Tim Seidinger · Robert Messerle · David Scheller-Kreinsen

Punktlandung Pflegegrad?

Analysen zur Reform der Pflegebegutachtung

WIdO e-Paper 14 (2026)

WIdO | Wissenschaftliches Institut der AOK

Impressum

Die vorliegende Publikation ist ein Beitrag des
Wissenschaftlichen Instituts der AOK (WIdO).

Punktlandung Pflegegrad?
Analysen zur Reform der Pflegebegutachtung
WIdO e-Paper 14 (2026)

Berlin, im September 2026

Tim Seidinger, Robert Messerle, David Scheller-Kreinsen

Wissenschaftliches Institut der AOK (WIdO)
im AOK-Bundesverband eGbR – Arbeitsgemeinschaft
von Körperschaften des öffentlichen Rechts
Rosenthaler Str. 31, 10178 Berlin

Geschäftsführender Vorstand:
Dr. Carola Reimann (Vorsitzende)
Jens Martin Hoyer (stellv. Vorsitzender)
https://www.aok.de/pp/impressum/
Registergericht Berlin (Charlottenburg), GsR 634 B

Aufsichtsbehörde:
Senatsverwaltung für Wissenschaft, Gesundheit und
Pflege (SenWGP) Oranienstraße 106, 10969 Berlin

Redaktionelle Bearbeitung: Tatjana Schwarz, Melanie Hoberg
Satz: Tatjana Schwarz
Grafik: AOK Medien GmbH
Titelfoto: iStock.com/Jacob Wackerhausen

Copyright: © 2026 Wissenschaftliches Institut der AOK (WIdO).
Creative Commons: Namensnennung - 4.0 International (CC BY 4.0)
https://creativecom-mons.org/licenses/by/4.0/deed.de
Der Inhalt dieser Publikation darf demnach uneingeschränkt und in allen Formen genutzt, geteilt und wiedergegeben werden, solange der Urheber und die Quelle angemessen angegeben werden.

DOI: https://doi.org/10.4126/FRL01-006529343

E-Mail: wido@wido.bv.aok.de
Internet: www.wido.de

Inhalt

Zusammenfassung/Abstract **4**

1 Einleitung **5**

2 Hintergrund **6**

2.1 Das Begutachtungsinstrument 6

2.2 Punktverteilung des NBA an den Schwellenwerten im Jahr 2024 8

2.3 Internationale Evidenz und Theorie 10

3 Methodik **12**

4 Ergebnisse **14**

4.1 Empirischer Nachweis von Punktwerthäufungen/Bunching 14

4.2 Identifikation mechanischer systembedingter Effekte – Die Rolle des Begutachtungsinstruments 14

4.3 Quantifizierung des anreiz- und verhaltensbedingten Bunchings 16

4.4 Potenzielle finanzielle Auswirkungen 17

5 Diskussion **19**

6 Fazit **22**

Anhang **23**

Abbildungsverzeichnis **27**

Tabellenverzeichnis **28**

Literaturverzeichnis **29**

Zusammenfassung/Abstract

Zusammenfassung

Zur finanziellen Stabilisierung der Sozialen Pflegeversicherung (SPV) sieht der Referentenentwurf zum Pflegeneuordnungsgesetz (PNOG, Referentenentwurf vom 5. Juni 2026) höhere Punkteschwellen bei der Einstufung in Pflegegrade vor. Offizielle Schätzungen und Gutachten gehen von erheblichen, teils zweistelligen Milliarden-Einsparungen bis 2040 aus. Diese Berechnungen berücksichtigen jedoch meist nicht, dass sich die Anreize und damit das Verhalten der beteiligten Akteure im Begutachtungsverfahren an neue Regeln anpassen können.

Der vorliegende Beitrag untersucht solche verhaltensökonomischen Anpassungseffekte (Bunching) an den Schwellenwerten für die Pflegegrade. Dabei wird analysiert, wie häufig Bewertungen nahe an den jeweiligen Schwellenwerten auftreten, inwieweit diese Häufungen systembedingt sind und welche finanziellen Auswirkungen daraus zu erwarten sind. Die Ergebnisse belegen, dass reine Ceteris-Paribus-Betrachtungen, die von unverändertem Anreiz und Verhalten ausgehen, unzureichend sind. Auch an den neuen Schwellenwerten sind Anpassungsreaktionen zu erwarten. Daher ist davon auszugehen, dass die Häufungen nicht an den ursprünglichen Grenzen bleiben, sondern sich an die neuen Grenzen verlagern. Dadurch könnte das tatsächliche Einsparpotenzial der Reform in der mittelfristigen Perspektive um bis zu 50 Prozent bzw. mehrere Milliarden Euro geringer ausfallen als bislang angenommen. Dieser Effekt sollte bei der fiskalischen Konsolidierungsplanung der SPV berücksichtigt werden. Vor dem Hintergrund dieser Ergebnisse erscheint es außerdem ratsam, das Instrument der Pflegebegutachtung insgesamt zu überarbeiten, statt einzelne Aspekte herauszugreifen.

