

Ausblicke der Nutzenbewertung des IQWiG

Stefan Lange

**Institut für Qualität und Wirtschaftlichkeit
im Gesundheitswesen**

17. Jahrhundert: Van Helmont (belgischer Arzt und Chemiker)

“Let us take out of the hospitals . . . 200 or 500 poor people, that have fevers, pleurisies. Let us divide them into halves, let us cast lots, that one half of them may fall to my share, and the other to yours; I will cure them without bloodletting and sensible evacuation; but you do, as ye know. . . . We shall see how many funerals both of us shall have.”

Doll R. Sir Austin Bradford Hill and the progress of medical science. *BMJ* 1992; 305: 1521-6.

„Versuch es doch einmal 10 Tage lang mit Deinen Knechten!“, bat er ihn. „Lass uns nur pflanzliche Nahrung zu essen und Wasser zu trinken geben. Dann vergleiche unser Aussehen mit dem der jungen Leute, die von den Speisen des Königs essen. Je nachdem, was Du dann siehst, verfahr mit Deinen Knechten. Der Aufseher nahm ihren Vorschlag an und machte mit ihnen eine 10-tägige Probe.“ (Buch Daniel, Kapitel 1, 12-14)

Daniel und seine Freunde am babylonischen Hof: 1,1-21



§ 135, SGB V (Erlaubnisvorbehalt)

- (1) **Neue Untersuchungs- und Behandlungsmethoden** dürfen in der **vertragsärztlichen und vertragszahnärztlichen Versorgung** zu Lasten der Krankenkassen nur erbracht werden, wenn der **Gemeinsame Bundesausschuss ... Empfehlungen** abgegeben hat über
 1. die **Anerkennung des diagnostischen und therapeutischen Nutzens** der neuen Methode sowie deren **medizinische Notwendigkeit und Wirtschaftlichkeit ...**

§ 137c, SGB V (Verbotsvorbehalt)

- (1) Der Gemeinsame Bundesausschuss nach § 91 überprüft ... Untersuchungs- und Behandlungsmethoden, die zu Lasten der gesetzlichen Krankenkassen **im Rahmen einer Krankenhausbehandlung** angewandt werden oder angewandt werden sollen, daraufhin, **ob sie** für eine ausreichende, zweckmäßige und wirtschaftliche Versorgung der Versicherten unter Berücksichtigung des allgemein anerkannten Standes der medizinischen Erkenntnisse **erforderlich sind**. ...
- (2) ... darf die Methode nicht mehr zu Lasten der gesetzlichen Krankenkassen erbracht werden; die Durchführung klinischer Studien bleibt unberührt.

§ 31, SGB V (Verbotsvorbehalt)

- (1) Versicherte haben Anspruch auf Versorgung mit apothekenpflichtigen Arzneimitteln, **soweit die Arzneimittel nicht nach § 34 oder durch Richtlinien nach § 92 Abs. 1 Satz 2 Nr. 6 ausgeschlossen sind, ...**

GKV-Wettbewerbsstärkungsgesetz (2007)

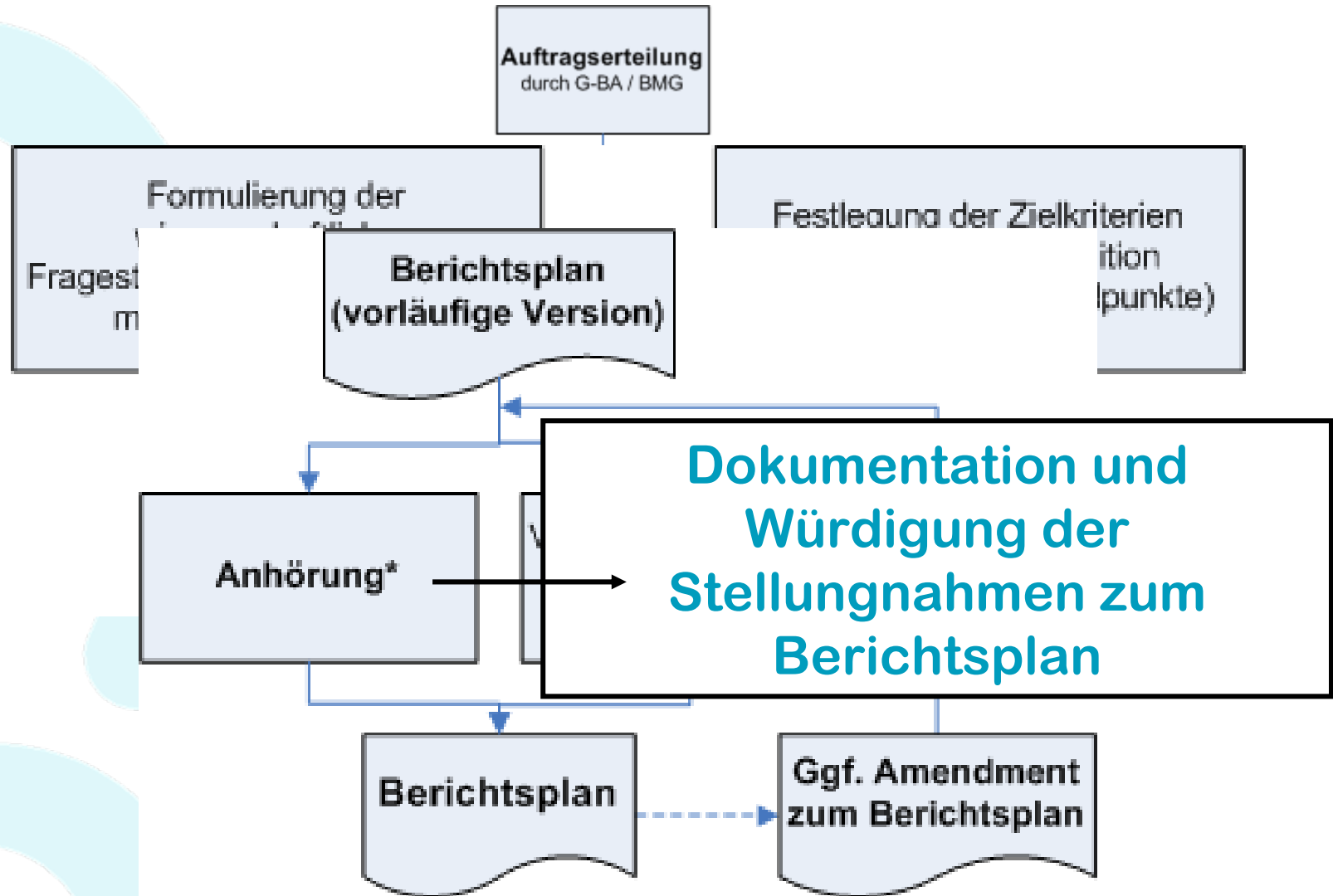
Änderung § 139a (4)

