



Anne Hecksteden · Ralf Kellner

Wissenschaftlicher Umgang mit den kleinen Fallzahlen in der Spitzensportforschung



Anne Hecksteden · Ralf Kellner

Wissenschaftlicher Umgang mit den kleinen Fallzahlen in der Spitzensportforschung

SPORTVERLAG *Strauß*

Schriftenreihe des Bundesinstituts für Sportwissenschaft 2021 | 02

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Herausgeber:

Bundesinstitut für Sportwissenschaft (BISp)
Graurheindorfer Str. 198
53117 Bonn
www.bisp.de

Ansprechpartner:

Dr. Marc Wonneberger
Graurheindorfer Straße 198
53117 Bonn
marc.wonneberger@bisp.de
www.bisp.de
Tel.: +49 228 99 640 9041
Fax: +49 228 99 640 9008

Hecksteden, Anne; Kellner, Ralf

Wissenschaftlicher Umgang mit den kleinen Fallzahlen in der Spitzensportforschung

Sportverlag Strauß, Hellenthal – 1. Aufl. 2021.
(Schriftenreihe des Bundesinstituts für Sportwissenschaft, Bd. 2021,02).
ISBN 978-3-86884-550-1

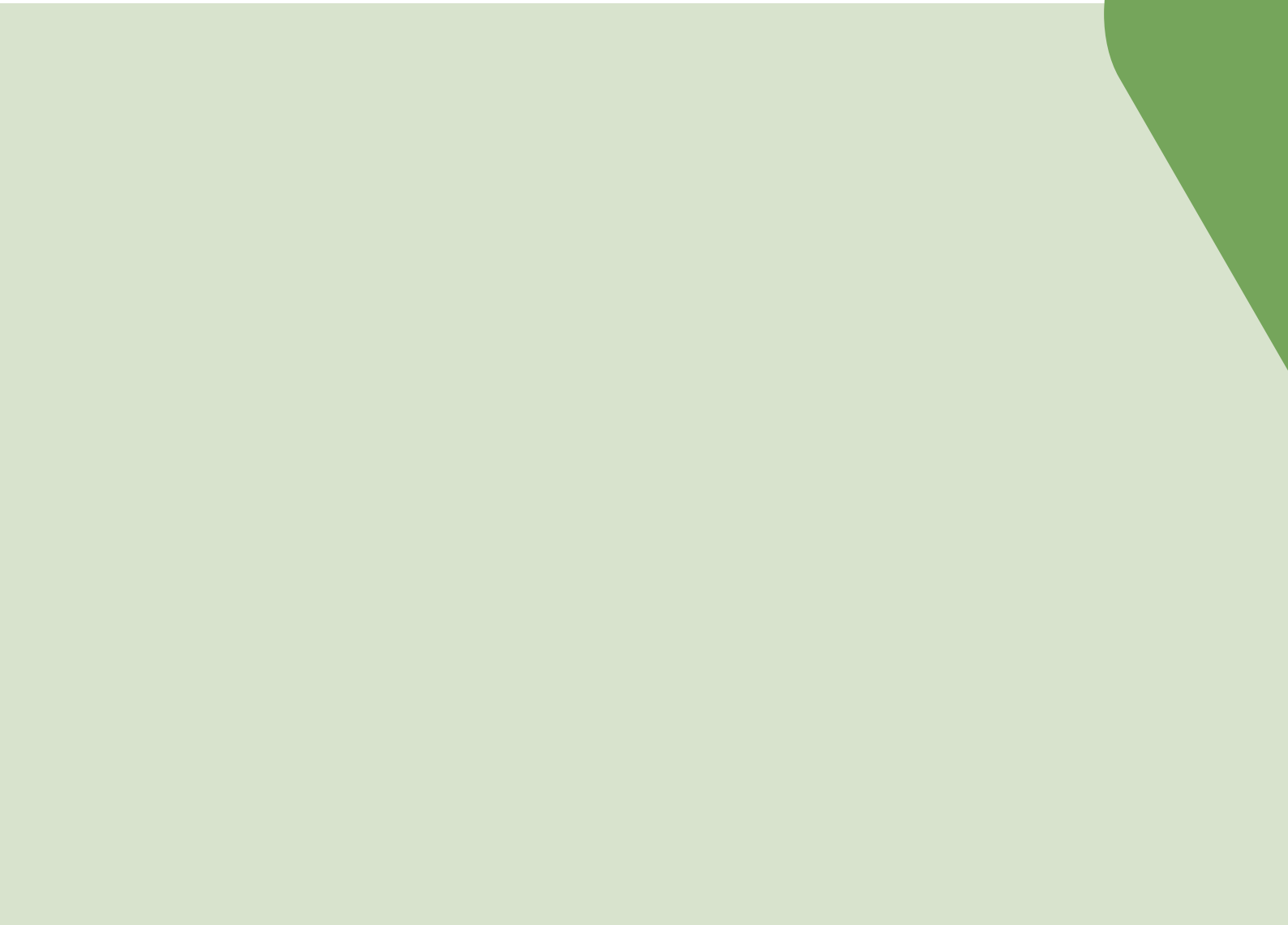
Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt.
Kein Teil dieses Werkes darf ohne ausdrückliche Genehmigung des Bundesinstituts für Sportwissenschaft digital oder analog vervielfältigt werden.
Die Empfehlungen und Erhebungsinstrumente in diesem Manual wurden von den Autoren und dem Verlag (Bundesinstitut für Sportwissenschaft) sorgfältig erwogen und geprüft. Dennoch kann keine Garantie übernommen werden, dass Daten fehlinterpretiert bzw. unsachmäßige Interventionen erfolgen. Eine Haftung der Autoren, des Herausgebers bzw. des Verlages und seiner Beauftragten für Personen-, Sach- und Vermögensschäden ist ausgeschlossen.

© SPORTVERLAG Strauß

Neuhaus 12 – 53940 Hellenthal
Fon (+49 2448) 247 00 40 - Fax (+49 2448) 919 56 10
E-Mail: info@sportverlag-strauss.de
<http://www.sportverlag-strauss.de>

Satz & Layout:	Sportverlag Strauß, Hellenthal
Umschlag:	Mike Hopf, Berlin
Umschlagfoto:	Gettyimages (Aleksandar Velasevi)
Herstellung:	Customized Business Services GmbH, Ferdinand-Jühlke-Straße 7, 99095 Erfurt
Printed in Germany	

Inhaltsverzeichnis



Inhaltsverzeichnis

	Abbildungsverzeichnis	7
1	Vorwort	11
2	Zusammenfassung	15
3	Problemstellung	19
3.1	Spitzensportforschung: Kleine Effekte und kleine Fallzahlen	21
3.2	Die Bedeutung der Fallzahl: Eine Auffrischung	22
3.2.1	Die Unterscheidung von Signal und Rauschen	22
3.2.2.	Das Grundprinzip: Zufällige Fehler mitteln sich aus	22
3.2.3	Die Rolle der Fallzahl speziell im Rahmen der Hypothesentestung	22
3.3	Kleine Fallzahl – Gravierende Probleme.	22
3.3.1	Bekannt problematisch: Negatives Ergebnis und geringe Power	22
3.3.2	Geringe Power und positives Ergebnis	23
3.4	Deshalb: Spezielles Instrumentarium & Replikationsversuche	24
4	Verbesserungsmöglichkeiten	25
4.1	Prävention, Pflicht & Kür: Strukturierter Umgang mit kleinen Fallzahlen	27
4.1.1.	Prävention: Benötigte und realisierbare Fallzahlen angleichen	27
4.1.2	Pflicht und Kür: Stufenweise Anpassung an kleine Fallzahlen	27
4.2	Projektstrategie	29
4.3	Transparente Rechtfertigung der Fallzahl	30
4.4.	Ab welcher Größe ist ein Effekt (praktisch) relevant?	30
4.4.1	Effekte als Vielfache der Variabilität – Effektstärken	31
4.4.2	Einschätzung der Sportpraxis	32
5	Instrumente für kleine Fallzahlen	33
5.1	Das Fundament: Methodische Grundprinzipien	35
5.2	Datenaufnahme	37
5.2.1	Grundüberlegungen	37
5.2.2	Spezielle Studiendesigns	38
5.2.3	Single-Subject Designs und Individualisierung	39
5.3	Datenanalyse	41
5.3.1	Bezug von Datenaufnahme und Datenanalyse	41
5.3.2	Generelle Aspekte mit besonderer Bedeutung bei kleinen Fallzahlen	42
5.3.3	Hierarchische Datenstruktur und gemischte Modelle	43
5.3.4	Resampling und Schrumpfschätzer	44
5.3.5	Strukturierte Berücksichtigung von Vorwissen & Bayesianische Statistik	45
5.3.6	Aktuelle Entwicklungen und „Magnitude Based Inference“	48
6	Von Einzelinstrumenten zum Studienplan	51
6.1	Vorbemerkungen – Wahlfreiheit und Transparenzpflicht	53
6.2	Typische Szenarien im Leistungs- und Spitzensport	53
6.2.1	Spitzensport Szenario	53

6.2.2	Pilotstudien Szenario	53
6.2.3	Big data scenario	54
7	Danksagungen	57
8	Literatur	61
9	Anhang.....	69
9.1	Statistische Software.....	71
9.1.1	Statistischer Code: R und R Studio	71
9.1.2	Python	71
9.1.3	JASP	71
9.2	Spezielle Anwendungen	71
9.2.1	Berechnungshilfen für Studien mit Single Subject Designs	71
9.2.2	Visualisierung von Daten und Ergebnissen	72
9.2.3	Fallzahlberechnung.....	72
9.3	Bayesianische Statistik	72
9.4	Gemischte Modelle	72

Abbildungsverzeichnis

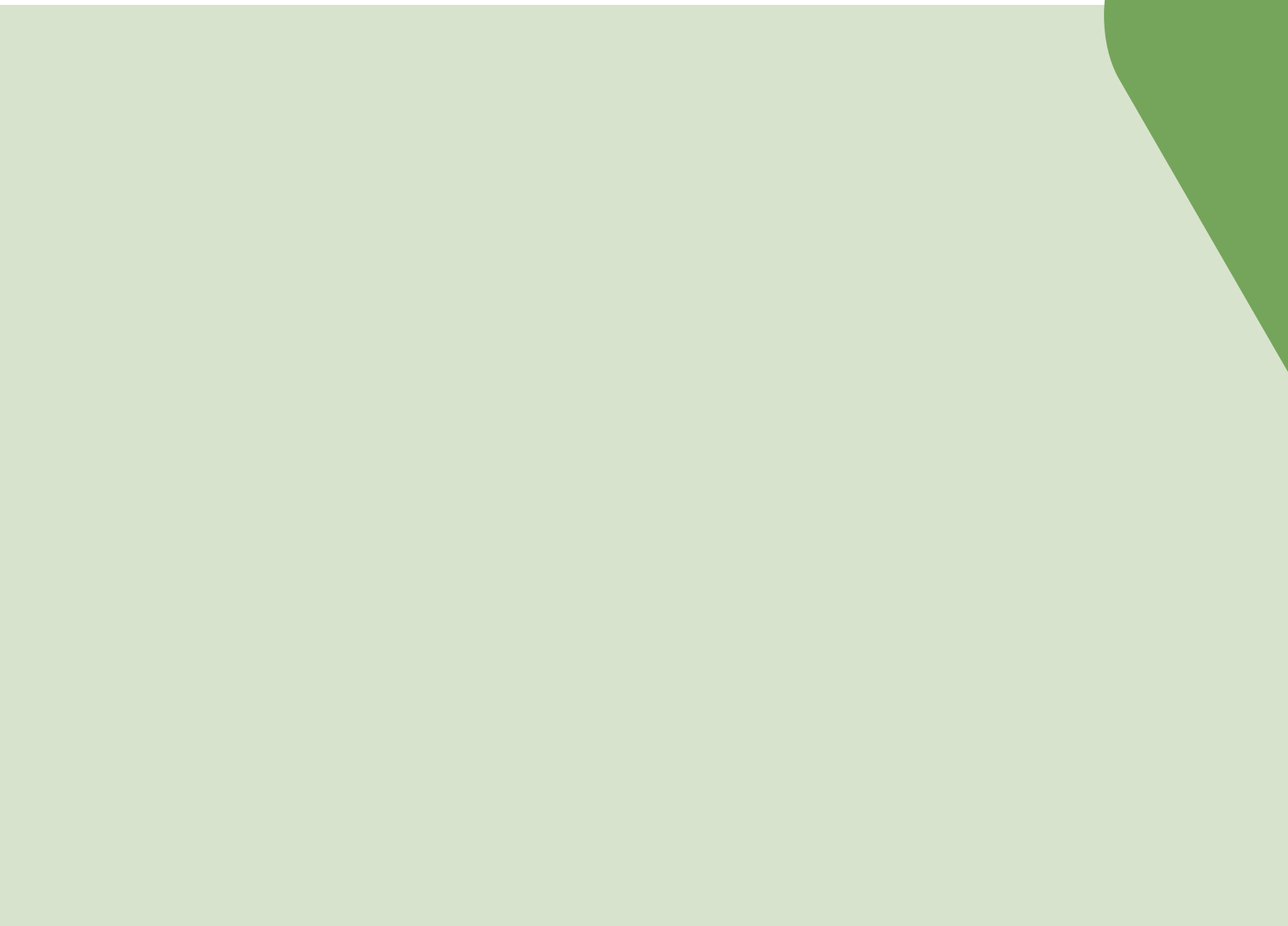


Abb. 1	28
Abb. 2	55
Abb. 3	73
Abb. 4	74

Vorwort

Diese Handreichung zum wissenschaftlichen Umgang mit den kleinen Fallzahlen im Leistungs- und Spitzensport wurde im Auftrag des Bundesinstituts für Sportwissenschaft erstellt. Die wesentlichen Teile einschließlich der Abbildungen 1 und 2 (in englischer Sprache) wurden international begutachtet publiziert (1).

Ziel der Expertise ist es, allen in diesem Bereich tätigen Personen ein Verständnis für die Problematik kleiner Fallzahlen und das zum Umgang mit den daraus entstehenden Herausforderungen verfügbare Instrumentarium zu ermöglichen. Im Interesse einer schlüssigen Darstellung werden auch Aspekte besprochen, die manchen Lesern vielleicht selbstverständlich erscheinen. Der Fokus der Expertise liegt auf (quasi-) experimentellen Studien und Single-Subject Designs. Eine Übertragung der Grundprinzipien auf andere Studientypen ist aber möglich. Wir hoffen, mit dieser Expertise zu einer höheren Genauigkeit und Verlässlichkeit der Ergebnisse wissenschaftlicher Studien im Leistungs- und Spitzensport und damit zu einem größeren Nutzen für die Sportpraxis beizutragen.

Ziel der Expertise ist es explizit nicht, eine statistische Beratung (oder noch besser die Zusammenarbeit mit einem ausgebildeten Statistiker innerhalb des Projektkonsortiums) zu ersetzen. Eine solche Zusammenarbeit ist fast immer vorteilhaft. Insbesondere bei der Fallzahlplanung für Studien mit limitierter Fallzahl, der erstmaligen Anwendung hierarchischer Studiendesigns und bei anspruchsvollen statistischen Optimierungsoptionen ist die Zusammenarbeit mit einem ausgebildeten Statistiker dringend empfehlenswert. Diese Überlegung sollte auch bei der Finanzplanung von Studien bedacht werden.

Abwägungen zwischen methodischen Optionen und die in den Abbildungen 1 und 2 vorgenommene Strukturierung sind, mit wenigen explizit kenntlich gemachten Ausnahmen, als begründete persönliche Einschätzung zu verstehen. Im Einzelfall liegt es im Ermessen des verantwortlichen Wissenschaftlers eine Entscheidung vorzunehmen und zu begründen.

Zusammenfassung

Kennzeichnend für den Leistungs- und Spitzensport ist eine hohe Leistungsdichte innerhalb einer kleinen Gruppe von Athleten. Für die Forschung ergibt sich daraus die Herausforderung mit (sehr) wenigen potenziellen Studienteilnehmern (sehr) kleine, aber doch relevante Effekte nachzuweisen. Diese Konstellation führt regelmäßig zu einem Missverhältnis zwischen (eigentlich) benötigter und realisierbarer Fallzahl.

Kleine Fallzahlen sind, unabhängig vom Ergebnis der Studie, mit erheblichen Problemen und Limitationen behaftet. Eine Durchführung sollte daher nur in Betracht gezogen werden, wenn benötigte und realisierbare Fallzahlen auch bei Ausschöpfung aller realistischer Instrumente nicht zur Deckung zu bringen sind und gleichzeitig ein dringlicher Forschungsbedarf besteht. Eine suboptimale Fallzahl ist dann zu rechtfertigen, wenn angenommen werden kann, dass die Ergebnisse der Studie dennoch zur Beantwortung der konkreten Forschungsfrage beitragen.

Zum „Pflichtprogramm“ von Studien mit kleiner Fallzahl gehören die optimale Umsetzung von methodisch-statistischen Standards, eine nachvollziehbare Kalkulation der benötigten und eine Begründung der realisierten Fallzahl sowie die explizite Berücksichtigung der bekannten „Nebenwirkungen“ bei der Interpretation der Ergebnisse. Die Fallzahlkalkula-

tion sollte eine praxisbezogene Abschätzung der Grenze relevanter Unterschiede beinhalten.

Aufbauend auf diesem „Pflichtprogramm“ existieren zahlreiche Optionen für die weitere Optimierung. Zentrale Instrumente sind die Aggregation von Einzelfallstudien („Single-Subject Designs“) und die strukturierte Berücksichtigung von Vorwissen (z. B. als informative Priorverteilung im Rahmen Bayesianischer Analysen). Eine kluge Studienplanung sollte das strategische Ziel des Projekts (generalisierbarer Erkenntnisgewinn oder Entscheidungsfindung innerhalb des untersuchten Settings), die Gründe für das Missverhältnis von benötigter und realisierbarer Fallzahl und vorteilhafte, aufeinander aufbauende Kombinationen einzelner Instrumente berücksichtigen. Die Einbindung eines ausgebildeten Statistikers bei der Ausgestaltung einer solchen „Kür“ ist dringlich empfehlenswert.

Durch das Zusammentreffen der Flexibilität des Instrumentariums mit der Instabilität von Effekten bei kleinen Fallzahlen sind ein detaillierter Datenaufnahme und -analyseplan und äußerste Transparenz bei der Kommunikation der Ergebnisse von großer Bedeutung; idealerweise in Form einer Registrierung des Studienplans und einer Veröffentlichung von anonymisierten Rohdaten und statistischem Code.

Problemstellung

3.1 Spitzensportforschung: Kleine Effekte und kleine Fallzahlen

Kernaufgabe der angewandten Forschung im Leistungs- und Spitzensport ist es, die Grundlagen für evidenzbasierte Entscheidungen im Trainingsprozess zu schaffen. Vor dem Hintergrund der Leistungsdichte müssen wissenschaftlich begründete Einschätzungen dabei hohen Anforderungen genügen. Nur so können Leistungsreserven systematisch erkannt, erschlossen und genutzt werden.

Dabei ist die Fallzahl von zentraler Bedeutung. Allgemein gilt: Je kleiner der nachzuweisende Effekt bzw. je höher die Anforderungen an die Präzision deskriptiver Indikatoren, desto größer die benötigte Fallzahl. Für die Forschung im Spitzensport ergibt sich daraus unmittelbar ein grundlegendes Dilemma: Durch Leistungsdichte und Optimierungsdruck werden selbst kleine Effekte relevant bzw. es steigen die Anforderungen an Genauigkeit und Verlässlichkeit von Indikatoren, die zur Entscheidungsfindung im Trainingsprozess herangezogen werden sollen. Gleichzeitig ist die potenziell verfügbare Anzahl repräsentativer Studienteilnehmer klein. Über diese die Spitzensportforschung kennzeichnende Konstellation hinaus tragen weitere Faktoren zu einer Diskrepanz zwischen benötigter und verfügbarer Fallzahl bei: Komplexe, multifaktoriell bedingte und/oder probabilistische Zielgrößen wie Regenerationsbedarf, Leistungsfähigkeit, Trainingseffektivität oder Verletzungsrisiko erhöhen die benötigte Fallzahl. Dagegen schränken Akzeptanz und Vereinbarkeit von Studienvorgaben mit dem Trainingsalltag die realisierbare Stichprobengröße weiter ein. Ähnliches gilt für anspruchsvolle Studiendesigns, die mit einer geringeren Anzahl von Teilnehmern auskommen.

Es ergibt sich die Frage, wie in der Trainings- und Wettkampfforschung im Spitzensport mit dieser kritischen Kombination aus hohen Anforderungen an Ergebnisse von Studien und kleinen Fallzahlen umgegangen werden sollte. Der aktuelle methodisch-statistische Goldstandard für experimentelle Studien in den Lebenswissenschaften ist die randomisierte,

kontrollierte Studie (RCT), die mittels Kovarianzanalyse (oder vergleichbarer Verfahren aus der Gruppe der generalisierten linearen Modelle) ausgewertet wird. Integraler Bestandteil dieses Ansatzes ist es, die benötigte Fallzahl basierend u. a. auf Studiendesign, Effektstärke, Anzahl der Zielp Parameter und gewählter statistischer Methode zu berechnen und dann auch einzuhalten. Bei einer „kleinen“ Studie („Studie mit unzureichender Teststärke“, „unterpowerete Studie“) liegt die tatsächlich realisierte Fallzahl unter dem berechneten Wert. Bedingt durch die im Spitzensport unabänderlich kleinen Fallzahlen werden RCTs (ebenso wie andere Studientypen) häufig entweder mit „pragmatisch“ begründeten kleinen Fallzahlen oder auf niedrigerem Leistungsniveau durchgeführt. In letzterem Fall muss die Übertragbarkeit in den Spitzensport jeweils kritisch hinterfragt werden. In ersterem Fall ist eine adäquate Wahrscheinlichkeit, relevante Effekte auch tatsächlich in der Studie nachzuweisen (Teststärke, „Power“) nicht gegeben und die Qualität der Ergebnisse entsprechend eingeschränkt. Dabei sind die mit einer kleinen Fallzahl einhergehenden Probleme gravierender als vielfach angenommen und können die Aussagekraft sowohl negativer als auch positiver Ergebnisse erheblich einschränken (siehe Kapitel 3.3).

Um diese Limitationen zumindest teilweise zu überwinden sind (wenn eine ausreichende Erhöhung der Fallzahl nicht möglich ist) angepasste Ergänzungen oder Alternativen des „RCT Standards“ notwendig. Erfreulicherweise steht dafür inzwischen ein umfangreiches, methodisch-statistisches Instrumentarium zur Verfügung (Abbildung 1). Eine große Rolle spielen dabei einerseits die sinnvolle Kombination von Einzelinstrumenten innerhalb des Studien- oder Projektplans ([40], Abbildung 1), andererseits die Berücksichtigung von Restriktionen der Studiendurchführung und des Anwendungshorizonts der Ergebnisse ([89], Abbildung 2). Dabei müssen die Perspektiven von logischer Stringenz einerseits und bestmöglicher, pragmatisch-athletengerechter Lösung von Problemstellungen aus der Sportpraxis andererseits immer wieder neu abgestimmt werden. Die Erfolgsprinzipien des RCT – Kontrolle, Randomisierung und Verblindung – sollten dabei, wenn irgend möglich, berücksichtigt werden.

3.2 Die Bedeutung der Fallzahl: Eine Auffrischung

3.2.1 Die Unterscheidung von Signal und Rauschen

Eine grundlegende Herausforderung der empirischen Forschung besteht darin, den wahren Effekt einer unabhängigen Variablen bzw. die Verteilung eines Merkmals („Signal“) vor dem Hintergrund diverser Varianzquellen („Rauschen“) möglichst genau zu charakterisieren. Aus der Sicht der Sportpraxis bedeutet das die Ableitung präziser, verlässlicher und somit im Trainingsalltag hilfreicher Einschätzungen – trotz der unvermeidlichen Unsicherheit einzelner Messwerte und Beobachtungen. Ein Großteil der methodisch-statistischen Instrumente dient daher einer Minimierung, Kontrolle und rechnerischen Abgrenzung von Sekundär- und Fehlervarianz. Von zentraler Bedeutung ist dabei eine angemessene Zahl von Beobachtungen (d. h. in der Regel eine ausreichend große Zahl von Teilnehmern).

