification is where enthusiasm for prediction can do the most quiet damage. A poor session plan is noticed the following week. A 13-year-old released on the strength of a score is rarely followed up by anyone, so the error never travels back to the system that made it, and the system learns nothing.

I use measurement and value it. What I want to ask is narrower and, I think, harder to answer well, namely what a responsible tool should say about a young person whose development has barely begun, and what it should decline to say. The answer I arrive at is modest. A tool can tell a coach a good deal about where an athlete is today, and little that deserves trust about where that athlete will be at 25.

Paul Forrest
Founder, Athleet.AI
25 January 2023

TWO PAGES THAT CIRCULATE ON THEIR OWN

Executive summary

Predictive tools learn the pathway that produced their data, and in youth sport that pathway is already bent by relative age, early maturation and early selection. Pointed at 12-year-olds, such tools are likely to reproduce those distortions with more confidence and less visibility than a human selector. A prediction that is acted upon also helps to create the outcome it forecast. My answer is that AI in talent systems should describe where an athlete is now, with uncertainty and maturation context shown, and should issue no verdict on where that athlete will end up.

The evidence on how athletes develop and the literature on machine learning in sport have grown up largely apart. The first shows that elite development is mostly non-linear and that early success is a weak and sometimes misleading guide to later success. The second shows fast growth in predictive modelling, trained overwhelmingly on adult, male, professional samples. This paper brings the two together. It asks what a model inherits when it is pointed at children, and what a club should require before any such model is allowed near a selection meeting.

1

Most elite athletes did not follow a straight line from junior to senior success.

In a survey of 256 elite Australian athletes across 27 sports, 83.6% reported non-linear trajectories and fewer than 7% showed pure junior to senior linearity.¹

PEER-REVIEWED A model that assumes smooth progression is therefore assuming the exception.

2

The factors that predict junior success tend to run opposite to those that predict adult world-class success.

Higher-performing juniors started their main sport earlier and progressed faster, while adult world-class athletes started later, did more varied practice as children and progressed more slowly at first.² **REVIEW** A companion meta-analysis reports the same pattern across 6,096 athletes.³ **REVIEW**

3

Talent identification has no agreed set of predictors to learn from.

A systematic review found high variability in the factors that separated more and less skilled athletes, and a sample skewed towards physical measures and male athletes.⁴ **REVIEW** Reviewers in the field advise against selection at younger ages.⁵ **REVIEW**

4

Training data in talent systems carry the biases of the selection that produced them.

Relative age effects appear consistently across sports and are strongest in adolescent males at representative level.⁶ **REVIEW** Athletes who were deselected or left usually vanish from the records, and I have found no study that measures how much this distorts a talent model. **EVIDENCE GAP**

5

A prediction that is acted upon changes the outcome it predicts.

Sport psychology already describes how the expectations of coaches, parents and athletes can make early advantages self-sustaining,⁷ **PEER-REVIEWED** and machine learning research describes the same loop formally for any model whose outputs drive decisions.

⁸ **PEER-REVIEWED** In a talent system the two loops are likely to reinforce each other.

6

Maturation confounds almost every physical measure taken in adolescence.

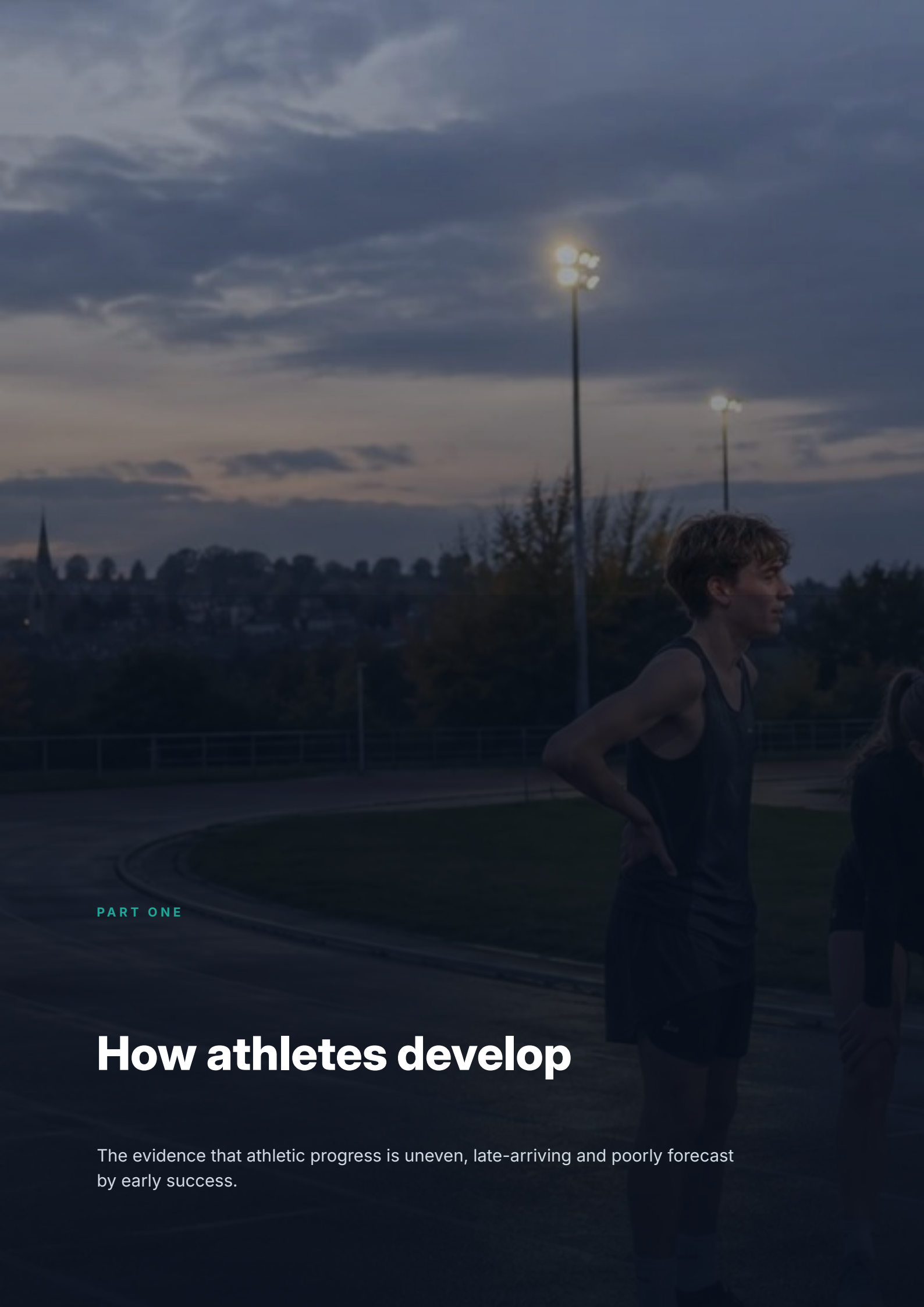
Maturity offset can be estimated from simple anthropometry, but only as an estimate with error.⁹