Abstract

To strengthen the financial sustainability of Germany's Social Long-Term Care Insurance (SPV), the draft Long-Term Care Reform Act (PNOG, draft of 5 June 2026) proposes raising the score thresholds used for care level assessments. Official estimates and expert reports project substantial savings through 2040, in some cases amounting to tens of billions of euros. However, these projections generally fail to take into account that incentives – and thus the behaviour of the stakeholders involved in the assessment process – may adapt to new rules.

This paper examines such behavioural adjustment effects (bunching) around the thresholds that determine care grade eligibility. It analyses how frequently assessment scores cluster near these thresholds, to what extent such concentrations are inherent to the assessment system, and what fiscal implications may result from them.

The findings demonstrate that purely ceteris paribus projections, which assume unchanged incentives and behaviour, are insufficient. Behavioural responses are likely to occur under the new thresholds as well. As a result, score clustering can be expected to shift toward the new cut-off points rather than remaining at the old thresholds. This may reduce the reform's actual savings potential, in the medium term, by up to 50 per cent, equivalent to several billion euros, compared with current estimates. These effects should therefore be taken into account when assessing the fiscal consolidation potential of the SPV. In light of these findings, it also seems advisable to review the care assessment process as a whole, rather than focusing on individual aspects.

1 Einleitung

Die Finanzlage der Sozialen Pflegeversicherung (SPV) hat sich in den vergangenen Jahren spürbar zugespitzt. Die Problematik wird inzwischen auch vom Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (SVR Wirtschaft 2026) aufgegriffen. Bei Ausgaben von 74 Milliarden Euro im Jahr 2025 (GKV-Spitzenverband 2026) wird für 2027 ein Defizit von 7,5 Milliarden Euro erwartet (BMG 2026). Zur Bewältigung dieser Herausforderung werden seit Jahren unterschiedliche Vorschläge diskutiert. Diese reichen von Strukturänderungen bis hin zu kurzfristigen Kostensenkungsmaßnahmen.

Zuletzt ist dabei auch der Zugang zur Pflegeversicherung in den Fokus gerückt. Hintergrund ist unter anderem die Beobachtung, dass ein wesentlicher Treiber für die steigende Pflegeinzidenz und -prävalenz der 2017 eingeführte neue Pflegebedürftigkeitsbegriff ist (Nolting et al. 2026; Rothgang et al. 2025). Dessen erklärtes Ziel war es, den Zugang zu Pflegeleistungen zu erleichtern und die Inanspruchnahme von Leistungen der Pflegeversicherung auszuweiten. In diesem Zuge wurde das Neue Begutachtungsassessment (NBA) etabliert, das den Zugang zur Inanspruchnahme der SPV reguliert. Das NBA bewertet den Pflegebedarf von Versicherten auf Basis eines strukturierten Interviews und eines standardisierten Fragebogens. Die so ermittelte Punktzahl dient als Basis für die Einstufung in einen der fünf Pflegegrade. Um den jeweiligen Pflegegrad zu erhalten, muss eine bestimmte Punktgrenze – der jeweilige Schwellenwert – überschritten werden.

Im Rahmen der Gesetzgebung zum Pflegeneuordnungsgesetz (PNOG, Referentenentwurf vom 5. Juni 2026) wird vorgeschlagen, die Schwellenwerte und Modulschwellen anzuheben, um die Ausgabendynamik zu begrenzen. Das Bundesministerium für Gesundheit (BMG) beziffert die erwarteten Finanzwirkungen aus der Anpassung der Begutachtungssystematik im PNOG für das Jahr 2027 mit 1,3 Milliarden Euro. Für das Jahr 2030 beziffert das BMG die Finanzwirkungen sogar auf 4,2 Milliarden Euro, womit diese Maßnahme den größten Einzelposten des Gesetzes darstellt. Auch der Bericht des IGES im Auftrag des Spitzenverbandes Bund der Krankenkassen (GKV-Spitzenverband) (Nolting et al. 2026) geht von einer hohen Wirkung auf administrative Pflegebedürftigkeit und SPV-Ausgaben aus. Die Fach-Arbeitsgruppen „Versorgung“ und „Finanzierung“ im Zukunftspakt Pflege (Zukunftspakt Pflege 2025) haben das mittelfristige Einsparpotenzial auf 0,1 bis 0,2 Beitragssatzpunkte (ca. 2 bis 4 Milliarden Euro) beziffert. Im Frühjahresgutachten des SVR Wirtschaft (SVR Wirtschaft 2026) wird für 2040 sogar eine Einsparung von 0,7 Beitragssatzpunkten (ca. 14 Milliarden Euro) projiziert.

Jedoch stellt sich die Frage, ob derartige Projektionen und Schätzungen, die von unverändertem Anreiz und Verhalten ausgehen (*Ceteris-Paribus*-Fortschreibungen), die tatsächlichen Effekte einer Schwellenwertanpassung hinreichend abbilden. Verhaltensökonomische Erkenntnisse legen nahe, dass Schwellenwerte nicht nur passiv Verteilungen strukturieren, sondern aktiv Anreize setzen. Vor diesem Hintergrund untersuchen wir erstens (1), ob es Hinweise auf Anpassungseffekte gibt. Zweitens (2) analysieren wir, ob diese Effekte durch die Begutachtungssystematik selbst hervorgerufen werden. Drittens (3) modellieren wir mittels unterschiedlicher Ansätze das Ausmaß möglicher Anpassungseffekte.