3.2.2. Das Grundprinzip: Zufällige Fehler mitteln sich aus

Eine umfassende Darstellung der theoretischen Grundlagen würde den Rahmen dieser Expertise sprengen und kann bei Bedarf Lehrbüchern der Methodologie und Statistik entnommen werden. Die Grundüberlegung ist allerdings intuitiv zugänglich: Messfehler, biologische Tag-zu-Tag Variabilität und andere Arten von zufälligem „Rauschen“ heben sich bei ausreichend großer Fallzahl gegenseitig auf. In anderen Worten: Die zufälligen (!) Abweichungen nach oben und unten „mitteln sich aus“. Dadurch approximiert die zentrale Tendenz der beobachteten Werte (z. B. deren Mittelwert) den wahren Wert. Es leuchtet ein, dass die Wirksamkeit dieses Prinzips von der Fallzahl abhängt. Der Nachweis von kleinen Effekten bzw. sehr präzise Schätzwerte erfordern daher eine hohe Fallzahl. Dieses Grundprinzip besitzt über die

statistischen Verfahren und Systeme hinweg Gültigkeit, es gibt aber durchaus Unterschiede in der Effizienz.

3.2.3 Die Rolle der Fallzahl speziell im Rahmen der Hypothesentestung

Speziell im Rahmen der Hypothesentestung ist die Fallzahl entscheidend für die Kontrolle der Wahrscheinlichkeit von falsch positiven und falsch negativen Ergebnissen. Integraler Bestandteil ist daher eine Fallzahlberechnung im Rahmen der Studienplanung basierend u. a. auf dem kleinsten Effekt, der nicht übersehen werden sollte („kleinster relevanter Unterschied“, siehe Kapitel 4.4). Die Untersuchung genau dieser Fallzahl (n) soll sicherstellen, dass das Risiko sowohl eines falsch positiven (β) als auch eines falsch negativen (β) Ergebnisses in einem akzeptablen Rahmen bleibt. Üblicherweise wird β mit 5% und β mit 20% angesetzt. $1-\beta$ wird als die Teststärke oder „Power“ der Studie bezeichnet. Bleibt die realisierte Fallzahl unter dem berechneten Wert für n steigt die Wahrscheinlichkeit, einen relevanten Effekt zu übersehen über das als akzeptabel eingeschätzte β hinaus an. Die Teststärke ($1-\beta$) sinkt entsprechend ab, es fehlt „Power“. (Anmerkung: Diese Ausführungen gelten für die in der Leistungssportforschung üblichen Unterschiedshypothesen.)

3.3 Kleine Fallzahl – Gravierende Probleme

3.3.1 Bekannt problematisch: Negatives Ergebnis und geringe Power

Wenn das Signifikanzniveau für die primäre Hypothese einer Studie nicht erreicht wird (z. B. $\alpha \geq 5\%$) ist eine kleine Fallzahl bzw. eine geringe Teststärke (z. B. eine a priori festgelegte Power von $< 80\%$) offensichtlich problematisch. In diesem Fall ist das Ergebnis in keine Richtung mit der zuvor festgelegten Sicherheit interpretierbar.

tierbar. Ähnliches gilt auch außerhalb der Logik statistischer Signifikanz. Durch die starke Abhängigkeit des Standardfehlers von der Anzahl der Beobachtungen sind auf kleinen Fallzahlen basierende Intervallschätzer (z. B. Konfidenzintervalle) in der Regel weit. Dadurch wird nicht nur eine genaue Einschätzung des zu erwartenden Effekts erschwert, häufig können sogar weder relevant negative noch positive Effekte ausgeschlossen werden.

3.3.2 Geringe Power und positives Ergebnis

Tatsächlich sind kleine Fallzahlen bzw. eine a priori zu geringe Teststärke jedoch zumindest ebenso problematisch, wenn die Ergebnisse einer Studie scheinbar „die Latte überqueren“ und das Signifikanzniveau erreichen (21, 53).

3.3.2.1 Niedriger positiver prädiktiver Wert

Die Wahrscheinlichkeit, dass signifikante Ergebnisse echten (d. h. in größeren Studien reproduzierbaren) positiven Befunden entsprechen sind bei Studien mit kleiner Fallzahl geringer. Mit anderen Worten: Der positive prädiktive Wert (PPW) von Studien mit kleiner Fallzahl ist niedrig. Das gilt selbst dann, wenn andere Verzerrungen (z. B. „Publikationsbias“) keine Rolle spielen (21, 53).

3.3.2.2 Überschätzung der Größe des Effekts

Darüber hinaus wird selbst bei reproduzierbaren positiven Befunden kleiner Studien die Stärke des Effekts im Mittel überschätzt (21, 53). Diese als „Fluch des Siegers“ („winner’s curse“) bezeichnete Verzerrung erklärt sich durch die Kombination von zwei Faktoren: (a) Bei kleinen Fallzahlen sind die beobachteten Effekte instabil. D. h. ein einzelner oder einige wenige Datenpunkte können zu erheblichen Änderungen der Ergebnisse führen. (b) Studien mit geringer Fallzahl / Teststärke liefern nur für die (zufällig) größten beobachteten Effektstärken signifikante Ergebnisse. Gemeinsam füh-

ren diese Faktoren dazu, dass der „siegreiche“ Wissenschaftler, dessen kleine Studie zu einem signifikanten Ergebnis führt, mit einiger Wahrscheinlichkeit den „Fluch“ einer nicht reproduzierbaren Effektstärke zu tragen hat.

Noch eine weitere, selten explizit berücksichtigte Problematik trägt zur Überschätzung der wahren Effekte in kleinen Studien bei. Es handelt sich um die fallzahlabhängige, systematische Überschätzung der wahren Effekte durch empirisch gewonnenen Effektschätzer („indirect bias“) (86). Das Ausmaß dieser Verzerrung ist umgekehrt proportional zur Fallzahl und daher in großen Stichproben vernachlässigbar klein, bei kleinen Fallzahlen jedoch unter Umständen erheblich (37).

3.3.2.3 Proteus-Phänomen

In der Summe führen die genannten Verzerrungen zum sog. Proteus-Phänomen: Selbst wenn die positiven Ergebnisse von Studien mit kleiner Fallzahl in größeren Folgestudien reproduziert werden können, liegen die Effektstärken dann in der Regel deutlich niedriger (21). Dasselbe gilt analog für die bei Anwendung in der Sportpraxis erwartbaren Effekte. Die Berücksichtigung der negativen Begleiterscheinungen kleiner Fallzahlen unabhängig vom Ergebnis der Studie ist daher essenziell sowohl für die Planung von Folge- bzw. Replikationsstudien als auch für die Einschätzung der praktischen Relevanz und die adäquate Entscheidungsfindung in der Sportpraxis.

3.3.2.4 Weitere mögliche Verzerrungen

Das Proteus-Phänomen und die ihm zugrunde liegenden Verzerrungen sind unmittelbare Begleiterscheinungen kleiner Fallzahlen. Darüber hinaus sind kleine Studien aber auch anfälliger für andere Probleme (z. B. Publikationsbias und „flexible Datenanalyse“ bzw. „p-hacking“). Daraus folgt die Forderung nach einem verbindlichen, detaillierten Plan für Datenaufnahme und -analyse, idealer Weise in publizierter Form.

3.4 Deshalb: Spezielles Instrumentarium & Replikationsversuche

In der Zusammenschau ergibt sich aus den negativen Begleiterscheinungen kleiner Fallzahlen (neben der Notwendigkeit eines spezifischen Instrumentariums zu deren Minimierung) die herausragende Bedeutung von Replikationsstudien (53, 109). Dabei sind für die Generierung praxistauglicher, robuster Wissensbestände nicht nur direkte Replikationsversuche sondern auch eine gewisse Pluralität der Zugänge in ei-

nem Forschungsfeld wichtig („Triangulation“) (64). Die Bedeutung von Replikationsstudien ist theoretisch herleitbar, kann aber auch an praktischen Beispielen illustriert werden (z. B. Kandidateng-Studien im Kontext der Trainingsresponse [13]). Leider sind Replikationsstudien trotz ihrer unstrittigen Bedeutung im Vergleich zu „innovativen“ Arbeiten immer noch deutlich schwerer zu finanzieren und zu publizieren. Glücklicherweise hat in dieser Beziehung ein Umdenken eingesetzt. So ermutigen einige Zeitschriften mit Sportbezug inzwischen ausdrücklich zur Einreichung von Replikationsstudien unabhängig von deren Ergebnis.

Die Problematik kleiner Fallzahlen in der Leistungs- und Spitzensportforschung

- › In der Leistungs- und Spitzensportforschung müssen kleine Effekte mit kleinen Fallzahlen untersucht werden. Dadurch entsteht regelmäßig ein Missverhältnis zwischen benötigter und realisierbarer Fallzahl.
- › Studien mit suboptimaler Fallzahl sind mit erheblichen Problemen und Limitationen behaftet. Das gilt unabhängig von Ergebnis und Art des statistischen Zugangs. Insbesondere ist das Risiko von falsch positiven und falsch negativen Ergebnissen erhöht (bzw. deren prädiktiver Wert gemindert) und die Größe des tatsächlichen Effekts wird im Mittel überschätzt.
- › Anpassungen von Datenaufnahme und Datenanalyse können zu Vermeidung des o.g. Missverhältnisses und/oder einer Minderung der negativen Auswirkungen beitragen. Dieses „Instrumentarium für kleine Fallzahlen“ wird in den nächsten Abschnitten vorgestellt.
- › Replikationsversuche bilden die zweite Säule des wissenschaftlichen Umgangs mit kleinen Fallzahlen.

Verbesserungsmöglichkeiten

Zum Verständnis des Instrumentariums für den wissenschaftlichen Umgang mit kleinen Fallzahlen, das in Kapitel 5 vorgestellt wird, sind einige Grundüberlegungen und -konzepte hilfreich. Dazu gehört neben der klaren Priorisierung der Instrumente (Kapitel 4.1), auch eine explizite strategische Ausrichtung des Projekts zwischen Grundlagenforschung und unmittelbarem Anwendungsbezug (Kapitel 4.2), der transparente Umgang mit kalkulierter und realisierter Fallzahl (Kapitel 4.3) und – in engem Bezug dazu – die Beurteilung der (sportpraktischen!) Relevanz von Effekten („smallest worthwhile difference“, Kapitel 4.4).

4.1 Prävention, Pflicht & Kür: Strukturierter Umgang mit kleinen Fallzahlen

4.1.1. Prävention: Benötigte und realisierbare Fallzahlen angleichen

Die erste und wichtigste Option im wissenschaftlichen Umgang mit kleinen Fallzahlen ist der Versuch die Fallzahl zu erhöhen. Auch wenn es selbstverständlich erscheinen mag kann nicht genug betont werden, dass diese Option Priorität besitzt und mit Nachdruck verfolgt werden sollte. Eine vertrauensvolle Zusammenarbeit mit der Sportpraxis, die gemeinsame Abwägung der Erfordernisse von Trainingsalltag und wissenschaftlicher Methodik, Incentives in unterschiedlicher Form, aber auch ein kritisches Hinterfragen der Notwendigkeit bzw. Sinnhaftigkeit einer Untersuchung der konkreten Fragestellung an Spitzenathleten können die Rekrutierung erleichtern. Andererseits können gezielte Anpassungen von Studiendesign, Methodik und Datenanalyse die benötigte Fallzahl reduzieren. Die Durchführung einer Studie mit niedriger Teststärke sollte ausschließlich dann in Betracht gezogen werden, wenn es nachvollziehbar nicht möglich ist eine der Fallzahlkalkulation entsprechende Menge adäquater Teilnehmer zu rekrutieren. Eine nachvollziehbare,

spezifische Begründung sollte in diesem Fall Bestandteil des Studienplans bzw. von Veröffentlichungen der Ergebnisse sein (vgl. Kapitel 4.3).

4.1.2 Pflicht und Kür: Stufenweise Anpassung an kleine Fallzahlen

Wenn eine suboptimale Fallzahl unvermeidlich ist, werden Studien häufig mit einer pragmatisch (z. B. durch die Größe des Kaders) begründeten Fallzahl, im Übrigen aber unverändert durchgeführt. Dieses Vorgehen ist in aller Regel verbesserungsfähig. Die Überlegung dahinter lautet: Wenn es nicht möglich ist, die Fallzahl an den übrigen Elementen des Studienplans auszurichten, dann sollte dieser an die erzielbare Fallzahl und die damit verbundenen Restriktionen angepasst werden.

4.1.2.1 Pflichtprogramm

Als erster Schritt ist es dabei wichtig zu beachten, dass einige generelle Aspekte von Design, Methodik und Datenanalyse bei Studien mit kleinen Fallzahlen eine höhere Relevanz besitzen (siehe Seite 23 ff und Abb. 1). Zu diesen Aspekten zählt auch ein vor Beginn der Datenaufnahme umfassend ausformulierter Studienplan einschließlich Festlegung von Hauptzielparameter(-n) und Details der Datenauswertung (21). Der optimalen Umsetzung dieser Standards sollte daher besonderes Augenmerk gewidmet werden. In gewisser Weise handelt es sich dabei um das „Pflichtprogramm“ im Umgang mit kleinen Fallzahlen.

4.1.2.2 Kür

Aufbauend auf diesen Grundlagen können spezielle Ergänzungen üblicher Methoden, aber auch grundsätzlichere Änderungen der methodisch-statistischen Herangehensweise – sozusagen als „Kür“ – dazu beitragen, ein Maximum an Information aus einer begrenzten Anzahl von Teilnehmern herauszuholen.

Dabei kann es vorkommen, dass aussagekräftige Ergebnisse allein auf Basis der

in der Studie erhobenen Daten dennoch nicht zu erwarten sind. In diesen Fällen kann die geplante Berücksichtigung von Daten aus früheren Studien (z. B. als externe Kontrollgruppe) oder von Vorwissen (z. B. als informative Priorverteilung in eine Bayesianischen Analyse) zur Verstärkung des aktuellen Datensatzes genutzt werden. Die sich dadurch ergebenden Limitationen sind in der Interpretation zu berücksichtigen. Da diese Limitationen vor allem die Ableitung generalisierbarer Schlussfolgerungen betreffen, bietet sich die Einbindung von Vorwissen insbesondere für die Entscheidungsfindung innerhalb der untersuchten Rahmenbedingungen an (89).

4.1.2.3 Ausweidlösungen

Schließlich sollten, wenn die bisher genannten Optionen an ihre Grenzen stoßen, auch Methoden außerhalb des üblichen Rahmens (quasi-) experimenteller oder beobachtender empirischer Studien in Betracht gezogen werden. Dies gilt insbesondere dann, wenn durch die notwendige Modellkomplexität oder die Seltenheit von Ereignissen die eigentlich erforderliche Fallzahl um Größenordnungen verfehlt wird. Dadurch sinkt zwar vordergründig der Evidenzgrad (20), durch die negativen Auswirkungen einer zu kleinen Fallzahl (vgl. Kapitel 3.3) kann die prädiktive Validität eines solchen Vorgehens dennoch überlegen sein. Ein Beispiel für eine solche alternative Herangehensweise ist die Aggregation subjektiver Einschätzungen im Sinne der „Schwarmintelligenz“ (97, 108). Auch an dieser Stelle können die impliziten Wissensbestände der Sportpraxis synergistisch in der angewandten Spitzensportforschung genutzt werden. Abbildung 1 illustriert das Konzept einer stufenweisen Anpassung des Studienplans an kleine Fallzahlen



Abbildung 1

4.2 Projektstrategie

Eine klare und explizite Projektstrategie ist gerade bei limitierter Fallzahl wichtig (wenn man also nicht „alles“ haben kann). Die Entwicklung einer solchen Projektstrategie beginnt mit der Festlegung des strategischen Ziels. Also der Entscheidung, ob der Fokus der Studie (tatsächlich) auf einem generalisierbaren Erkenntnisgewinn liegt oder doch eher auf der Unterstützung von Entscheidungen unter Rahmenbedingungen, die weitgehend den Studienbedingungen entsprechen (oder sogar für die Teilnehmer der Studie selbst). Generalisierbare Schlussfolgerungen (z. B. „Wirkt Intervention X?“) stehen für Wissenschaftler habituell im Fokus. Für Trainer und Betreuer steht dagegen die Entscheidungsfindung für konkrete Athleten oder Mannschaften im Vordergrund („Sollte Intervention X von Athleten, die den Studienteilnehmern ähnlich sind (oder sogar von ihnen selbst), genutzt werden?“). Die Unterscheidung dieser beiden Perspektiven hat erhebliche Auswirkungen auf die optimale Gestaltung des Studienplans (89).

Die komplementäre Frage betrifft den Wissensstand vor Beginn der Studie: Muss der zu untersuchende Effekt noch grundsätzlich belegt werden oder betrifft die Forschungsfrage eher Relevanz und Praktikabilität vor dem Hintergrund spezieller Rahmenbedingungen oder Zielparameter. Im ersten Fall wird der ohnehin niedrige positive prädiktive Wert einer Studie mit niedriger Fallzahl durch eine niedrige Vor-

Studien-Wahrscheinlichkeit weiter herabgesetzt. Aus diesem Grund ist die Gefahr falsch positiver Ergebnisse bei Studien, die versuchen eine grundlegend neue Hypothese mit niedriger Teststärke zu verifizieren („Pilotstudien“), besonders hoch. Insbesondere wenn weitere Probleme hinzukommen (z. B. Publication bias, flexible/multiple Datenanalyse, selektive Publikation von Zielparametern) besteht die Gefahr, dass solche Studien eher zur Verwirrung als zu nachhaltigem Wissenszuwachs beitragen. Diese Überlegung sollte bei der Planung und Interpretation von Pilotstudien unbedingt beachtet werden (vgl. Abbildung 2). *(Anmerkung.: Der Begriff „Pilotstudie“ wird in der Literatur unterschiedlich verwendet. Im Kontext klinischer Studien handelt es sich in der Regel um vorbereitende Machbarkeitsstudien, die keine Hypothesentestung beinhalten (98). Hier wird „Pilotstudie“ dagegen, wie in der Leistungssportforschung üblich, im Sinn einer Untersuchung neuer Hypothesen mit begrenzten Ressourcen („try-out“) verwendet (z. B. [41, 42].)*

Im Gegensatz zu der Auswertung der Daten einzelner Studien und der nachfolgenden Interpretation der Ergebnisse gründen sich Entscheidungen idealerweise in einer präzisen Zusammenfassung des zum jeweiligen Zeitpunkt verfügbaren Wissensstandes. Neben Daten bzw. Informationen aus der aktuellen Studie gehört dazu auch eventuelles Vorwissen. Daraus folgt, dass bei Studien, die in erster Linie der Entscheidungsfindung unter den jeweils untersuchten

Projektstrategie

- › Ist das strategische Ziel die Unterstützung von Entscheidungen unter (oder nahe) den konkreten Rahmenbedingungen der Studie bietet es sich an, ggf. vorhandenes Vorwissen zur Verstärkung des Datensatzes aus der kleinen Studie zu nutzen. Bayesianische Verfahren bieten hierzu die Möglichkeit (89).
- › Ergebnisse früherer Studien aber auch Wissensbestände aus der Praxis können so sowohl zur Rechtfertigung als auch zur Ergebnisqualität kleiner Studien im Elitebereich beitragen. Die gezielte Durchführung von Vorstudien z. B. auf etwas niedrigerem Leistungsniveau kann wesentlicher Teil der Projektstrategie sein.
- › Dieses Prinzip gilt insbesondere auch für Einzelfallstudien ($n=1$) und ist dann eine theoriegestützte und im Leistungs- und Spitzensport erprobte Möglichkeit der Individualisierung (7, 45, 47, 93).

Rahmenbedingungen dienen, die Ergebnisse nicht ausschließlich die Auswertung der aufgenommenen Daten widerspiegeln müssen. Vielmehr kann und sollte strukturiert erhobenes Vorwissen mit einfließen (89). Bayesianische Verfahren bieten hierzu die Möglichkeit. Im Rahmen einer längerfristigen Projektstrategie kann es dabei sinnvoll sein, empirisches Vorwissen gezielt zu generieren. Eine Möglichkeit sind vorbereitende Studien auf etwas niedrigerem Leistungsniveau. Deren Ergebnisse können dann mit an wenigen Spitzenathleten erhobenen Daten spezifiziert werden (47).

4.3 Transparente Rechtfertigung der Fallzahl

Aus den in Kapitel 3.3 dargelegten Aspekten ergibt sich die Notwendigkeit eines transparenten Umgangs mit der Fallzahlproblematik, und zwar unabhängig vom Ergebnis der Studie. Diese Forderung ist sowohl im Sinne der Be- und Verwertung einzelner Studien als auch im Interesse einer nachhaltigen Entwicklung der Spitzensportforschung insgesamt.

Letztlich geht es dabei um die Beantwortung der Frage, ob anzunehmen ist, dass die Ergebnisse der Studie (trotz ihrer kleinen Fallzahl) zur Beantwortung der konkreten Forschungsfrage beitragen (58). In zweiter Linie auch, ob die kleine Fallzahl plausibel gerechtfertigt werden kann und angemessene Anpassungen des Studienplans vorgenommen wurden. Zu einer solchen Rechtfertigung einer kleinen Fallzahl gehören daher (a) eine nachvollziehbare Schätzung der eigentlich notwendigen/wünschenswerten Fallzahl; Im Kontext der angewandten Forschung im Leistungs- und Spitzensport als a priori Fallzahlkalkulation auf Basis expliziter Überlegungen zur Größe relevanter Effekte (46, 58, 61) (vgl. Kapitel 4.3), (b) eine spezifische Begründung der realisierten Fallzahl und (c) eine Erläuterung relevanter Entscheidungen im Zuge der Studienplanung. Für Letzteres ist wichtig zu bedenken, dass beide Seiten des Verhältnisses von notwendiger und realisierbarer Fallzahl durch die Gestaltung des Studienplans beeinflusst werden

können. Die Teststärke (bzw. der Standardfehler von Effektschätzern und damit die Weite von deren Konfidenzintervall) wird von vielen Faktoren beeinflusst, darunter die Größe relevanter Effekte (vgl. Kapitel 4.4), Details des Studiendesigns und der verwendeten statistischen Verfahren, aber auch praktische Aspekte der Datenaufnahme wie die Reliabilität der verwendeten Messverfahren und die Standardisierung der Messbedingungen. Andererseits beeinflussen der Studienplan – aber auch in der Trainingspraxis direkt verwertbare Ergebnisse (z. B. Leistungsdiagnostik, individuelle Responseschätzer) – die Teilnahmebereitschaft.

Die Optimierung der Teststärke bei kleinen Fallzahlen erfordert prinzipiell vertieftes Fachwissen in der Methodologie und Statistik. Dies gilt umso mehr, wenn, wie im Fall der Spitzensportforschung, noch praktische Beschränkungen (Akzeptanz, Reisetätigkeit, Periodisierung, eng getakteter Trainingsalltag etc.) hinzukommen. Um zu vermeiden, dass „Power Reserven“ ungenutzt bleiben ist eine statistische Beratung (noch besser eine echte Kooperation bei der Ausarbeitung des Studienplans) daher dringend empfehlenswert. Ähnliches gilt für die Fallzahlkalkulation. Ein selten diskutierter Aspekt in diesem Zusammenhang ist, ob Spitzensportler als finite Population betrachtet werden können bzw. sollten. Da eine kleine Fallzahl mehr Information über eine kleine als über eine (unendlich) große Population liefert, liegt in diesem Fall (*ceteris paribus*) die benötigte Fallzahl etwas niedriger (1, 24). Relevant wird dieser Unterschied allerdings nur, wenn mehr als etwa 20% der Gesamtpopulation untersucht werden. Aktuelle Softwarepakete ermöglichen Fallzahl-schätzungen für anspruchsvolle, hierarchische Studiendesigns auch mit nur rudimentären Programmierkenntnissen. Ein Beispiel ist das R-Paket „Superpower“ (59) und die dazugehörige Shiny-App. Diese Berechnungshilfen ersetzen jedoch nicht eine methodisch-statistische Beratung.

