

MEINUNG

## Gibt es eine Datenkrise der experimentellen Wissenschaften?

*Der Kenntnisstand der Wissenschaft entwickelt sich rasant. Die Zahl der jährlichen Veröffentlichungen explodiert. So gab es im Jahr 2017 aus dem Bereich Wissenschaft und Technik 2,4 Millionen Publikationen, 750.000 mehr als zehn Jahre zuvor [1]. Die Publikationszahl wächst jährlich um etwa vier Prozent. Wie hoch ist die Qualität in der Quantität? Die Frage nach der Reproduzierbarkeit bereitgestellter Daten und damit der Zuverlässigkeit des Erkenntnisgewinns ist nicht neu und beschäftigt seit Jahren die Wissenschaftler/-innen, Zeitschrifteneditor/-innen und Förderinstitutionen. Trotzdem ist die Frage nach einer Daten- und Reproduktionskrise eher zu verneinen, allerdings benötigen wir Standards und kritische Rezeption.*

Veröffentlichungen stellen Erkenntnisse aus Experimenten, Studien oder theoretischen Betrachtungen bereit. Ihr Primäranliegen ist das Verbreiten neuer Ergebnisse, Schlüsse und daraus abgeleiteter Hypothesen. Wichtig ist hier anzumerken, dass Wissenschaft einen internationalen und im Ansatz demokratischen Prozess nutzt, der im Wesentlichen auf Ehrlichkeit und Vertrauen baut. Die Gesamtheit des Wissens resultiert aus diesem globalen Vorgang. Deshalb sind Normen und Regeln international festzulegen. Den Editor/-innen und Gutachter/-innen der Journale kommt im offenen oder anonymen Peer-Review-

Verfahren eine wesentliche Rolle bei der Festlegung der Standards und ihrer Kontrolle zu.

### Experimentelles Design als Kompromiss

Wissenschaftliches Arbeiten beginnt mit Hypothesen und Zielsetzungen und dem Ausarbeiten des experimentellen oder theoretischen Vorgehens, hier als „experimentelles Design“ bezeichnet. Diese Arbeiten setzen umfangreiche Investitionen an Personal, Sachmitteln und Zeit voraus. Sie entscheiden auch maßgeblich über berufliche Karrieren und wissenschaftliches Renommee. Nebenbedingungen wie begrenzte

Projektlaufzeiten und Fristen für Berichtspflichten, Verfügbarkeit von Mitarbeitenden, leistungsorientierte Mittelzuweisungen, persönliche monetäre Zuwendungen, erfolgsabhängige Anstellungen und Beförderungen und Erfolgsaussichten für Folgeanträge werden zu mitentscheidenden Faktoren beim Design der wissenschaftlichen Projekte und bestimmen den Druck zur Veröffentlichung. Diese Gemengelage widerstreitender Kräfte und Kriterien aus Primäranliegen und Nebenbedingungen führt dazu, dass das experimentelle Design als Kompromiss entsteht und nicht unbedingt mit dem Ziel der höchsten Präzision und maximalen Aussagekraft.

### Krise der Reproduzierbarkeit

Wissenschaft verlangt Wiederholbarkeit der Experimente oder Studien mit ähnlichem Ergebnis. Die Angaben zu den Methoden und die Darstellung der Ergebnisse müssen hinreichend detailliert sein, um die Experimente nachvollziehen und reproduzieren zu können. In einer Befragung erklärten mehr als 70 Prozent einer Gruppe von 1576 befragten Wissenschaftler/-innen, dass ihnen die Reproduktion einzelner Ergebnisse aus der Literatur misslungen sei [2]. Die Hälfte der Befragten gab an, dass Ergebnisse aus eigenen