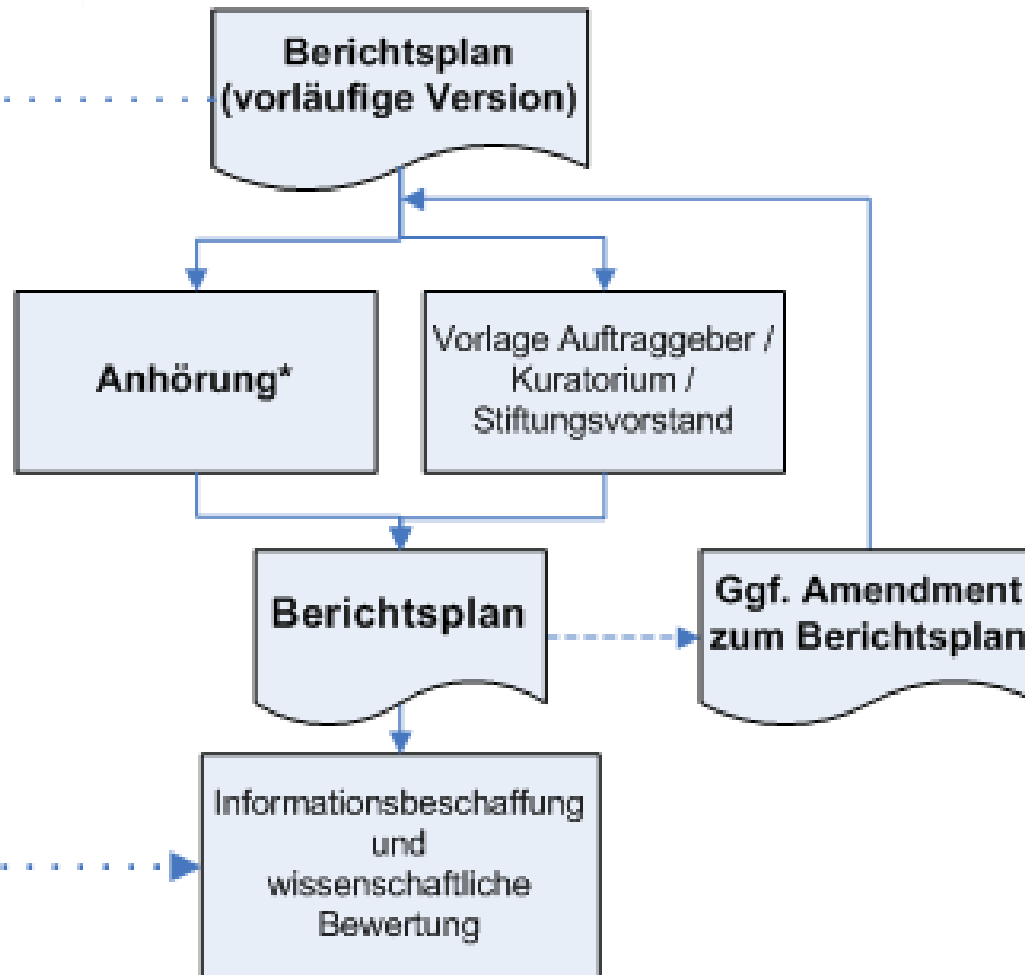
"(4) Das Institut hat zu gewährleisten, dass die Bewertung des medizinischen Nutzens nach den **international anerkannten Standards der evidenzbasierten Medizin ... erfolgt. ...**"