4.4. Ab welcher Größe ist ein Effekt (praktisch) relevant?

Im Rahmen der Fallzahlkalkulation muss festgelegt werden, ab welcher Größe man einen

Effekt mit einiger Sicherheit ($1-\beta$) nicht übersehen möchte (58, 60, 61). Ähnliches gilt für die akzeptable Weite eines Konfidenzintervalls. Auch diese kann nur unter Bezug auf die Mindestgröße relevanter Unterschiede sinnvoll bestimmt werden (58, 60). Dabei ist die Frage, ab welcher Größe ein bestimmter Effekt relevant wird, eine der anspruchsvollsten Festlegungen bei der Planung einer Studie. Das liegt auch daran, dass es sich um eine artifizielle Dichotomie handelt, der in der Realität eher ein Kontinuum von winzigen, sicher irrelevant kleinen bis zu großen, offensichtlich relevanten Effekten entspricht.

4.4.1 Effekte als Vielfache der Variabilität – Effektstärken

Eine Möglichkeit zur Quantifizierung und Bewertung von Effekten ist die Bildung eines Quotienten aus Mittelwert und Standardabweichung ([standardisierte] Effektstärke [26]). Damit wird der Effekt nicht mehr in der ursprünglichen Einheit, sondern als dimensionsloses Vielfaches der Variabilität ausgedrückt. Trotz des einfachen Prinzips sind im Detail Unterschiede möglich, insbesondere bei der Frage welche Standardabweichung verwendet wird (25, 67).

Bei der Effektstärke handelt es um ein Maß des Signal-Rausch-Verhältnisses, damit ist der Bezug zur benötigten Fallzahl direkt erkennbar. Für die Relevanz in der Sportpraxis ist jedoch der zu erwartende Effekt bedeutsamer. Eine weitere Problematik von Effektstärken als Indikator der (praktischen) Relevanz ist die Kontextabhängigkeit der Streuung, die den Nenner des Quotienten bildet. Daher kann die Effektstärke z. B. durch strikte Standardisierungsmaßnahmen erhöht werden (mit der Folge einer niedrigeren, benötigten Fallzahl), ohne dass diese einen Einfluss auf den im Mittel erwartbaren Effekt hätten (88).

4.4.1.1 Bewertung der Effektstärke

Die Bewertung der Effektstärke folgt in der Regel der Einteilung nach Cohen (26). Dabei gelten

Effekte jeweils ab einer Effektstärke (Cohens d) von 0,2, 0,5 und 0,8 als klein, mittel bzw. groß. Häufig wird bei der Fallzahlplanung die Grenze eines kleinen Effekts ($d = 0,2$) als „smallest effect size of interest“ verwendet. Die Attraktivität dieser Einteilung basiert zumindest zum Teil auf der Verfügbarkeit einer anerkannten Referenz (26). Allerdings entsprechen die von Cohen vorgeschlagenen Grenzen sicherlich nicht in jedem Kontext den Grenzen der tatsächlichen (sport-) praktischen Relevanz (60). Speziell im Leistungssport sind diese in erster Linie von der Leistungsdichte abhängig; Im Fall von Surrogatmarkern oder bei nur indirektem Bezug zu den „harten Outcomes“ des Leistungssports, auch vom Grad des Zusammenhangs mit Leistung und Leistungsentwicklung. *(Anmerkung: Die Effektstärke nach Cohen und die genannte Interpretation beziehen sich auf den Vergleich von zwei Mittelwerten in einer t-Test Situation. Bei anderen statistischen Modellen hat auch die Effektstärke eine andere Interpretation. Bei der ANOVA (F-Test) gelten z. B. Effekte jeweils ab einem f Wert von 0,1, 0,25 bzw 0,4 als klein, mittel und groß. Bei komplexen Modellen ist eine Bestimmung der Fallzahl nur simulativ möglich.)*

Eine Abwandlung der Effektstärke nach Cohen ist die von Hopkins für die Bewertung von Leistungsentwicklungen im Sport entwickelte „smallest worthwhile difference“ (52). Die Leistungsdichte wird dabei durch den Bezug zur Standardabweichung der Leistung in der jeweiligen Bezugsgruppe berücksichtigt. Bei einer pauschalen Anwendung des speziellen Werts von 0,2 (16) aus der Ursprungspublikation (52) geht dieser Vorteil allerdings verloren. Zudem darf nicht vergessen werden, dass normativ festgelegt werden muss, was ein relevanter Zuegewinn an Wettkampferfolgen ist. Die Festsetzung von Hopkins et al. (eine zusätzliche Medaille in zehn Wettkämpfen [52]) ist sicherlich nicht auf jeden Kontext übertragbar. Zusammenfassend lässt sich sagen, dass die von Hopkins vorgeschlagene Art der Berechnung der „smallest worthwhile difference“ sich für Leistungswerte eignet. Der konkrete Wert von 0,2 ist jedoch nicht generell übertragbar und sollte für den einzelnen Anwendungsfall hergeleitet und begründet werden. *(Anmerkung: Alternativ wird in der Ursprungspublikation ein Bezug zur intrain-*

dividuellen Variabilität der Leistung mit einem Faktor von 0,3 vorgeschlagen [52]. Dieser Ansatz hat sich aber nicht durchgesetzt. Einen Überblick zu Grenzwerten mit Bezug zur intraindividuellen Streuung bietet [46]).

Von der Beurteilung der praktischen Relevanz abgegrenzt werden sollte die Abschätzung einer realistisch erwartbaren Effektstärke (z. B. auf Basis von früheren Studien, in-house-Reliabilität, geplanten Messwiederholungen und/oder praktischer Erfahrung) (vgl. Kapitel 4.4).

4.4.2 Einschätzung der Sportpraxis

Bei der Festlegung der „smallest worthwhile difference“ nach Hopkins (52) wird der Bezug zur Relevanz in der Sportpraxis durch die Fest-

legung eines relevanten Zugewinns an Wettkampferfolgen gewährleistet (52). Wenn leistungsfähige prädiktive Modelle verfügbar sind, kann diese Logik auch auf die Einschätzung der Relevanz von Unterschieden in den Einfluss- und Bedingungsfaktoren von Leistung und Leistungsentwicklung angewendet werden. Aktuell ist diese Voraussetzung allerdings nicht gegeben. Eine Alternative bieten Expertenmeinungen (58). Deren Erfassung und Aggregation sollten als Teil des Gesamtprojekts nachvollziehbar und geplant erfolgen und transparent kommuniziert werden. Zusammen mit einer Abschätzung der zu erwartenden Streuung (z. B. über eine in-house-Reliabilitätsstudie oder literaturbasiert) kann eine informative Grundlage für die Fallzahlberechnung hergestellt werden.

Transparente Begründung von kleinstem relevantem Unterschied und Fallzahl

- ▶ Eine kleine Fallzahl ist dann zu rechtfertigen, wenn angenommen werden kann, dass die Ergebnisse der Studie dennoch zur Beantwortung der konkreten Forschungsfrage beitragen (58).
- ▶ Der Fallzahlkalkulation sollte eine Abschätzung der Größe von sportpraktisch relevanten Unterschieden des jeweiligen Zielparameters im konkreten Anwendungsfeld zugrunde gelegt werden. Pauschale Grenzwerte sind wenig hilfreich (z. B. kleiner Effekt nach Cohen d. h. $d \geq 0,2$ [26] oder kleinster relevanter Unterschied nach Hopkins d. h. $\geq 20\%$ der Standardabweichung zu Studienbeginn [52]).
- ▶ Eine normative Komponente (z. B. Festlegung eines relevanten Zugewinns an Wettkampferfolgen (52) ist immer notwendig.
- ▶ Eine Einbindung von Praxis-/ bzw. Expertenwissen bei der Beurteilung der sportpraktischen Relevanz von Unterschieden bzw. Effekten ist sinnvoll und statthaft. Die Art der Erhebung sollte transparent beschrieben werden (z. B. Befragung der Bundestrainer [43]).
- ▶ Beobachtete Effektstärken sind für die Beurteilung der Relevanz problematisch, da sie stark durch Standardisierungsmaßnahmen und andere Einflüsse auf die Streuung beeinflusst werden (88).
- ▶ Die Fallzahlberechnung für hierarchische Studiendesigns ist anspruchsvoll und sollte durch einen ausgebildeten Statistiker begleitet oder durchgeführt werden. Das R-Paket „Superpower“ und die dazugehörige Shiny-App eignen sich für eigene Abschätzungen (59).

Instrumente für kleine Fallzahlen

5.1 Das Fundament: Methodische Grundprinzipien

Das nachfolgende Instrumentarium für kleine Fallzahlen schließt an die methodischen und statistischen Grundlagen der Planung, Durchführung und Auswertung von Studien an. Eine systematische Darstellung dieser Grundlagen kann Lehrbüchern entnommen werden. Eine Aufbereitung speziell für den Fall von Trainingsstudien ist in der jüngeren Literatur verfügbar (40) und kann bei Bedarf als Ausgangspunkt verwendet werden. Von besonderem Interesse ist die randomisierte, kontrollierte Studie (randomised controlled trial; RCT) einschließlich der Datenauswertung mittels Kovarianzanalyse (ANCOVA) oder gemischter linearer Modelle („mixed modelling“) (40). Der RCT ist aktuell der Referenzstandard für die Untersuchung

von Interventionseffekten. Die Erfolgsprinzipien des RCTs – Vergleich mit einer Kontrollbedingung, Randomisierung, Verblindung und die Kontrolle von Sekundär und Störvarianz durch Standardisierung, Stratifizierung und die Berücksichtigung von Kovariaten sollten wo immer möglich auch unter den schwierigen Bedingungen der Datenaufnahme im Spitzensport Berücksichtigung finden. Es steht jedoch außer Frage, dass das nicht immer in vollem Umfang möglich sein wird (z. B. wegen fehlender Akzeptanz bei Trainern und/oder Athleten oder der Offensichtlichkeit der Intervention).

Die folgenden Vignetten bieten einen kurzen Überblick über die wesentlichen methodischen Grundprinzipien, die die paradigmatische Stellung der randomisierten, kontrollierten Studie begründen. Eine umfassende Diskussion kann der in Haupttext und Vignetten angegebenen Literatur entnommen werden.

Die Erfolgsprinzipien des RCTs

Ziel von (quasi-) experimentellen Längsschnittstudien ist eine Beurteilung des Effekts der untersuchten Intervention bzw. der unabhängigen Variablen. Im Idealfall sollen dabei kausale und über die Studie hinaus generalisierbare Schlussfolgerungen ermöglicht werden. Die methodischen Grundlagen dafür werden in der idealtypischen, randomisierten, kontrollierten Studie beispielhaft umgesetzt und bilden die Grundlage der paradigmatischen Bedeutung dieses Studiendesigns.

Eine große Rolle für Verständnis und Umsetzung spielen bekannte und unbekannte Einflussfaktoren des Effekts der unabhängigen Variablen (Moderatoren, Kovariaten) (84).

Die Erfolgsfaktoren des RCTs sollten wo immer möglich auch unter den schwierigen Bedingungen der Datenaufnahme im Spitzensport Berücksichtigung finden.

Vergleich zu einer Kontrollbedingung

Die Kontrollgruppe oder Kontrollbedingung einer Längsschnittstudie dient der Abschätzung, was ohne die Intervention passiert wäre. Dafür werden die Veränderungen unter Intervention mit jenen unter der Kontrollbedingung verglichen. Diverse andere Erklärungen für beobachtete Veränderungen in der Interventionsgruppe (z. B. Zeit- und Lerneffekte) können so eliminiert werden. Eine Kontrollbedingung ist daher eine **wichtige Voraussetzung für kausale Schlussfolgerungen**.

„Kontrolle“ kann in unterschiedlicher Form implementiert werden, z. B. als Kontrollgruppe in einer randomisierten kontrollierten Studie, als Kontrollbedingung in einem randomisierten Cross-Over Design oder auch als wiederholte Messungen während der Baselinephase in einem Multiple Baseline Design. Die Kontrollbedingung kann je nach Kontext und Fragestellung eine Fortsetzung des bisherigen Verhaltens, ein Placebo bzw. eine Sham-Intervention oder auch eine etablierte Art der Intervention (bisherige best-practice) sein, zu der die innovative Intervention in Vergleich gesetzt werden soll. Historische Kontrolldaten aus früheren Studien oder simulierte Kontrollen aus Reliabilitätsdaten sind mit erheblichen Limitationen behaftet (5, 46) und kommen nur in Betracht, wenn eine gleichzeitige Kontrolle nicht umsetzbar ist.

Besonderheiten im Leistungs- und Spitzensport

Für Interventionen, die im Spitzensport untersucht werden, sind in der Regel deutliche Hinweise auf eine leistungsförderliche Wirkung verfügbar. Diese den Athleten in der Kontrollgruppe vorzuenthalten ist daher häufig nicht adäquat und vermittelbar. Dies sollte kein Grund sein auf das Element der Kontrolle innerhalb der Studie zu verzichten. Je nach Reversibilität und Zeitverlauf des Interventionseffekts und der Möglichkeit eines zuverlässigen Absetzens kommen insbesondere Varianten von Cross-Over oder Multiple Baseline Design in Frage. Auch eine „waiting control“ ist möglich, bei der die Teilnehmer der Kontrollgruppe die Intervention nach Studienende erhalten.

Randomisierung

Die zufällige Zuordnung von Teilnehmern zu einer Gruppe oder Untersuchungsbedingung vermeidet systematische Verzerrungen und trägt zur Effektivität einer Verblindung bei. Randomisierung ist damit eine **wichtige Voraussetzung für kausale Schlussfolgerungen**.

Eine annähernde Gleichverteilung von Kovariaten über die Gruppen oder Untersuchungsbedingungen hinweg („baseline balance“) ist nicht das primäre Ziel der Randomisierung (84). Dennoch ist es im Interesse der Effizienz sinnvoll eine Stratifizierung für bei Studienbeginn verfügbare Kovariaten vorzunehmen (z. B. Geschlecht und Leistungsniveau) (84). Bekannte Kovariaten, bei denen das nicht möglich ist, können und sollten im Rahmen einer Kovarianzanalyse (ANCOVA) berücksichtigt werden (91). Dabei sollten die einzuschließenden Kovariaten bereits im Studienplan festgelegt und nicht von der Signifikanz der Unterschiede zu Studienbeginn abhängig gemacht werden (6, 11, 85). Bei Designs mit mehreren Studienphasen (z. B. Cross-Over) wird deren Reihenfolge randomisiert um Zeit- und Reihenfolgeeffekte zu kontrollieren (92). Das konkrete Vorgehen ist detailliert und reproduzierbar zu beschreiben (84).

Zusammenfassen lässt sich diese Grundlogik der Zuteilung von Studienteilnehmern zu Gruppen oder Bedingungen als: „Stratifiziere was du kannst und randomisiere den Rest.“ („Balance what you can and randomize what you can't.“) (84).

Anmerkung: Ein angemessenes Verhältnis zwischen der Anzahl von Stratifizierungskriterien (bzw. möglichen Kombinationen von Strata) und Teilnehmern sollte dabei gewahrt bleiben. In keinem Fall sollte die Anzahl der Zellen größer sein als die Fallzahl.

Besonderheiten im Leistungs- und Spitzensport

In der Regel sind die Teilnehmer kleiner Studien im Leistungssport (im Gegensatz zu klinischen Studien) bei Studienbeginn bekannt. Wenn Kovariaten verfügbar sind, kann dieser Umstand für ein paarweises Matching bzw. eine mehrdimensionale Stratifizierung genutzt werden. Randomisiert wird dann innerhalb der Paare oder Subgruppen.

Anmerkung: Von der randomisierten Zuordnung der Teilnehmer muss deren randomisierte Auswahl abgegrenzt werden. Diese dient der Vermeidung einer Stichprobenverzerrung („selection bias“) und trägt damit zur Generalisierbarkeit der Ergebnisse bei. Zufallsstichproben spielen in der Leistungs- und Spitzensportforschung wegen der kleinen Grundgesamtheit und der speziellen Rahmenbedingungen allerdings nur eine untergeordnete Rolle.

Standardisierung

Die systematische Konstanthaltung bekannter Einflussfaktoren dient der Begrenzung der Sekundärvarianz. Standardisierung kann sich auf einzelne Messwerte (Testverfahren, Untersuchungsablauf, Abstand zur letzten Trainingseinheit etc.) oder die Veränderung über den Studienzeitraum von

Längsschnittstudien beziehen. In letzterem Fall werden die Intervention selbst und/oder Moderatoren der Interventionswirksamkeit standardisiert.

Besonderheiten im Leistungs- und Spitzensport

Im Rahmen von wissenschaftlichen Studien, insbesondere unter Laborbedingungen, ist in der Regel eine deutlich strengere Standardisierung möglich als im Trainingsalltag. Dadurch steigt die beobachtete Effektstärke. Diese Gesetzmäßigkeit kann zur Steigerung der Teststärke und Minderung der benötigten Fallzahl genutzt werden, trägt allerdings auch zur Überschätzung der in der Sportpraxis erwartbaren Effektstärken bei (88). Eine detaillierte und vollständige Beschreibung der Standardisierungsmaßnahmen ist daher von entscheidender Bedeutung für die Beurteilung von Ergebnissen hinsichtlich zu erwartender Relevanz in der Sportpraxis.

Verblindung

Die Geheimhaltung der Zuordnung einzelner Teilnehmer zu einer bestimmten Gruppe oder Bedingung dient der Vermeidung von Verzerrungen durch eine (meist) unbewusste Erwartungshaltung. Im Idealfall betrifft die Verblindung alle an Datenaufnahme und Datenauswertung beteiligten Personen. Ist nur dem Teilnehmer nicht bekannt zu welcher Gruppe oder Bedingung er gehört, handelt es sich um eine einfache Verblindung. Sind auch die unmittelbar mit der Studiendurchführung befassten Mitglieder des Forschungsteams verblindet, handelt es sich um eine Doppelblind-Studie. Wenn eine umfassende Verblindung aller Beteiligten nicht möglich ist sollte eine weitestmögliche Verblindung dennoch angestrebt werden. Die Möglichkeiten dazu sind vielfältig (z. B. die externe, anonymisierte Auswertung von Proben-, Bild- oder Videomaterial).

In der Pharmaforschung erfolgt die Verblindung der Intervention klassischerweise durch identisch aussehende Scheinpräparate ohne Wirkstoff (Placebo). Pharmafirmen sind verpflichtet für zugelassene Präparate Placebos zu produzieren. Bei nicht pharmakologischen Studien werden Scheininterventionen („sham interventions“) verwendet (z. B. das Setzen und Vernähen von Hautschnitten ohne weitergehenden Eingriff bei chirurgischen Studien [14]).

Beachtet werden sollte die Bedeutung einer Randomisierung für die Zuverlässigkeit der Verblindung (84) da nicht-zufällige Entscheidungen gerade in kleinen Stichproben häufig erraten werden können.

Besonderheiten im Leistungs- und Spitzensport

Einige typische Interventionen im Sport (z. B. Trainingsinterventionen) sind kaum zu verblinden, insbesondere bei dem großen Vorwissen von Trainern und Athleten im Spitzenbereich. Bei anderen Interventionsformen sind kreative Lösungen möglich. Ein gutes Beispiel ist die Studie von Broatch et al. zur Kaltwasserimmersion (15). Die inhärente Wirkungslosigkeit der möglichen Placebos sollte jedoch kritisch hinterfragt werden. Ein Beispiel für eine zumindest fragliche Wirkungslosigkeit sind kalorienfreie Süßstoffe als Placebo für zuckerhaltige Getränke (75).

5.2 Datenaufnahme

5.2.1 Grundüberlegungen

Wesentliches Ziel von Anpassungen der Datenaufnahme an eine kleine Fallzahl ist eine Optimierung des Signal-Rausch-Verhältnisses.

Dadurch wird die Abgrenzung kleiner Effekte (Signal) von zufälliger Streuung (Rauschen) erleichtert. Genauer gesagt zielen die Anpassungen im Sinne des Max-Kon-Min Prinzips (33, 68) auf eine Minimierung des Standardfehlers bzw. eine Maximierung der beobachteten Effektstärke durch: (a) Die Erfassung des Signalmaximums (Primärvarianz), (b) die Kontrolle weiterer

Einflussfaktoren (Sekundärvarianz), und (c) die Minimierung zufälliger Streuung (Fehlervarianz). Hinzu kommt (d) Die Erhöhung der Anzahl von Datenpunkten, mit denen im Rahmen der Datenanalyse gearbeitet werden kann. Teil dieser Strategie sollte auch sein, einen Verlust von Information im Rahmen der Datenverarbeitung und -analyse zu vermeiden. Das bedeutet in erster Linie, dass kontinuierliche Merkmale nicht dichotomisiert oder anderweitig klassifiziert werden sollten (z. B. anhand der beobachteten Prä-Post Differenz im Rahmen einer „Responderanalyse“ [44, 81, 82]).

5.2.1.1 Zufällige Streuung minimieren

Der Kampf gegen die zufällige Streuung beginnt auf der Ebene des Messfehlers mit einer Priorisierung der Reliabilität von Messinstrumenten und/oder Mehrfachmessungen mit nachfolgender Mittelwertbildung (40). *(Anmerkung: Im Kontext (gemischter) Modelle kann und sollte in der Regel mit den Rohwerten gearbeitet werden.)* Eine stringente Standardisierung von Testabläufen, Interventionen und Rahmenbedingungen begrenzt die Auswirkungen bekannter oder vermuteter Störgrößen auf die gemessenen Werte und deren Veränderungen im Längsschnitt. Dabei sollte eine bewusste Abwägung der erzielten Erhöhung von Effekt- und Teststärke mit der externen Validität bzw. der Anwendbarkeit der Ergebnisse in der Praxis erfolgen. Da im Trainingsalltag eine umfassende und gezielte Standardisierung nicht möglich (oder zumindest nicht verhältnismäßig) ist, wäre dort allein wegen des dadurch höheren Nenners schon mit niedrigeren Effektstärken zu rechnen (88).

5.2.1.2 Maximum des Signals abgreifen

Durch ein durchdachtes und wohlinformiertes Timing von Tests kann sichergestellt werden, dass das Maximum eventueller Effekte erfasst wird (40). Beispielsweise sollte der Post-Test in Trainingsstudien durchgeführt werden nachdem die akuten Effekte der letzten Trainingseinheit abgeklungen sind, aber bevor ein Detraining einsetzt. Gleichzeitig muss die Dauer der

Trainingsphase ausreichend lang sein. Dafür ist zumindest eine begründbare Abschätzung der zeitlichen Verläufe akuter und chronischer Effekte notwendig, wobei diese Zeitverläufe sich zwischen verschiedenen Parametern unterscheiden können. Falls die Kinetik der bedeutsamen Effekte unbekannt ist kann es sinnvoll sein diese durch Vorstudien abzuschätzen. Diese Vorstudien können in der Regel auf niedrigerem Leistungsniveau durchgeführt werden. Daher ist diese Option besonders dann attraktiv, wenn das Missverhältnis zwischen notwendiger und realisierbarer Fallzahl durch die Exklusivität der Zielgruppe bedingt ist.