Die im Folgenden dargestellten Analysen liefern deutliche Hinweise auf anreiz- und verhaltensbedingte Anpassungseffekte. Im Einklang mit internationalen Studien (Hernández-Pizarro et al. 2020; Ishihara und Takahashi 2022) legen die Ergebnisse nahe, dass reine Ceteris-Paribus-Betrachtungen die Wirkungen einer Schwellenwertanpassung nur unzureichend erfassen. Stattdessen ist zu erwarten, dass sich bestehende Anreize und Verhaltensanpassungen auf die neuen Schwellenwerte verlagern und weiterhin die Verteilung der Punktwerte beeinflussen. Durch diese Reaktionen könnte es zu einer Reduktion der erwarteten Einsparpotenziale von bis zu 50 Prozent kommen, was mittelfristig Auswirkungen in Milliardenhöhe hätte. Entsprechende Schätzungen sollten bei der möglichen Entlastungswirkung für die SPV berücksichtigt werden.

2 Hintergrund

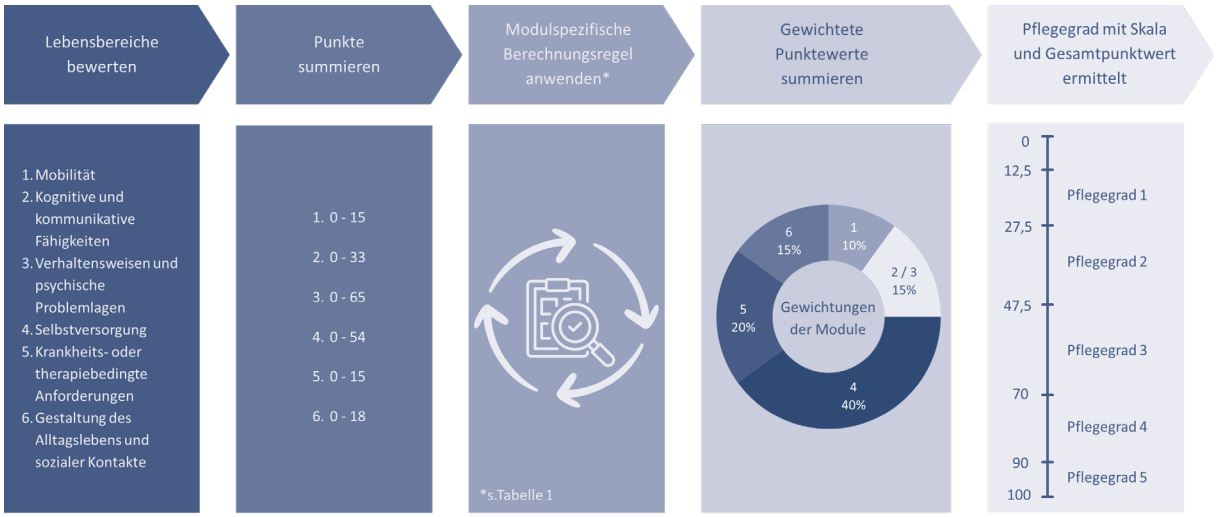
2.1 Das Begutachtungsinstrument

Um Leistungen der SPV in Anspruch nehmen zu können, ist ein Nachweis über die Pflegebedürftigkeit erforderlich. Dieser wird durch eine Begutachtung festgestellt. Versicherte der Pflegeversicherung können einen Antrag auf diese Begutachtung jederzeit formlos per Anruf, E-Mail oder Brief bei ihrer zuständigen Pflegekasse stellen. Anschließend beauftragt die Pflegekasse den regionalen Medizinischen Dienst (MD). Dieser entsendet im Regelfall innerhalb von vier Wochen eine qualifizierte Person zur Begutachtung. Dabei handelt es sich meist um Pflegefachkräfte oder Ärztinnen und Ärzte. Auf Basis des standardisierten Begutachtungsinstruments erstellen sie eine fachliche Einschätzung in Form eines schriftlichen Gutachtens. Im Folgenden werden die zentralen Komponenten dieses Verfahrens kurz dargestellt. Eine ausführliche Beschreibung kann z. B. Jürchott et al. (2026) oder den Richtlinien des MD entnommen werden.

Seit dem Zweiten Pflegestärkungsgesetz (PSG II) aus dem Jahr 2017 erfolgt die Feststellung der Pflegebedürftigkeit nach § 15 SGB XI über das „Neue Begutachtungsassessment“ (NBA). Die grundsätzliche Bewertungslogik ist in Abbildung 1 dargestellt. Das NBA misst den Grad der individuellen Selbstständigkeit anhand von sechs zentralen Lebensbereichen (Modulen), welche somatische, psychische und kognitive Kapazitäten abbilden. Innerhalb der Module 1 bis 4 sowie Modul 6 bewerten die Begutachtenden die Selbstständigkeit auf einer vierstufigen Skala; Modul 5 erfasst die Häufigkeit und den Umfang spezifischer krankheits- oder therapiebedingter Hilfebedarfe.