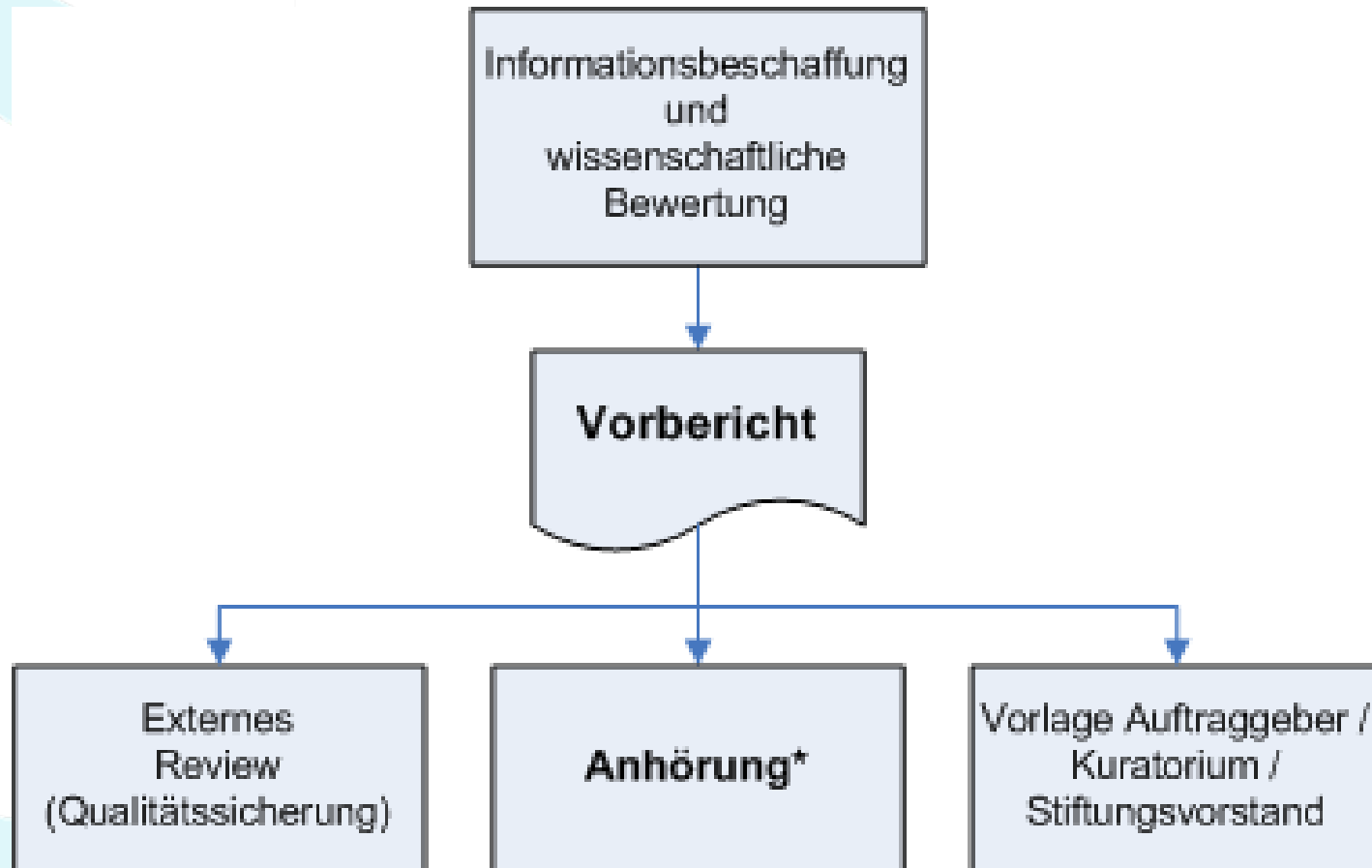
GKV-Wettbewerbsstärkungsgesetz (2007)