5.2.2 Spezielle Studiendesigns

Ein wesentliches Merkmal von Designs für Studien mit kleiner Fallzahl sind unterschiedliche Formen der Wiederholung auf Ebene des einzelnen Teilnehmers. Die Grundlogik dabei ist, dass durch Gewinnung mehrerer Datenpunkte pro Teilnehmer die notwendige Anzahl von Beobachtungen nicht mehr allein über die Fallzahl erreicht werden muss. Außerdem fungiert in einigen dieser Designs jeder Teilnehmer als seine bzw. ihre Kontrolle. Dadurch steigt nicht nur die Zahl der Teilnehmer, an denen der Interventionseffekt beobachtet werden kann. Darüber hinaus erhöht das perfekte „Matching“ (die gleichen Personen erhalten Test und Kontrollbedingung) die Teststärke im Vergleich zu einem Parallelgruppendesign (92).

5.2.2.1 Reversible Effekte – Cross-Over Designs

Cross-over Studien bestehen aus einer Interventions- und einer Kontrollphase mit einer Auswaschphase dazwischen. Die Reihenfolge von Interventions- und Kontrollbedingung für den einzelnen Teilnehmer wird randomisiert festgelegt. Randomisierte Cross-over Studien sind eine Sonderform der randomisierten kontrollierten Studie (RCT) (40, 92), die insbesondere dann in Betracht gezogen werden kann, wenn die Anzahl der einzelnen Teilnehmer limitiert ist. Wichtigste Voraussetzung ist die Reversibilität des Interventionseffekts. Das wiederholte

Cross-Over Design geht noch weiter und beinhaltet jeweils mindestens zwei Interventions- und Kontrollperioden mit den entsprechenden Auswaschphasen (44, 81, 90). Neben einer weiteren Erhöhung der Teststärke ermöglicht die Wiederholung des Vergleichs von Interventions- und Kontrollbedingung auf Ebene des einzelnen Teilnehmers hier zusätzlich eine valide Beurteilung der interindividuellen Unterschiede in der Wirksamkeit („individuelle Response“) (44, 80-82). Ein wesentlicher Nachteil ist die deutliche Verlängerung der Einbindung jedes einzelnen Teilnehmers und deren mögliche Konsequenzen für die Adhärenz, Compliance und – im speziellen Fall des Leistungssports – der Vereinbarkeit mit der periodisierten Trainingsplanung. Die Möglichkeit von Carry-Over-Effekten durch über die Auswaschphase verbleibende Interventionseffekte sollte ebenfalls bedacht werden. Eine umfassende Darstellung der Planung und Auswertung von Cross-Over Studien wurde von Senn publiziert (92).

5.2.2.2 Nicht reversible Effekte – Repeated Testing und Multiple Baseline Design

Wenn Interventionseffekte nicht (sicher) reversibel sind (z. B. bei Interventionen mit ausgeprägtem Lern- oder epigenetischen Effekten) oder wenn eine Cross-over Studie aus praktischen Gründen nicht in Frage kommt (z. B. wegen der Studiendauer oder einer fraglichen Compliance mit dem Absetzen der Intervention in der Auswaschphase) können andere Formen der Wiederholung auf Ebene des einzelnen Teilnehmers einen zumindest teilweisen Ausgleich bieten. Für länger dauernde (Trainings-) Interventionen kommt eine Ergänzung des Parallelgruppendesigns (RCT) mit Eingangs- und Abschlusstest durch ein wiederholtes Testen über die Interventionsphase hinweg in Betracht (46). Eine standardisierte Baseline -Phase mit wiederholten Messungen ermöglicht die Berechnung individueller Baseline-Referenzbereiche und trägt somit zusätzlich zur Abgrenzung kleiner Veränderungen durch eine nachfolgende Intervention bei (44, 45). Im Kontext von Einzelfallstudien wird dieses Design auch als „AB Design“ bezeichnet. Wird der Zeitpunkt des Über-

gangs von der Baseline zur Interventionsphase variiert ergibt sich das Multiple Baseline Design (55). Können Veränderungen in der abhängigen Variablen jeweils nach dem Übergang von der Baseline in die Interventionsphase beobachtet werden, trägt dies zur Plausibilität eines kausalen Zusammenhangs bei. Ein wesentlicher Nachteil des Multiple Baseline Designs ist die notwendige Anzahl von Messungen. Wenn Effekte auf Leistung und Beanspruchung untersucht werden, kann hier die Verwendung von Trainings- („training is testing“) oder Monitoringdaten sinnvoll sein.

(Anmerkung: Im Vorgriff auf die Darstellung der datenanalytischen Optionen sei darauf hingewiesen, dass eine zum Studiendesign komplementäre Ausgestaltung der Datenanalyse hier unbedingt erforderlich ist. Insbesondere muss die Datenanalyse mit der Vermischung intra- und interindividueller Varianz in der hierarchischen Datenstruktur umgehen können (4, 44, 46, 82, 83).)

5.2.3 Single-Subject Designs und Individualisierung

Studiendesigns, die den Vergleich zu einer Kontrollbedingung und wiederholte Beobachtungen auf Ebene des einzelnen Teilnehmers verbinden, können als Aggregation kontrollierter, ggf. randomisierter Einzelfallstudien betrachtet werden (4, 80, 82, 83). Sie werden deshalb auch als „Single-Subject Designs“ bezeichnet. Wie bereits erläutert bietet die Aggregation solcher „Ein-Personen Studien“ die Möglichkeit mit wenigen Teilnehmern eine ausreichende Anzahl von Datenpunkten für gruppenstatistische Analysen zu generieren. Die wichtigsten Beispiele sind die oben bereits angesprochenen Designtypen wiederholtes Cross-Over für reversible Effekte und das Multiple-Baseline Design für nicht-reversible Effekte oder wenn praktische Erwägungen gegen ein Cross-Over sprechen. Die Aggregation von Datenpunkten mehrerer Teilnehmer ist dabei entscheidend für eine Generalisierbarkeit über die Studie selbst hinaus. Die Ergebnisse einer isolierten Ein-Personen Studie besitzen dagegen speziell für das untersuchte Individuum Gültigkeit. Im Spitzensport werden Einzelfallstudien bereits seit Jahrzehnten

als bedeutsames Instrument betrachtet (8, 55, 56, 73). Im Kern steckt dahinter die Überlegung, dass Spitzensportler extreme, positive Ausreißer sind, zu deren Charakterisierung eine gruppenbasierte Herangehensweise weder geeignet noch durchführbar ist (49, 99). Die Elemente von Kontrolle, experimenteller Manipulation, Wiederholung und gezielter Standardisierung (eventuell auch von Randomisierung und Verblindung), wie sie in Single-Subject Designs flexibel implementiert werden können, stärken dabei die Genauigkeit und (innere und prädiktive) Validität der Ergebnisse (4, 8, 55, 73, 82, 83, 89, 90). Darüber hinaus bieten Single-Subject Designs die Möglichkeit zu Analysen auf individueller Ebene z. B. im Sinne eines Vergleichs der individuellen Response auf unterschiedliche Interventionsformen (4, 44, 80-82). Bei Designs ohne oder mit nur einem intraindividuellen Vergleich zwischen Intervention und Kontrolle (z. B. RCT oder einfaches Cross-Over) ist eine Beurteilung der individuellen Response nicht

möglich, da eine Wiederholung auf der Ebene des Interaktionseffekts fehlt („no repetition – no interaction“) (44, 80, 82). Letztlich handelt es sich bei Einzelfallstudien um ein „Ausprobieren nach wissenschaftlichen Standards“, das eine evidenzbasierte Individualisierung u. a. von Belastungssteuerung und Regenerationsmanagement ermöglicht (44, 55, 80, 82). *(Anmerkung: Wenn die Ergebnisse der einzelnen „Ein-Personen Studien“ in der Betreuung des jeweiligen Athleten genutzt werden sollen, kann es sinnvoll sein, dafür andere analytische Verfahren zu verwenden wie bei den gruppenstatistischen Analysen (4, 82, 89). Vgl. Kapitel 4.2 und 5.3 insbesondere 5.3.5)*

Einen Überblick über zwei wesentliche Grundtypen der Single-Subject Designs bieten die folgenden Vignetten. Weitere Single-Subject Designs können als (limitierte) Varianten der beiden Grundtypen verstanden werden. Einen Überblick bieten z. B. (8, 55). Weiterführende Ressourcen finden sich im Anhang.

Single-Subject Designs

Studiendesigns, die auf Ebene des einzelnen Teilnehmers den Vergleich zu einer Kontrollbedingung und die Wiederholung von Beobachtungen verbinden werden als aggregierte Einzelfallstudien oder „Single-Subject Designs“ bezeichnet (4, 80, 82, 83). Durch die intraindividuelle Wiederholung ergeben sich mehrere Beobachtungen pro Teilnehmer. Dadurch sinkt die notwendige Fallzahl. Gleichzeitig werden Analysen auf Ebene des einzelnen Teilnehmers möglich (z. B. valide Beurteilung der individuellen Response [44, 82]). Dabei kann es angebracht sein, für Analysen auf Gruppen- und Einzelfallebene unterschiedliche statistische Herangehensweisen zu verwenden (47, 89).

Single-Subject Designs sind vielfältig und erlauben eine flexible Umsetzung der methodischen Grundelemente. Die wichtigsten Grundtypen sind das wiederholte Cross-Over Design und das Multiple Baseline Design. Die Auswahl richtet sich im Wesentlichen nach der Reversibilität des Interventionseffekts, der Durchführbarkeit einer Auswaschphase und dem verfügbaren Zeitfenster.

Wiederholtes Cross-Over Design

Eine oder mehrere Interventionsbedingungen werden je mindestens zweimal angewendet, getrennt jeweils von einer Auswaschphase. So kann die Effektivität einer oder mehrerer Interventionen auf Gruppen- und Einzelfallebene beurteilt werden.

Wesentliche Voraussetzung ist die Reversibilität des Interventionseffekts innerhalb einer praktikablen Auswaschphase.

Wesentliche Nachteile sind die lange Studiendauer und die Möglichkeit von Carry-Over Effekten.

Besonderheiten im Leistungs- und Spitzensport

Die Verkettung von Studienphasen ist insbesondere vor dem Hintergrund einer periodisierten Trainingsplanung problematisch. Bei länger dauernden (z. B. Trainings-) Interventionen wird es schwer, ausreichend lange Zeitfenster unter konstanten Rahmenbedingungen zu finden. Das wiederholte Cross-Over Design eignet sich daher im Kontext des Sports insbesondere für die Untersuchung akuter Effekte (36, 107). Die Rückmeldung seiner individuellen Response an den Athleten bzw. den Trainer ist eine gute Möglichkeit das Aufwand-Nutzen-Verhältnis einer Studienteilnahme für die Praxispartner zu verbessern (47).

Multiple Baseline Design

Das Multiple Baseline Design ruht auf zwei Pfeilern: (a) Wiederholte Messungen der abhängigen Variable(-n) über eine Baseline- und die anschließende Interventionsphase hinweg und (b) die Variation des Zeitpunkts für den Übergang von der Baseline- in die Interventionsphase. Treten Änderungen in abhängigen Variablen jeweils mit festem zeitlichem Bezug nach dem Beginn der Interventionsphase auf, trägt dies zur Plausibilität eines kausalen Effekts (im Vergleich zum AB Design ohne Variation des Beginns der Interventionsphase) bei.

Eine Reversibilität des Effekts ist nicht notwendig, da die Studie nur eine Interventionsphase beinhaltet und mit dieser endet.

Die individuelle Response der Teilnehmer kann abgeschätzt werden, wenn auch mit Limitationen (46, 82).

Ein wesentlicher Nachteil ist die Anzahl der notwendigen Messzeitpunkte.

Besonderheiten im Leistungs- und Spitzensport

Insbesondere bei Leistungstests oder invasiven Messungen kann die Anzahl der Messzeitpunkte problematisch werden. Die Nutzung von Trainingsdaten („training as testing“) oder/und Daten aus dem Monitoring kann hier hilfreich sein.

5.3 Datenanalyse

5.3.1 Bezug von Datenaufnahme und Datenanalyse

Ziel der Datenanalyse sind unverzerrte und präzise Antworten auf die jeweiligen Forschungsfragen. Dazu ist es gerade bei kleinen Fallzahlen notwendig, das Studiendesign (bzw. Datenstruktur) und die realisierte Fallzahl effizient zu nutzen und somit in Wert zu setzen (40).

Analog zur Datenaufnahme lassen sich auch bei der Datenanalyse erwartbare Aspekte (Pflichtprogramm) und optionale Optimierungsmöglichkeiten (Kür) unterscheiden.

5.3.1.1 Studienplanung und Registrierung

Eine detaillierte und verbindliche Planung trägt entscheidend zur Qualität der Ergebnisse bei. Diese generelle Regel gilt in besonderem Maße für Studien mit kleiner Fallzahl (21). Dazu gehört auch, dass Datenaufnahme und Datenanalyse in dieser klaren Reihenfolge durchgeführt werden. Zwischenanalysen mit Rückwirkungen auf die weitere Datenaufnahme (z. B. eine Fortsetzung der Aufnahme weiterer Teilnehmer bis ein signifikantes Ergebnis erzielt wird) sollten bei kleinen Studien aufgrund der oben bereits diskutierten Instabilität der Ergebnisse unbedingt vermieden werden (21). Dies gilt auch für die Aggregation von Einzelfallstudien. Hier

muss klar zwischen einer prospektiven Studie und retrospektiven Analysen unterschieden werden.

Ebenso wichtig ist eine konkrete Festlegung des Plans für die Datenanalyse bereits im Rahmen der Studienplanung (also vor Beginn der Datenaufnahme), im Gegensatz zu einer „flexiblen Analyse“ bei der verschiedene Varianten ausprobiert und positive Ergebnisse selektiv publiziert werden („p-Hacking“). Bei einer „flexiblen Analyse“ ist zwar die Chance höher ein signifikantes und somit publizierbares Ergebnis zu erhalten, allerdings auf Kosten einer (noch) höheren Wahrscheinlichkeit falsch positiver Ergebnisse. Diese „Alphafehlerkumulation“ führt zu einer weiteren Absenkung des positiven prädiktiven Werts. Dabei kann ein Auswertungsplan durchaus auf Unwägbarkeiten der Datenaufnahme Bezug nehmen (z. B. Einsatz parametrischer bzw. nicht parametrischer Verfahren je nach Ergebnis der Voraussetzungsprüfung). Eine Auswahl von zu berücksichtigenden Kovariaten in Abhängigkeit von signifikanten Unterschieden bei Studieneinschluss („baseline balance“) ist dagegen kritisch zu beurteilen (84, 91). Diese sollten vor Beginn der Datenaufnahme theoriegeleitet festgelegt werden. Der Analyseplan sollte auch eine explizite Festlegung der Hauptzielparameter und Angaben zum Umgang mit multiplen Testungen (z. B. Bonferroni Korrektur) enthalten (3).

Die kritische Bedeutung eines detaillierten und verbindlichen Plans für Datenaufnahme und Datenanalyse macht eine Registrierung von Studien mit kleiner Fallzahl nahezu unverzichtbar. Eine Möglichkeit sind Studienregister z. B. bei der Open Science Foundation (<https://osf.io/prereg/>) oder auf www.clinicaltrials.gov. Die Registrierung bei einer wissenschaftlichen Zeitschrift („registered report“) bietet nicht nur eine gewisse Sicherheit bezüglich der (ergebnisunabhängigen) Veröffentlichung sondern auch frühzeitige Rückmeldungen von externen Gutachtern, die ggf. noch bei der Durchführung berücksichtigt werden können. „Registered reports“ werden zunehmend auch von sportwissenschaftlichen Zeitschriften angeboten. Eine weitere Möglichkeit ist die eigenständige Veröffentlichung des Studienprotokolls. Diese Option kommt aber in der Regel nur für besonders

relevante Studien in Betracht. (Anmerkung: Die Notwendigkeit einer verbindlichen und detaillierten Festlegung des Studienplans (incl. Datenanalyse) vor Beginn der Datenaufnahme bezieht sich auf die primäre Fragestellung einer Studie. Explorative Sekundäranalysen sind auch bei kleinen Datensätzen statthaft, müssen aber zwingend eindeutig als solche gekennzeichnet werden.)

5.3.2 Generelle Aspekte mit besonderer Bedeutung bei kleinen Fallzahlen

Ähnlich wie bei der Datenaufnahme gibt es auch bei der Datenanalyse Aspekte, die zwar allgemeine Gültigkeit besitzen, aber bei kleinen Fallzahlen von besonderer Bedeutung sind und daher konsequent umgesetzt werden sollten. Zu diesem „Pflichtprogramm“ gehört z. B. die einseitige Hypothesentestung, wenn dies der (gerichteten) Hypothese entspricht (40).

5.3.2.1 Missing data

Datenlücken („missing data“) und der Umgang mit ihnen sind bei kleinen Fallzahlen besonders relevant (65). Zum einen ist auch der Ausschluss einzelner Studienteilnehmer ein bedeutsamer Verlust, zum anderen ist die Leistungsfähigkeit gängiger Verfahren zum imputieren („aufüllen“) von Datenlücken von der Größe des Datensatzes insgesamt abhängig. Es sei angemerkt, dass hierarchische Studiendesigns und gemischte Modelle relativ robust gegenüber Datenlücken sind, vorausgesetzt diese sind zufällig. Ein verständlicher Überblick findet sich bei Bennett (10) und Bell (9), spezielle Überlegungen im Zusammenhang mit kleinen Fallzahlen werden in (65) besprochen

5.3.2.2 Rohdatendarstellung

Kleine Fallzahlen sind mit vielen Nachteilen verbunden. Ein Vorteil ist dagegen die Möglichkeit einer übersichtlichen und informativen Visualisierung von Rohdaten. Dieser Vorteil sollte sowohl im Sinne der Ergebniskommunikation

als auch im Sinne der Transparenz genutzt werden. Klassische Möglichkeiten sind Einzelfallprofile (z. B. Prä-Post Verläufe) und Scatterplots. Bei Studiendesigns mit zahlreichen Wiederholungen auf individueller Ebene und ggf. unter verschiedenen Bedingungen ermöglichen Variabilitätsplots einen guten Überblick. Abbildung 3 illustriert ein Beispiel aus einer Studie zur Effektivität der Kaltwasserimmersion bei professionellen Fußballspielern. Wenn die Voraussetzungen inferenzstatistischer Verfahren („Hypothesentests“) nicht erfüllt sind sollte erwogen werden, ausschließlich Visualisierungen (18, 23) und deskriptive Verfahren (79) zu verwenden.

5.3.2.3 Verteilungsform / Datentransformation

Eine unter Umständen sehr effektive, aber nur selten konsequent genutzte Optimierungsmöglichkeit ist der differenzierte Umgang mit der Verteilungsform. Prinzipiell sind Testverfahren, die eine Verteilungsannahme machen (häufigste Annahme: Normalverteilung), in der Regel effizienter – d. h. man benötigt (*ceteris paribus*) weniger Teilnehmer bzw. Datenpunkte. Eine Anwendung solcher „parametrischen“ Verfahren ist daher gerade bei Studien mit kleiner Fallzahl sinnvoll, vorausgesetzt die Voraussetzungen für deren Anwendung sind gegeben. Allerdings ist die übliche Prüfung auf Abweichungen von der Normalverteilung (z. B. durch einen Shapiro-Wilds Test) bei kleinen Stichproben doppelt problematisch: Durch die geringe Teststärke führen selbst erwartbare und visuell deutlich erkennbare Abweichungen von der Normalverteilung nicht zu einem signifikanten Testergebnis. Andererseits können die durch Abweichungen zwischen beobachteter und theoretischer Verteilungsform bedingten Einbußen an Teststärke nicht durch eine höhere Fallzahl ausgeglichen werden.

Durch Transformation von Rohdaten kann die Übereinstimmung zwischen der beobachteten Verteilung der Rohwerte und der von parametrischen Verfahren angenommenen, theoretischen Verteilung verbessert werden. Dadurch können die Teststärke bzw. die Präzision von Schätzwerten häufig deutlich erhöht werden (40, 45, 46). Ein typisches Szenario sind

Parameter mit einem Baseline-Mittelwert nahe und einem harten Limit bei Null – aber einem Potential für Auslenkungen mit deutlich höheren Werten. Typische Beispiele sind das Muskelenzym Kreatinkinase (CK) und der Entzündungsparameter C reaktive Protein (CRP) (45). Eine log Transformation reduziert die Rechtsschiefe der beobachteten Verteilung (also den „Schwanz“ hoher Werte) und damit die Diskrepanz zur symmetrischen Normalverteilung (45). Im Gegensatz dazu haben in homogenen (Trainings-) Gruppen hochtrainierter Athleten Leistungsdaten häufig eine Linksschiefe aus „schlechten Tagen“. Diese kann durch eine log Transformation der Differenz zum Maximum (Betrag) reduziert werden. In anderen Fällen sind die adäquate theoretische Verteilung und/oder angemessene Transformationsschritte weniger klar und sollten im Rahmen einer statistischen Beratung geklärt werden.

Neben der Annäherung an eine bestimmte Verteilungsform der Rohdaten kann eine Datentransformation auch dazu beitragen, dass andere Voraussetzungen parametrischer Testverfahren besser erfüllt werden (z. B. Varianzhomogenität und Normalverteilung der Residuen) (28). Andererseits muss die zusätzliche Komplexität in Datenanalyse und Interpretation der Ergebnisse adäquat berücksichtigt werden (12, 34). Eine lesbare Einführung bietet (28). In bestimmten Fällen kann durch die Verwendung gemischter Modelle eine Datentransformation vermieden werden (siehe unten).

5.3.3 Hierarchische Datenstruktur und gemischte Modelle

Wenn sich in einem Datensatz mehrere Beobachtungen einer übergeordneten Einheit zuordnen lassen, spricht man von einer hierarchischen Datenstruktur. Bei Studiendesigns mit Mess- und/oder Interventionswiederholung ist diese übergeordnete Einheit der einzelne Teilnehmer. Dabei kommt es zu einer Vermischung von inter- und intraindividuelle Streuung, die in einfachen statistischen Verfahren nicht vorgesehen ist (quasi eine Mischform aus verbundener und unverbundener Stichprobe). Gemischte Modelle erweitern das allgemeine

lineare Modell durch die Möglichkeit eine solche „hierarchische“ Datenstruktur abzubilden und bieten sich daher zur Auswertung bei intervall- oder verhältnisskalierten Zielvariablen an (4, 46). Zufällige Effekte repräsentieren dabei Quellen zufälliger Streuung um die systematischen festen Effekte (z. B. Intervention vs. Kontrolle), kontinuierliche Kovariaten können ebenfalls berücksichtigt werden. Im Kontext der hier besprochenen Designs wird in der Regel der jeweilige Teilnehmer als zufälliger Effekt modelliert, dadurch werden interindividuelle Unterschiede in den Rohwerten berücksichtigt (das individuelle „Niveau“). Falls das Studiendesign eine Wiederholung der Intervention beinhaltet (z. B. ein wiederholtes Cross-Over) kann auch die interindividuelle Streuung der Interventionseffekte valide analysiert werden („individuelle Response“). Dafür wird die Interaktion

zwischen Teilnehmer und Interventionsbedingung als weiterer zufälliger Effekt in das Modell aufgenommen (44). Darüber hinaus können gemischte Modelle für die Analyse nicht normalverteilter Rohdaten ohne Transformation (d. h. in der ursprünglichen Einheit) (62) und kleiner Datensätze mit „missing data“ genutzt werden (siehe jeweilige Abschnitte). Wichtige Voraussetzung ist eine Mindestgröße des Datensatzes bzw. eine Mindestanzahl an Freiheitsgraden (29).

Empfehlenswerte, praxisorientierte Einführungen in gemischte Modelle finden sich in (35, 39). Spezifische Empfehlungen für die Anwendung bei Studien mit Cross-Over (92), wiederholtem Cross-Over (4, 82, 83) oder Messwiederholungsdesign (46) können der Literatur entnommen werden. Zumindest bei der erstmaligen Anwendung ist eine statistische Beratung dringend empfehlenswert.