Die für diese Module auf Basis eines Fragebogens ermittelten Rohpunkte werden über modulspezifische Berechnungsregeln in gewichtete Punkte überführt (s. Tabelle 1). Mit einer Gewichtung von 40 Prozent ist das Modul Selbstversorgung besonders ausschlaggebend für den Gesamtscore. Eine Besonderheit gibt es bei den Modulen 2 und 3: Hier wird bei der Berechnung jeweils nur das Modul berücksichtigt, das einen höher gewichteten Punktwert hat. Die Summe aller gewichteten Einzelwerte ergibt einen Gesamtscore (bis zu 100 Punkte), der die finale Basis für die Einstufung in einen der fünf Pflegegrade bildet. Aufgrund der prozentualen Modulgewichtungen verändert sich der Gesamtscore in Schritten von jeweils 1,25 Punkten.

Abbildung 1: Bewertungslogik nach dem Neuen Begutachtungsassessment (NBA)



Quelle: eigene Darstellung nach § 15 SGB XI

© WIdO 2026

Tabelle 1: Überblick über rohe und gewichtete Punkte

Module	Grad der Beeinträchtigung					
	Punkte-Typ	keine (0)	geringe (1)	erhebliche (2)	schwere (3)	vollständige (4)
1. Mobilität (5 Items)	roh	0 - 1	2 - 3	4 - 5 *[4 - 6]*	6 - 9 *[7 - 9]*	10 - 15
	gewichtet	0	2,5	5	7,5	10
2. kognitive und kommunikative Fähigkeiten (11 Items)	roh	0 - 1	2 - 5	6 - 10	11 - 16	17 - 33
	gewichtet	0	3,75	7,5	11,25	15
3. Verhaltensweisen und psychische Problemlagen (13 Items)	roh	0	1 - 2	3 - 4	5 - 6	7 - 65
	gewichtet	0	3,75	7,5	11,25	15
4. Selbstversorgung (13 Items)	roh	0 - 2 *[0 - 3]*	3 - 7 *[4 - 9]*	8 - 18 *[10 - 24]*	19 - 36 *[25 - 39]*	37 - 54 *[40+]*
	gewichtet	0	10	20	30	40
5. krankheits- oder therapiebedingte Anforderungen (16 Items)	roh	0	1	2 - 3	4 - 5	6 - 15
	gewichtet	0	5	10	15	20
6. Gestaltung des Alltags und sozialer Kontakte (6 Items)	roh	0 *[0 - 1]*	1 - 3 *[2 - 3]*	4 - 6	7 - 11	12 - 18
	gewichtet	0	3,75	7,5	11,25	15

In eckigen Klammern abweichender Vorschlag des Expertenbeirats aus dem Jahr 2013.

Quelle: eigene Darstellung nach § 15 SGB XI

© WIdO 2026

Der aktuelle Reformvorschlag im PNOG sieht vor, die Zugangsschwellen für Neuanträge und Höherstufungen in den Pflegegraden 1 bis 3 anzuheben, um das System finanziell zu stabilisieren und die Einstufungskriterien an die ursprünglichen Empfehlungen des Expertenbeirats aus dem Jahr 2013 anzupassen (BMG 2013). Die konkrete Verschiebung der Intervalle für die gewichtete Gesamtpunktzahl ist in Tabelle 2 dargestellt.

Die Unterschiede betreffen die Pflegegrade 1 bis 3. Konkret sind bei jedem Schwellenwert zwei Punktwerte betroffen: 12,5 und 13,75 führen nicht mehr zum Anspruch auf Pflegegrad 1. Bei den Werten 27,5 und 28,75 erfolgt eine Einstufung in Pflegegrad 1 statt 2, und bei 47,5 bzw. 48,75 Punkten wird Pflegegrad 2 statt 3 zugewiesen. Zusätzlich ist vorgesehen, die Schwellenwerte innerhalb der Module 1, 4 und 6 entsprechend den ursprünglichen Empfehlungen anzupassen. Dabei geht es um die Schwellenwerte zur Umrechnung der Rohpunkte in gewichtete Punkte. Beispielsweise werden in Modul 4 nun 25 statt 19 Rohpunkte notwendig, um 30 gewichtete Punkte zu erhalten (s. Tabelle 1).[1]

[1] Die folgenden Modellierungen konzentrieren sich aus Gründen der Datenverfügbarkeit auf die Anpassung der Pflegegrad-Schwellenwerte außerhalb der Module. Es wird angenommen, dass die daraus gewonnenen Erkenntnisse übertragbar sind, da Häufungen auch in der Verteilung der Rohpunkte innerhalb der Module zu beobachten sind.