PEER-REVIEWED Bio-banding shows that sport can group and judge young athletes by biological as well as chronological age,¹⁰ **REVIEW** but in my observation neither practice is yet standard in the data products offered to academies. **PRACTICE**

What follows for a club or governing body

The immediate step is to separate two uses of data that are often bundled together. Describing an athlete's current capacities, with maturation context and error bands, is useful and defensible. Ranking children by predicted future potential, or letting a score shape a deselection, rests on evidence that does not exist for this age group and on data that are known to be skewed. The first use deserves investment. The second should wait for evidence that nobody has yet produced.

I recommend that clubs and governing bodies bar ranked or predictive outputs for athletes below mid-adolescence, require any tool to show its uncertainty and its maturation adjustment, name a person who is accountable for every selection decision a score informs, and keep records of those who leave so that the system can one day learn from its misses. **ATHLEET.AI**



PART ONE

How athletes develop

The evidence that athletic progress is uneven, late-arriving and poorly forecast by early success.

CHAPTER ONE

The 14-year-old who would have been cut

Talent systems make their most consequential decisions at the age when the evidence for those decisions is weakest, and predictive tools risk making those decisions faster and harder to question.

The case for algorithmic talent identification is easy to state and superficially persuasive. Human selectors are inconsistent, they favour what they can see on the day, and they carry biases they cannot always name. A model, the argument runs, would weigh the same evidence the same way every time, and could draw on more of it. My own view is that the first half of that argument is largely right and the second half assumes something that the development literature does not support, which is that the evidence a model would weigh contains a reliable signal about adult performance in the first place.

1.1 A composite, and why it matters

ILLUSTRATIVE SCENARIO

The late grower in an invented county pathway

Sam is an invented composite. At 13, Sam trains with a regional team-sport pathway run by a fictional county association. Sam was born in late summer, is among the smallest in the squad, and had started the sport only two years earlier after several seasons of swimming and athletics. The pathway has recently trialled a vendor dashboard that combines sprint times, a jump test, match minutes and a coach rating into a single "potential index". Sam sits in the bottom quartile. At the end-of-season review, the index is one of four items on the page, and it is the only one expressed as a number. Sam is released. Two years later, after a growth spurt and a season in a local club, Sam is faster than most of the retained group, but nobody in the pathway is looking any more, and the dashboard has no record that Sam ever improved.

Nothing in that scenario requires a bad model or a careless coach. The index might have been built competently on historical data, and the review panel might have treated it as one input among several. The difficulty is structural. Every feature the index used was confounded by relative age and maturation, the historical data it learned from contained almost no athletes like Sam who were released and later flourished, and the release decision removed the one thing (continued exposure to good coaching) that might have let Sam's development show. I return to each of these mechanisms in later chapters, since together they form the core of the argument.

1.2 Two literatures that rarely meet

There is a growing body of work applying machine learning to performance prediction and injury risk in sport. A systematic review of AI in team sports identified 58 studies using 11 AI techniques across 12 sports, with a pooled sample that was 97% male and 76% professional.¹¹ **REVIEW** A broader survey of AI for team sports covers match outcome prediction, tactical decisions, player investment and injury prediction, and notes weaknesses in the models as well as strengths.¹² **REVIEW** Neither review is concerned with children, and neither claims to be.

A separate and well-developed body of work shows that athlete development is non-linear and that junior success is a weak guide to senior success.¹ **PEER-REVIEWED** The two literatures rarely cite each other. In my reading of both, almost nobody has asked what happens when predictive tools, trained largely on the data of athletes who were selected and survived, are pointed at 12-year-olds. **EVIDENCE GAP**

The practical setting makes the question more pressing. Wearable and GPS monitoring is normal in professional sport and, in my observation, is moving steadily into academies and school-age programmes, where the same dashboards appear with younger names on them. **PRACTICE** Consumer generative AI entered the conversation more abruptly. OpenAI released ChatGPT as a research preview on 30 November 2022,¹³ **INDUSTRY** and within weeks I had heard coaches describe using it to draft sessions and plans. **PRACTICE** None of this yet amounts to algorithmic selection of children, though it lowers the cost of building something that looks like it.

CHAPTER TWO

What the development evidence shows

The best available evidence describes elite development as uneven, frequently interrupted and poorly predicted by early results.

Any predictive tool embeds a theory of development, whether its builders state one or not. If the tool treats a high score at 12 as evidence of a high ceiling at 25, it assumes that performance rank is reasonably stable across adolescence. This chapter sets out what the evidence says about that assumption.