## EINFLUSS DES EXPERIMENTELLEN DESIGNS AUF DIE AUSSAGEKRAFT

Das experimentelle Design entscheidet über die Wahrscheinlichkeit, dass die Schlussfolgerung aus einem Experiment richtig ist, und die Zuverlässigkeit, dass ein vorhandener Unterschied gefunden wird. Diese beiden Ziele sind teils gegenläufige Desiderata, wie hier beispielhaft erörtert. Es wurden Experimente durchgeführt und die erhaltenen Daten ausgewertet. Ziel des Experiments war es, real vorhandene Unterschiede zu finden oder die Hypothese eines Unterschieds zurückzuweisen, beispielsweise dass ein Pharmakon die Abwehrkraft eines Individuums stärkt. Je nach Art der Auswertung steigt hierbei die Wahrscheinlichkeit, den Unterschied zu finden (= Aussagekraft), aber auch die statistische Fehlerwahrscheinlichkeit, dass der Unterschied gar nicht existiert (Fehler 1. Art). Grünliche Schattierung markiert

gute Werte, rötliche Schattierung ungünstige Werte. Die Aussagekraft, d. h. man findet den vorhandenen Unterschied, ist umso höher je näher der Wert bei 1 als maximalem Wert liegt. Fehler 1. Art beschreibt formal die Irrtumswahrscheinlichkeit, d. h. wie wahrscheinlich es ist, dass aus den Daten ein Unterschied gefolgert wird, den es gar nicht gibt. Je kleiner der Wert ist, desto geringer ist die Wahrscheinlichkeit eines Fehlers. Er bedeutet statistisch, dass die Nullhypothese eines fehlenden Unterschieds zurückgewiesen wird und somit ein signifikanter Unterschied gefolgert wird, obwohl er tatsächlich nicht vorliegt. In anderen Worten: Die bereitgestellten Daten unterstützen nicht die Schlussfolgerung. Der Rechnung liegen ein nominale Fehler 1. Art von 0.05 und eine nominale Aussagekraft von 0.8 zugrunde [5].

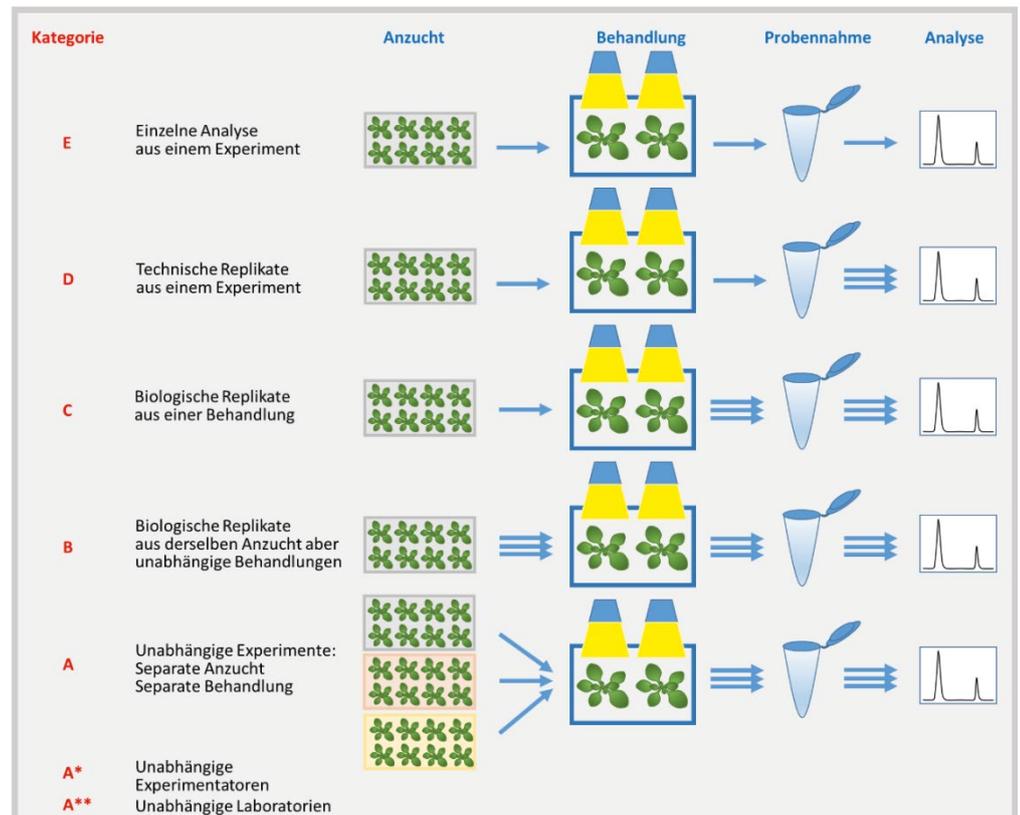
Strategie im experimentellen Design	Fehler 1. Art	Aussagekraft
Darstellen des besseren von zwei Experimenten	0,0975	0,96
Zwei von zwei durchgeführten Experimenten sind signifikant	0,0025	0,64
Darstellen des besten von drei Experimenten	0,143	0,992
Drei von drei durchgeführten Experimenten sind signifikant	0,000125	0,512
Zwei von drei durchgeführten Experimenten sind signifikant	0,00725	0,896