Tabelle 2: Übersicht über die Schwellenwerte der gewichteten Gesamtpunktzahl

	Kein PG	PG 1	PG 2	PG 3	PG 4	PG 5
gesetzliche Regelung (Stand August 2026	0 - <12,5	12,5 - <27	27 - <47,5	47,5 - <70	70 - <90	90 - 100
Vorschlag des Expertenbeirats aus dem Jahr 2013	0 - <15	15 - <30	30 - <50	50 - <70	70 - <90	90 - 100

Quelle: eigene Darstellung © WIdO 2026

2.2 Punktverteilung des NBA an den Schwellenwerten im Jahr 2024

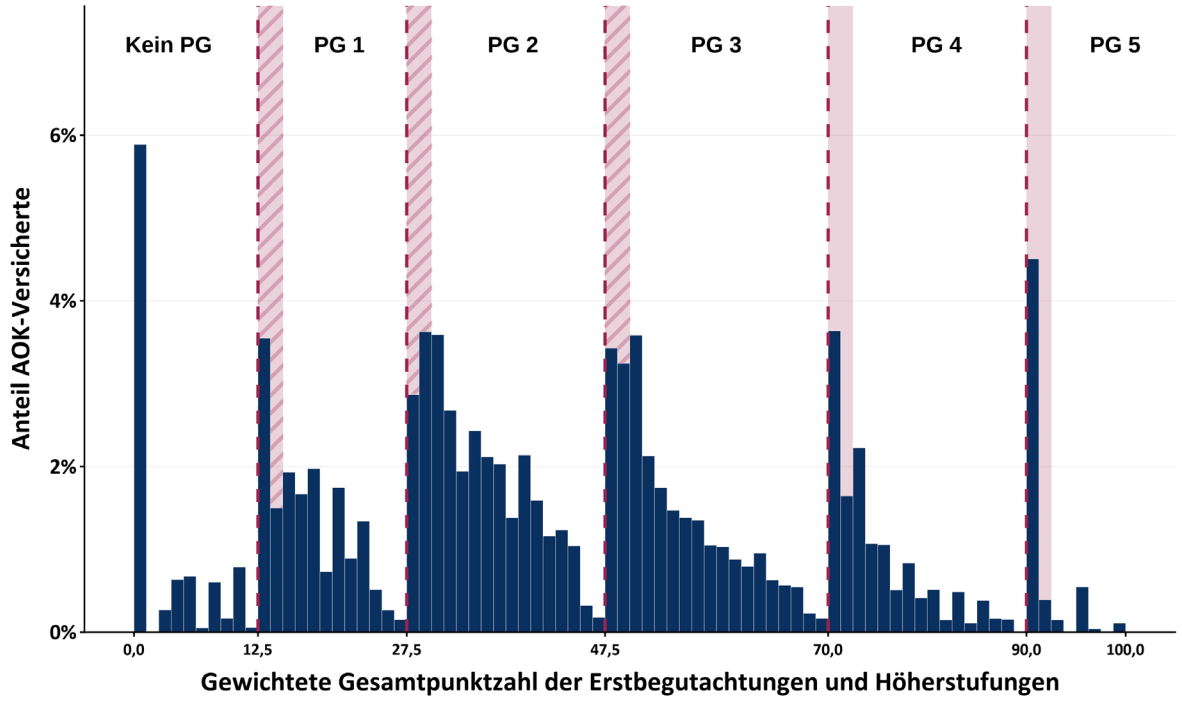
Abbildung 2 zeigt die empirische Verteilung der Gesamtpunktzahl aller Erstbegutachtungen und Höherstufungen von AOK-Versicherten im Jahr 2024. Die vertikalen roten Abschnitte markieren einen Bereich von 2,5 Punkten über den bestehenden Schwellenwerten. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen (PG 1 bis PG 3). Die genauen Fallzahlen in diesen Grenzbereichen sind in Tabelle 3 aufgeführt. Insgesamt fallen 20,5 Prozent der Erstbegutachtungen und 15,5 Prozent der Höherstufungen in diese Bereiche.

Tabelle 3: Übersicht über mögliche Einstufungsänderungen

Anpassung durch Schwellenwertänderung	Erstbegutachtung	Anteil an allen Erstbeg. (*n* = 573.461)	Anteil innerhalb des Pflegegrades	Höherstufung	Anteil an allen Höherst. (*n* = 539.247)	Anteil innerhalb des Pflegegrades
	(1)	(2)	(3)	(4)	(5)	(6)
kein PG anstatt PG1	47.303	8,8 %	32,2 %	7.318	1,3 %	24,8 %
PG1 anstatt PG2	43.389	8,0 %	22,9 %	27.510	4,8 %	18,9 %
PG2 anstatt PG3	19.902	3,7 %	27,9 %	53.970	9,4 %	25,5 %
insgesamt	**110.594**	**20,5 %**		**88.798**	**15,5 %**	

Quelle: MD-Begutachtungsdaten 2024 für AOK-Versicherte. Anpassung der Schwellenwerte nur auf Ebene der Pflegegrade. © WIdO 2026

Abbildung 2: Verteilung der gewichteten Punkte bei Erstbegutachtungen und Höherstufungen von AOK-Versicherten



Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen.