2.1 Trajectories that bend

The most informative single study remains the analysis of the Australian National Athlete Development Survey by Gulbin and colleagues. They asked 256 elite athletes from 27 sports to describe their competitive history across junior and senior levels and grouped the answers into three broad trajectories.¹ [PEER-REVIEWED](#) Only 16.4% followed a pure ascent. The majority, 57.4%, followed what the authors call a mixed descent, and overall 83.6% of the athletes reported non-linear development, with pure junior to senior linearity in fewer than 7% of cases.¹ [PEER-REVIEWED](#) Linear progress was the exception.

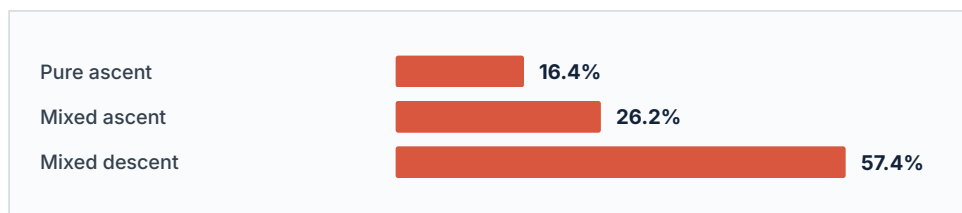


FIGURE 1 Share of 256 elite athletes by development trajectory. Source: Gulbin et al. (2013).

The study has limits that a careful reader should keep in view. It is retrospective, it relies on athletes' own accounts, and it samples only those who reached the elite level, so it can say how successful athletes got there without saying how many similar young athletes fell away. Even so, the direction of the finding is hard to dismiss. The authors also report that athletes in sports measured in centimetres, grams or seconds were less likely to experience a descending trajectory than athletes in other sports (43% against 70%),¹ [PEER-REVIEWED](#) which suggests that the degree of non-linearity varies by sport and that a model built for one sport may not transfer to another.

2.2 Opposite predictors at junior and senior level

A second line of evidence concerns what early participation predicts. Barth and colleagues meta-analysed 71 study reports covering 262 athlete samples, 685 effect sizes and 9,241 athletes.² [REVIEW](#) Compared with national-class peers, adult world-class athletes had done more coach-led practice in other sports during childhood and ad-

olescence, started their main sport later, accumulated less main-sport practice and progressed more slowly at first. Among juniors the pattern reversed, so that higher-performing juniors had started earlier, practised their main sport more, done less in other sports and progressed faster.² **REVIEW** The authors graded the quality of evidence as moderate to high for all predictors except peer-led play.² **REVIEW**

Güllich, Macnamara and Hambrick reached the same conclusion in a meta-analysis of 51 study reports, 477 effect sizes and 6,096 athletes, including 772 of the world's top performers.³ **REVIEW** Their interpretation is that varied early practice is associated with slower initial progress but more sustainable long-term development.³ **REVIEW**

The implication for prediction is uncomfortable. If early specialisation, early start and fast early progress are the features that distinguish good juniors, then a model trained to recognise good juniors will learn to reward exactly the profile that the adult evidence associates with lower odds of world-class performance. I would not put it more strongly than that, since these are associations at group level that say little about any one child, but a model optimised on junior outcomes is optimising for the wrong target if the aim is adult excellence.

2.3 Models of development and their critics

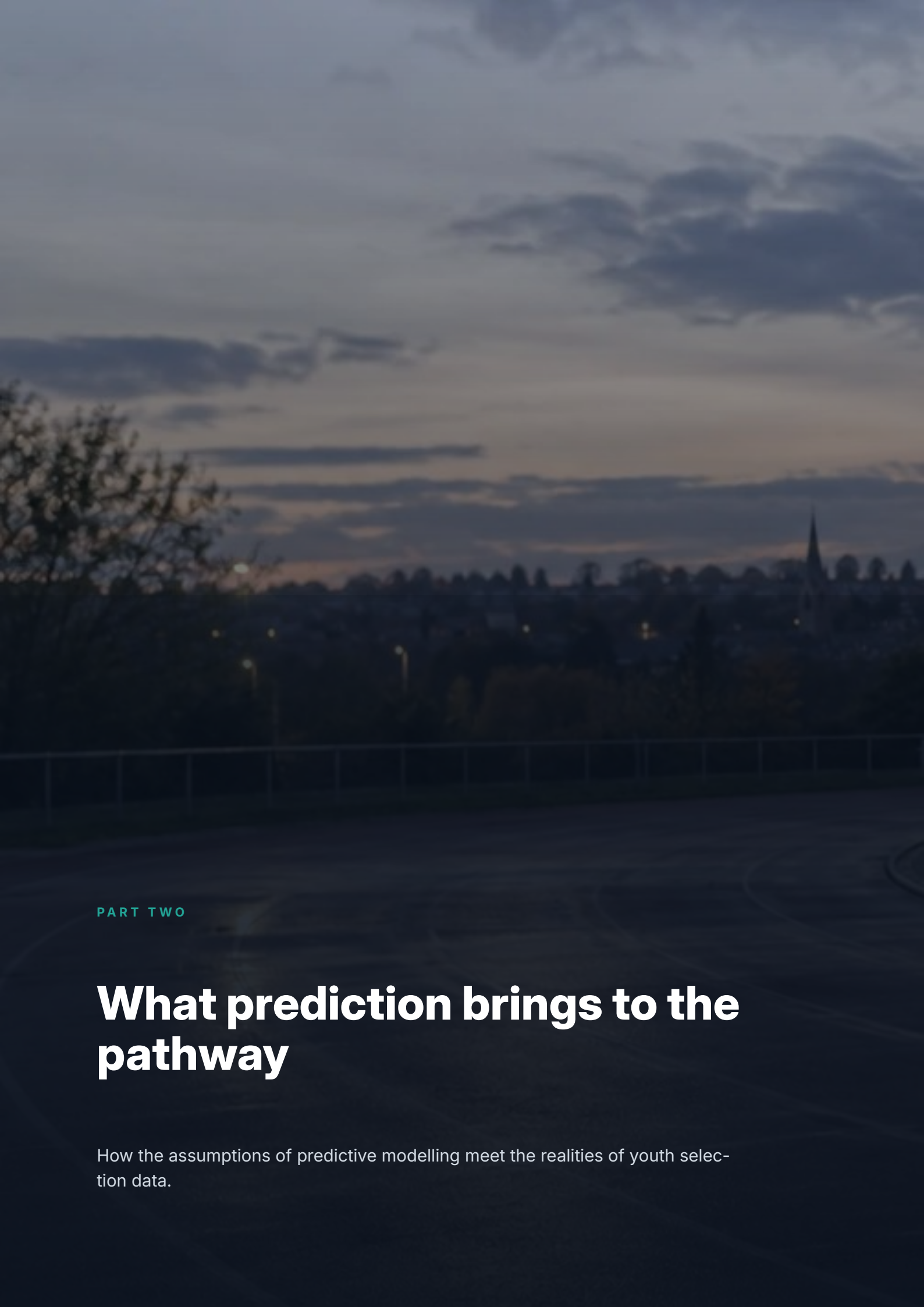
Governing bodies have long tried to turn this complexity into usable guidance. The Long-Term Athlete Development model set out stages of development and proposed periods in which particular capacities are especially trainable.¹⁴ **PRACTICE** The model was taken up by a variety of national governing bodies in the United Kingdom, yet Ford and colleagues concluded that it lacks an empirical base, that it is one-dimensional and generic, and that it should be treated as a work in progress.¹⁵ **REVIEW** The Youth Physical Development model, proposed partly in response, treats the development of strength, movement skill and other qualities as ongoing across childhood and adolescence and ties training emphasis to growth and maturation.¹⁶ **REVIEW**