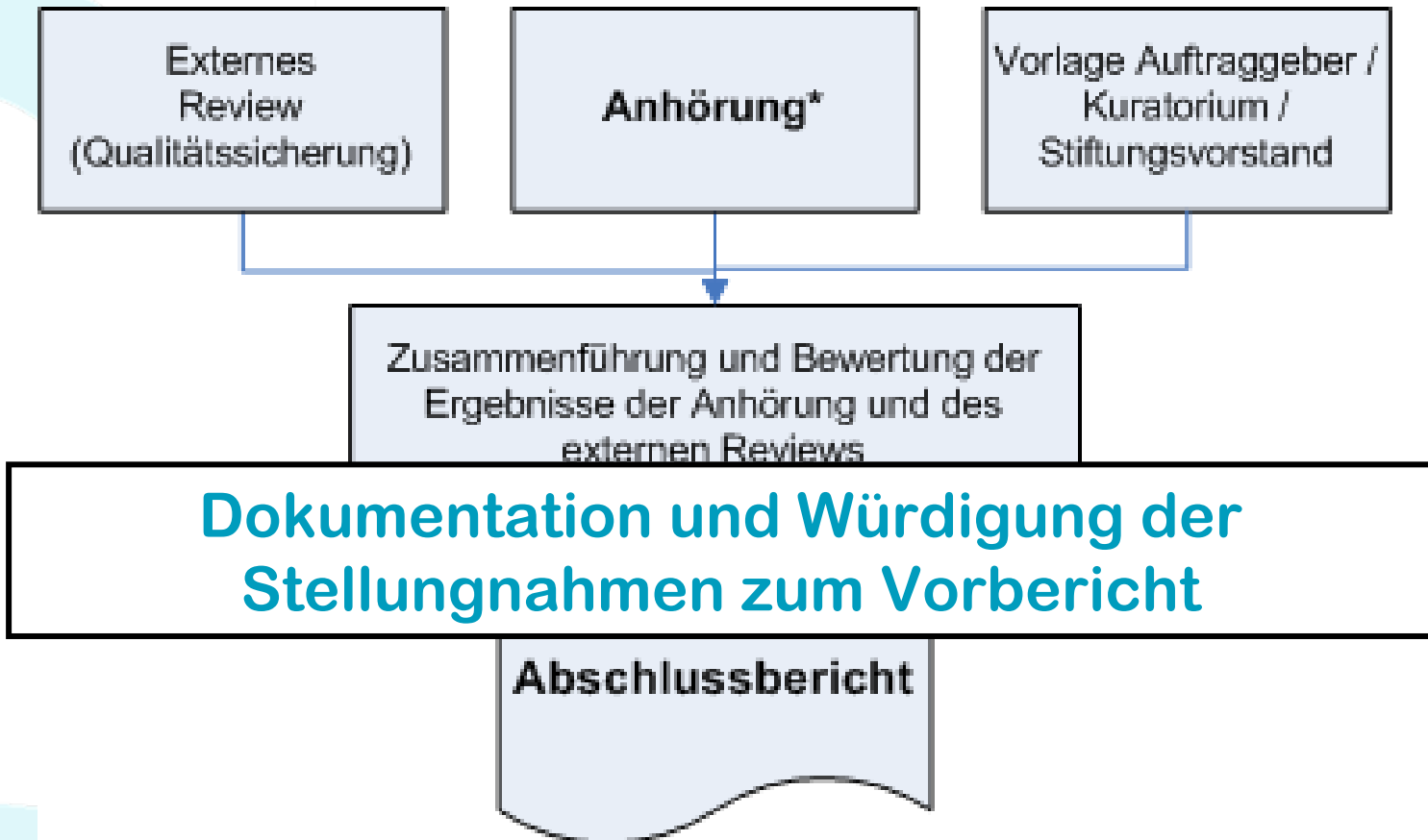
Änderung § 35b (1)

"... Das Institut bestimmt auftragsbezogen über die Methoden und Kriterien für die Erarbeitung von Bewertungen nach Satz 1 **auf der Grundlage der in den jeweiligen Fachkreisen anerkannten internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie.** Das Institut gewährleistet bei der auftragsbezogenen Erstellung von Methoden und Kriterien und der Erarbeitung von Bewertungen **hohe Verfahrenstransparenz und eine angemessene Beteiligung** der in § 35 Abs. 2 und § 139a Abs. 5 Genannten. Das Institut veröffentlicht die jeweiligen Methoden und Kriterien im Internet."