Quelle: MD-Begutachtungsdaten 2024 für AOK-Versicherte

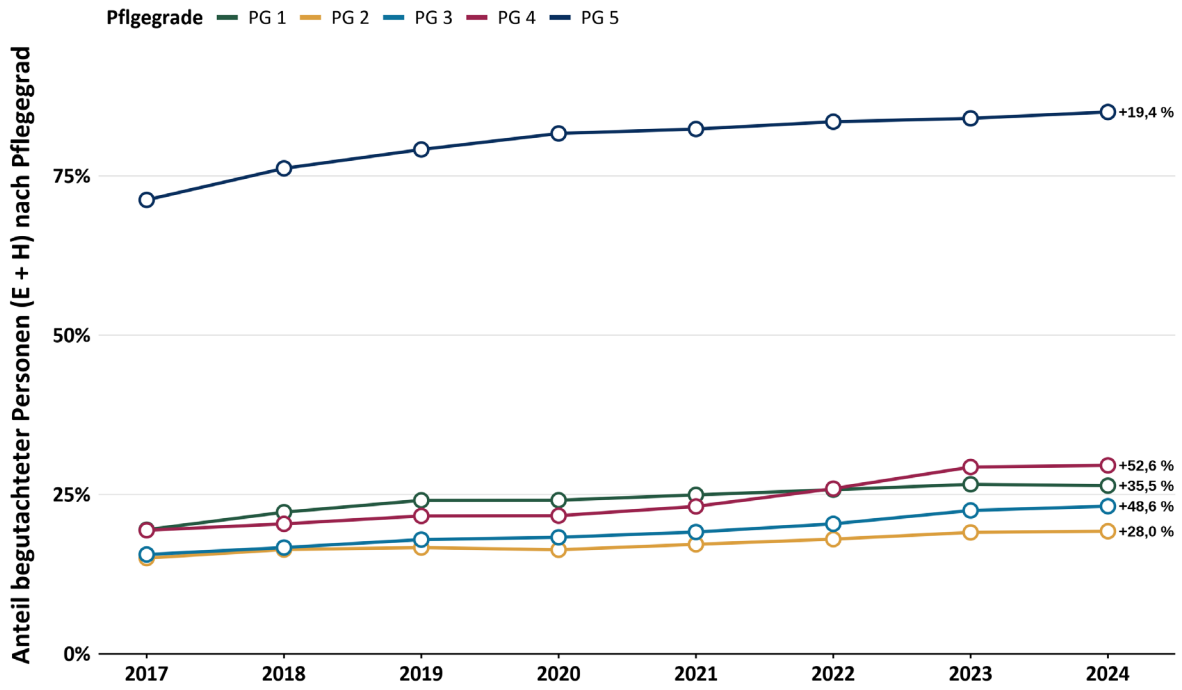
© WIdO 2026

Zudem verdeutlicht Abbildung 2 eine auffällige Häufung von Fallzahlen direkt oberhalb der gesetzlichen Schwellenwerte, während sich knapp darunter kaum Beobachtungen finden. Dieses Phänomen wird in der internationalen Literatur als „Bunching" bezeichnet (s. z. B. Kleven (2016)). Dieses Bunching spiegelt sich auch in den Spalten (3) und (6) von Tabelle 3 wider und illustriert, warum eine Schwellenwertanpassung *prima vista* einen so großen Anteil an Pflegebedürftigen beträfe und mit so hohen Einsparerwartungen verbunden ist.

Eine zeitliche Betrachtung seit Einführung des neuen Systems zeigt, dass das Ausmaß des Bunchings über die Jahre zugenommen hat. In Abbildung 7 im Anhang ist die Verteilung der gewichteten Punkte von Erstbegutachtungen und Höherstufungen für die Jahre 2017 und 2024 dargestellt. Besonders deutlich ist die gesteigerte Häufung am Schwellenwert zu Pflegegrad 1 erkennbar. Aber auch oberhalb der anderen Schwellenwerte gab es eine Zunahme. Abbildung 3 stellt für die Jahre 2017 bis 2024 den prozentualen Anteil der Pflegebedürftigen eines jeweiligen Pflegegrades dar, die maximal 2,5 Punkte über dem Schwellenwert liegen. Für die Pflegegrade 1 bis 3 entspricht dies dem von einer Schwellenwerterhöhung gemäß PNOG betroffenen Anteil. Zwischen 2017 und 2024 nahm dieser für die Pflegegrade 1 bis 3 zwischen 27,2 Prozent und 47,9 Prozent zu. Diese Entwicklung lässt zwei Rückschlüsse zu: Entweder haben sich die Charakteristika der Begutachteten über die Jahre systematisch verändert, und/oder die am Begutachtungsprozess beteiligten Akteure reagieren mit zunehmender Vertrautheit des Instruments anders auf die Anreize im Begutachtungsprozess.

Für den Pflegegrad 5 ist ein deutlicher Niveauunterschied in der Häufung direkt oberhalb der Schwellenwerte zu erkennen. Während die Pflegegrade 1 bis 4 im Jahr 2024 Werte zwischen 20 und 40 Prozent aufweisen, liegt der Anteil bei den Personen mit PG 5 bei fast 80 Prozent. Dies dürfte methodisch dadurch mitbedingt sein, dass mit Erreichen des höchsten Pflegegrades eine weitere Differenzierung nach oben hin keine weiteren Auswirkungen zeigt und das Bewertungsspektrum im Wesentlichen ausgereizt ist.

Abbildung 3: Prozentualer Anteil der Pflegebedürftigen, die maximal 2,5 Punkte über dem Schwellenwert liegen, gemessen an allen Pflegebedürftigen innerhalb des jeweiligen Pflegegrades



Die Werte rechts geben die prozentuale Änderung zwischen 2017 und 2024 wieder.