The broader reviews point the same way. The Great British Medalists project reviewed evidence on the performer, the environment and practice, rated the quality of that evidence, and called for studies of the complex interactions that the authors expect exist across those domains.¹⁷ **REVIEW** Johnston and colleagues found only 20 studies meeting their criteria for a systematic review of talent identification, with 60% focused on physical profiles, 65% on male samples and 60% on athletes aged between ten and 20, and a high degree of variability in which factors discriminated between more and less skilled athletes.⁴ **REVIEW** Till and Baker describe talent as a complex and largely misunderstood phenomenon lacking a strong research base, and they do not recommend talent identification and selection at younger ages.⁵ **REVIEW**

What a machine learning system would need, in other words, is a stable and well-understood relationship between measurable childhood features and adult outcomes. The literature describes something closer to its opposite, and it is candid about how little of the relationship has been measured at all.

JUNIOR LABELS

A model's target variable is a choice. Training on "selected for the under-15 squad" teaches the model to predict under-15 selection, which the evidence suggests is a different quantity from adult excellence.



PART TWO

What prediction brings to the pathway

How the assumptions of predictive modelling meet the realities of youth selection data.

CHAPTER THREE

What predictive AI assumes

Supervised learning rests on assumptions that are reasonable in many settings and weak in youth sport, and a model that cannot state them cannot be checked against them.

A supervised model learns a mapping from inputs to a labelled outcome using historical cases, then applies that mapping to new cases. For the mapping to transfer, three conditions need to hold at least approximately. The relationship between inputs and outcome must be reasonably stable over time, the historical cases must represent the population the model will be used on, and the outcome label must mean what the users think it means. Each is well understood in the machine learning literature, and each is fragile in talent identification.

3.1 Where the field stood

Most published AI work in sport targets adults. The review by Claudino and colleagues found artificial neural networks, decision trees, support vector machines and Markov processes to be the most frequent techniques, with football, basketball, handball and volleyball the most frequent sports.¹¹ REVIEW The authors report good performance metrics for the main techniques, and they also call for prospective evaluation to establish predictive performance.¹¹ REVIEW That caveat carries extra weight once the population changes from professionals in their mid-20s to children whose bodies are changing month by month.

Beal and colleagues divide team-sport computing problems into prediction, tactics, player investment and injury.¹² REVIEW Player investment is the nearest analogue to talent identification in that survey, but it concerns adults with performance histories, where past output is a reasonable if imperfect guide to near-term output. I have not found published, prospectively validated evidence that machine learning can forecast adult elite status from measures taken before or around puberty. EVIDENCE GAP

3.2 Three assumptions under strain

The first assumption is stationarity. Machine learning researchers use the term concept drift for changes in the relationship between inputs and target over time, and a substantial literature exists on detecting and adapting to it.¹⁸ REVIEW Talent pathways drift constantly. Coaching philosophies change, competition formats change, selection ages move, and the sport itself evolves, so that the profile that produced senior internationals a decade ago may not be the profile that will produce them a decade from now. A model trained on the athletes of the past learns the pathway of the past. (Adult professional data drift too, but there the lag between input and outcome is months and can be monitored, whereas in youth talent work it is often ten years or more.)

The second assumption is representativeness. The records a club or governing body holds are, almost by definition, records of those it selected. The next chapter develops this point, since it is the one I think most tool builders underrate.

The third assumption is a stable and meaningful label. Obermeyer and colleagues showed how a widely used health algorithm, which used health care costs as a proxy for health need, produced racially biased results because the proxy did not measure the same thing for different groups.¹⁹ PEER-REVIEWED The parallel in sport is direct, in my view. Labels such as "signed a professional contract" or "selected for a national age-group squad" are proxies for talent, and they depend partly on the same biases (relative age, maturation, access) that a fair tool would hope to correct.