IQWiG



NICE

AHRQ

Auftrag

Auftrag

Scope

**Berichts-
plan**

Protokoll

**Vor-
bericht**

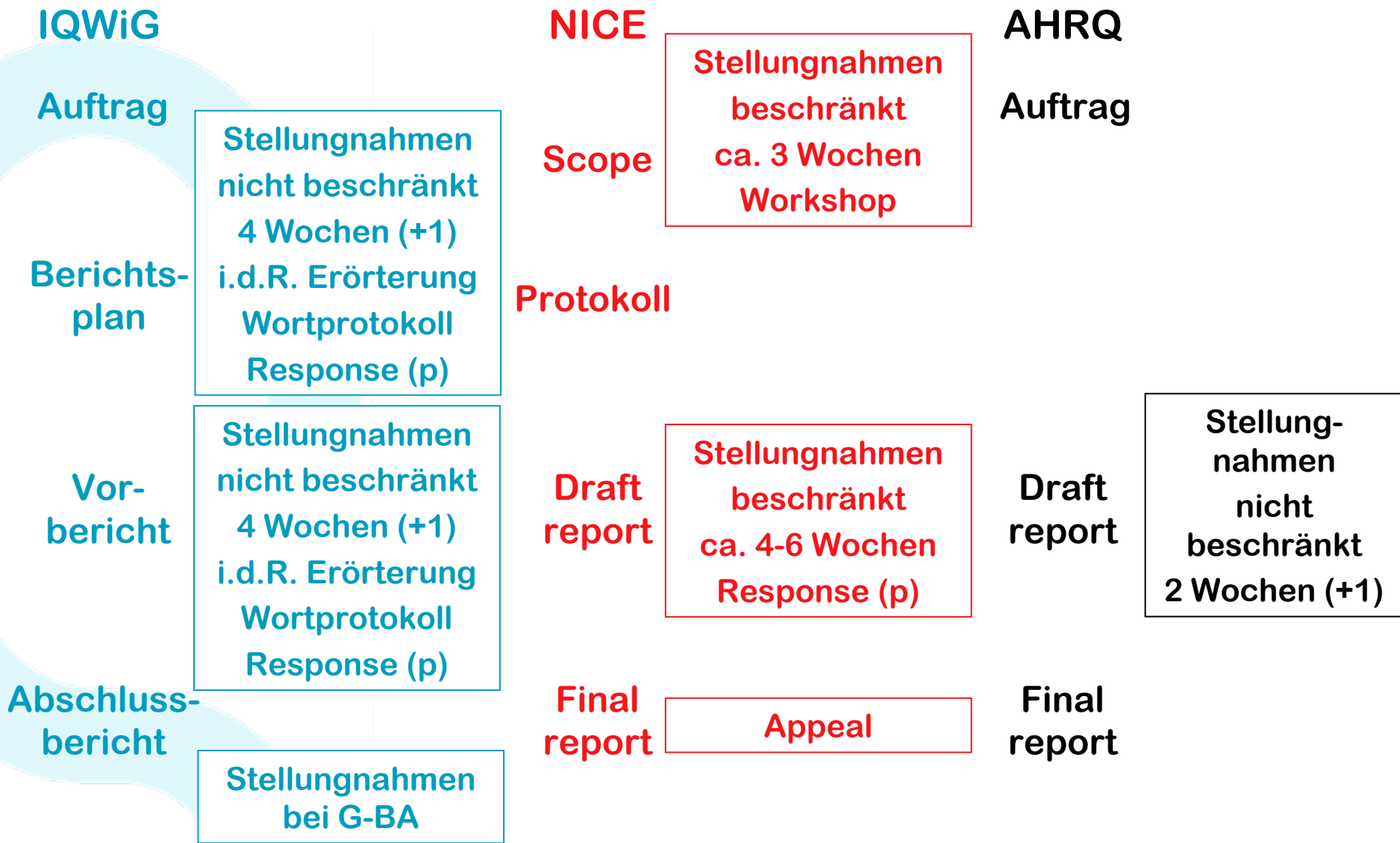
**Draft
report**

**Draft
report**

**Abschluss-
bericht**

**Final
report**

**Final
report**



Ginkgohaltige Präparate bei Alzheimer Demenz

"Die **Schlussfolgerungen** des IQWiG beruhen wesentlich auf den **beiden in der Ukraine durchgeführten Studien des Herstellers**. ... Die in keinem Studienregister angemeldeten Studien wurden dem IQWiG jeweils erst bekannt gegeben, als die Datenauswertung

Informationen aus Wortprotokoll der Erörterung

gegenüber auf seine Anfrage nach abgeschlossenen Studien hin aber verschwiegen und **erst nach Veröffentlichung des ersten IQWiG-Vorberichts** Anfang 2007 als fertiger Studienbericht überreicht. Die zu diesem Zeitpunkt fast abgeschlossene, bis heute unveröffentlichte Studie **SCHWABE 2008** wird weiterhin verschwiegen und gegenüber dem IQWiG **erstmalig in der Anhörung zum zweiten Vorbericht** Anfang 2008 erwähnt, wiederum als die **Ergebnisse feststehen.**"

arznei-telegramm 2009; 40(3): 29-30.

Memantin bei Alzheimer Demenz

"Es liegt kein Nutzenbeleg für die Behandlung von moderater bis schwerer Alzheimer Demenz (AD) mit Memantin vor. Für Patienten mit mittelschwerer Alzheimer Demenz gibt es jedoch für einzelne Endpunkte Hinweise auf einen Nutzen. **Die Aussagen stehen jedoch unter dem Vorbehalt, dass ein großer Teil der erhobenen Daten nicht in die Bewertung einfließen konnte, da sie nicht oder nur unvollständig zur Verfügung standen.**"

Peinemann F et al. Negative pressure wound therapy: Potential publication bias caused by lack of access to unpublished study results data. BMC Med Res Methodol 2008; 8: 4.

"Full-text publications were available on 30% of the 1649 patients in the 19 completed or discontinued RCTs (495 analysed patients in 10 published RCTs vs. 1154 planned patients in 9 unpublished RCTs). Of the 14 conference abstracts on these 19 trials, 6 abstracts (43%) were later published as full-text articles.."

Analyses of cost-effective BMD scanning and treatment strategies for generic alendronate, and the cost-effectiveness of risedronate and strontium ranelate in those people who would be treated with generic alendronate

Dr Matt Stevenson

February 2008



Concomitant use of bisphosphonates and acid suppressants versus current use of bisphosphonates without acid suppressant use and risk of fracture. Extract from Servier submission to NICE

[Contents of table in confidence, Servier]

"Commercial-in-confidence"!