Versuchen in Einzelfällen nicht wiederholbar waren. Ein besonders eklatantes Beispiel wurde 2012 für präklinische Studien in der Onkologie berichtet, bei denen nur etwa 11 Prozent der ursprünglich erzielten Daten durch nachfolgende Ergebnisse bestätigt werden konnten [3].

Aus den eigenen Erfahrungen als Editor verschiedener Zeitschriften und den Diskussionen in *Editorial Boards* zeigt sich, dass in Veröffentlichungen eine Zunahme vager statistischer Angaben zu beobachten ist; beispielsweise mit der Stichprobenzahl von  $n = 3$  ohne Angabe der Art der Erhebung. Der Eindruck einer gezielten Verschleierung entsteht. Die exakte Beschreibung ist aber essentiell, um die Ergebnisse angemessen zu bewerten. Je nach gewähltem experimentellen Design steigt entweder die Fehlerwahrscheinlichkeit bei gleichzeitig zunehmender Zuverlässigkeit des Auffindens eines vorhandenen Unterschieds (siehe Kasten), z. B. wenn man nur das Beste von drei Ergebnissen präsentiert. Demgegenüber sinkt die Fehlerwahrscheinlichkeit, aber gleichzeitig fällt auch die Wahrscheinlichkeit, einen vorhandenen kleineren Unterschied zu erkennen, wenn Signifikanz in allen drei durchgeführten Experimenten gefordert wird. In diesem Fallbeispiel liegt die Irrtumswahrscheinlichkeit bei 0,000125.

### Umfassende Transparenz als erster Schritt

Die genaue und hinreichend umfassende Beschreibung des experimentellen Designs und der Datenlage ist der wichtige erste Schritt zur Klarstellung und Transparenz. Eine eindeutige Begrifflichkeit ist zu nutzen, die technische und biologische Replikate von unabhängigen Experimenten unterscheidet. Abbildung 1 zeigt mögliche Designs eines Labor-experiments in der Biologie anhand des Beispiels von Pflanzen, die nach ihrer Anzucht einer Behandlung und nachfolgenden Analyse unterzogen werden. Technische Replikate messen denselben Parameter in einer



**ABB. 1** Kategorisierung experimenteller Designs. Einzelne Analysen (E) und technische Replikate (D) werden biologischen Fragestellungen in der Regel nicht gerecht. Auch biologische Replikate aus einer Behandlung (C) bergen Risiken unerkannter Fehler. Unabhängige Behandlungen aus einer Anzucht (B) kann adäquat sein. Ideal ist die Kategorie (A), bei der die maximal im Labor mögliche Unabhängigkeit der Wiederholungen gewählt wird. Das Einbeziehen unabhängiger Experimentatoren (A\*) und unterschiedlicher Laboratorien (A\*\*) erhöht weiter die Sicherheit zuverlässiger Daten. Ein einzelner Pfeil zeigt eine einfache Durchführung, drei Pfeile unabhängige mehrfache Wiederholung dieses Schritts. Die im Text genannten Nebenbedingungen verhindern meist die Umsetzung eines solchen Standards.

Probe mehrmals und sind somit nur methodisch interessant. Sie zeigen die Genauigkeit der Methode. Biologische Replikate sind Messungen in unterschiedlichen Proben aus derselben Anzucht. Sie bergen die Gefahr, dass nicht evident gewordene Abweichungen im Bodensubstrat, bei der Bewässerung oder bei der Beleuchtung zu abnormalen und daher nicht reproduzierbaren Ergebnissen führen. Da diese Pflanzen gemeinsam unter denselben Bedingungen aufgewachsen sind, die aber nicht vollständig von den Experimentator/-innen erfasst sind, liefern sie gleiche Ergebnisse. Unter leicht veränderten Bedingungen, die dem Experimentator bzw. der Experimentatorin ebenfalls unbewusst sind, sind sie nicht reproduzierbar.