Assumption	What supervised learning needs	What youth sport tends to provide
Stationarity	A stable link between inputs and outcome	Pathways, formats and selection ages that change over a decade
Representative sample	Training cases like the cases scored	Records dominated by selected and retained athletes
Meaningful label	An outcome that measures the target	Proxy labels such as squad selection or contract
Independent outcome	Predictions that do not alter results	Selection decisions that change training and opportunity
Comparable inputs	Features measured on a common scale	Physical tests confounded by age and maturation

TABLE 1 Assumptions of supervised prediction set against typical conditions in youth talent systems. Source: author's analysis drawing on Lu et al. (2019), Obermeyer et al. (2019) and Perdomo et al. (2020).

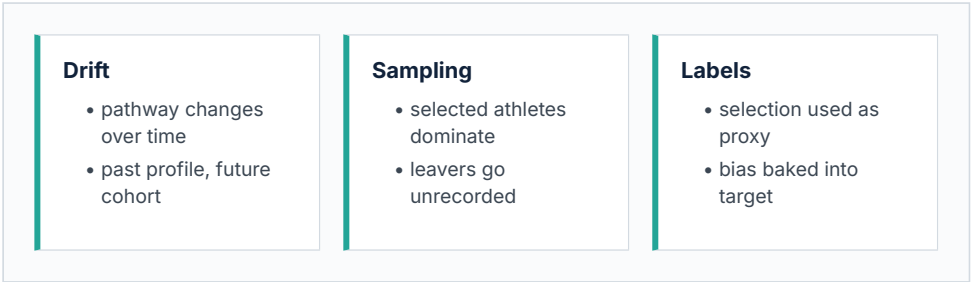


FIGURE 2 Three ways a talent model can inherit error. Source: author's framework.

CHAPTER FOUR

Survivorship in the training set

A talent dataset is a record of past selection decisions, and a model trained on it learns those decisions along with whatever they got wrong.

Survivorship bias is the error of drawing conclusions from cases that passed a filter while ignoring those that did not. In talent systems, the filter is selection itself, applied repeatedly from the first trial to the senior squad. It is rarely neutral. My argument in this chapter is that survivorship is a built-in property of talent data, and that relative age effects show how large the inherited distortion can be.

4.1 Data from those who stayed

A typical pathway mostly records the athletes it kept. It holds test results, match data and coach assessments for athletes who were in the programme, and it usually stops recording when an athlete is released, drops out or moves to another sport. A model trained to predict who reaches the senior level therefore learns from two groups, those who were retained and succeeded and those who were retained and did not. The athlete who was released at 13 and flourished at 19 is largely absent, and so the model has little opportunity to learn that such athletes exist. **PRACTICE**

This matters more given the development evidence. If most elite athletes followed non-linear trajectories,¹ **PEER-REVIEWED** and if adult world-class athletes were typically slower to progress early,² **REVIEW** then the athletes most likely to be released early include a disproportionate share of those with late-arriving potential. The dataset under-represents precisely the group the model most needs to recognise. I know of no study that quantifies this effect for a talent model, and it would be hard to design one, since the missing athletes are missing. **EVIDENCE GAP**

4.2 Relative age as inherited bias

Relative age effects offer the best measured example of a selection bias that a model would inherit. Copley and colleagues re-examined 38 studies published between 1984 and 2007, containing 253 independent samples across 14 sports and 16 countries.⁶ **REVIEW** They found relative age effects to be consistently present, with small effect sizes overall that increased with the relative age difference. Age category, skill level and sport context moderated the effect, and the authors identified adolescent males aged 15 to 18, at regional and national representative level in highly popular sports, as most at risk.⁶ **REVIEW**

Two features of that finding are relevant to model design. The effect is strongest at the representative levels that generate most talent data, and the authors attribute relative age effects partly to physical differences and selection practices within annual age groups.⁶ **REVIEW** A model trained on representative squads will therefore see a popu-

LEAVERS

Recording what happens to released athletes is cheap compared with building a model. Without those records, no system can learn how often its early judgements were wrong.

lation in which being born early in the selection year is associated with being selected, and if the label is selection, the model can learn birth date (or any feature correlated with it, such as height or sprint speed) as a marker of talent.

Removing date of birth from the inputs does not solve this, because the physical features that remain carry the same information. A model that cannot see age can still see size. The more honest approach is to make age and maturation explicit, adjust for them, and report the adjustment, which is the subject of the final chapter of Part two.

I would add one caveat. Copley and colleagues called for more evidence on female and more culturally diverse contexts,⁶ REVIEW so the size of the effect outside the populations they studied was less certain at the time of their review. That uncertainty is itself a reason for caution when a tool is sold across sports and sexes on the strength of evidence from one group.

CHAPTER FIVE

The prediction that makes itself true

In talent systems a forecast is also an intervention, because the athlete labelled as promising receives more of the things that produce improvement.

Ordinary forecasting assumes the forecast does not change the weather. Talent identification violates that assumption as a matter of routine, since the purpose of identifying talent is to treat the identified athletes differently. This chapter examines what that loop does to a predictive tool.

5.1 Pygmalion, Galatea and Matthew

Hancock, Adler and Côté proposed a theoretical model to explain relative age effects through the behaviour of social agents.⁷ [PEER-REVIEWED](#) They argue that parents influence relative age effects through Matthew effects, in which initial advantages attract further advantages, that coaches do so through Pygmalion effects, in which higher expectations lead to better treatment and better performance, and that athletes themselves contribute through Galatea effects, in which their own expectations shape their effort and results.⁷ [PEER-REVIEWED](#) The model was offered as a framework for testing, and the authors are explicit that the field had until then been largely descriptive.⁷

[PEER-REVIEWED](#)

Read alongside the development evidence, the model describes a mechanism by which an early judgement can become self-confirming. An athlete selected at 12 gets better coaching, more competition and better facilities, and is likely to develop a more confident self-image as well. The athlete released at 12 gets less of each. When the two are compared at 18, the difference in outcome partly reflects the difference in treatment, and a retrospective analysis will record the early judgement as accurate.