Quelle: eigene Berechnung auf Basis von MD-Begutachtungsdaten von 2017 bis 2024 für AOK-Versicherte

© WIdO 2026

2.3 Internationale Evidenz und Theorie

Schwellenwerte beeinflussen die Anreize und damit das Verhalten von Individuen. Dies ist in der internationalen Literatur klar belegt. Kleven (2016) gibt hierzu eine Übersicht. Ursprünglich wurde das Bunching im steuerlichen Kontext untersucht. Seitdem wurde die Literatur auf zahlreiche Bereiche ausgeweitet. Dazu gehören beispielsweise die Bewertung von Testergebnissen (Diamond und Persson 2016; Dee et al. 2019) oder die Beurteilung von Anträgen im Rahmen einer Invalidenversicherung (Garcia-Gomez et al. 2025). Im Kontext der Pflegeversicherung wurden ähnliche Phänomene bereits für Spanien (Hernández-Pizarro et al. 2020) und Japan (Ishihara und Takahashi 2022) dokumentiert und analysiert.

Zu den möglichen Gründen für das beobachtete Muster in Deutschland gehören z. B. (1) Selektionseffekte: der Fragebogen inklusive des Bewertungsverfahrens zur Feststellung der Pflegebedürftigkeit ist frei zugänglich und kann auf verschiedenen Webseiten ohne viel Aufwand ausgefüllt werden. Dementsprechend werden wahrscheinlich primär Personen einen Pflegegrad beantragen, die über der Schwelle zu Pflegegrad 1 liegen. Analog kann dies auf die Höherstufungen übertragen werden. Personen werden diese tendenziell eher beantragen, wenn sie sich (kurz) über einem höheren Schwellenwert einschätzen. (2) Demgegenüber steht die Möglichkeit, dass das Bunching durch angepasste Anreize und angepasstes Verhalten zustande kommt. Die Motive hierfür können unterschiedlich sein. Eine Möglichkeit ist, dass das Begutachtungsinstrument und die daran geknüpften Leistungsansprüche die Bedarfe der Pflegebedürftigen nicht richtig abbildet. Personen, die kurz unter oder kurz über einem Schwellenwert liegen, haben wahrscheinlich sehr ähnliche Bedarfe, weisen allerdings eine Differenz in den Leistungsansprüchen von bis zu 800 Euro monatlich in den Pflegesachleistungen auf. Entsprechend könnten die Höherstufungen durch die Begutachtenden als Ausgleich für ein Bewertungsinstrument erfolgen, das den tatsächlichen Versorgungsaufwand unvollständig abbildet (Roy 2025). Auf den ersten Blick kann ein solcher Befund überraschen. Schließlich liegt die Begutachtung in Deutschland beim MD, einer Institution, die unabhängig von dem Ergebnis der Begutachtung vergütet wird. Erfahrungen aus Spanien

zeigen jedoch, dass dies allein Verhaltensanpassungen nicht verhindert. Auch dort haben die bewertenden Fachkräfte keinen finanziellen Vorteil durch höhere Einstufungen und stufen Antragstellende dennoch überproportional häufig oberhalb der Schwellen ein. Für Spanien wird dieses Ergebnis mit prosozialem Verhalten der Begutachtenden erklärt (Hernández-Pizarro et al. 2020). Alternativ könnte das Verhalten auch durch die Absicht begründet sein, die Anzahl an Widerspruchsgutachten und erneuter Begutachtungen nach kurzer Zeit zu vermeiden, da innerhalb eines Monats Widerspruch gegen die Begutachtung eingelegt werden kann. Für die AOK-Versicherten zeigt unsere Analyse der Begutachtungsdaten, dass die Widerspruchsquote innerhalb der Pflegegrade mit zunehmender Punktzahl ansteigt und kurz vor dem Schwellenwert ihr Maximum erreicht. Anschließend fällt sie um mehrere Prozentpunkte ab (s. Abbildung 10 im Anhang). Außerdem gibt es keine Karenzzeit bis ein neuer Antrag auf Pflegebegutachtung gestellt werden darf. Bei knappen zeitlichen Ressourcen seitens der Begutachtenden, könnte dies ein Anreiz sein, Personen, die knapp unterhalb einer Schwelle liegen, hochzustufen.

Zur Quantifizierung des Bunching-Umfangs gibt es zahlreiche unterschiedliche Methoden (Saez 2010; Chetty et al. 2011; Kleven 2016). Sowohl Hernández-Pizarro et al. (2020) als auch Ishihara und Takahashi (2022) schätzen das finanzielle Ausmaß der verhaltensbedingten Verzerrung mit Hilfe einer kontrafaktischen Verteilung (vgl. Abschnitt 3). Mit dieser Methode kommen Hernández-Pizarro et al. (2020) auf einen Anteil von bis zu 40 Prozent der Beobachtungen an einem Schwellenwert, die durch Verhaltensanpassungen verändert wurden. Ishihara und Takahashi (2022) schätzen, dass sich in Japan als Folge des Bunchings die durchschnittlichen monatlichen Ausgaben pro Pflegebedürftigen um bis zu 15,5 Prozent erhöht haben.

