



**Prof. Dr. med. Thomas Danne**  
**Präsident der DDG**  
**Vorstandsvorsitzender diabetesDE**  
Kinderkrankenhaus auf der Bult  
Janusz-Korczak-Allee 12  
30173 Hannover

Hamburg  
14.07.2009

## **Stellungnahme zum Bericht des IQWiG und der Beschlussvorlage des GBA zum Einsatz von langwirksamen Insulinanaloga zur Behandlung des Diabetes mellitus Typ 2**

Sehr geehrter Herr Professor Danne,

mit Schreiben vom 23.6.2009 baten Sie mich um eine Stellungnahme zur Validität der im IQWiG-Bericht verwendeten Methodik. Ich komme dieser Bitte gern nach, beschränke mich aber in der folgenden Diskussion auf diejenigen methodischen Aspekte, die zu einer anderen Bewertung des Zusatznutzens von langwirksamen Insulinanaloga zur Folge haben könnten.

### **Datenbasis**

Die Nutzenbewertung des IQWiG beruht auf den Ergebnissen einer Reihe von mit Hilfe einer Literaturrecherche mit Unterstützung durch die Hersteller identifizierten randomisierten Studien. Ich sehe hier keinen Grund zu der Annahme, dass die Studienauswahl nicht hinreichend vollständig sei, um Schlussfolgerungen zu erlauben.

In den Bericht gehen die Ergebnisse von 18 Studien ein. Die Studien untersuchen eine Reihe unterschiedlicher Endpunkte von Relevanz für die Nutzenbewertung. Allerdings decken sie nicht das gesamte Spektrum der im vorliegenden Fall interessierenden Nutzenendpunkte (die ja auch Schadensendpunkte mit einbeziehen) ab. Die Nutzenbewertung kann deshalb nur vorläufig, nicht abschließend durchgeführt werden. Die sich aus der beschränkten Datenverfügbarkeit ergebende Vorläufigkeit von Sicherheits- und Nutzenbewertungen ist aber kein für diese Fragestellung spezifisches, sondern ein grundsätzliches Problem im gegenwärtigen System der Arzneimittelprüfung, das Herstellerverpflichtungen überwiegend nur bis zur Zulassung vorsieht.

Im vorliegenden Fall liegt der Fokus der nutzenrelevanten klinischen Prüfungen auf der Untersuchung der Hypoglykämien. Da an dieser Stelle allgemein der bedeutsame Zusatznutzen der Insulinanaloga sowohl quantitativ als auch qualitativ gesehen wird, ist diese Fokussierung angemessen.

**Die Datenlage ist damit trotz der aufgeführten prinzipiellen Vorbehalte und Datenlücken insgesamt und insbesondere im Vergleich zu vielen anderen Nutzenbewertungen als gut zu bezeichnen.**

### Prospektivität

In randomisierten Studien ist die strikte Prospektivität, die die nachträgliche Anpassung der Untersuchungsmethoden und Kriterien an die Ergebnisse verhindern soll, in der Regel durch die Festlegungen im Studienprotokoll vor Beginn der Datenerhebung gewährleistet. Das Einfließen der Ergebnisse in zusammenfassende Meta-Analysen, die die vorhandene Evidenz bündeln sollen, kann im Protokoll naturgemäß nicht geregelt sein. Für Meta-Analysen müssen daher eigene Protokolle verfasst werden, die vor Erstellung der Meta-Datenbasis und vor Beginn von Auswertungen erstellt werden sollten. Im Fall von IQWiG-Nutzenbewertungen erfüllt im Idealfall der Berichtsplan diese Funktion.

In der Praxis hat sich jedoch herausgestellt, dass die strikte Form der Prospektivität nicht zu realisieren ist. Weder ist zum Zeitpunkt der Berichtsplanerstellung die Datenqualität noch die technische Auswertbarkeit der Daten hinreichend bekannt noch ist die Durchdringung der Designbesonderheiten und das Verständnis des Zusammenwirkens der Faktoren in hinreichender Tiefe entwickelt, als dass die Analyse bis ins kleinste Detail vorher festgelegt werden könnte. Stattdessen sind im Analyseprozess laufend kleine Entscheidungen zu treffen, die Einfluss auf das quantitative und qualitative Ergebnis haben können. Hier sind transparente Begründung und wissenschaftliche Redlichkeit gefragt.

Die Notwendigkeit, nicht prospektiv festgelegte Analyseschritte durchzuführen, allein devalidiert eine Meta-Analyse indes noch nicht. In diesem Punkt folge ich der Argumentation des IQWiG in der Anhörung zum Berichtsplan. **Die im Verlauf erfolgten Festlegungen sind allerdings mögliche Einfallstore für Verfälschungen und bedürfen deshalb der eingehenden Prüfung.**

### Verfahrensbedingte potentielle Verfälschungsquellen

Im vorliegenden Verfahren sind mir an zwei Stellen im Berichtsplan vorher nicht vorgesehene methodische Vorgehensweisen aufgefallen, die ergebnisrelevant sind und die nicht oder nicht überzeugend begründet sind. Beide benachteiligen die Insulinanaloga, indem sie den Nutzenbeleg erschweren.