Hierarchische Datensätze und gemischte Modelle

- ▶ Wenn sich mehrere Beobachtungen einer übergeordneten Einheit zuordnen lassen, spricht man von einer hierarchischen Datenstruktur. Im Kontext von Studiendesigns für kleine Fallzahlen ist die übergeordnete Einheit der einzelne Teilnehmer.
- ▶ Zur Analyse eignen sich insbesondere gemischte lineare Modelle. Die ID des Teilnehmers ist dabei ein zufälliger Effekt.
- ▶ Detaillierte Hilfestellung für die Analyse von Studien mit unterschiedlichen hierarchischen Designs können der Literatur entnommen werden (Cross-over (92), wiederholtes Cross-over (4, 82, 83), wiederholtes Testen im Zuge einer ununterbrochenen Trainingsphase (46), andere Single-Subject Designs (55))
- ▶ Zumindest bei erstmaliger Planung einer Studie mit hierarchischem Design ist eine statistische Beratung dringend empfehlenswert.

5.3.4 Resampling und Schrumpfschätzer

Zwei wesentliche Limitationen der Ergebnisse von Studien mit kleiner Fallzahl sind Verzerrung (englisch „bias“) und Varianz. Verzerrung ist der Unterschied zwischen dem empirischen Ergebnis und dem wahren Wert, wie er mit einer sehr großen Fallzahl angenähert würde. Varianz ist die Variation der Ergebnisse von Stichprobe zu Stichprobe. Resamplingtechniken und

Schrumpfschätzer bieten die Möglichkeit diese Limitationen zu begrenzen und gleichzeitig Ihr Ausmaß abzuschätzen.

5.3.4.1 Resampling

Resampling-Techniken können verwendet werden, um die Verzerrung zu reduzieren und die Varianz abzuschätzen (d. h. wie „robust“ die Ergebnisse sind) (74, 76, 100). Als generelle

Regel kann gelten, dass bei kleinen Fallzahlen im Originaldatensatz das Jackknife Resampling anderen Optionen (Bootstrapping) vorzuziehen ist (74). Wie der Name „Resampling“ bereits andeutet werden aus dem ursprünglichen Datensatz wiederholt Stichproben gezogen. Diese „Resample“ werden jeweils mit den üblichen Verfahren analysiert und die Ergebnisse anschließend als Mittelwert zusammengefasst (74). Die Standardabweichung der Ergebnisse der Resample vermittelt einen Eindruck der Varianz. Bei der klassischen Jackknife Methode werden die einzelnen Resample jeweils durch Auslassen einer Beobachtung aus dem ursprünglichen Datensatz gebildet. Bei 10 Teilnehmern werden also 10 Resamples mit jeweils 9 Individuen gebildet und analysiert. Es gibt aber auch Varianten, die im Einzelfall sinnvoll sein können (45).

5.3.4.2 Schrumpfschätzer

Die mit kleinen Fallzahlen (im Mittel) einhergehende Verzerrung der Effektschätzer nach oben („indirect bias“, siehe Kapitel 3.3) kann durch fallzahlabhängige Schrumpffaktoren berücksichtigt werden (86): Im einfachsten Fall beträgt der Schrumpffaktor $n/n+1$ (86). Im konkreten Fall sollten Schrumpffaktoren jedoch durch einen Statistiker bestimmt werden. Schrumpfschätzer können auch bei der Analyse kleiner Subgruppen sinnvoll eingesetzt werden. In diesem Fall werden die Effektschätzer der Subgruppe in Richtung des Gesamtergebnisses verschoben. Interessanterweise bedeutet das, dass Daten bzw. Information von außerhalb der jeweiligen Subgruppe in die Analyse einbezogen wird. Im Rahmen der vorherrschenden, frequentistischen Statistik handelt es sich dabei um eine (zumindest in der Sportwissenschaft) wenig genutzte Optimierungsoption. Dagegen ist die Möglichkeit einer systematischen Verknüpfung der aktuellen Daten mit Vorwissen ein zentrales Merkmal der Bayesianischen Statistik (45, 101).

5.3.5 Strukturierte Berücksichtigung von Vorwissen & Bayesianische Statistik

In der Statistik werden zwei grundlegende Systeme unterschieden: (a) die zumindest in den Lebenswissenschaften dominierende frequentistische Statistik und (b) die Bayesianische Statistik. Die Bayesianische Statistik bietet gerade bei kleinen Fallzahlen einige Vorteile (37, 38, 101). Ein wesentlicher davon ist die Möglichkeit Vorwissen über den jeweiligen Effekt (z. B. die Wirksamkeit einer Intervention) strukturiert und logisch nachvollziehbar als „Prior“ in die Analyse einzubeziehen. Dabei wird der „Prior“ in Abhängigkeit von der Evidenz aus den Daten der neuen Studie abgeändert („Bayesian Updating“). Die intuitive Attraktivität dieses Prinzips ist seit langem bekannt. Durch aktuelle Software wird die praktische Umsetzung erheblich erleichtert.

5.3.5.1 Informative Priorverteilung

Eine informative Priorverteilung beschreibt den Bereich der Werte (z. B. für die Wirksamkeit einer Intervention), die, entsprechend dem Wissensstand vor Beginn der Studie, plausibel erwartet werden können. Durch diesen zusätzlichen Input erhält die Datenanalyse einen (Informations-) Vorsprung, der gerade bei kleinen Fallzahlen relevant wird (7, 45, 93). Dabei hängt der Wert des Priors von der Anwendbarkeit bzw. Übertragbarkeit des Vorwissens auf die Rahmenbedingungen der aktuellen Studie ab. Daher sollten Priorverteilungen so spezifisch sein wie möglich (Sportart, Niveau, Zielparameter, Zeitverlauf etc.). Die Faktoren, die bei der Abschätzung der Übertragbarkeit zu beachten sind unterscheiden sich nicht von der Bildung einer begründeten Einschätzung in anderem Kontext z. B. bei der generellen Rechtfertigung einer Studie an Spitzenathleten oder in einer bestimmten Sportart. Auch die möglichen Informationsquellen sind prinzipiell dieselben: Empirisch begründbare Priorverteilungen können aus früheren/publizierten wissenschaftlichen Arbeiten abgeleitet werden. Idealerweise sind diese schon in einer Metaanalyse systematisch

zusammengefasst oder dieser Schritt ist Teil der Projektstrategie. Analog dazu kann es auch Teil der Projektstrategie sein, gezielt passendes empirisches Vorwissen zu erzeugen z. B. durch Vorstudien auf geringfügig niedrigerem Leistungsniveau. Je nach Fragestellung kommen auch Routinedaten als Grundlage für eine informative Priorverteilung in Frage (45). Theoretischer Nachteil der numerischen Verwendung empirischer Daten zur Charakterisierung der Priorverteilung ist die Überlagerung mit zufälliger Streuung (z. B. Messfehler) (86). Alternativ kann aus subjektiven Einschätzungen ohne expliziten Bezug zu empirischen Ergebnissen („Expertenmeinung“) eine Priorverteilung gebildet werden („expert elicitation“) (45, 101).

Wichtig ist zu beachten, dass die Priorverteilung, auch wenn sie auf empirischer Grundlage gebildet wird, letztlich eine subjektive Einschätzung widerspiegelt. Eine zwingende Herleitung ist nicht möglich (wohl aber eine abwägende Begründung). Das entspricht der Pluralität des wissenschaftlichen Meinungsbilds, führt aber zu einer der wesentlichen Limitation Bayesianischer Analysen mit informativer Priorverteilung: Die Ergebnisse sind nur unter der Voraussetzung gültig, dass die Priorverteilung als plausibel akzeptiert wird. Ein pragmatisches Kriterium für eine konsistente Priorverteilung ist, dass man sie als Grundlage einer Wette auf das Studienergebnis verwenden würde (87). Die Durchführung der Analysen mit mehreren Priorverteilungen, die die plausibel begründbare Bandbreite abdecken, erlaubt es den Einfluss des Priors auf die Ergebnisse abzuschätzen. Da die relative Bedeutung von Prior und aktuellen Daten u. a. von der Anzahl der Datenpunkte

abhängt, ist eine solche Sensitivitätsanalyse insbesondere bei sehr kleinen Datensätzen bedeutsam. Weitere Erläuterungen hierzu einschließlich einer interaktiven Software finden sich in (32).

5.3.5.2 Möglichkeiten zur nachvollziehbaren Ableitung einer Priorverteilung

Die üblichste Form eines Priors ist die Priorverteilung (101, 105). Dabei müssen Verteilungsform und Verteilungsparameter so festgelegt werden, dass sie den vor Studienbeginn plausiblen Wertebereich charakterisieren. In Box 1 wird dies an einem Beispiel illustriert, dabei dient eine Metaanalyse als Grundlage. Ein Beispiel für die Ableitung einer Priorverteilung aus Routinedaten im Leistungssport findet sich in (45). Praktikable Möglichkeiten zur Ableitung einer Priorverteilung aus Praxis-/Erfahrungswissen („expert elicitation“) werden in (101) erläutert. Eine Alternative zur Priorverteilung sind simulierte Datensätze („data priors“, [37]). Dabei werden die simulierten Datenpunkte zusammen mit den im Rahmen der Studie empirisch erhobenen mit den üblichen statistischen Verfahren ausgewertet. Dadurch kann der Einsatz simulationsbasierter Verfahren umgangen werden. Erwähnenswert ist, dass Bayesianische Verfahren auch dann angewendet werden können, wenn adäquates Vorwissen nicht verfügbar ist oder die Analysen bewusst auf die im Rahmen der Studie erhobenen Daten beschränkt werden sollen. In diesem Fall werden nicht informative Priorverteilungen („flat priors“, „diffuse priors“) verwendet.

Box 1: Beispiel für die Ableitung einer informativen Priorverteilung aus empirischem Vorwissen

Ziel: Priorverteilung für die Wirksamkeit einer Ganzkörper-Kaltwasserimmersion auf die Wiederherstellung der Sprintleistungsfähigkeit bei Fußballspielern.

Grundlage: Metaanalyse von Poppendieck et al. (71): „Cooling and performance recovery of trained athletes: a meta-analytical review“

Erster Schritt: Abschätzung des zu erwartenden Wertebereichs: Über alle Kühlmethoden, Leistungsparameter und Zeitabstände hinweg lag der Vorteil der Kaltwasserimmersion gegenüber der

Kontrollbedingung bei 2,4% der Ausgangsleistung. Für Sprintleistungen war die Wirksamkeit höher als für Ausdauer und Kraftleistungen (2,6% der Ausgangsleistung), ebenso (bei Einteilung nach der Kühlmethode) für die Ganzkörper-Kaltwasserimmersion (5,1%, Hedge's g 0,62 [0,45 0,8]). Für die Kombination aus beiden Kriterien scheint daher eine Erwartung von 5% der Ausgangsleistung plausibel. Ein negativer Effekt ist unwahrscheinlich.

Zweiter Schritt: Verteilungsform festlegen: Weder theoretisch noch empirisch (z. B. [48]) ergeben sich klare Hinweise auf eine relevante Abweichung von der Normalverteilung.

Dritter Schritt: Übersetzung der Erwartung in Verteilungsparameter

Die Erwartung „Vorteil von 5% der Ausgangsleistung, negativer Effekt unwahrscheinlich“ (s.o.) lässt sich für die Normalverteilung in „Mittelwert 5%, zwei Standardabweichungen weg von der Null“ übersetzen.

Ergebnis: Eine plausible, empirisch begründbare Priorverteilung ist $N(5, 2,5)$

Limitationen: (a) Insbesondere bei Schritt 3 ist eine gewisse Subjektivität nicht zu vermeiden. Dies ist bei allen Arten der Herleitung der Fall und lässt sich durch die Berücksichtigung einer Bandbreite von Priorverteilungen in der weiteren Analyse kompensieren (Sensitivity Analysis [32]). (b) Aus empirischen Daten oder Ergebnissen abgeleitete Priorverteilungen neigen dazu, sowohl die zentrale Tendenz als auch die Streuung zu überschätzen (86, 87).

5.3.5.3 Weitere Vorteile Bayesianischer Verfahren

Neben der Möglichkeit Vorwissen in die Analyse einzubeziehen haben eine Bayesianische Herangehensweise noch weitere Vorteile. Relevant sind insbesondere: (a) die geringere Abhängigkeit des Bayesianischen Wahrscheinlichkeitsbegriffs („subjektive Einschätzung“) von großen Fallzahlen, im Vergleich zur fre-

quentistischen „relativen Häufigkeit bei einer großen Anzahl von Wiederholungen“ und (b) die intuitivere Interpretation von Bayesianischen Intervallschätzern (Vertrauensintervall, „credible interval“), die sich insbesondere für die Entscheidungsfindung in der Praxis besser eignet als das frequentistische Konfidenzintervall (89). Box 2 bietet einen Überblick über die Vorteile Bayesianischer Verfahren bei Studien mit kleiner Fallzahl.

Box 2: Vorteile Bayesianischer Analysen bei Studien mit kleiner Fallzahl

- Geringere Abhängigkeit von der Anzahl der Beobachtungen (theoretisch und im angewandten Kontext)
- „Vorsprung“ (d. h. geringere benötigte Fallzahl) durch die Berücksichtigung von Vorwissen in der Datenanalyse (z. B. Wirksamkeit der Intervention in Vorstudien auf niedrigerem Leistungsniveau)
- Intuitiv informative Vertrauensintervalle für die Beurteilung der praktischen Relevanz von Effekten.
- Eignet sich besonders für sehr kleine Fallzahlen bis hin zur Beurteilung einzelner Athleten ($n=1$; „Individualisierung“)

5.3.5.4 Bayesianische Verfahren – Herausforderungen und Implementierung

Bayesianische Analysen werden aktuell im Spitzensport vor allem im Kontext der Individualisierung verwendet, vermutlich weil bei der minimalen Fallzahl ($n=1$) der Effekt der informativen Priorverteilung am stärksten zum Tragen kommt. Beispiele sind das Regenerationsmanagement (7, 45, 47) aber auch die statistischen Grundlagen des Biologischen Athletenpasses (93, 94). Hinzu kommen komplexe Anwendungen für die Auswertung von Zeitreihen (z. B. [51]).

Zu den wesentlichen Gründen für die zurückhaltende Nutzung zählen mutmaßlich der Sprung in ein ungewohntes statistisches System und die „Angreifbarkeit“ informativer Priorverteilungen z. B. im Begutachtungsprozess. Die noch vor wenigen Jahren hohen technischen Hürden für die simulationsbasierten Analysen sind mit heutiger Software für den Anwender kaum noch relevant. Auch gemischte Modelle können analog zu einer frequentistischen Herangehensweise spezifiziert und im Anschluss Bayesianisch geschätzt werden.

Eine umfassende Einführung in die Bayesianische Statistik (57) und deren Anwendung speziell bei kleinen Fallzahlen (101) übersteigt den Rahmen dieser Expertise. Ein leicht zugänglicher Überblick findet sich bei Wagenmakers (105) und van de Schoot (101). Ein aktuelles Online-Buch (<https://www.bayesrulesbook.com/>) und das begleitende R Paket bieten einen kostenlosen, praxisorientierten Einstieg. Ein exzellentes, anwendungsnahes Lehrbuch wurde von Kruschke veröffentlicht (57).

Das R Paket `brms` eignet sich zur Bayesianischen Analyse hierarchischer Datensätze (19, 69). Es sind nur grundlegende Kenntnisse in der Bedienung von R bzw. R Studio notwendig. Die ebenfalls kostenlose Software JASP bietet eine graphische Nutzeroberfläche und ermöglicht damit einen Einstieg in Bayesianische Analysen gänzlich ohne Programmierkenntnisse (72, 102, 104). Ausgewählte Beispiele für die Anwendung Bayesianischer Verfahren im Spitzensport sind (45, 66).

5.3.6 Aktuelle Entwicklungen und „Magnitude Based Inference“

5.3.6.1 Aktuelle Entwicklungen in Methodologie und Statistik

Die Methodologie und Statistik ist ein aktives Forschungsfeld und in den letzten Jahren wurden anspruchsvolle Ansätze für Studien mit kleiner Fallzahl entwickelt. Ein prominentes Beispiel ist das ERC geförderte Projekt IDEAL, das sich mit der Optimierung von Design und Analyse klinischer Studien bei seltenen Erkrankungen befasst (27, 50). Die Ergebnisse von IDEAL wurden umfassend (d. h. in der Regel einschließlich des statistischen Codes) publiziert (27). Allerdings erfordert die sinnhafte Anwendung dieser anspruchsvollen Verfahren zwingend die Zusammenarbeit mit einem ausgebildeten Statistiker (z. B. über universitätsinterne Zentren für klinische Studien). Auch unabhängig von technisch-mathematischen Herausforderungen erscheint die direkte Übertragbarkeit in die Spitzensportforschung teilweise fraglich z. B. wegen noch kleinerer Fallzahlen oder einer geringeren Akzeptanz aufwändiger Studiendesigns.

5.3.6.2 Entwicklungen aus der Sportwissenschaft

Von Fachwissenschaftlern beteiligter Disziplinen (z. B. Sportwissenschaft und Sportmedizin) entwickelte statistische Verfahren sollten zwingend zunächst von ausgebildeten Statistikern validiert und möglichst auch in einer entsprechenden Fachzeitschrift publiziert werden. Dies wird inzwischen auch von einigen sportwissenschaftlichen Zeitschriften gefordert. Ziel ist zu überprüfen, ob und unter welchen Umständen die postulierten Eigenschaften der Methode tatsächlich gewährleistet sind.

Die Kontroverse um das von Hopkins und Batterham für kleine Fallzahlen entwickelte Verfahren „Magnitude Based Inference“ („MBI“) (52) illustriert diesen Punkt eindringlich (77, 106), ist jedoch nicht das einzige Beispiel (30, 31, 96). MBI war zeitweilig in der Sportwissenschaft

sehr populär (17), wird aber von Statistikern kritisch beurteilt (77, 106). Dabei werden die kritischen Beurteilungen sowohl von theoretischen Überlegungen als auch von umfangreichen Simulationen gestützt (77, 106). Letztere widerlegen insbesondere die von den Entwicklern postulierten geringeren Fehlerraten bei kleinen Fallzahlen (77). Eine verständliche, nicht-technische Darstellung findet sich bei Sainani (78). Auf dieser Grundlage wird MBI inzwischen von einigen hochrangigen sportwissenschaftlichen Zeitschriften als Grundlage für die Schlussfolgerungen einer Studie explizit ausgeschlossen. Unstrittig ist dagegen die Forderung nach einer Beurteilung der Bedeutsamkeit eines Effekts anhand von Effektstärken oder Konfidenzintervallen in Relation zu einem inhaltlich begründbaren, kleinsten relevanten Unterschied (77). Eine empfehlenswerte frequentistische Alternative zu MBI ist die Nicht-Unterlegenheits- oder Äquivalenztestung („non-inferiority“ oder „equivalence testing“ [2]). Dabei wird das Fehlen

eines Unterschieds relevanter Größe durch einzelne oder eine Kombination einseitiger Hypothesentests untersucht (2, 60). Grundlage der Entscheidungsfindung ist dann folgende Aussage: „Wenn ein Effekt relevanter Größe abgelehnt werden kann, dann ist es plausibel so zu handeln, als ob ein eventueller Effekt zu klein ist um bedeutsam zu sein.“ (2, 60). Eine voll bayesianische Umsetzung der Grundidee von MBI wurde von Mengersen et al. publiziert (66).

Von Neuentwicklungen abzugrenzen ist die Übertragung anerkannter Verfahren aus anderen Bereichen und deren (erstmalige) Anwendung in der Sportwissenschaft. Hier kann, ähnlich wie z. B. bei Messinstrumenten, auf die bekannten und bereits publizierten Eigenschaften des Verfahrens verwiesen werden. Eine explizite Zuordnung zu einem der Fälle (Neuentwicklung oder Transfer) ist in Studienplan und Veröffentlichung vorzunehmen. Die Einbindung eines Statistikers ist sicherlich auch im Fall des Methodentransfers sinnvoll.

Aktuelle Entwicklungen und „MBI“

- › Anspruchsvolle Verfahren aus der aktuellen statistischen Literatur erfordern in der Umsetzung (trotz publiziertem Code) zwingend die Zusammenarbeit mit einem ausgebildeten Statistiker. Die Übertragbarkeit auf die Rahmenbedingungen des Spitzensports sollte jeweils kritisch hinterfragt werden.
- › Von Fachwissenschaftlern entwickelte statistische Verfahren sollten vor der Anwendung von ausgebildeten Statistikern überprüft und publiziert werden.
- › „Magnitude Based Inference“ ist trotz ihrer sinnvollen Grundidee keine valide inferenzstatistische Methode. Alternativen für die Umsetzung sind Nicht-Unterlegenheits- oder Äquivalenztestung (frequentistisch) (2, 60) oder eine Bayesianische Analyse (66).

Von Einzelinstrumenten zum Studienplan

6.1 Vorbemerkungen – Wahlfreiheit und Transparenzpflicht

Aus angewandter Perspektive gehören zu den Aspekten, die bei der Zusammenstellung des Studienplans berücksichtigt werden sollten, insbesondere: (a) Der Grund bzw. die Gründe für die kleine Fallzahl. Dieser Aspekt ist entscheidend für die Ausgestaltung der Datenaufnahme. (b) Das strategische Ziel des Projekts (generalisierbarer Erkenntnisgewinn oder Entscheidungsfindung innerhalb des spezifischen Settings der Studie). Dieser Aspekt ist entscheidend für die Herangehensweise bei der Datenanalyse und insbesondere für die Möglichkeit Vorwissen bei der Analyse zu berücksichtigen (vgl. Kapitel 4.2 und 5.3.5) (c) Die vorhandene methodisch-statistische Expertise. Das „Pflichtprogramm“ in den Kapiteln 3.1. und 3.2 sollte durch die studienverantwortlichen Fachwissenschaftler umsetzbar sein (vgl. Checkliste in Abbildung 4). Bei den weiterführenden Optimierungsmöglichkeiten ist die Einbindung eines ausgebildeten Statistikers dringlich empfehlenswert. Dies betrifft insbesondere deren erstmalige Implementierung.

6.2 Typische Szenarien im Leistungs- und Spitzensport

Neben den Einzelinstrumenten entscheidet deren sinnvolle Kombination innerhalb eines Studienplans über die Validität und Präzision der Ergebnisse und damit letztlich über deren Wert für die Sportpraxis. In Anbetracht der Vielzahl zu berücksichtigender Einflussfaktoren und gegenseitiger Abhängigkeiten der Instrumente ist es kaum möglich einen praktikablen „Goldstandard“ für Studien mit kleiner Fallzahl zu empfehlen. Allerdings können mutmaßlich drei typische Szenarien unterschieden werden:

6.2.1 Spitzensport Szenario

Die kleine Zahl potenzieller Studienteilnehmer wird durch die verständliche Zurückhaltung bezüglich der Teilnahme an wissenschaftlichen Studien ohne direkten Nutzen für die teilnehmenden Athleten weiter geschmälert. Gleichzeitig ist das primäre, strategische Ziel der Studie häufig die Entscheidungsfindung innerhalb des konkreten Settings der Studie (oder sogar für die teilnehmenden Athleten selbst). In letzterem Fall sollten die Ergebnisse eine ausgewogene und präzise Zusammenfassung des aktuellen Wissensstandes darstellen und nicht notwendiger Weise ausschließlich die in der Studie aufgenommenen Daten widerspiegeln (89).

Vor dem Hintergrund dieser Konstellation sollte versucht werden ein Maximum an Information aus jedem einzelnen Teilnehmer „herauszuholen“. Daher sind Studiendesigns mit Wiederholung und Kontrolle auf individueller Ebene („Single-Subject Designs“) hier besonders geeignet. Je nach Zielparаметer, Reversibilität und Zeitverlauf von Veränderungen, der umsetzbaren Studiendauer und der Akzeptanz von Auswaschphasen kommen insbesondere Cross-over und wiederholtes Cross-over Design, ein multiple Baseline Design oder auch beobachtende Designs mit Messwiederholung in Betracht.