5.2 Performative prediction

Machine learning research has recently formalised the same problem. Perdomo and colleagues describe predictions as performative when they influence the outcome they aim to predict, and they note that when this is ignored it shows up as distribution shift that is routinely addressed by retraining.⁸ [PEER-REVIEWED](#) Their framework concerns the conditions under which repeated retraining settles on a stable point, meaning that predictions are calibrated against the outcomes that follow from acting on them. That is a technical result, and I do not want to stretch it. Its relevance here is conceptual. In a talent system, retraining a model on outcomes produced by its own earlier selections will tend to confirm the model, whether or not its original judgements were sound. My inference, which goes beyond what the paper itself addresses, is that a talent model can become more confident over time without becoming more accurate. [ATHLEET.AI](#)

The human version of this loop has existed for as long as there have been selectors. What an algorithmic score adds is harder to see. A number looks objective, it travels easily between meetings and documents, and it is rarely accompanied by the hesitation a coach would express in conversation. In my experience of review panels, a single numerical item on a page attracts attention out of proportion to its evidential weight. **PRACTICE**

A score that shapes a deselection has become part of the treatment. Once a prediction changes which athletes receive coaching and competition, later outcomes can no longer be used as independent evidence that the prediction was right. Any evaluation of a talent tool has to account for this, or it will measure the tool's influence and call it accuracy.

5.3 Accountability when the loop closes

Two practical questions follow. The first is who is responsible when a score contributes to a child's release. Vendors will say the tool only informs a decision, and clubs will say they relied on the best available evidence, so the responsibility can end up resting with nobody in particular. I think this is one of the least discussed risks in the area. **PRACTICE** The second question is how an organisation could ever learn that its tool was wrong, given that it rarely tracks the athletes it lets go. I return to both in Part three.

CHAPTER SIX

Maturation as a confounder

Adolescent athletes differ in biological age by years within a single birth cohort, and almost every physical measure a talent model might use moves with that difference.

Relative age concerns when in the selection year a child was born. Maturation concerns how far through puberty the child has progressed, which can vary widely among children of the same chronological age. The two effects overlap but are distinct, and for model design the second is arguably the larger problem, since it affects speed, power, strength and size directly.

6.1 Estimating maturity

Mirwald and colleagues developed equations that estimate maturity offset, meaning the time before or after peak height velocity, from simple anthropometric measures such as height, sitting height, leg length and body mass alongside age.⁹ **PEER-REVIEWED** The method is non-invasive and widely used in sport, which makes it attractive for any data system. It is also an estimate with an error band, derived from particular longitudinal samples, and its outputs should be treated as approximate.⁹

PEER-REVIEWED

For a predictive tool this creates a layered uncertainty. A physical test result is measured with some error; the maturity estimate used to interpret it carries its own error; and the relationship between adjusted current performance and adult performance is, as Part one showed, weak and poorly characterised. Stacking these uncertainties does not remove them, and a tool that reports a single potential score has, in effect, hidden all three.

The practical consequence for any data system is that the timing of a test matters as much as its result. Growth does not wait for the testing calendar. Two athletes tested in the same week may sit on opposite sides of their growth spurt, one probably gaining speed and power quickly and the other yet to start, and a snapshot comparison will favour whichever is further along. Repeated measurement helps, since a trend over several seasons says more about development than a single score, though it also gives a model more data to over-interpret. In my own coaching the most useful maturity information has often been the simplest, a height measured regularly and plotted, which lets coach and athlete see the growth spurt arrive and adjust their expectations. **PRACTICE**