Erst das unabhängige Anziehen der Pflanzen mit nachfolgender Behandlung vermeidet dieses Risiko (Abbildung 1).

In der Befragung zur Reproduzierbarkeit gab ein Drittel der Forscher/-innen an, Maßnahmen zur Erhöhung der Reproduzierbarkeit getroffen zu haben [2]. Hierzu zählt als eine gängige Praxis, Schlüsselexperimente von anderen Experimentator/-innen im eigenen Labor reproduzieren zu lassen bzw. eine Auswertung einfach/doppelt verblindet durchzuführen.

### Fachspezifische Standards als zweiter Schritt

Im Kompromiss zwischen Primärkriterium bester Wissenschaft und Sekundärkriterien durch andere

Zwänge erscheint es wichtig, sich in den Disziplinen auf Mindeststandards für die wünschenswerte Sorgfalt zu verständigen. Dies könnte nach Abbildung 1 je nach Art des Experiments die Kategorie A sein, nämlich die Experimente mit unabhängig angezogenem biologischem Material zu wiederholen. Abweichungen von diesem Standard müssten begründet werden. Hier lassen sich verschiedene Gründe vorstellen, beispielsweise wenn anstelle der exakten Reproduktion des Experiments anderweitig erhaltene Ergebnisse zusammengekommen die Schlussfolgerungen auf den vorliegenden Zusammenhang oder beteiligten Mechanismus unterstützen. Standards für Laborversuche unterscheiden sich von denen für Freilandversuche. Zentral bleibt die Forderung nach unabhängigen Wiederholungen der experimentellen Behandlungen [4].

Zusätzliche ethische Aspekte ergeben sich beispielsweise bei der Durchführung von Tierversuchen in der Grundlagen- und präklinischen Forschung. Die Zahl der eingesetzten Tiere darf weder zu klein noch zu groß sein. Experimente mit einer zu kleinen Zahl an Versuchstieren sind bedeutungslos, da sie unzuverlässige Ergebnisse liefern; eine zu große Tierzahl ist ethisch nicht vertretbar. Unter solchen Bedingungen kann sich in statistischen Modellierungen [5] ergeben, dass Einzelexperimente mit angemessener Tierzahl gegenüber unabhängig wiederholten Experimenten mit zu geringer Tierzahl zu bevorzugen sind. Frommlet und Heinze [5] werteten Publikationen aus, die teils ohne Replikationen und teils mit unklaren Beschreibungen in der Zeitschrift „Immunity“ erschienen waren und als repräsentativ auch für andere hochrangige Zeitschriften angesehen werden können.

### Die Wichtigkeit vieler paralleler Maßnahmen

Das Konzept der *Open Science* strebt nach einer vollständigen Verfügbarkeit

wissenschaftlicher Daten, maximaler Transparenz ihrer Erhebung und freier Verfügbarkeit der Veröffentlichungen. In dieser Bewegung wurden neue Publikationsformate aufgelegt, die durch zahlreiche weitere Maßnahmen die Transparenz flankieren. Weitere Ebenen des wissenschaftlichen Systems sind einzubeziehen. Hier seien aus meiner Sicht zehn Beispiele aus verschiedenen Aktionsfeldern genannt:

- (1) Materialien und Methoden müssen umfassend und klar beschrieben werden und dürfen nicht durch Begrenzungen der Gesamtwortzahlen in den Veröffentlichungen beschnitten werden. Der Hinweis auf die Nutzung eines fertigen *kits*, in dem alle Komponenten durch eine Firma vorgegeben sind, ist ohne nähere Beschreibung ungenügend.
- (2) Wichtig erscheint, dass die Widerlegung von Hypothesen, landläufig als negative Ergebnisse beschrieben, veröffentlichbar wird und Anerkennung findet.
- (3) Die Veröffentlichung von Ergebnissen, die Daten reproduzieren, sollte erleichtert werden. Ein Ansatz ist, dies in speziellen Zeitschriften zu ermöglichen. Besser erscheint mir eine breitere Anerkennung und damit auch eine Veröffentlichung in Standardjournals. Der Link zu den vorangegangenen Studien sollte mit DOIs beispielsweise schon im Abstract erleichtert werden.
- (4) Die Zugänglichkeit zu Originaldaten verstärkt die Transparenz für den Leser bzw. die Leserin. Ich frage mich seit längerem, ob hier helfen könnte, die publizierten Daten im Supplement mit ihrem Datum der Erhebung aus dem Laborbuch zu versehen. Daraus würde besser erkennbar, ob es sich um technische, biologische Replikate oder vollständig unabhängige Experimente handelt.
- (5) Zu begrüßen ist, dass Zeitschriften vermehrt dazu übergehen, statistische Plausibilitätsprü-

fungen vorzunehmen und dabei Expert/-innen einzubinden.