Die Berücksichtigung von Vorwissen bietet eine Möglichkeit, den kleinen Datensatz zu ergänzen. Falls notwendig kann dieses Vorwissen im Rahmen vorbereitender Studien mit weniger exklusiven Teilnehmern gezielt generiert werden. Die passende Herangehensweise an die Datenanalyse sind in diesem Fall gemischte Modelle (4, 82), die im Rahmen Bayesianischer Verfahren mit informativer Priorverteilung geschätzt werden (89, 101). Als positiver Nebeneffekt kann bei dieser Herangehensweise auch die individuelle Response des einzelnen Athleten beurteilt werden (44, 47), so dass ein direkter Nutzen für den Teilnehmer entsteht.

6.2.2 Pilotstudien Szenario

Hier ist die Verfügbarkeit von Ressourcen (z. B. Verbrauchsmaterialien, Geräte- und Arbeitszeit) der limitierende Faktor. In diesem Fall muss

versucht werden, die Anzahl notwendiger Messungen durch eine konsequente Begrenzung und Kontrolle von Sekundär- und Störvarianz zu minimieren. Schlüsselinstrumente sind dabei eine strikte und umfassende Standardisierung, Messmethoden mit hoher Reliabilität und eng gefasste Einschlusskriterien. Falls die Voraussetzungen erfüllt sind (vgl. Kapitel 5.2.2), kann durch ein Cross-over ein Matching zwischen Interventions- und Kontrollbedingung erreicht werden. Ziel der genannten Instrumente ist eine Maximierung der beobachteten Effektstärke, und zwar in erster Linie durch eine Minimierung der Variabilität (also des Nenners). Dieses Vorgehen erhöht zwar die Teststärke, schränkt aber die Übertragbarkeit der Ergebnisse auf Praxisbedingungen (externe Validität) ein (88). Resampling und Schrumpfschätzer können genutzt werden, um eine Verzerrung der Ergebnisse durch die kleine Fallzahl zu reduzieren und deren Robustheit abzuschätzen. Nicht vergessen werden sollte, dass die generell mit kleinen Fallzahlen verknüpften Probleme – namentlich ein niedriger positiver prädiktiver Wert (21) und eine tendenzielle Überschätzung der wahren Effektstärke (86) – durch die in innovativen Pilotstudien typischer Weise niedrige Vorstudien-Wahrscheinlichkeit und die gezielte Minimierung der Streuung verstärkt werden. Diese Überlegung muss zwingend bei der Interpretation der Ergebnisse beachtet werden.

6.2.3 Big data scenario

In diesem Fall entsteht das Ungleichgewicht zwischen benötigter und realisierbarer Fallzahl in erster Linie durch die Komplexität des benötigten Modells. Wenn viele Einflussfaktoren (und ggf. deren Interaktionen) berücksichtigt werden müssen um eine akzeptable Vorhersagegenauigkeit zu erreichen, dann ist selbst eine auf den ersten Blick hohe Fallzahl schnell viel zu klein (38). Dieses Problem ist besonders gravierend, wenn seltene Ereignisse (z. B. Sportverletzungen, Olympiasiege) erklärt oder vorhergesagt

werden sollen und besonders relevant, wenn es um die Nutzung von Modellen und Vorhersagen bei der Entscheidungsfindung in der Praxis geht. In diesem Fall bräuchte man enorme Fallzahlen („Big Data“) wie sie im Sport in aller Regel unerreichbar sind. Wichtig ist zu beachten, dass der limitierende Faktor die Anzahl der Teilnehmer (oder Ereignisse) ist, die jeweils eine einzigartige Kombination von Einflussfaktoren repräsentieren. Diese Beschränkung kann nicht durch die Berücksichtigung von noch mehr erklärenden Variablen überwunden werden. Hier liegen Grenzen des Missverhältnisses von benötigter und realisierter Fallzahl, jenseits derer statistisches Lernen (z. B. „künstliche Intelligenz“) zum Scheitern verurteilt ist. Die Erforschung der genetischen Hintergründe komplexer Phänotypen liefert hier ein illustratives Beispiel (13, 103). Durch eine Reduktion der Allgemeingültigkeit (z. B. Berücksichtigung nur eines Geschlechts, Altersgruppe, Spielposition etc.) und damit auch der Modellkomplexität kann die benötigte Fallzahl reduziert werden. Außerdem sollte die angenommene kausale Struktur explizit gemacht (54) und bei der Auswahl der erklärenden Variablen berücksichtigt werden (70). Darüber hinaus kann Vorwissen genutzt werden um aus Rohdaten prädiktiver Variablen informative Merkmale zu generieren („feature engineering“). Hier können Synergien aus menschlicher und künstlicher Intelligenz genutzt werden. Auf Seite der Modellierung können anspruchsvolle Verfahren wie z. B. Bayesian additive regression trees (BART) dazu beitragen, die für komplexe Modelle benötigte Fallzahl graduell zu reduzieren (22, 63, 95). Eine Zusammenarbeit mit Statistik bzw. Bioinformatik ist dabei unbedingt notwendig. Der Schlüssel sind jedoch große, multizentrische Studien (13). Wenn diese nicht durchführbar oder nicht ausreichend sind, sollten andere Zugangswege in Betracht gezogen werden (z. B. die Aggregation subjektiver Einschätzungen im Sinne der Schwarmintelligenz [97, 108]). Abbildung 2 fasst die beschriebenen Szenarien in Form eines Entscheidungsbaums zusammen.

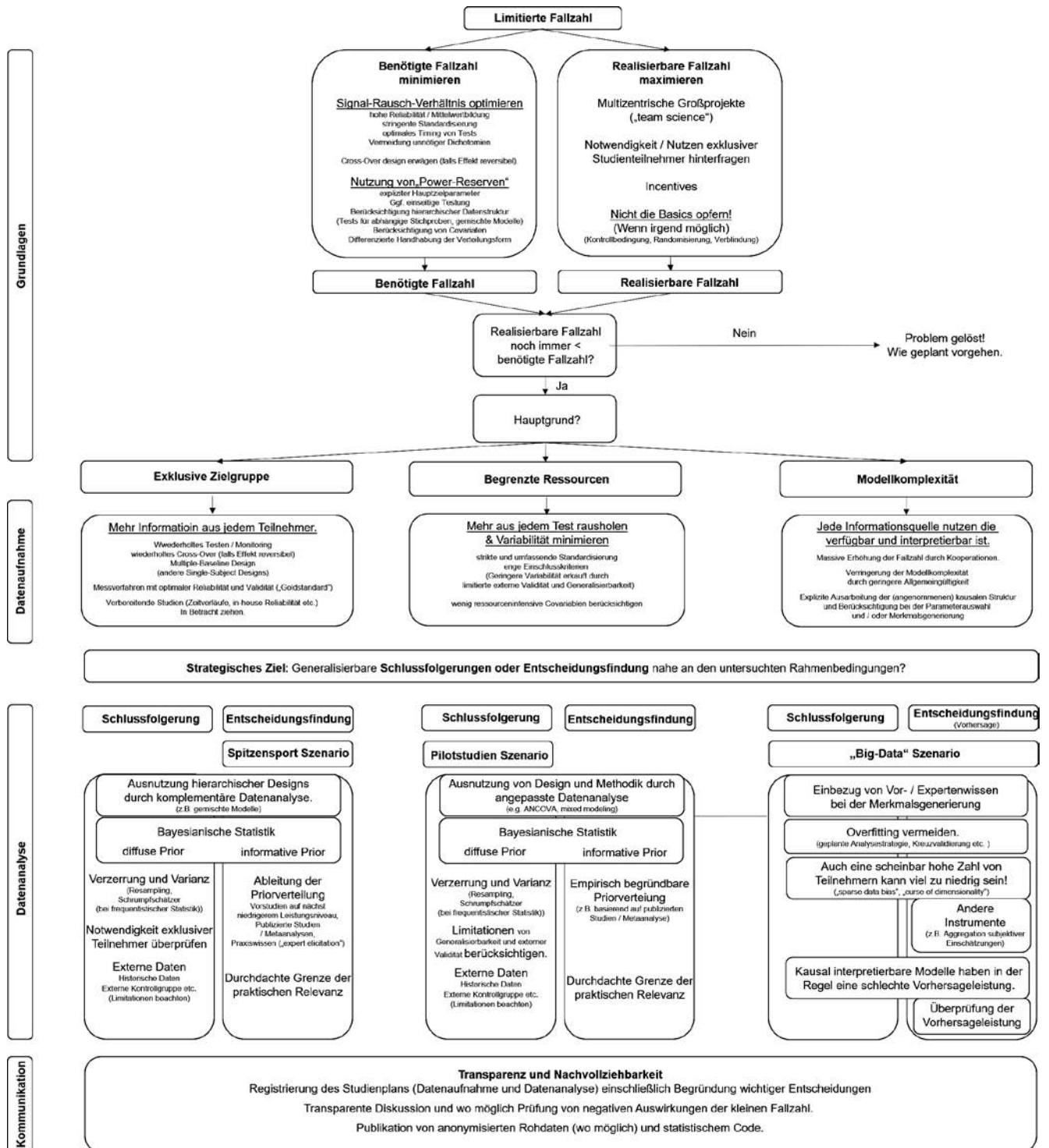


Abbildung 2

Danksagungen

Unser Dank gilt den vielen Kollegen, die durch konstruktive Kritik und hilfreiche Hinweise zur Entstehung dieser Expertise beigetragen haben. Besonders hervorzuheben sind neben den Mitgliedern des wissenschaftlichen Beirats insbesondere Prof. Dr. Tim Meyer, Dr. Sascha Schwindling, Dr. Sabrina Forster und Dr. Tanja Stenger. Gleichrangig gilt unser Dank den Wissenschaftskordinatoren der Verbände für wertvolle Beiträge und Rückmeldungen aus der Perspektive der Sportpraxis.

Wissenschaftlicher Beirat (alphabetisch)

Prof. Dr. Lars Donath
PD Dr. Oliver Faude
Prof. Dr. Olaf Hoos
Prof. Dr. Nicole Heussen
PD Dr. Werner Pitsch
Dr. Kerstin Schmidt

Literatur

1. Estimation and Sample Size Determination for Finite Populations <https://www.alnap.org/system/files/content/resource/files/main/finite.pdf>. (Accessed 2021/06/20)
2. Aisbett J, Lakens D, and Sainani K. Magnitude Based Inference in Relation to One-sided Hypotheses Testing Procedures. Preprint 2020. doi: 10.31236/osf.io/pn9s3
3. Althouse AD. Adjust for Multiple Comparisons? It's Not That Simple. *Ann Thorac Surg* 101: 1644-1645, 2016.
4. Araujo A, Julious S, and Senn S. Understanding Variation in Sets of N-of-1 Trials. *PLoS One* 11: e0167167, 2016.
5. Atkinson G, and Batterham AM. True and false interindividual differences in the physiological response to an intervention. *Exp Physiol* 100: 577-588, 2015.
6. Austin PC, Manca A, Zwarenstein M, Juurlink DN, and Stanbrook MB. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *J Clin Epidemiol* 63: 142-153, 2010.
7. Barth V, Kasbauer H, Ferrauti A, Kellmann M, Pfeiffer M, Hecksteden A, and Meyer T. Individualized Monitoring of Muscle Recovery in Elite Badminton. *Front Physiol* 10: 778, 2019.
8. Bates BT. Single-subject methodology: an alternative approach. *Med Sci Sports Exerc* 28: 631-638, 1996.
9. Bell ML, and Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res* 23: 440-459, 2014.
10. Bennett DA. How can I deal with missing data in my study? *Aust N Z J Public Health* 25: 464-469, 2001.
11. Berger VW. Testing for baseline balance: can we finally get it right? *J Clin Epidemiol* 63: 939-940; author reply 940-932, 2010.
12. Bland JM, Altman DG, and Rohlff FJ. In defence of logarithmic transformations. *Stat Med* 32: 3766-3768, 2013.
13. Bouchard C. Exercise genomics – a paradigm shift is needed. *Br J Sports Med* 49: 1492-1496, 2015.
14. Brim RL, and Miller FG. The potential benefit of the placebo effect in sham-controlled trials: implications for risk-benefit assessments and informed consent. *J Med Ethics* 39: 703-707, 2013.
15. Broatch JR, Petersen A, and Bishop DJ. Postexercise cold water immersion benefits are not greater than the placebo effect. *Med Sci Sports Exerc* 46: 2139-2147, 2014.
16. Buchheit M. Chasing the 0.2. *Int J Sports Physiol Perform* 11: 417-418, 2016.
17. Buchheit M. The Numbers Will Love You Back in Return-I Promise. *Int J Sports Physiol Perform* 11: 551-554, 2016.
18. Buja A, Cook D, Hofmann H, Lawrence M, Lee EK, Swayne DF, and Wickham H. Statistical inference for exploratory data analysis and model diagnostics. *Philos Trans A Math Phys Eng Sci* 367: 4361-4383, 2009.
19. Bürkner P. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software* 80: 2017. doi: 10.18637/jss.v080.i01
20. Burns P, Rohrich R, and Chung K. The Levels of Evidence and their role in Evidence-Based Medicine. *Plast Reconstr Surg* 128: 305-310, 2011.
21. Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, and Munafò MR. Power failure: why small sample size under-

- mines the reliability of neuroscience. *Nat Rev Neurosci* 14: 365-376, 2013.
22. Chipman HA, George E, and McCulloch RE. BART: Bayesian additive regression trees. *The Annals of Applied Statistics* 4: 266-298, 2010.
 23. Chowdhury NR, Cook D, Hofmann H, Majumder M, Lee E, and Toth AL. Using visual statistical inference to better understand random class separations in high dimension, low sample size data. *Comput Stat* 30: 293-316, 2015.
 24. Cochran WG. *Sampling Techniques*. New York: John Wiley & Sons, Inc, 1977.
 25. Cohen J. A power primer. *Psychol Bull* 112: 155-159, 1992.
 26. Cohen J. *Statistical power analysis for the behavioral sciences*. NJ: Hillsdale, 1988.
 27. IDeAL consortium webpage, statistical code for download https://www.ideal.rwth-aachen.de/?page_id=806#toggle-id-11. [November 26, 2020].
 28. Curran-Everett D. Explorations in statistics: the log transformation. *Adv Physiol Educ* 42: 343-347, 2018.
 29. Dang Q, Mazumdar S, and Houck PR. Sample Size and Power Calculations Based on Generalized Linear Mixed Models with Correlated Binary Outcomes. *Comput Methods Programs Biomed* 91: 122-127, 2008.
 30. Dankel SJ, and Loenneke JP. Authors' Reply to Tenan et al.: „A Method to Stop Analyzing Random Error and Start Analyzing Differential Responders to Exercise“. *Sports Med* 50: 435-437, 2020.
 31. Dankel SJ, and Loenneke JP. A Method to Stop Analyzing Random Error and Start Analyzing Differential Responders to Exercise. *Sports Med* 50: 231-238, 2020.
 32. Depaoli S, Winter SD, and Visser M. The Importance of Prior Sensitivity Analysis in Bayesian Statistics: Demonstrations Using an Interactive Shiny App. *Front Psychol* 11: 608045, 2020.
 33. Kerlinger F. *Foundations of behavioural research*. New York: Rinehart and Winston, 1966.
 34. Feng C, Wang H, Lu N, and Tu XM. Log transformation: application and interpretation in biomedical research. *Stat Med* 32: 230-239, 2013.
 35. Galecki A, and Burzykowski T. *Linear Mixed-Effects Models Using R*. New York: Springer, 2013.
 36. Goltz FR, Thackray AE, King JA, Dorling JL, Atkinson G, and Stensel DJ. Interindividual Responses of Appetite to Acute Exercise: A Replicated Crossover Study. *Med Sci Sports Exerc* 50: 758-768, 2018.
 37. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol* 35: 765-775, 2006.
 38. Greenland S., Mansournia M.A., and Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ* 2016. doi: 10.1136/bmj.i1981
 39. Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, and Inger R. A brief introduction to mixed effects modelling and multi-model inference in ecology. *PeerJ* 6: e4794, 2018.
 40. Hecksteden A, Faude O, Meyer T, and Donath L. How to Construct, Conduct and Analyze an Exercise Training Study? *Front Physiol* 9:1007, 2018.
 41. Hecksteden A, Grutters T, and Meyer T. Association between postexercise hypotension and long-term training-induced blood

- pressure reduction: a pilot study. *Clin J Sport Med* 23: 58-63, 2013.
42. Hecksteden A, Grutters T, and Meyer T. Associations between Acute and Chronic Effects of Exercise on Indicators of Metabolic Health: A Pilot Training Trial. *PLoS One* 8: e81181, 2013.
 43. Hecksteden A, Heinze T, Faude O, Kindermann W, and Meyer T. Validity of Lactate Thresholds in Inline Speed Skating. *J Strength Cond Res* 29: 2497-2502, 2015.
 44. Hecksteden A, Kellner R, and Donath L. Dealing with small samples in football research. *Science and Medicine in Football*, 2021, in press.
 45. Hecksteden A, Kraushaar J, Scharhag-Rosenberger F, Theisen D, Senn S, and Meyer T. Individual response to exercise training – a statistical perspective. *J Appl Physiol* 118: 1450-1459, 2015.
 46. Hecksteden A, Pitsch W, Julian R, Pfeiffer M, Kellmann M, Ferrauti A, and Meyer T. A New Method to Individualize Monitoring of Muscle Recovery in Athletes. *Int J Sports Physiol Perform* 12: 1137-1142, 2017.
 47. Hecksteden A, Pitsch W, Rosenberger F, and Meyer T. Repeated testing for the assessment of individual response to exercise training. *J Appl Physiol* 124: 1567-1579, 2018.
 48. Hecksteden A, Skorski S, Egger F, Buder F, Kellner R, and Meyer T. Bayesianische Methoden in Spitzensportforschung und -praxis. In: *Sports, Medicine and Health Summit. Hamburg (online): 2021.*
 49. Hecksteden A, Skorski S, Schwindling S, Hammes D, Pfeiffer M, Kellmann M, Ferrauti A, and Meyer T. Blood-Borne Markers of Fatigue in Competitive Athletes – Results from Simulated Training Camps. *PLoS One* 11: e0148810, 2016.
 50. Hecksteden A, and Theisen D. Editorial based on the article: „Skeletal Muscle Signature of a Champion Sprint Runner“ by Scott Trappe and colleagues. *J Appl Physiol* 118: 1447-1448, 2015.
 51. Hilgers R, Bogdan M, Burman C, Dette H, Karlsson M, König F, Male C, Mentré F, Moltenberghs G, and Senn S. Lessons learned from IDeAl – 33 recommendations from the IDeAl-net about design and analysis of small population clinical trials. In: *Orphanet Journal of Rare Diseases* 11; 13(1): 77, 2018.
 52. Hopker J, Griffin J, Brookhouse J, Peters J, Schumacher YO, and Iljukov S. Performance profiling as an intelligence-led approach to antidoping in sports. *Drug Test Anal* 12: 402-409, 2020.
 53. Hopkins WG. How to Interpret Changes in an Athletic Performance Test. sportsci.org/jour/04/wghtestshtm: 2004. (accessed: 14/06/2013)
 54. Ioannidis JP. Why replication has more scientific value than original discovery. *Behav Brain Sci* 41: doi:10.1017/S0140525X18000729, 2018.
 55. Kalkhoven JT, Watsford ML, and Impellizzeri FM. A conceptual model and detailed framework for stress-related, strain-related, and overuse athletic injury. *J Sci Med Sport* 23: 726-734, 2020.
 56. Kinugasa T, Cerin E, and Hooper S. Single-subject research designs and data analyses for assessing elite athletes' conditioning. *Sports Med* 34: 1035-1050, 2004.
 57. Kinugasa T, Miyanaga Y, Shimojo H, and Nishijima T. Statistical evaluation of conditioning for an elite collegiate tennis player using a single-case design. *J Strength Cond Res* 16: 466-471, 2002.
 58. Kruschke J. *Doing Bayesian Data Analysis – A Tutorial with R, JAGS, and Stan.* Academic Press, 2015.

59. Lakens D. Sample Size Justification. Preprint <https://psyarxiv.com/9d3yf/>: 2021.
60. Lakens D, and Caldwell AR. Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science* 4: 1-14, 2021.
61. Lakens D, Scheel C, and Isager P. Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science* 1: 259-269, 2018.
62. Lenth RV. Some practical guidelines for effective sample size determination. *The American Statistician* 55: 187-193, 2001.
63. Lo S, and Andrews S. to transform or not to transform: using generalized mixed models to analyse reaction time data. *Front Psychol* 6: 1171, 2015.
64. Logan BR, Sparapani R, McCulloch RE, and Laud PW. Decision making and uncertainty quantification for individualized treatments using Bayesian Additive Regression Trees. *Stat Methods Med Res* 28: 1079-1093, 2019.
65. Munafò, MR, and Smith GD. Repeating experiments is not enough – Verifying results requires disparate lines of evidence. *Nature* 553: 399-401, 2018.
66. McNeish D. Missing data methods for arbitrary missingness with small samples. *Journal of Applied Statistics* 44: 24-39, 2017.
67. Mengersen KL, Drovandi CC, Robert CP, Pyne DB, and Gore CJ. Bayesian Estimation of Small Effects in Exercise and Sports Science. *PLoS One* 11: e0147311, 2016.
68. Morris SB. Estimating effect sizes from pre-test-posttest-control group designs. *Organ Res Methods* 11: 364-386, 2008.
69. Krumwiede, N. Taking the Mystery Out of Research: Approaches to Evaluating Experimental Research. *Orthopaedic Nursing* 21: 65, 2002.
70. Nalborczyk L, Batailler C, Loevenbruck H, Vilain A, and Bürkner P. An Introduction to Bayesian Multilevel Models Using brms: A Case Study of Gender Effects on Vowel Variability in Standard Indonesian J Speech Lang Hear Res 62: 1225-1242, 2019.
71. Piccininni M, Konigorski S, Rohmann JL, and Kurth T. Directed acyclic graphs and causal thinking in clinical risk prediction modeling. *BMC Med Res Methodol* 20: 179, 2020.
72. Poppendieck W, Faude O, Wegmann M, and Meyer T. Cooling and performance recovery of trained athletes: a meta-analytical review. *Int J Sports Physiol Perform* 8: 227-242, 2013.
73. Quintana D, and Williams D. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: A non-technical guide using JASP. *BMC Psychiatry* 18: 178, 2018.
74. Reboussin DM, and Morgan TM. Statistical considerations in the use and analysis of single-subject designs. *Med Sci Sports Exerc* 28: 639-644, 1996.
75. Rodgers JL. The Bootstrap, the Jackknife, and the Randomization Test: A Sampling Taxonomy. *Multivariate Behav Res* 34: 441-456, 1999.
76. Romo-Romo A, Aguilar-Salinas CA, Brito-Cordova GX, Gomez Diaz RA, Vilchis Valentin D, and Almeda-Valdes P. Effects of the Non-Nutritive Sweeteners on Glucose Metabolism and Appetite Regulating Hormones: Systematic Review of Observational Prospective Studies and Clinical Trials. *PLoS One* 11: e0161264, 2016.
77. Rothpearl A. The jackknife technique in statistical analysis. *Chest* 95: 940, 1989.