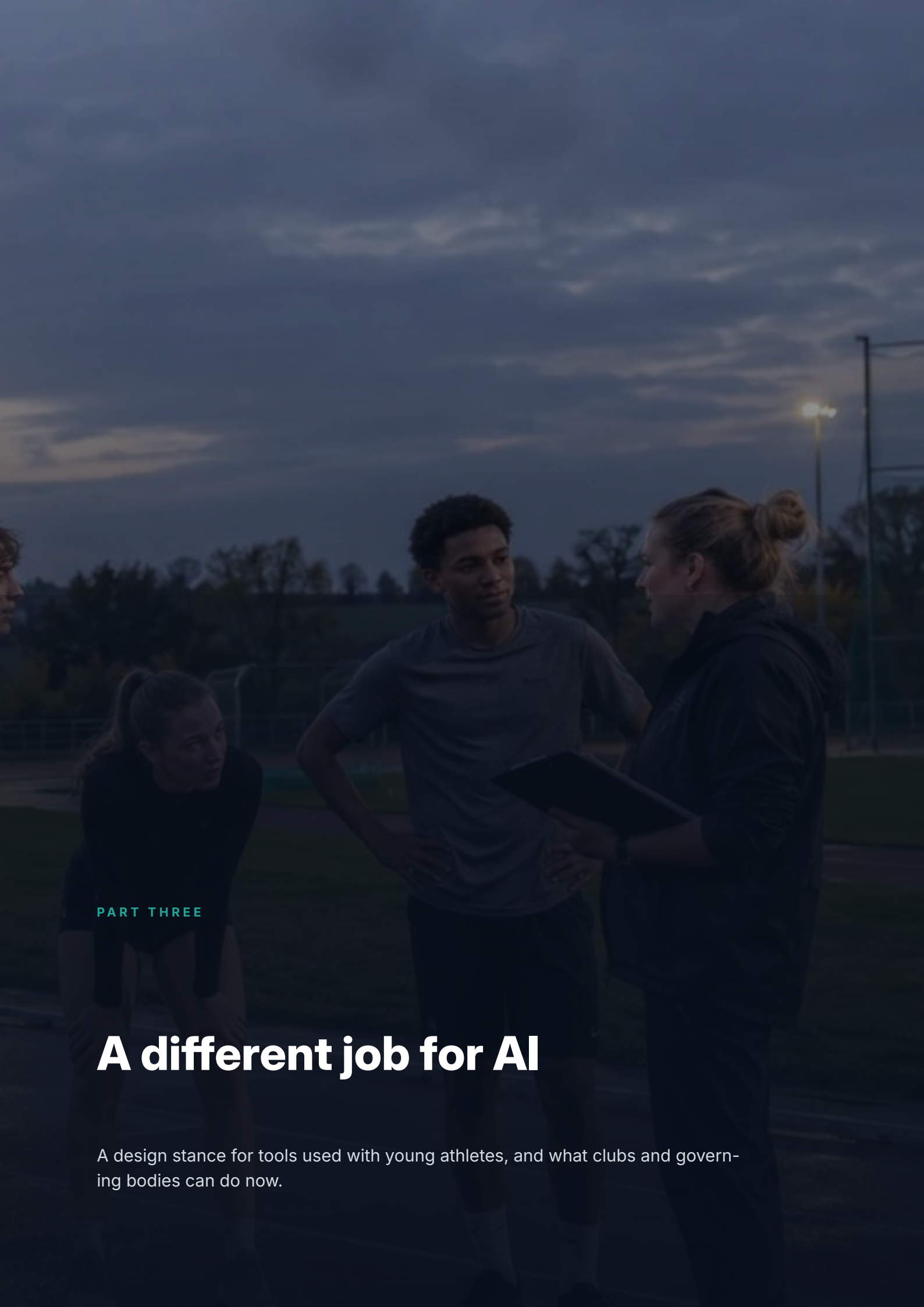
6.2 What bio-banding teaches

Bio-banding is the practice of grouping young athletes by maturity status, usually estimated from the percentage of predicted adult height they have reached.¹⁰ **REVIEW** Cumming and colleagues describe applications to competition, talent identification

and strength and conditioning, and they present bio-banding as a way to reduce the maturity-related mismatches that age-group sport creates between early and late developers.¹⁰ **REVIEW**

Three lessons carry over to model design, in my view. Maturation status should be an explicit, visible part of any assessment and should not be left for a model to infer from correlated features. Comparisons should be made within maturity bands as well as within age groups. And an assessment made during the adolescent growth spurt should be treated as provisional, since relative standing can change as peers catch up. The Youth Physical Development model reinforces the second point by tying training emphasis to growth and maturation throughout childhood and adolescence.¹⁶ **REVIEW**

A late-maturing athlete with average sprint times for chronological age may be performing well for biological age. A model that sees only chronological age will score that athlete below an early maturer with better times, and will be right about the present and possibly wrong about everything else. **ATHLEET.AI**

A photograph of a coach, a young man, and a young woman on a sports field at dusk. The coach, a woman with blonde hair in a bun, is holding a clipboard and talking to the young man. The young man is standing with his hands on his hips. The young woman is in a starting crouch. The background shows a sports field with a goalpost and a light pole under a cloudy sky.

PART THREE

A different job for AI

A design stance for tools used with young athletes, and what clubs and governing bodies can do now.

CHAPTER SEVEN

Describing the athlete in front of you

The honest role for AI in talent systems is to describe where an athlete is now, with its uncertainty and context shown, and to leave judgements about the future with accountable people.

The preceding chapters point towards a firm conclusion about what AI should not be asked to do with young athletes. They also leave a good deal that it can do well. Current capacities can be measured, tracked, adjusted for age and maturation and shown with their uncertainty, and that information can help coaches plan training and hold better conversations with athletes and parents. The design stance I propose treats description as the product and prediction as out of scope for the youngest athletes.

ATHLEET.AI

7.1 What a responsible tool reports

1	Current status	Show measured capacities today, with the date and conditions of each test.
2	Uncertainty	Show measurement error and the error in any maturity estimate beside each result.
3	Maturation context	Show estimated maturity status and compare within maturity bands as well as age groups.
4	Change over time	Show each athlete's own trend across several seasons.
5	No verdict	Produce no ranking, potential score or selection recommendation below mid-adolescence.

FIGURE 3 A reporting sequence for tools used with young athletes. Source: author's framework.

The first four steps are ordinary good practice in athlete monitoring, and none of them requires advanced machine learning or data that most pathways do not already collect. The fifth is different in kind, since it limits what the tool may say. A tool built this way still helps a selector, because it gives a fuller and fairer picture of the athlete and of how that athlete is changing. It simply declines to compress that picture into a number that claims to know the future. **ATHLEET.AI**

Where older athletes are concerned, and where predictive outputs are used, I would require the builder to state the training population, the label, the evaluation design and the steps taken to account for relative age, maturation and survivorship. A builder that cannot answer those questions has not yet built a tool fit for selection.

ATHLEET.AI

7.2 Implications for clubs and governing bodies

My recommendations are as follows. **ATHLEET.AI**

- Bar ranked or predictive outputs, including potential scores, for athletes below mid-adolescence, and treat any such output for older adolescents as advisory and provisional.
- Require every tool used in selection to display maturity status and measurement uncertainty alongside each result, and to report how relative age was handled.
- Name an accountable person for every selection or deselection decision that a score informs, and record the reasons in words as well as numbers.
- Keep simple follow-up records for released and departing athletes, so that the system can measure its own misses and correct its tools and its judgement.