- (6) Interessant ist die Idee, die Autor/-innen aufzufordern, die Grenzen ihrer eigenen Veröffentlichung zu beschreiben. Dies ist mir bei *iScience* begegnet.
- (7) Übersichtsartikel sind gehalten, Bewertungen und Kritik von Ergebnissen vorzunehmen und davon Abstand zu nehmen, eine unkommentierte Zusammenstellung anzufertigen. Solche unkritische Wiedergabe birgt die Gefahr, dass sich schwache Evidenz in der Literatur festsetzt. Hier kann allerdings auch eine Gegenposition bezogen werden, nämlich dass Bewertungen einen *Mainstream* zementieren könnten.
- (8) Das Training zur Datenbewertung, die Sensibilisierung gegenüber Fehlermöglichkeiten und die Kritikfähigkeit an Veröffentlichungen müssen gefördert werden. Das betrifft die Studierenden, *Early Career Researcher*, aber auch die gesamte wissenschaftliche Community. Der Dialog sollte lauter geführt werden.
- (9) Die Laufzeit von Forschungsprojekten sollte verlängert werden, um den Druck schnell zu veröffentlichen abzumildern. In einzelnen Formaten sind vier Jahre Förderung bereits etabliert, dies sollte Standard beispielsweise auch für Einzelsachmittelzuwendungen der DFG werden. Es gibt zahlreiche weitere Gründe für diese Forderung, die allerdings Gegenstand eines anderen Meinungsartikels sein sollte.
- (10) Im Rahmen der *Scientific Literacy* ist der Prozess des wissenschaftlichen Erkenntnisgewinns der interessierten Öffentlichkeit zu vermitteln. Dies ist eine heikle Mission, weil zu erklären ist, dass es im Sinne dieses Artikels zwar zuverlässigen und exzellenten Erkenntnisfortschritt gibt, aber auch fragwürdige Studien. Im *Web of Science* sind bereits über 225.000 Publikationen

über COVID-19 gelistet. Es drängt sich die Frage auf, welcher Anteil dieser doch in kurzer Zeit entstandenen Arbeiten wesentliche Erkenntnisse vermittelt.

Auch wenn ich weiter oben kritisch zu sehende Vorgehensweisen und Entwicklungen geschildert habe, so gibt es aus meiner Sicht keine grundsätzliche Daten- und Reproduktionskrise, denn die analytischen Möglichkeiten erlauben es, in nie dagewesener Qualität und mit hoher Präzision wertvolle Daten zu biologi-

schen Vorgängen zu generieren. Allerdings bedarf es einer kritischen Selbstkontrolle im Wissenschaftssystem, um valide Erkenntnisse von schwachen Evidenzen und fehlerhaften Befunden zu trennen. Alle Akteur/-innen sind gehalten, ihren Beitrag zur Verbesserung der Daten- und Reproduktionslage zu leisten. Dieser Meinungsartikel legt den Fokus auf Datenzuverlässigkeit, Reproduzierbarkeit und Standards – essentielle Prämissen für glaubwürdige Wissenschaft. Andere Themen der „Datenkrise“ wie die Datenflut und

Plagiate müssen zukünftigen Abhandlungen vorbehalten sein.

### Literatur

- [1] K. E. White (2019). National Science Foundation InfoBrief, NSF 19–317.
- [2] M. Baker (2016). Nature 533, 452–454.
- [3] C. G. Begley, L. M. Ellis (2012). Nature 483, 530–533.
- [4] A. Rogers et al. (2021). Journal of Experimental Botany 72, 5270–5274.
- [5] F. Frommlet, G. Heinze (2021). Laboratory Animals 55, 65–75.