- Die **Segmentierung** der Meta-Analysen nach Therapieschemata und Tageszeit der Einnahme sowie teilweise der Studiendauer
- Die ad hoc Ex-post-Festlegung einer zuungunsten der Insulinanaloga **verschobenen Entscheidungsgrenze für unverblindete Studien mit subjektiven Endpunkten** auf der Basis einer einzigen Veröffentlichung.

Beide Vorgehensweisen sind unüblich. Die nachträgliche Einführung dieser beiden neuen Methodiken im laufenden Verfahren weckt Zweifel an der Neutralität der Berichterstatter.

Im einzelnen:

### Segmentierte Analysen

Ziel der vom IQWiG auf der Basis der publizierten Studiendaten, ergänzt durch Abschlussberichte, durchgeführten eigenen Analysen ist eine zusammenfassende Bewertung des Nutzens der untersuchten Insulinanaloga im Vergleich zu Humaninsulin bzw. untereinander. Dazu werden Meta-Analysen aggregierter Daten vorgenommen und in einem zweiten Schritt mit einzelnen von den Herstellern unterbreiteten Meta-Analysen auf Individualdatenbasis abgeglichen.

Zur Methodik studienübergreifender Analysen: **Meta-Analysen** aggregierter wie individueller Daten erlauben zunächst die Ermittlung eines Gesamt-Mittels als **globalem Effektschätzer**. Dieser Schätzer hat grundsätzlich eine **größere Präzision**, der zugehörige Test eine **größere Power** als die entsprechenden Schätzer und Tests aus den Einzelstudien. Zusätzlich erlaubt eine Random-Effect-Meta-Analyse die quantitative Abschätzung und korrekte Berücksichtigung der **Heterogenität** der Studienergebnisse. Will man die Heterogenität aufklären, so erlauben einem **Meta-Regressionen**, den Einfluss von Studiencharakteristika wie im vorliegenden Fall Therapieschemata, Tageszeit der Einnahme, Studiendauer u.ä. zu ermitteln und damit die Heterogenität zu reduzieren/die Power weiter zu erhöhen. **Meta-Analysen mit individuellen Patientendaten** erlauben es darüber hinaus, auch die Einflüsse von Patientencharakteristika zu quantifizieren und damit den unterschiedlichen Patient Mix zu kontrollieren.

Das IQWiG ist bei den Hypoglykämien diesen Weg jedoch nicht gegangen. Stattdessen hat das IQWiG statt einer großen Meta-Analyse pro Vergleich (gegebenenfalls unter Einbeziehung von Kovariaten), jeweils eine Serie kleiner Meta-Analysen durchgeführt, die getrennt für je eine Kombination von Therapie-Schema, Tageszeit der Einnahme sowie Tageszeit und Schwere der Hypoglykämie vorgenommen wurden. Die so vorgenommene Segmentierung der Analyse reduziert die Power der einzelnen Tests erheblich, führt entsprechend zu einer erheblichen Verbreiterung der Konfidenzintervalle und zu einem Mangel an Signifikanzen. Man kann dieses Vorgehen vergleichen mit dem eines Statistikers, der bei einer randomisierten Studie die primäre Analyse auslässt und stattdessen das Analysekollektiv in viele kleine Subgruppen aufteilt und jede dieser Subgruppen einzeln testet. Dieser Statistiker wird kaum ein signifikantes Ergebnis erzielen.

Es sei darauf hingewiesen, dass die in der Fachliteratur vorzufindenden Meta-Analysen zu Insulinpräparaten den Weg der Segmentierung nicht gegangen sind. Entsprechend ist in diesen Meta-Analysen die Power größer und ein Vorteil von Glargin deutlicher sichtbar als in der IQWiG-Analyse. Das IQWiG gibt leider keine Begründung für das abweichende Vorgehen.

Selbst wenn man zu der Überzeugung gelangen sollte, dass sich die Glargin-Effekte zwischen den Therapieschemata/nach Tageszeit so stark unterscheiden, dass eine zusammenfassende Bewertung nicht vertretbar ist, wäre die richtige Auswertungsmethode nicht eine segmentierte Analyse, sondern ein Meta-Regressionsmodell. Hier besteht wiederum eine Parallele zur Auswertung einzelner randomisierter Studien: eine korrekte Subgruppenanalyse betrachtet nicht etwa die Subgruppen einzeln (dann wären die kleineren Subgruppen benachteiligt), sondern besteht aus stratifizierten Analysen bzw. Interaktionstests, die deutlich mehr Power als die segmentierten Tests haben.