78. Sainani KL. The Problem with „Magnitude-Based Inference“. *Med Sci Sports Exerc* 50: 2166-2176, 2018.
79. Sainani KL, Lohse KR, Jones PR, and Vickers A. Magnitude-based Inference is not Bayesian and is not a valid method of inference. *Scandinavian Journal of Science and Medicine in Sports* 29: 1428-1436, 2019.
80. Scheel AM, Tiokhin L, Isager PM, and Lakens D. Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspect Psychol Sci* 16: 755-755, 2021.
81. Senn S. Applying results of randomised trials to patients. N of 1 trials are needed. *BMJ* 317: 537-538, 1998.
82. Senn S. Individual response to treatment: is it a valid assumption? *BMJ* 329: 966-968, 2004.
83. Senn S. Mastering variation: variance components and personalised medicine. *Stat Med* 35: 966-977, 2015.
84. Senn S. Sample size considerations for n-of-1 trials. *Stat Methods Med Res* 28: 372-383, 2017.
85. Senn S. Seven myths of randomisation in clinical trials. *Stat Med* 32: 1439-1450, 2013.
86. Senn S. Testing for baseline balance in clinical trials. *Stat Med* 13: 1715-1726, 1994.
87. Senn S. Transposed conditionals, shrinkage, and direct and indirect unbiasedness. *Epidemiology* 19: 652-654; discussion 657-658, 2008.
88. Senn S. Trying to be precise about vagueness. *Statistics in Medicine* 26: 1417-1430, 2007.
89. Senn S. U is for Unease: Reasons for Mistrusting Overlap Measures for Reporting Clinical Trials. *Statistics in Biopharmaceutical Research* 3: 302-309, 2011.
90. Senn S. You May Believe You Are a Bayesian But You Are Probably Wrong. *RMM – Special Topic: Statistical Science and Philosophy of Science* 2: 48-66, 2011.
91. Senn S, Rolfe K, and Julious SA. Investigating variability in patient response to treatment—a case study from a replicate cross-over study. *Stat Methods Med Res* 20: 657-666, 2011.
92. Senn SJ. Covariate imbalance and random allocation in clinical trials. *Stat Med* 8: 467-475, 1989.
93. Senn SJ. *Cross-over Trials in Clinical Research*. John Wiley & Sons, Chichester, United Kingdom 2. edition: 2002.
94. Sottas PE, Baume N, Saudan C, Schweizer C, Kamber M, and Saugy M. Bayesian detection of abnormal values in longitudinal biomarkers with an application to T/E ratio. *Biostatistics* 8: 285-296, 2007.
95. Sottas PE, Robinson N, Rabin O, and Saugy M. The athlete biological passport. *Clin Chem* 57: 969-976, 2011.
96. Tan YV, and Roy J. Bayesian additive regression trees and the General BART model. *Stat Med* 38: 5048-5069, 2019.
97. Tenan MS, Vigotsky AD, and Caldwell AR. Comment on: „A Method to Stop Analyzing Random Error and Start Analyzing Differential Responders to Exercise“. *Sports Med* 50: 431-434, 2020.
98. Tetlock PE, Mellers BA, and Scoblic JP. Bringing probability judgments into policy debates via forecasting tournaments. *Science* 355: 481-483, 2017.
99. Thabane L, Ma J, Chu R, Cheng J, Ismaila A, Rios LP, Robson R, Thabane M, Giangregorio L, and Goldsmith CH. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol* 10: 1, 2010.

100. Trappe S, Luden N, Minchev K, Raue U, Jemiolo B, and Trappe T. Skeletal Muscle Signature of a Champion Sprint Runner. *J Appl Physiol* 118: 1460-1466, 2015.
101. Tukey LW. Bias and confidence in not quite large samples *Abstr Annu Math Stat* 29: 614, 1958.
102. Van de Schoot R editor. *Small sample size solutions: A guide for applied researchers and practitioners*. Routledge, 2020.
103. van Doorn J, van den Bergh D, Bohm U, Dablander F, Derks K, Draws T, Etz A, Evans NJ, Gronau QF, Haaf JM, Hinne M, Kucharsky S, Ly A, Marsman M, Matzke D, Gupta A, Sarafoglou A, Stefan A, Voelkel JG, and Wagenmakers EJ. The JASP guidelines for conducting and reporting a Bayesian analysis. *Psychon Bull Rev* 28: 813-826, 2021.
104. Vlahovich N, Hughes DC, Griffiths LR, Wang G, Pitsiladis YP, Pigozzi F, Bachl N, and Eynon N. Genetic testing for exercise prescription and injury prevention: AIS-Athlome consortium-FIMS joint statement. *BMC Genomics* 18: 818, 2017.
105. Wagenmakers EJ, Love J, Marsman M, Jamil T, Ly A, Verhagen J, Selker R, Gronau QF, Dropmann D, Boutin B, Meerhoff F, Knight P, Raj A, van Kesteren EJ, van Doorn J, Smira M, Epskamp S, Etz A, Matzke D, de Jong T, van den Bergh D, Sarafoglou A, Steingroever H, Derks K, Rouder JN, and Morey RD. Bayesian inference for psychology. Part II: Example applications with JASP. *Psychon Bull Rev* 25: 58-76, 2018.
106. Wagenmakers EJ, Morey R, and Lee MD. Bayesian Benefits for the Pragmatic Researcher. *Current Directions in Psychological Science* 25: 2016. doi: 10.1177/0963721416643289
107. Welsh AH, and Knight EJ. „Magnitude-based inference“: a statistical review. *Med Sci Sports Exerc* 47: 874-884, 2015.
108. Wievelhove T, Thase C, Glahn M, Hessel A, Schneider C, Hottenrott L, Meyer T, Kellmann M, Pfeiffer M, and Ferrauti A. Repeatability of the Individual Response to the Use of Active Recovery the Day After High-Intensity Interval Training: A Double-Crossover Trial *Int J Sports Phys Perf* epub ahead of print, 2021.
109. Woolley AW, Chabris CF, Pentland A, Hashmi N, and Malone TW. Evidence for a collective intelligence factor in the performance of human groups. *Science* 330: 686-688, 2010.
110. Zwaan RA, Etz A, Lucas RE, and Donellan MB. Making replication mainstream. *Behav Brain Sci* 41: 2018. doi:10.1017/S0140525X17001972

Anhang

Im Folgenden wurde eine subjektive und zeitgebundene Auswahl von Anmerkungen und (Software-) Empfehlungen zusammengestellt. Diese konzentriert sich auf Bereiche, die für den wissenschaftlichen Umgang mit kleinen Fallzahlen besonders relevant und / oder in der Leistungssportforschung noch nicht generell verbreitet sind.

Wichtig ist zu beachten, dass intuitive Visualisierungen und Berechnungshilfen methodisch-statistische Expertise in der Konzeption und Durchführung von Datenaufnahme und Datenanalyse nicht ersetzen können.

9.1 Statistische Software

9.1.1 Statistischer Code: R und R Studio

Die kostenlose Software R ermöglicht die Umsetzung aller in diesem Buch beschriebenen datenanalytischen Verfahren. (Download unter: <https://cran.r-project.org/bin/windows/base/>) Neben den grundlegenden, in R direkt verfügbaren, Funktionalitäten („Base R“) ermöglichen vielfältige Pakete die flexible Umsetzung auch anspruchsvoller und innovativer Verfahren. Dies gilt auch für die flexible Visualisierung von Rohdaten und Ergebnissen (s.u.). Weitere Vorteile im Vergleich zu gängigen Statistikprogrammen mit graphischer Nutzeroberfläche sind die lückenlose Nachvollziehbarkeit der Datenanalyse anhand des statistischen Codes bzw. die Möglichkeit veröffentlichten Code anderer Autoren zu nutzen, weiterzuentwickeln und / oder deren Analysen zu verifizieren. **Diese Kombination aus Flexibilität und Transparenz ist gerade bei kleinen Fallzahlen von zentraler Bedeutung** (s.o.).

Die ebenfalls kostenlose Nutzeroberfläche R Studio (Download unter: <https://www.rstudio.com/products/rstudio/>), die in der Regel präzise Dokumentation der Pakete aber auch frei verfügbare Codebeispiele aus Nutzerforen erleichtern die Nutzung von R. Unterstützung für den Einsteiger bieten auch die zahlreichen „Cheat Sheets“ für verbreitete Pakete und Auf-

gaben (Download z. B. unter: <https://www.rstudio.com/resources/cheatsheets/>). Dennoch ist beim Umstieg von Statistikprogrammen mit graphischer Nutzeroberfläche eine Lern- und Einarbeitungsphase unvermeidlich. Einen schnellen Einstieg bei bereits vorhandenen Statistikkenntnissen bietet das „R Cookbook“ von Paul Teetor in seiner zweiten Auflage. Ebenfalls empfehlenswert und in deutscher Sprache ist „R für Dummies“ von Andrie de Vries (dritte Auflage). Einen umfassendere Einführung bietet „Discovering Statistics using R“ von Andy Field et al. Interaktive Online-Kurse (z. B. [datacamp.com](https://www.datacamp.com)) können den Einstieg in die konkrete Umsetzung erleichtern.

9.1.2 Python

Python kann bei vorhandenen Kenntnissen als Alternative zu R genutzt werden. Allerdings ist Python bisher in der Sportwissenschaft kaum verbreitet, was die Kommunikation und Zusammenarbeit erschweren kann.

9.1.3 JASP

JASP ist eine kostenlose statistische Software mit graphischer Nutzeroberfläche, die es erlaubt eine beschränkte Bandbreite von Analysen sowohl frequentistisch als auch Bayesianisch umzusetzen. JASP eignet sich daher insbesondere für den Einstieg in Bayesianische Verfahren bei (bisher) fehlenden R Kenntnissen. Download unter: <https://jasp-stats.org/download/>.

9.2 Spezielle Anwendungen

9.2.1 Berechnungshilfen für Studien mit Single Subject Designs

Für die Konzeption und Analyse von Studien mit Single-Subject Designs sind im Netz diverse Kalkulatoren (z. B. unter <http://www.singlecaseresearch.org/>), statistischer Code (z. B. aus dem IDEAL Projekt <https://www.idealproject.org/>).

ideal.rwth-aachen.de/?page_id=806#toggle-id-11) und andere Berechnungshilfen verfügbar (Überblick z. B. in Chen, L. T., Peng, C. Y. J., & Chen, M. E. (2015). Computing tools for implementing standards for single-case designs. *Behavior Modification*, 39(6), 835-869). Während publizierter statistischer Code transparent modifiziert und angewendet werden kann, sollten Online-Kalkulatoren, Spreadsheets etc. nur in begründeten Ausnahmefällen verwendet werden. Dies gilt insbesondere vor dem Hintergrund des umfassenden Spektrums verfügbarer R Pakete.

9.2.2 Visualisierung von Daten und Ergebnissen

Die R Pakete ggplot und plotly bieten umfassende Möglichkeiten zur Visualisierung von (Roh-) Daten und Ergebnissen. Mit plotly sind auch interaktive Grafiken möglich. „Cheat sheets“ erleichtern zu Beginn die Anwendung (ggplot: <https://github.com/rstudio/cheatsheets/blob/master/data-visualization-2.1.pdf>, plotly: https://images.plot.ly/plotly-documentation/images/r_cheat_sheet.pdf)

9.2.3 Fallzahlberechnung

Eine gängige, kostenlose Software zur Fallzahlberechnung ist GPower. GPower hat eine grafische Nutzeroberfläche, deckt aber nur ein

begrenztes Spektrum von Szenarien. Für die anspruchsvollen, hierarchischen Studiendesigns, wie sie in Kapitel 5.2.2 besprochen werden ist eine Fallzahlschätzung nur simulativ möglich. Hier eignet sich das R Paket Superpower. Die zugehörige Shiny-App erlaubt einen intuitiven Einstieg.

9.3 Bayesianische Statistik

Einführende Literatur- und einige Softwareempfehlungen finden sich in Kapitel 5.3.5.4.

Anmerkung: Simulationsbasierte Analysen stellen im Vergleich zu anderen statistischen Verfahren erheblich höhere Anforderungen an die Computerhardware. Insbesondere Laptops können hier bereits bei kleinen Datensätzen an Grenzen stoßen.

9.4 Gemischte Modelle

Einführende Literatur zu gemischten Modellen und weiterführende Literatur zu deren Anwendung in bestimmten Szenarien findet sich am Ende von Kapitel 5.3.3..

Anmerkung: Gemischte Modelle können sowohl Bayesianisch als auch frequentistisch geschätzt werden. Ihre Anwendung ist somit nicht an eines der beiden statistischen Systeme gebunden.

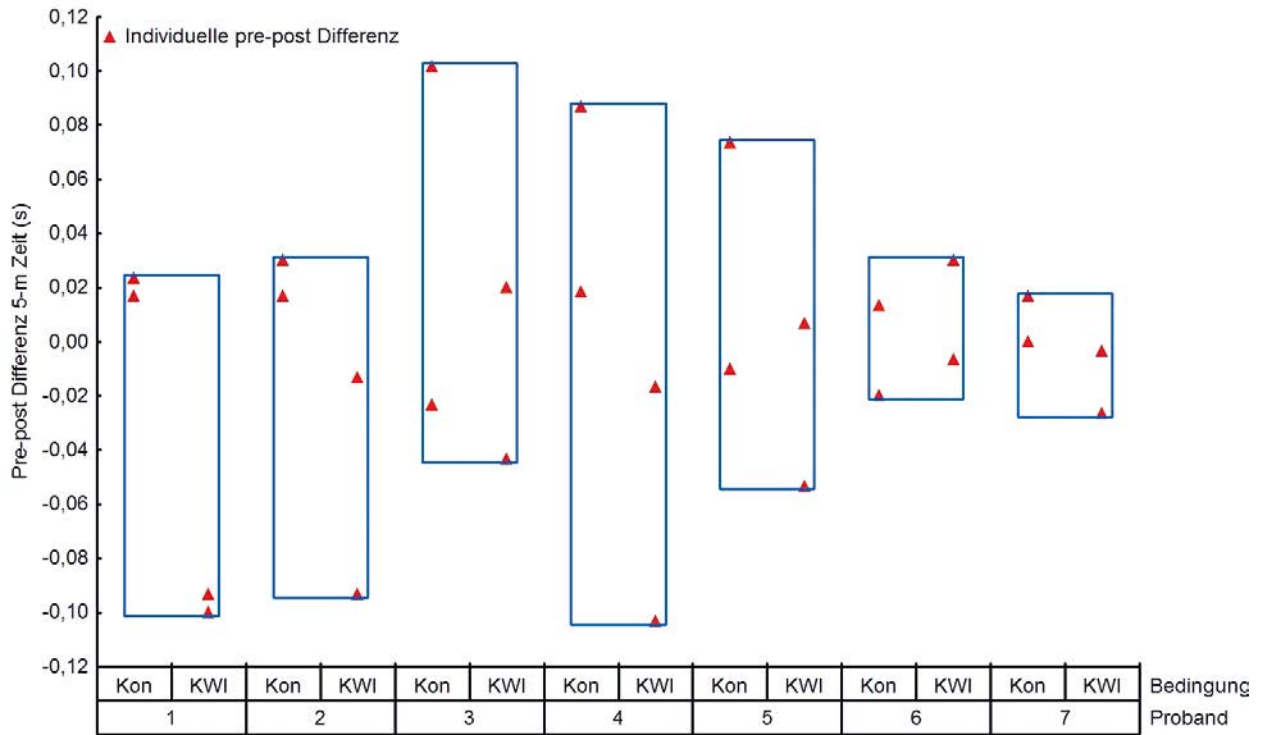


Abbildung 3

Checkliste „Pflichtprogramm“

Planung und Vorbereitung

- Methodische Standards implementiert wo immer möglich (Kontrolle, Randomisierung, Verblindung, Standardisierung, Berücksichtigung von Kovariaten)
- Kleinster relevanter Unterschied festgelegt (spezifisch für die Rahmenbedingungen der geplanten Studie)
- Basale Optionen zur Minimierung der benötigten Fallzahl erwogen und verbindlich festgelegt
 - Datenaufnahme: Reliabilität, Mehrfachmessungen, Standardisierung, Cross-Over etc.
 - Datenanalyse: Tests für verbundene Stichproben, einseitige Testung, Kovariaten, keine vermeidbaren Dichotomien etc.
- Benötigte Fallzahl berechnet (statistische Beratung empfohlen)
- Basale Maßnahmen zur Erhöhung der realisierbaren Fallzahl erwogen (etwas niedrigeres Leistungsniveau, Incentives etc.)
- Ist anzunehmen, dass die Studie vermutlich trotz dem Missverhältnis von berechneter und realisierbarer Fallzahl und den damit verbundenen Limitationen zur Beantwortung der Forschungsfrage beiträgt?
- Verbindlicher Studienplan ausgearbeitet und hinterlegt (einschließlich Hauptzielparameter und detaillierter Festlegung der Datenanalyse)

Datenaufnahme

- Studienplan abarbeiten - Flexibilität wo unvermeidlich (Abwägung in enger und offener Zusammenarbeit von Wissenschaft und Praxis)
- Zwischenanalysen mit Rückwirkung auf die Datenaufnahme vermeiden

Datenanalyse

- Analyseplan für Hauptfragestellung abarbeiten (ggf. Transformation von Rohdaten, Variationen vermeiden (Alphafehlerkumulation))

Interpretation und Kommunikation

- Berücksichtigung der bekannten Nebenwirkungen und Limitationen kleiner Fallzahlen bei der Interpretation der Ergebnisse
- Bezug zum kleinsten relevanten Unterschied
- Explorative Sekundäranalysen als solche kennzeichnen
- Transparenz wo immer möglich (Einzelfall- und/oder Rohdatendarstellung, Veröffentlichung von anonymisierten Rohdaten und statistischem Code)

Abbildung 4

Schriftenreihe des Bundesinstituts für Sportwissenschaft im **SPORTVERLAG Strauß**

- Band 2009,01**
Stahl, Silvester
Selbstorganisation von Migranten im deutschen Vereinssport. Ein Forschungsbericht zu Formen, Ursachen und Wirkungen. 148 Seiten, 21 cm, kt.
978-3-86884-510-5 € 16,80
- Band 2009,02**
Kämpfe, Astrid
Homo Sportivus Oeconomicus. Intrinsische und extrinsische Motivation im Verlauf von Spitzensportkarrieren. 403 Seiten, 21 cm, kt.
978-3-86884-511-2 € 28,80
- Band 2010,01**
Schmiege, Peter / Voriskova, Sarka / Marquardt, Gesine / Glasow, Nadine
Bauliche Voraussetzungen für den paralympischen Sport. Forschungsprojekt im Auftrag des Bundesinstituts für Sportwissenschaft Bonn. 304 Seiten, 21 cm, kt.
978-3-86884-512-9 € 24,80
- Band 2010,02**
Faude, Oliver / Wegmann, Melissa / Krieg, Anne / Meyer, Tim
Kälteapplikationen im Spitzensport. Eine Bestandsaufnahme der wissenschaftlichen Evidenz. 144 Seiten, 21 cm, kt.
978-3-86884-513-6 € 15,80
- Band 2010,03**
Fahrig, Stephan
Zur Interaktionsproblematik im Riemenzweier der Sportart Rudern. 174 Seiten, 21 cm, kt.
978-3-86884-514-3 € 17,80
- Band 2011,01**
Schlesiger, Günter
Sportplätze. Sportfreianlagen: Planung – Bau – Ausstattung – Pflege. 324 Seiten, 29,7 cm, kt.
978-3-86884-515-0 **vergriffen**
- Band 2011,02**
Wirth, Klaus
Exzentrisches Krafttraining. Auswirkungen auf unterschiedliche Maximal- und Schnellkraftparameter. 714 Seiten, 21 cm, kt.
978-3-86884-516-7 € 42,80
- Band 2011,03**
Braun, Sebastian
Ehrenamtliches und freiwilliges Engagement im Sport. Sportbezogene Sonderauswertung der Freiwilligensurveys von 1999, 2004 und 2009. 80 Seiten, 21 cm, kt.
978-3-86884-517-4 € 12,80
- Band 2011,04**
Ott, Peter
Planung und Bau von Tanzsportanlagen. Eine Orientierungshilfe des Bundesinstituts für Sportwissenschaft. 52 Seiten, 29,7 cm, kt.
978-3-86884-518-1 € 21,80
- Band 2011,05**
Wyns, Bernhard
Betreibermodelle öffentlicher Bäder. Wirkung von Betreibermodellen auf die Betriebsführung, Personalwirtschaft und Sportnachfrage. 490 Seiten, 21 cm, kt.
978-3-86884-519-8 € 58,00
- Band 2011,06**
Krause, Daniel
Zur Darstellungsperspektive von Bildschirminstruktion und -feedback beim visuomotorischen Imitationslernen. Laborexperimentelle Untersuchungen zu den Effekten verschiedener Faktoren im Kontext der Darstellungsperspektivenauswahl: Bild-Bewegungsebenen-Disparität, Modell-Betrachter-Disparität, Perspektivenanzahl und Selbstkontrolle. 386 Seiten, 21 cm, kt.
978-3-86884-520-4 € 29,80
- Band 2012,01**
Wirth, Klaus / Schlumberger, Andreas / Zawieja, Martin / Hartmann, Hagen
Krafttraining im Leistungssport. Theoretische und praktische Grundlagen für Trainer und Athleten. 148 Seiten, 30 cm, kt. 2., korr. Aufl. 2013.
978-3-86884-521-1 € 29,80
- Band 2012,02**
Kalbe, Ute / Krüger, Oliver / Wachtendorf, Volker / Berger, Wolfgang
Umweltverträglichkeit von Kunststoff- und Kunststoffrasenbelägen auf Sportfreianlagen. Erfassung von potentiellen Schadstoffen bei Sportböden auf Kunststoffbasis (Kunststoff- und Kunststoffrasenbeläge) auf Sportfreianlagen, unter Berücksichtigung von Alterungs- und Verschleißprozessen. 302 Seiten, 21 cm, kt.
978-3-86884-522-8 € 34,80
- Band 2013,01**
Daumann, Frank / Römmelt, Benedikt
Qualitätsmanagement im Bundessportfachverband. Qualitätsrelevante Stakeholder in Bundessportfachverbänden – Eine qualitative Studie als Basis für die Implementierung eines Qualitätsmanagementsystems.
536 Seiten, 21 cm, kt.
978-3-86884-523-5 € 36,00
- Band 2013,02**
Stadtmann, Tobias
Optimierung von Talentselektion und Nachwuchsförderung im Deutschen Basketball Bund aus trainingswissenschaftlicher Sicht. 428 Seiten, 21 cm, kt.
978-3-86884-524-2 € 39,80
- Band 2013,03**
Braun, Sebastian
Freiwilliges Engagement von Jugendlichen im Sport. Eine empirische Untersuchung auf Basis der Freiwilligensurveys von 1999 bis 2009. 108 Seiten, 21 cm, kt.
978-3-86884-525-9 € 12,80

Band 2014,01

Zastrow, Hilke / Schlapkohl, Nele / Raab, Markus
DeMaTra – Ein Messplatztraining für taktische Kompetenzen im Handball. 164 Seiten, 21 cm, kt.
 978-3-86884-526-6 € 24,80

Band 2014,02

Radtke, Sabine / Doll-Tepper, Gudrun
Nachwuchsgewinnung und -förderung im paralympischen Sport. Ein internationaler Systemvergleich unter Berücksichtigung der Athleten-, Trainer- und Funktionärsperspektive. 392 Seiten, 21 cm, kt.
 978-3-86884-527-3 € 38,80

Band 2014,03

Kleinert, Jens
Toolbox Beziehungsarbeit: Zur Beziehungsqualität in der sportpsychologischen Betreuung. Beitrag Qualitätssicherung in der Sportpsychologie. 68 Seiten, 30 cm, kt.
 978-3-86884-528-0 € 21,80

Band 2014,04

Brand, Ralf u.a.
Leitfaden zur Qualitätssicherung für die sportpsychologische Betreuung im Leistungssport. Beitrag Qualitätssicherung in der Sportpsychologie. 96 Seiten, 30 cm, kt.
 978-3-86884-529-7 € 28,80

Band 2015,01

Eßig, Natalie / Lindner, Sara / Magdolen, Simone / Siegmund, Loni
Leitfaden Nachhaltiger Sportstättenbau – Kriterien für den Neubau nachhaltiger Sporthallen. 424 Seiten, 21 cm, kt.
 978-3-86884-530-3 € 49,80