Governing bodies are well placed to set these as conditions for pathway funding or accreditation, since individual clubs have limited influence over vendors. The white-label dashboards now reaching academies are rarely built with these conditions in mind, in my observation, and procurement is the most practical point at which to insist on them. **PRACTICE**

None of these steps prevents a club from using data. They change what the data are allowed to say, and they create the feedback that talent systems have mostly lacked.

7.3 Conclusion and open questions

Talent identification asks for a forecast about a system, the developing athlete, whose course the evidence describes as non-linear, whose early signals point in the wrong direction as often as the right one, and whose future is shaped by the decision that the forecast itself informs. Predictive AI does not change those facts. Pointed at young athletes and trained on the records of past selection, it is more likely to entrench existing biases than to correct them, and to do so behind a number that looks neutral. The more useful contribution is descriptive, and it is achievable with tools that already exist.

Several questions remain open, and I do not think the evidence yet answers them. Should any AI tool used with under-16s be barred outright from producing ranked or predictive outputs, or is a mid-adolescent threshold defensible for some sports and not others? How would a model account for athletes who leave a pathway and return years later, when most systems do not record the leaving? Who, in law and in practice, is accountable when an algorithmic score shapes a deselection? And would a talent system that tracked its released athletes for a decade discover that its early judgements were better, or worse, than it currently assumes? **EVIDENCE GAP**

END MATTER

References

1. Gulbin, J., Weissensteiner, J., Oldenziel, K. and Gagné, F. (2013) Patterns of performance development in elite athletes. *European Journal of Sport Science* 13(6), 605-614.
2. Barth, M., Güllich, A., Macnamara, B.N. and Hambrick, D.Z. (2022) Predictors of junior versus senior elite performance are opposite: a systematic review and meta-analysis of participation patterns. *Sports Medicine* 52(6), 1399-1416.
3. Güllich, A., Macnamara, B.N. and Hambrick, D.Z. (2022) What makes a champion? Early multidisciplinary practice, not early specialization, predicts world-class performance. *Perspectives on Psychological Science* 17(1), 6-29.
4. Johnston, K., Wattie, N., Schorer, J. and Baker, J. (2018) Talent identification in sport: a systematic review. *Sports Medicine* 48(1), 97-109.
5. Till, K. and Baker, J. (2020) Challenges and [possible] solutions to optimizing talent identification and development in sport. *Frontiers in Psychology* 11, 664.
6. Cobley, S., Baker, J., Wattie, N. and McKenna, J. (2009) Annual age-grouping and athlete development: a meta-analytical review of relative age effects in sport. *Sports Medicine* 39(3), 235-256.
7. Hancock, D.J., Adler, A.L. and Côté, J. (2013) A proposed theoretical model to explain relative age effects in sport. *European Journal of Sport Science* 13(6), 630-637.
8. Perdomo, J.C., Zrnic, T., Mendler-Dünner, C. and Hardt, M. (2020) Performative prediction. *Proceedings of the 37th International Conference on Machine Learning*, PMLR 119, 7599-7609.
9. Mirwald, R.L., Baxter-Jones, A.D.G., Bailey, D.A. and Beunen, G.P. (2002) An assessment of maturity from anthropometric measurements. *Medicine and Science in Sports and Exercise* 34(4), 689-694.
10. Cumming, S.P., Lloyd, R.S., Oliver, J.L., Eisenmann, J.C. and Malina, R.M. (2017) Bio-banding in sport: applications to competition, talent identification, and strength and conditioning of youth athletes. *Strength and Conditioning Journal* 39(2), 34-47.
11. Claudino, J.G., Capanema, D.O., de Souza, T.V., Serrão, J.C., Machado Pereira, A.C. and Nassis, G.P. (2019) Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. *Sports Medicine - Open* 5, 28.
12. Beal, R., Norman, T.J. and Ramchurn, S.D. (2019) Artificial intelligence for team sports: a survey. *Knowledge Engineering Review* 34, e28.
13. OpenAI (2022) ChatGPT: optimizing language models for dialogue. OpenAI blog, 30 November 2022. Available at: openai.com/blog/chatgpt.
14. Balyi, I. and Hamilton, A. (2004) *Long-Term Athlete Development: Trainability in Childhood and Adolescence*. Victoria: National Coaching Institute British Columbia and Advanced Training and Performance.
15. Ford, P., De Ste Croix, M., Lloyd, R., Meyers, R., Moosavi, M., Oliver, J., Till, K. and Williams, C. (2011) The Long-Term Athlete Development model: physiological evidence and application. *Journal of Sports Sciences* 29(4), 389-402.
16. Lloyd, R.S. and Oliver, J.L. (2012) The Youth Physical Development model. *Strength and Conditioning Journal* 34(3), 61-72.
17. Rees, T., Hardy, L., Güllich, A., Abernethy, B., Côté, J., Woodman, T., Montgomery, H., Laing, S. and 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.
18. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. and Zhang, G. (2019) Learning under concept drift: a review. *IEEE Transactions on Knowledge and Data Engineering* 31(12), 2346-2363.
19. Obermeyer, Z., Powers, B., Vogeli, C. and Mullainathan, S. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464), 447-453.

END MATTER

About the author

Paul Forrest is the founder of Athleet.AI. He advises boards and leadership teams on the adoption and governance of artificial intelligence, drawing on a background in business process reengineering and on hands-on work building machine learning systems, including small, domain-specific language models. He has served as a special adviser to UK government on the Global Entrepreneur Programme and works as a board member and business adviser.

He is also a qualified athletics coach who coaches sprint, hurdle and endurance athletes as a volunteer, and much of his thinking about how people learn, decide and improve comes from the track.

This paper is part of the Athleet.AI Working Papers, AI in Coaching and Athlete Development. The views expressed are the author's own.

athleet.ai

Predicting the Unpredictable

Athleet.AI Working Papers, AI in Coaching and Athlete Development
Paper one of six · 25 January 2023