Die beschriebenen methodischen Wege wurden vom IQWiG nicht einmal versuchsweise besprochen. Man mag hier die größere Komplexität der Analysen als Hinderungsgrund vermuten. Was aber vollends unverständlich ist, ist, dass es keine Versuche des IQWiG gibt, die segmentierten Analysen in irgendeiner Form aufeinander zu beziehen. Das IQWiG zählt lediglich Signifikanzen und würdigt jede einzelne überkritisch, was bei suboptimaler Power problematisch ist. Das IQWiG scheint den Begriff der supportiven Evidenz nicht zu kennen. Dieses Argument wird in Zulassungsverfahren verwendet, wenn mehrere schwer miteinander zu kombinierende Studien vorliegen, von denen nur eine im Hauptendpunkt signifikant ist und die anderen zwar in die richtige Richtung weisen, aber den Hauptendpunkt verfehlt haben. In diesem Fall kann der Gesamteindruck als Ausgleich für eine fehlende zweite signifikante Studie akzeptiert werden. Eine analoge Betrachtungsweise würde im vorliegenden Fall bereits ohne formale Durchführung einer großen Meta-Analyse auf supportive Evidenz für einen Vorteil von Glargin gegenüber NPH Insulin bei der Hypoglykämie-Häufigkeit erkennen.

### **Entscheidungsgrenze**

Im Ergebnis der segmentierten Analysen erkennt das IQWiG nur an einer Stelle einen Beleg, an anderen Stellen nur Hinweise (die eigentlich als supportive Evidenz gewertet werden müssten). Es fällt in einigen Fällen schwer, diesen Einzelbewertungen zu folgen. Besonders schwer nachzuvollziehen ist die IQWiG-Argumentation bei der Studie 4016. Hier zeigt sich ein deutlicher Effekt zugunsten von Glargin bei den symptomatischen durch Blutzuckermessung bestätigten nächtlichen Hypoglykämien. Das obere Konfidenzintervall selbst der segmentierten Meta-Analyse liegt hier deutlich unter 1. An dieser Stelle hat das IQWiG im nachhinein eine neue Entscheidungsgrenze von 0,75 unter Berufung auf eine Meta-analyse von Wood et al (2007) eingeführt, die sich bemüht, den mittleren Effekt mangelnder Verblindung auf die Effektschätzer für subjektive Endpunkte ganz allgemein zu beziffern. Im vorliegenden Fall liegt der Wert von 0,75 noch innerhalb des Konfidenzbereichs. Damit entfällt der sonst zu konstatierende Beleg.

Hierzu ist anzumerken, dass eine solche Ad-hoc-Grenzziehung in voller Kenntnis des zu bewertenden Ergebnisses äußerst problematisch ist. Zudem ist der konkrete Grenzwert methodisch zweifelhaft. Die Arbeit von Wood et al konstatiert ausdrücklich die Heterogenität in den gefundenen Effekten. Zudem dürfte ein durch Selbstmessung untermauertes symptomatisches Ereignis wohl kaum mit der subjektiven Bewertung einer Befindlichkeit in einen Topf geworfen werden. Man darf im Übrigen gespannt sein, ob das IQWiG diese angeblich universelle Grenze auch in Zukunft bei allen unverblindet ermittelten Nutzenendpunkten zur Anwendung bringen wird. Es stünde damit weltweit ziemlich allein da. Bisher werden weltweit Untersuchungen im spezifischen Anwendungsfeld zur Bezifferung eines Bias für erforderlich gehalten.

### **Zusammenfassung**

Die in Frage stehende Beschlussvorlage für den GBA basiert auf einer hinreichenden Datenbasis zur Bewertung des Zusatznutzens der Insulinanaloga im Hinblick auf die Hypoglykämien. Die methodische Herangehensweise des IQWiG ist jedoch an zwei wichtigen Stellen im internationalen Vergleich unüblich und nicht oder nicht ausreichend begründet und wohl auch nicht begründbar: Die vorgenommene Segmentierung reduziert die Power in nicht akzeptabler Weise. Die nachträgliche Einführung einer neuen Nachweisgrenze für unverblindete Studien ist nicht zu rechtfertigen. Der Bericht wertet aufgrund dieser methodischen Unzulänglichkeiten eigentlich vorhandene Belege ungerechtfertigt zu Hinweisen ab und ignoriert im übrigen die in größerem Umfang vorhandene supportive

Evidenz. Durch diese Vorgehensweise entsteht eine Spannung zwischen dem Gesamtbild, das sich aus den zahlreichen Analysen ergibt, und dem formalen Ergebnis des IQWiG. Aufgrund des vorliegenden Materials bin ich geneigt, auch ohne eine an sich wünschenswerte zusammenfassende Meta-Analyse auf den hinreichenden Nachweis eines relevanten Zusatznutzens zu erkennen.

Auf die Bewertung der klinischen Relevanz dieses Zusatznutzens für den Patienten möchte ich hier verzichten. Als Methodiker obliegt es mir, eine Einschätzung zur Evidenzlage und zur Plausibilität von Effektschätzungen abzugeben. Die Bewertung der klinischen Relevanz möchte ich Ihrem Urteil überlassen.

Mit freundlichen Grüßen

Prof. Dr. Karl Wegscheider