*Karl-Josef Dietz, Bielefeld*

DOI:10.11576/biuz-5392

## AUS DEM VBIO

# Mathematisch-naturwissenschaftliche Gesellschaften: Fort- und Weiterbildung von Lehrkräften ist keine Freizeitgestaltung

*Ein Positionspapier der in der Initiative Wissenschaft-verbundet zusammengeschlossenen mathematisch-naturwissenschaftlichen Fachgesellschaften fordert Politik wie Schulträger auf, Lehrkräften mathematisch-naturwissenschaftlicher Schulfächer qualifizierte Angebote zu schaffen und die Teilnahme daran während der regulären Arbeitszeit zu erleichtern.*

Mathematisch-naturwissenschaftliche Erkenntnisse sind nicht nur angesichts des Klimawandels und der Corona-Pandemie wichtiger denn je. Zugleich unterliegt die Wissenschaft einer hohen Dynamik. Dem ist bei der Fortentwicklung des mathematisch-naturwissenschaftlichen Schulunterrichts Rechnung zu tragen. Qualitativ hochwertige Fort- und Weiterbildungen für mathematisch-naturwissenschaftliche Lehrkräfte sind daher ein unverzichtbares Instrument zur stetigen Verbesserung des Unterrichts. Davon sind die fünf auf der Plattform *Wissenschaft-verbundet* zusammengeschlossenen mathematisch-naturwissenschaftlichen Gesellschaften überzeugt.

Der Dachverband der Geowissenschaften (DVGeo), die Deutsche Mathematiker-Vereinigung (DMV), die Deutsche Physikalische Gesell-

schaft (DPG), die Gesellschaft Deutscher Chemiker (GDCh) sowie der Verband Biologie, Biowissenschaften und Biomedizin in Deutschland (VBIO) vertreten deutschlandweit über 130.000 Mitglieder und haben nun ein Positionspapier zur Fort- und Weiterbildung von Lehrkräften verabschiedet. Es richtet sich sowohl an die Politik als auch an Bildungsbehörden und Schulträger und fordert diese auf, sachgerechte Angebote zu fördern und weiter zu entwickeln. Dabei muss den besonderen Bedürfnissen der mathematisch-naturwissenschaftlichen Schulfächer Rechnung getragen werden. Nötig sind fachliche und fachdidaktische Fortbildungsangebote sowie Formate, die beide Elemente verbinden. Der wissenschaftliche Fortschritt sowie die aktuellen Herausforderungen der Unterrichtsent-

wicklung z. B. in Bezug auf die Digitalisierung müssen berücksichtigt werden. Essenziell dafür ist eine hinreichende finanzielle als auch organisatorische Unterstützung. Insbesondere sind Lehrkräfte für Fort- oder Weiterbildungsmaßnahmen ausreichend und unkompliziert vom Unterricht freizustellen. Die mathematisch-naturwissenschaftlichen Gesellschaften betonen dabei ganz explizit, dass Fort- oder Weiterbildungen keineswegs nur in die Abendstunden oder die Wochenenden geschoben werden dürfen, sondern selbstverständlich zur und in die Gesamtarbeitszeit gehören.

Sie finden das Positionspapier im Wortlaut unter [https://wissenschaft-verbundet.de/gemeinsame-aktivitaeten/download/positionspapier\\_fortbildungen.pdf](https://wissenschaft-verbundet.de/gemeinsame-aktivitaeten/download/positionspapier_fortbildungen.pdf)

DOI:10.11576/biuz-5393



**ABB. 1** Qualitativ hochwertige Fort- und Weiterbildungen für mathematisch-naturwissenschaftliche Lehrkräfte sind ein unverzichtbares Instrument zur stetigen Verbesserung des Unterrichts. Foto: DPG / Anja Metzelthin



Verband | Biologie, Biowissenschaften  
& Biomedizin in Deutschland

**GEMEINSAM  
FÜR DIE**

**BIEWISSENSCHAFTEN**

### **Gute Gründe, dem VBIO beizutreten:**

- Werden Sie Teil des größten Netzwerks von Biowissenschaftlern in Deutschland
- Unterstützen Sie uns, die Interessen der Biowissenschaften zu vertreten
- Nutzen Sie Vorteile im Beruf
- Bleiben Sie auf dem Laufenden – mit dem VBIO-Newsletter und dem Verbandsjournal „Biologie in unserer Zeit“
- Treten Sie ein für die Zukunft der Biologie



[www.vbio.de](http://www.vbio.de)

**Jetzt beitreten!**