Band 2015,02

Gerland, Bernd Paul
Der Yips – eine erlernte Störung motorischer Leistungsvollzüge? Phänomenanalyse und Interventionsmöglichkeiten am Beispiel des Putt-Yips im Golf. 210 Seiten, 30 cm, kt.
 978-3-86884-531-0 € 42,80

Band 2016,01

Braun, Sebastian
Ehrenamtliches und freiwilliges Engagement von Älteren im Sport. Sportbezogene Sonderauswertung der Freiwilligensurveys von 1999, 2004 und 2009. 72 Seiten, 30 cm, kt.
 978-3-86884-532-7 **vergriffen**

Band 2016,02

Linz, Lothar / Ohlert, Jeannine
Sportpsychologische Vebandskonzeption des Deutschen Handballbundes 2016–2020. 64 Seiten, 30 cm, kt.
 978-3-86884-533-4 **vergriffen**

Band 2016,03

Klietsch, Heiko
Mediengerechte Sportanlagen. Mediale Raum-/Flächen- und Ausstattungsprogramme in Stadien und Arenen aus dem Bereich des Fernsehens, der Mess- und Kommunikationstechnik. 472 Seiten, 21 cm, kt.
 978-3-86884-534-1 € 56,80

Band 2016,04

Killing, Wolfgang / Böttcher, Jörg / Keil, Jan-Gerrit
Sportwissenschaftliche Aspekte des Hochsprungs. 2. korrigierte Aufl. 2017. 300 Seiten, 30 cm, kt.
 978-3-86884-535-8 € 49,00

Band 2016,05

Kalbe, Ute / Susset, Bernd / Bandow, Nicole
Umweltverträglichkeit von Kunststoffbelägen auf Sportfreianlagen. Modellierung der Stofffreisetzung aus Sportböden auf Kunststoffbasis zur Bewertung der Boden- und Grundwasserträglichkeit. 428 Seiten, 21 cm, kt.
 978-3-86884-536-5 € 49,80

Band 2016,06

Hohmann, Andreas / Singh, Ajit / Voigt, Lenard
Konzepte erfolgreicher Nachwuchstrainings (KerN). Abschlussbericht zum Forschungsprojekt „Langfristiger Leistungsaufbau im Nachwuchsleistungssport“. 336 Seiten, 30 cm, kt.
 978-3-86884-537-2 € 49,80

Band 2016,07

Kellmann, Michael / Kölling, Sarah / Hitzschke, Brit
Das Akutmaß und die Kurzskaala zur Erfassung von Erholung und Beanspruchung im Sport. Manual. 124 Seiten, 30 cm, kt.
 978-3-86884-538-9 € 33,00

Band 2017,01

Hohmann, Andreas / Pietzonka, Micha
Techniktraining zur Entwicklung der Spielfähigkeit im Fußball, Handball und Basketball. 568 Seiten, 30 cm, kt.
 978-3-86884-539-6 € 59,00

Band 2017,02

Breuer, Christoph / Wicker, Pamela / Orlowski, Johannes
Bundes- und mischfinanzierte Trainer im Sport – Standortbedingungen und Migrationsanalyse. 176 Seiten, 21 cm, kt.
 978-3-86884-540-2 € 19,80

Band 2017,03

Leistner, Philip / Koehler, Mark
Verbesserung der akustischen Eigenschaften von Sporthallenböden. 116 Seiten, 21 cm, kt.
 978-3-86884-541-9 € 19,80

Band 2017,04

Weber, Christoph
Theoretische und praktische Leistungsdiagnostik im Goalball. 204 Seiten, 30 cm, kt.
 978-3-86884-542-6 € 39,80

Band 2017,05

Loffing, Florian / Neugebauer, Judith / Hagemann, Jörg / Schorer, Jörg
Eye-Tracking im Spitzensport – Validität, Grenzen und Möglichkeiten. 136 Seiten, 30 cm, kt.
 978-3-86884-543-3 € 29,80

Band 2019,01

Daumann, Frank (Projektleitung) / Heinze, Robin / Kumpel, Jeremias / Barth, Michael / Diethold, Sascha-Gregor
Quality Performance Measurement in National Sport-Governing Bodies. 396 Seiten, 21 cm, kt.
 978-3-86884-544-0 € 39,80

- Band 2019,02**
Killing, Wolfgang
Evaluation sportwissenschaftlicher Unterstützungsleistungen im Spitzensport am Beispiel der Leichtathletik. 242 Seiten, 30 cm, kt.
978-3-86884-545-7 € 49,80
- Band 2019,03**
Krüger, Michael / Nielsen, Stefan / Becker, Christian / Rehmann, Lukas
Sportmedizin zwischen Sport, Wissenschaft und Politik – eine deutsche Geschichte. Ein Forschungsprojekt zur Geschichte der Sportmedizin. 404 Seiten, 30 cm, kt.
978-3-86884-546-4 € 64,80
- Band 2019/04**
Radtke, Sabine / Schäfer, Lisa
Inklusion im Nachwuchsleistungssport. Vereinbarkeit von Schule und paralympischem Leistungssport an Eliteschulen des Sports versus Regelschulen. 354 Seiten, 21 cm, kt.
978-3-86884-547-1 € 38,00
- Band 2020/01**
Kellmann, Michael / Kölling, Sarah
Das Akutmaß und die Kurzskala zur Erfassung von Erholung und Beanspruchung für Erwachsene und Kinder/Jugendliche. 216 Seiten, 30 cm, kt.
978-3-86884-548-8 € 39,80
- Band 2021,01**
Memmert, Daniel u. a.
Bekämpfung der Spielwettbewerbsmanipulation: Evaluierung der Präventionsmaßnahmen im organisierten Sport. Abschlussbericht zum Forschungsprojekt. 1. Aufl. 2021. 160 Seiten, 6 Abb., 69 Tab., 30 cm, kt.
978-3-86884-549-5 € 39,80
- ## Publikationen ohne Band-Nummerierung
- Beckmann, Jürgen / Wenhold, Franziska
Handlungsorientierung im Sport. Manual zur Handlungsorientierung im Sport (HOSP). 1. Aufl. 2009. 48 Seiten, 29,7 cm, kt.
978-3-86884-508-2 € 16,00
- Behringer, Michael / Heede, Andreas vom / Mester, Joachim
Krafttraining im Nachwuchsleistungssport unter besonderer Berücksichtigung von Diagnostik, Trainierbarkeit und Trainingsmethodik. Wissenschaftliche Expertise des BISP, Band II. 1. Aufl. 2010. 194 Seiten, 21 cm, kt.
978-3-86884-487-0 € 17,80
- Brand, Ralf / Ehrlenspiel, Felix / Graf, Karla
Wettkampf-Angst-Inventar (WAI). Manual zur komprehensiven Eingangsdiagnostik von Wettkampfangst, Wettkampfstärke und Angstbewältigungsmodus im Sport. 1. Aufl. 2009. 150 Seiten, 29,7 cm, kt.
978-3-86884-491-7 € 26,80
- Brejcha-Richter, Stanislav / Hillenbach, Elke / Klein, Klaus / Kukowka, Dorothea
40 Jahre Bundesinstitut für Sportwissenschaft. Bilanz und Ausblick. Dokumentation der Festveranstaltung. 1. Aufl. 2011. 64 Seiten, 21 cm, kt.
978-3-86884-455-9 € 12,00
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2007/2008. Analyse zur Situation der Sportvereine in Deutschland. 1. Aufl. 2009. 740 Seiten, 21 cm, kt.
978-3-86884-500-6 € 48,00
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2009/2010. Analyse zur Situation der Sportvereine in Deutschland. 1. Aufl. 2011. 596 Seiten, 21 cm, kt.
978-3-86884-456-6 € 38,80
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2011/2012. Analyse zur Situation der Sportvereine in Deutschland. 1. Aufl. 2013. 902 Seiten, 21 cm, kt.
978-3-86884-463-4 vergriffen
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2013/2014. Analyse zur Situation der Sportvereine in Deutschland. 1. Aufl. 2015. 864 Seiten, 21 cm, kt.
978-3-86884-581-5 € 48,80
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2015/2016. Bd. I. Analyse zur Situation der Sportvereine in Deutschland. 1. Aufl. 2017. 830 Seiten, zahlr. z.T. farb. Abb. und Tab., 21 cm, kt.
978-3-86884-595-2 € 48,00
- Breuer, Christoph (Hrsg.)
Sportentwicklungsbericht 2015/2016. Bd. II. Weiterführende Strukturanalysen. 1. Aufl. 2017. 648 Seiten, zahlr. z.T. farb. Abb. und Tab., 21 cm, kt.
978-3-86884-596-9 € 48,00
- Breuer, Christoph (Hrsg.)
Sportverbände, Sportvereine und ausgewählte Sportarten. Weiterführende Analysen der Sportentwicklungsberichte. 1. Aufl. 2013. 452 Seiten, 21 cm, kt.
978-3-86884-580-8 € 34,80
- Breuer, Christoph / Feiler, Svenja
Sportentwicklungsbericht 2011/2012. Analyse zur Situation der Sportvereine in Deutschland. Kurzfassung. 1. Aufl. 2013. 54 Seiten, 21 cm, kt.
978-3-86884-464-1 € 12,00
- Breuer, Christoph / Feiler, Svenja
Sport Development Report 2013/2014. Analysis of the situation of sports clubs in Germany. Abbreviated Version. 1. Aufl. 2015. 42 Seiten, 21 cm, kt.
978-3-86884-583-9 € 14,00
- Breuer, Christoph / Feiler, Svenja
Sportentwicklungsbericht 2013/2014. Analyse zur Situation der Sportvereine in Deutschland. Kurzfassung. 1. Aufl. 2015. 46 Seiten, 21 cm, kt.
978-3-86884-582-2 € 12,00
- Breuer, Christoph / Feiler, Svenja
Sport Development Report 2015/2016. Analysis of the situation of sports clubs in Germany. Abbreviated Version. 1. Aufl. 2017. 40 Seiten, 11 z.T. farb. Abb. 18 Tab. 21 cm, kt.
978-3-86884-586-0 € 14,00

- Breuer, Christoph / Feiler, Svenja
Sportentwicklungsbericht 2015/2016. Analyse zur Situation der Sportvereine in Deutschland. Kurzfassung. 1. Aufl. 2017. 40 Seiten, 11 z.T. farb. Abb. 18 Tab. 21 cm, kt. 978-3-86884-585-3 € 12,00
- Breuer, Christoph / Hallmann, Kirstin
Die gesellschaftliche Relevanz des Spitzensports in Deutschland. 1. Aufl. 2012. 48 Seiten, 21 cm, kt. 978-3-86884-467-2 € 12,00
- Breuer, Christoph / Hallmann, Kirstin / Ilgner, Michael
Akzeptanz des Spitzensports in Deutschland – Zum Wandel der Wahrnehmung durch Bevölkerung und Athleten. 1. Aufl. 2017. 52 Seiten, 7 farb. Abb. 21 Tab. 21 cm, kt. 978-3-86884-584-6 vergriffen
- Breuer, Christoph / Hallmann, Kirstin / Ilgner, Michael
Erfolgsfaktoren der Athletenförderung in Deutschland. 1. Aufl. 2015. 86 S. 17 Abb. 17 Tab. 21 cm. 978-3-86884-594-5 € 14,80
- Breuer, Christoph / Wicker, Pamela
Zur Situation der Sportarten in Deutschland. Eine Analyse der Sportvereine in Deutschland auf Basis der Sportentwicklungsberichte. 1. Aufl. 2011. 384 Seiten, 21 cm, kt. 978-3-86884-460-3 € 28,40
- Breuer, Christoph / Wicker, Pamela
Sportökonomische Analyse der Lebenssituation von Spitzensportlern in Deutschland. 1. Aufl. 2010. 58 Seiten, 21 cm, kt. 978-3-86884-477-1 € 12,00
- Bundesinstitut für Sportwissenschaft (Hrsg.)
BISp-Handreichung zur Expertise Krafttraining im Nachwuchsleistungssport. Zusammenfassung der wissenschaftlichen Expertisen des BISp erstellt durch die Arbeitsgruppen Ruhr-Universität Bochum, Universität Leipzig, Deutsche Sporthochschule Köln 1. Aufl. 2010. 46 Seiten, 21 cm, kt. 978-3-86884-488-7 vergriffen
- Elbe, Anne-Marie / Wenhold, Franziska / Beckmann, Jürgen
Fragebogen zur Leistungsorientierung im Sport. Sport Orientation Questionnaire (SOQ). 1. Aufl. 2009. 44 Seiten, 29,7 cm, kt. 978-3-86884-493-1 € 16,00
- Eskau, Andrea (Red.)
Nachwuchsrekrutierung und Nachwuchsförderung im Leistungssport der Menschen mit Behinderungen. BISp-Arbeitstagung. Bonn, 14. / 15. Mai. 1. Aufl. 2009. 202 Seiten, 29,7 cm, kt. 978-3-86884-507-5 € 24,00
- Eskau, Andrea (Red.)
Regenerationsmanagement und Sporttechnologie im Leistungssport der Menschen mit Behinderungen. 1. Aufl. 2013. 92 Seiten, 21 cm, kt. 978-3-86884-590-7 € 18,00
- Ferrauti, Alexander u.a.
Basketball Talente. Leitlinien und Empfehlungen auf dem Weg zur Spitze. 1. Aufl. 2015. 38 Seiten, 10,5 x 30 cm, kt. 978-3-86884-591-4 vergriffen
- Gänsslen, Axel / Schmehl, Ingo
Leichtes Schädel-Hirn-Trauma im Sport. Handlungsempfehlungen. 1. Aufl. 2015. 44 Seiten, 21 cm, kt. 978-3-86884-593-8 € 9,80
- Hahn, Andreas u.a. (Hrsg.)
Biomechanische Leistungsdiagnostik im Schwimmen. Erfahrungen im Leistungssport und Ableitungen für die Ausbildung von Studierenden. Beiträge zum dvs-Symposium Schwimmen 10.–12.09.2009 in Leipzig. 1. Aufl. 2010. 188 Seiten, 21 cm, kt. 978-3-86884-489-4 € 18,00
- Hartmann, Ulrich u.a.
Krafttraining im Nachwuchsleistungssport unter besonderer Berücksichtigung von Ontogenese, biologischen Mechanismen und Terminologie. Wissenschaftliche Expertise des BISp, Band I. 1. Aufl. 2010. 312 Seiten, 21 cm, kt. 978-3-86884-486-3 € 23,80
- Hillenbach, Elke (Red.)
BISp-Jahrbuch. Forschungsförderung 2011/12. 1. Aufl. 2012. 174 Seiten, 29,7 cm, kt. 978-3-86884-462-7 € 28,00
- Hillenbach, Elke (Red.)
BISp-Jahrbuch. Forschungsförderung 2014/15. 1. Aufl. 2016. 262 Seiten, 29,7 cm, kt. 978-3-86884-632-4 € 28,00
- Hillenbach, Elke (Red.)
BISp-Jahrbuch. Forschungsförderung 2015/16. 1. Aufl. 2017. 226 Seiten, 29,7 cm, kt. 978-3-86884-633-1 € 28,00
- Hillenbach, Elke (Red.)
BISp-Jahrbuch. Forschungsförderung 2016/17. 1. Aufl. 2018. 368 Seiten, 29,7 cm, kt. 978-3-86884-634-8 € 48,00
- Horn, Andrea / Neumann, Gabriele (Hrsg.)
BISp-Ratgeber Projektförderung. Von der Idee zum erfolgreichen Projekt. Möglichkeiten und Wege der Projektförderung beim Bundesinstitut für Sportwissenschaft. 2. Aufl. 2009. 60 Seiten, 29,7 cm, kt. 978-3-86884-482-5 € 12,00
- Kellmann, Michael / Weidig, Thorsten
Pausenverhaltensfragebogen (PVF)- Manual. 1. Aufl. 2010. 94 Seiten, 29,7 cm, kt. 978-3-86884-485-6 € 16,00
- Killing, Wolfgang / Hommel, Helmar
Bundestrainerforum „DLV-Kältekonzferenz“ 06.–07.12.2008 in Mainz. 1. Aufl. 2009. 76 Seiten, 21 cm, kt. 978-3-86884-496-2 € 14,80
- Killing, Wolfgang / Hommel, Helmar
Bundestrainerforum „DLV-Kraftkonzferenz“ 15.–16.11.2008 in Potsdam. 1. Aufl. 2009. 172 Seiten, 21 cm, kt. 978-3-86884-495-5 vergriffen
- Klein, Klaus / Koch, Thomas / Palmen, Michael / Weinke, Irina (Red.)
BISp-Report 2009/10. Bilanz und Perspektiven. 1. Aufl. 2010. 130 Seiten, 30 cm, kt. 978-3-86884-471-9 € 14,80

- Klein, Klaus / Koch, Thomas / Palmen, Michael / Koch, Irina (Red.)
BISp-Report 2010/11. Bilanz und Perspektiven. 1. Aufl. 2012. 126 Seiten, 30 cm, kt.
 978-3-86884-468-9 € 14,80
- Klein, Klaus / Palmen, Michael / Stell, Sabine / Streppelhoff, Robin (Red.)
BISp-Report 2012. Bilanz und Perspektiven. 1. Aufl. 2013. 86 Seiten, 30 cm, kt.
 978-3-86884-570-9 € 19,80
- Klein, Klaus u.a. (Red.)
BISp-Report 2014/15. Bilanz und Perspektiven. 1. Aufl. 2015. 102 Seiten, 30 cm, kt.
 978-3-86884-572-3 € 19,80
- Klein, Klaus u.a. (Red.)
BISp-Report 2015/16. Bilanz und Perspektiven. 1. Aufl. 2016. 110 Seiten, 30 cm, kt.
 978-3-86884-573-0 € 8,60
- Klein, Klaus u.a. (Red.)
BISp-Report 2016/17. Bilanz und Perspektiven. 1. Aufl. 2017. 172 Seiten, 30 cm, kt.
 978-3-86884-574-7 € 16,80
- Meyer, Tim / Ferrauti, Alexander / Kellmann, Michael / Pfeiffer, Mark
Regenerationsmanagement im Spitzensport. REGman – Ergebnisse und Handlungsempfehlungen. 1. Aufl. 2016. 122 Seiten, 30 cm, kt.
 978-3-86884-589-1 **vergriffen**
- Muckenhaupt, Manfred (Hrsg.)
Wissen im Hochleistungssport. Perspektiven und Innovationen. Veröffentlichung anlässlich des Internationalen Symposiums Informations- und Wissensmanagement im Hochleistungssport 24. Juni–26. Juni 2010 Heinrich-Fabrit-Institut Blaubeuren. 1. Aufl. 2011. 180 Seiten, 21 cm, kt.
 978-3-86884-475-7 € 26,00
- Müller, Elena (Red.)
Top-Forschung für den Sport. BISp-Symposium. Bonn, 15. April 2010. 1. Aufl. 2011. 154 Seiten, 21 cm, kt.
 978-3-86884-474-0 € 24,00
- Neumann, Gabriele (Hrsg.)
Sportpsychologische Betreuung des deutschen Olympia- & Paralympicteams 2008. Erfolgsbilanzen-Erfahrungsberichte-Perspektiven. 1. Aufl. 2009. 156 Seiten, 29,7 cm, kt.
 978-3-86884-492-4 **vergriffen**
- Neumann, Gabriele (Red.)
Talentiagnose und Talentprognose im Nachwuchssport. 2. BISp-Symposium: Theorie trifft Praxis. 1. Aufl. 2009. 444 Seiten, 21 cm, kt.
 978-3-86884-497-9 **vergriffen**
- Neumann, Gabriele / Stehle, Peter (Hrsg.)
Fußball interdisziplinär. Zur Optimierung der Prävention, Rehabilitation und Wiederverletzungsprophylaxe von Knie- und Sprunggelenksverletzungen. 1. Aufl. 2009. 86 Seiten, 29 cm, kt.
 978-3-86884-490-0 **vergriffen**
- Pilz, Gunter A. / Behn, Sabine / Harzer, Erika / Lynen von Berg, Heinz / Selmer, N.
Rechtsextremismus im Sport. In Deutschland und im internationalen Vergleich. 2. ergänzte Aufl. 2014. 204 Seiten, 29,7 cm, kt.
 978-3-86884-506-8 € 28,00
- Quade, Karl / Rebel, Mirjam / Müller, Elena (Red.)
Volleyball. BISp-geförderte Forschungsprojekte der Jahre 2005 bis 2009. 1. Aufl. 2009. 142 Seiten, 29,7 cm, kt.
 978-3-86884-483-2 **vergriffen**
- Shakhlina, Larissa J. G.
Medizinisch-biologische Grundlagen des sportlichen Trainings von Frauen. 1. Aufl. 2010. 302 Seiten, 21 cm, kt.
 978-3-86884-479-5 € 24,50
- Stehle, Peter (Hrsg.)
BISp-Expertise: „Sensomotorisches Training – Propriozeptives Training“ Band I. 1. Aufl. 2009. 128 Seiten, 21 cm, kt.
 978-3-86884-498-6 € 19,80
- Stehle, Peter (Hrsg.)
BISp-Expertise: „Sensomotorisches Training – Propriozeptives Training“ Band II. 1. Aufl. 2009. 96 Seiten, 21 cm, kt.
 978-3-86884-499-3 € 16,80
- Stoll, Oliver / Achter, Mathias / Jerichow, Mareike
Vom Anforderungsprofil zur Intervention. Eine Expertise zu einem langfristigen sportpsychologischen Beratungs- und Betreuungskonzept für den Deutschen Schwimm-Verband e.V. (DSV). 1. Aufl. 2010. 82 Seiten, 29,7 cm, kt.
 978-3-86884-473-3 € 22,80
- Weidig, Thorsten
Erfolgsfaktor Trainer. Das Trainerverhalten in Spiel- und Wettkampfpausen auf dem Prüfstand. 1. Aufl. 2010. 194 Seiten, 21 cm, kt.
 978-3-86884-478-8 € 18,80
- Wenhold, Franziska / Elbe, Anne-Marie / Beckmann, Jürgen
Fragebogen zum Leistungsmotiv im Sport. Achievement Motives Scale-Sport (AMS-Sport). 1. Aufl. 2009. 58 Seiten, 29,7 cm, kt.
 978-3-86884-494-8 € 16,00
- Wenhold, Franziska / Elbe, Anne-Marie / Beckmann, Jürgen
Volitionale Komponenten im Sport. Fragebogen zur Erfassung volitionaler Komponenten im Sport (VKS). 1. Aufl. 2009. 50 Seiten, 29,7 cm, kt.
 978-3-86884-509-9 € 16,00
- Wetterich, Jörg / Eckl, Stefan / Schabert, Wolfgang
Grundlagen zur Weiterentwicklung von Sportanlagen. 1. Aufl. 2009. 326 Seiten, 21 cm, kt.
 978-3-86884-505-1 € 28,90
- Wippert, P.- M. / Brückner, M. / Fließner, M.
Der Nationale Dopingpräventionsplan: eine Potentialeinschätzung. Forschungsbericht zur Evaluation des NDPP. 1. Aufl. 2014. 150 Seiten, 29,7 cm, kt.
 978-3-86884-592-1 € 38,00



PD. Dr. Anne Hecksteden ist Akademische Rätin am Institut für Sport- und Präventivmedizin der Universität des Saarlandes. Ihr Forschungsschwerpunkt ist die Individualisierung von Diagnostik und Intervention in Leistungs- und Gesundheitssport.



Prof. Dr. Ralf Kellner war im Projektzeitraum Inhaber des Lehrstuhls für Statistik und quantitative Methoden an der Universität des Saarlandes. Ein Forschungsschwerpunkt ist die Anwendung Bayesianischer Verfahren. Seit dem Sommersemester 2021 bekleidet Prof. Kellner den Lehrstuhl für Financial Data Analytics an der Universität Passau.

SPORTVERLAG *Strauß*
Neuhaus 12 · 53940 Hellental
info@sportverlag-strauss.de
www.sportverlag-strauss.de

ISBN 978-3-86884-550-1



9 783868 84501