Kurzwirksame Insulinanaloga zur Behandlung des Diabetes mellitus Typ 1

"Für Kinder und Jugendliche wurden **auch Langzeitstudien** identifiziert, die allerdings auf Grund fehlender vollständiger Publikationen sowie **nicht zur Verfügung gestellter Daten seitens des Sponsors dieser Langzeitstudien (Novo Nordisk)** nicht in die Nutzenbewertung eingeschlossen werden konnten. Derzeit können daher keine abschließenden Aussagen in Bezug auf den Nutzen der Insulinanaloga für Kinder und Jugendliche gemacht werden."

1958: Modell W, Houde RW: Factors influencing clinical evaluation of drugs. With special reference to the double-blind technique. JAMA (1958): 167: 2190-99.

"No experiment in any discipline can be pursued without controls. ... The control is a basis of comparison, and, thus, there is a control inherent in every judgment on a drug. The only important question about the control is, ..., whether it is sound as a basis for comparison on which the judgment is based."

"Only in the case of the disease in which an irrevocable or unquestionable characteristic course has been established, and particularly when the disease is rare, is the historical control justified."

„**Ziel** ist eine **möglichst hohe Strukturgleichheit** in diesen Gruppen hinsichtlich bekannter wie unbekannter Störgrößen (...), damit ein unter Umständen bei der Auswertung der Ergebnisse festzustellender **Unterschied zwischen den Gruppen tatsächlich auf die zu prüfende Therapie kausal zurückgeführt werden kann. ...**

Gäbe es eine Technik, die das Ziel der Strukturgleichheit genauso gut – oder eventuell sogar besser – erreicht wie die klassische randomisierte Studie, so spricht nichts dagegen, diese neue Methode einzusetzen. **Randomisierung ist kein Selbstzweck.** Derartige Modifikationen sind schon öfters diskutiert worden (zum Beispiel [14]), haben sich aber bisher auf breiter Front nicht durchgesetzt.“

Schuck P, Müller H, Resch K-L: Wirksamkeitsprüfung: „Doppelblindstudien“ und komplexe Therapien. Dt Ärztebl 2001; 98: Seite A-1942 / B-1826 / C-1555

Präzise Fragestellung (PICO)

**Systematisches (reproduzierbares, transparentes) Vorgehen
(Literaturrecherche und –bewertung)**

Einschätzung der Ergebnissicherheit als Basis für Empfehlung

Ergebnissicherheit

Qualitativ (Biaspotenzial)

Quantitativ (Präzision der Schätzung, Irrtumsniveau)

Größe (und Art) des Effekts

Royle P, Waugh N. Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system. Health Technol Assess 2003; 7(34): iii-x, 1-64.

20 HTAs zu "Effektivitätsfragestellungen"

NICE: zwischen 12/2000 und 12/2002

14 zu Arzneimitteln, davon 9 ausschließlich RCTs

andere Designs z.B. bei Zweit- und Drittlinientherapie hämatologischer Erkrankungen

6 zu anderen Interventionen, davon 2 ausschließlich RCTs

Royle P, Waugh N. Challenges in Using Nonrandomized Studies in Systematic Reviews of Treatment Interventions. Ann Intern Med. 2005;142:1112-1119.

78 HTAs zu "Wirksamkeits- und Effektivitätsfragestellungen"

AHRQ: zwischen 02/1999 und 09/2004

27 ausschließlich RCTs (davon 25 zu Arzneimitteln)

"No established guidelines address situations in which nonrandomized studies can or should be considered for inclusion in a systematic review or what study designs to consider."

Committee on Reviewing Evidence to Identify Highly Effective Clinical Services. Institute of Medicine (IOM). 2008. Knowing what works in health care: A roadmap for the nation. Eden J, Wheatley B, McNeil B, Sox H (Editors). Washington, DC: The National Academies Press. <http://www.nap.edu/catalog/12038.html>

"For intervention effectiveness questions for which RCTs form the highest level of evidence, it is essential to fully document the rationale for including nonrandomized evidence in a review. **Current practice does not meet this standard, however.**"

→ **Begründungspflicht für Einbezug von non-RCTs ist ein Standard!**

→ **Nutzenbewertung anhand "best-verfügbarer Evidenz" ist kein Standard!**

§ 35b, Absatz 1, SGB V

"Beim Patienten-Nutzen sollen insbesondere die Verbesserung des Gesundheitszustandes, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität ... berücksichtigt werden."

GKV-WSG 2007



„Da sich der Nutzen einer Maßnahme auf den Patienten beziehen soll, beruht diese Bewertung auf Ergebnissen wissenschaftlicher Untersuchungen zur Beeinflussung patientenrelevanter Endpunkte. Als ‚**patientenrelevant**‘ soll in diesem Zusammenhang verstanden werden, **wie ein Patient fühlt, seine Funktionen und Aktivitäten wahrnehmen kann oder überlebt** [44]. Dabei werden ... **Effekte der Interventionen** berücksichtigt, die eine Bewertung der Beeinflussung insbesondere folgender patientenrelevanter Zielgrößen zur Feststellung krankheits- und behandlungsbedingter Veränderungen erlauben:

- 1. Mortalität,**
- 2. Morbidität (Beschwerden und Komplikationen),**
- 3. gesundheitsbezogene Lebensqualität.“**

„Ergänzend können der interventions- und erkrankungs-
bezogene Aufwand und die **Zufriedenheit der Patienten mit der
Behandlung** berücksichtigt werden. Solche Aspekte kommen
allerdings **in der Regel nur als sekundäre Zielgrößen in
Betracht.**“

§ 34, Absatz 1, SGB V

„Von der Versorgung sind außerdem

- Arzneimittel ausgeschlossen, bei deren Anwendung eine Erhöhung der Lebensqualität im Vordergrund steht.“

■ Gesundheitsbezogene Lebensqualität

- „Die Regelung stellt sicher, dass die Finanzierung von Arzneimitteln, deren Einsatz im Wesentlichen durch die Art der persönlichen Lebensführung bedingt ist, durch die einzelne Patientin bzw. den einzelnen Patienten selbst zu gewährleisten ist.“



REFLECTION PAPER ON THE REGULATORY GUIDANCE FOR THE USE OF HEALTHRELATED QUALITY OF LIFE (HRQL) MEASURES IN THE EVALUATION OF MEDICINAL PRODUCTS



Guidance for Industry. Patient-Reported
Outcome Measures: Use in Medical
Product Development to Support Labeling
Claims. Draft Guidance. 2006



**Nutzenbewertung kann nicht geringere
Anforderungen setzen als Zulassung**

**REFLECTION PAPER ON THE REGULATORY GUIDANCE FOR
THE USE OF HEALTHRELATED QUALITY OF LIFE (HRQL)
MEASURES IN THE EVALUATION OF MEDICINAL PRODUCTS**



"In severe, life-threatening diseases, such as cancer, HRQL may provide an important information. In all cases, there must be confidence that the observed HRQL benefit is achieved without any reduction in efficacy. ...

In chronic non life threatening conditions requiring long term treatments, when the two drugs have similar efficacy and safety, the information on HRQL might be important for the choice of one medicinal product over the other in the current clinical practice."

„Sowohl Nutzen- als auch Schadenaspekte können eine **unterschiedliche Wichtigkeit für die Betroffenen** haben, Daher kann ... eine unterschiedliche Gewichtung der Zielgrößen vorgenommen werden. Das hat dann zur Folge, dass die Voraussetzung für die vergleichende Würdigung bestimmter (geringer gewichteter) Zielgrößen die vergleichende Würdigung anderer (höher gewichteter) Zielgrößen ist. **So kann z. B. eine geringere Ausprägung unerwünschter Effekte nur dann einen Zusatznutzen für die Patienten bedeuten, wenn gleichzeitig die höher gewichteten Zielgrößen in einem ähnlichen bzw. akzeptabel geringeren Ausmaß als dem der Vergleichsintervention günstig beeinflusst werden.** Dies setzt allerdings voraus, dass eine a priori Definition dessen, was noch als akzeptabel geringer angesehen werden kann (**Irrelevanzbereich**, siehe Abschnitt 6.4.5), vorliegt.“

Nichtmedikamentöse lokale Verfahren zur Behandlung der benignen Prostatahyperplasie

"(B) Alternativ wurde die folgende Herangehensweise vorgeschlagen: Die Rationale zur Entwicklung und Einführung der weniger invasiven (Alternativ-)Verfahren sei, dass diese bei vergleichbaren oder nur geringfügig schlechteren Ergebnissen hinsichtlich der Symptomatik zu geringeren Raten von unerwünschten Ereignissen bzw. Komplikationen führten. Ein Zusatznutzen liege daher in der **Kombination von höchstens irrelevanter Unterlegenheit des Alternativverfahrens bzgl. der Symptomatik und gleichzeitig nachgewiesenen Vorteilen hinsichtlich des Auftretens von Komplikationen.**"

Guidance for Industry. Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labeling
Claims. Draft Guidance. 2006



“ ... **extremely complex** concepts such as quality of life,
which is widely understood to be a multidomain concept ... ”

“The amount and kind of evidence that the FDA expects to
support a labeling claim measured by a PRO instrument is
the same as that required for any other labeling claim.”

Guidance for Industry. Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labeling
Claims. Draft Guidance. 2006



“Smallest difference that is considered clinically important; this can be a specified difference (the minimum important difference (MID)) or, in some cases, any detectable difference. The MID is used as a **benchmark to interpret mean score differences between treatment arms in a clinical trial.**”

“**Responder definition - Change in score that would be clear evidence that an individual patient experienced a treatment benefit.** Can be based on experience with the measure using a distribution-based approach, a clinical or nonclinical anchor, an empirical rule, or a combination of approaches.”

Nachweis?

Punktschätzer

→ Quantifizierung des Irrtumsniveaus unmöglich

Untere Konfidenz-Grenze

→ konservativ

Guidance for Industry. Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labeling
Claims. Draft Guidance. 2006



“If the MID is truly to be the smallest effect considered meaningful, however, it would be logical to establish the null hypothesis to rule out a difference less than or equal to the MID. This is rarely done, and would have major implications for sample size.”

Selektive Serotonin- und Noradrenalin-Wiederaufnahme- hemmer (SNRI) bei Patienten mit Depressionen

"Für die Einschätzung der Relevanz eines Gruppenunterschieds bezüglich stetiger Skalen sollten aus der Literatur bekannte 'Minimal Important Differences' (MID) für den Gruppenunterschied berücksichtigt werden. Da für keine der eingesetzten Skalen eine verwendbare MID für den Gruppenunterschied vorlag, wurde eine Berechnung von Cohen's d durchgeführt. **Für Cohen's d wurde eine Relevanzgrenze von 0,2 angenommen, Bei der Interpretation der Effektstärken wurde das Konfidenzintervall zu dieser Relevanzgrenze in Relation gesetzt.** Lag das Konfidenzintervall vollständig oberhalb der Relevanzgrenze wurde von einer relevanten Effektgröße ausgegangen und ein Nutzenbeleg oder -hinweis attestiert."

Guidance for Industry. Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labeling
Claims. Draft Guidance. 2006



"When clinical trials show small mean effect sizes, rather than considering results in terms of an MID, it may be more informative to examine the distribution of responses between treatment groups to more fully characterize the treatment effect and examine the possibility that the mean improvement reflects very different responses in subsets of patients. ... When defining a meaningful change on an individual patient basis (i.e., a responder), that definition is generally larger than the minimum important difference for application to group mean comparisons."

Guidance for Industry. Patient-Reported Outcome Measures:
Use in Medical Product Development to Support Labeling
Claims. Draft Guidance. 2006



“Because responses to PRO measures are subjective, representing a patient’s impression, **openlabel studies**, where patients and investigators are aware of assigned therapy, are **rarely credible.**”

“Every effort should be made to assure that patients are masked to treatment assignment throughout the trial.”

Abbildung 2 aus Publikation

Woods et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ* 2008; 336: 601-5

Langwirksame Insulinanaloga zur Behandlung des Diabetes mellitus Typ 2

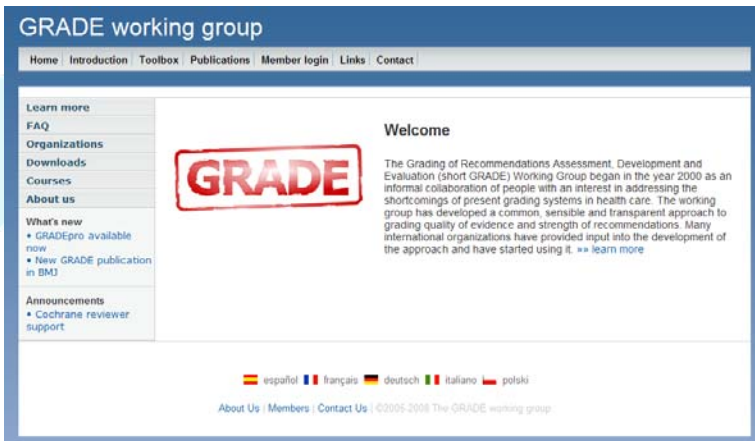
"Diese **mittlere Verzerrung** wird im Folgenden für nicht schwere und nicht schwere nächtliche Hypoglykämien **im Sinne einer adjustierten Entscheidungsgrenze** angewendet, um für mögliche systematische Verzerrung in den Studien zu kontrollieren. Nur wenn die obere Grenze des 95 %-Konfidenzintervalls für das Odds Ratio des Behandlungsunterschiedes unterhalb von 0,75 lag, so wurde der Behandlungseffekt als so groß angesehen, dass er nicht allein durch Bias erklärt werden kann."

„We suggest that a sufficiently extreme difference between the outcome ranges for treated and untreated patients might be defined by two rules: (a) that the conventionally calculated probability of the two groups of observations coming from the same population should be less than 0.01 and (b) that the estimate of the treatment effect (rate ratio) should be large.

...

We therefore suggest that rate ratios beyond 10 are highly likely to reflect real treatment effects, even if confounding factors associated with the treatment may have contributed to the size of the observed associations.“

Glasziou P, Chalmers I, Rawlins M, McCulloch P. BMJ 2007; 334: 349-52.



Heraufstufen von
Ergebnissen aus non-RCTs
bei Vorliegen starker (1
Stufe) bzw. sehr starker
Effekte (2 Stufen) auf einer
4-stufigen Skala.

Starker Effekt: Relatives Risiko 2

Sehr starker Effekt: RR 5

Punktschätzer? Konfidenzgrenze!

Die internationalen Standards sind klar, und das IQWiG folgt ihnen.

Im internationalen Vergleich arbeitet das IQWiG äußerst transparent und erlaubt eine extensive Mitwirkung der Öffentlichkeit.

Warum werden nicht **alle** Studienergebnisse zeitnah öffentlich zugänglich gemacht?

Studien mit einem Design hoher Verzerrungsanfälligkeit erfordern i.d.R. einen unverhältnismäßig hohen Aufwand, um ausreichend sicher interpretiert werden zu können.

Eine "vierte Hürde" (Erlaubnisvorbehalt) dürfte Manches einfacher machen und die Abläufe beschleunigen!