3 Methodik

Um die beobachteten Effekte zu analysieren und um bewerten zu können, ob in der Pflegebegutachtung ähnliche Verhältnisse wie beispielsweise in Japan oder Spanien vorliegen, folgt die Untersuchung einem dreistufigen Aufbau:

1. Prüfung der Existenz des Bunchings: Analytischer Nachweis von Punktwerthäufungen (Dichtetest)

Zunächst wird empirisch geprüft, ob die beobachteten und beschriebenen Häufungen auch im wissenschaftlichen Sinne als signifikante Häufungen oberhalb der Schwellenwerte identifiziert werden können. Ergänzend zur visuellen Evidenz erfolgt die analytische Bestimmung über den etablierten Dichtetest nach Cattaneo et al. (2018) (in Anlehnung an McCrary (2008)). Das Verfahren schätzt die Punktdichte links und rechts der Schwellenwerte separat ein, um statistische Sprünge an den Grenzen zu identifizieren. Während eine unbeeinflusste Verteilung stetig verläuft, indiziert ein signifikanter Dichtesprung an der Schwelle systematische Fallverschiebungen nach oben.

2. Identifikation der Ursachen: Identifikation systembedingter Effekte (Datensimulation)

Anschließend wird überprüft, inwieweit die Berechnungslogik der gewichteten Punktwerte zu einem rein mechanischen, systembedingten Bunching führt – also ob die beobachteten Häufungen unmittelbar aus dem Bewertungssystem selbst stammen. Hierzu wurde eine simulationsbasierte Analyse implementiert:

- **Datengenerierung:** Auf Basis simulierter Begutachtungsdaten der ungewichteten Modulpunkte (Rohpunkte) erfolgt die Überführung in gewichtete Gesamtpunkte mittels des gesetzlichen Gewichtungsverfahrens.
- **Empirische Fundierung:** Als Erwartungswerte für eine realitätsnahe Simulation dienen die empirischen Verteilungen der ungewichteten Modulpunkte.
- **Intermodulare Korrelationen:** Das Modell berücksichtigt explizit die wechselseitigen Abhängigkeiten zwischen den Modulen (z. B. die Korrelation zwischen Einschränkungen der Selbstversorgung in Modul 4 und Einschränkungen der Mobilität in Modul 1). Die zugrunde gelegte Korrelationsmatrix wurde präzise an die empirisch beobachteten Gegebenheiten angepasst (s. Abbildung 6 im Anhang).

3. Bestimmung des Umfangs: Quantifizierung des anreiz- und verhaltensbedingten Bunchings

Zur Quantifizierung des tatsächlichen Ausmaßes der Verschiebungen werden drei komplementäre Ansätze angewendet. Alle drei Methoden basieren auf dem Vergleich der empirischen Verteilung mit einer spezifischen kontrafaktischen Verteilung, die keine oder reduzierte verhaltensbedingte Reaktionen abbildet. Das heißt, es wird die tatsächlich beobachtete Punkteverteilung mit der hypothetischen Verteilung verglichen, die ohne anreiz- bzw. verhaltensbedingte Intervention eingetreten wäre. Dadurch soll bewertet werden, welcher Anteil des beobachteten Bunchings tatsächlich auf Anreize zurückzuführen ist.

a) **Simulationsbasierter Ansatz:** Die in Schritt 2 generierte Verteilung dient als kontrafaktische Verteilung. Die Differenz zur empirischen Verteilung isoliert und quantifiziert direkt den anreiz- und verhaltensgetriebenen Effekt.

b) **Ökonometrischer Ansatz (Excess-Mass-Schätzung):** Ergänzend wird eine klassische *Excess-Mass*-Schätzung (Saez 2010; Chetty et al. 2011) durchgeführt. Die Methode vergleicht ebenfalls die beobachtete Verteilung der Punktwerte mit einer kontrafaktischen Verteilung ohne Verhaltensanpassungen. Die zusätzliche Häufung von Beobachtungen an den Schwellenwerten (Excess Mass) misst das Ausmaß der Anpassungsreaktionen. Anders als bei den übrigen Ansätzen wird die kontrafaktische Verteilung dabei nicht direkt aus den Daten abgeleitet, sondern

mittels eines statistischen Modells geschätzt. Die Ergebnisse hängen daher von den gewählten Modellannahmen ab.[2]

c) **Vergangenheitsbasierter Ansatz:** Da das Ausmaß des Bunchings seit der Systemumstellung sukzessive zugenommen hat, wird die empirische Verteilung des Einführungsjahres 2017 als historische kontrafaktische Verteilung genutzt und auf das Fallzahlvolumen von 2024 hochgerechnet. Dieser Ansatz modelliert die hypothetischen Pflegegradverschiebungen unter der Bedingung einer noch minimalen System- und Bewertungserfahrung.

Als Teil der Quantifizierung erfolgt anschließend eine Abschätzung, welche finanzielle Bedeutung die identifizierten Effekte haben.

[2] Die ökonometrische Schätzung basiert auf zusätzlichen Annahmen und mehreren frei wählbaren Parametern wie Polynomgrad, Fensterbreite, Klassenbreite, die das Ergebnis beeinflussen und unsicherer machen. Die Wahl der Spezifikation ist im Kern eine Abwägung zwischen Anpassungsgüte und Überanpassung. Ein flexibleres Modell bildet die Verteilung außerhalb der Schwellen genauer ab. Es läuft jedoch Gefahr, zufällige Schwankungen mitzumodellieren oder einen Teil der Häufung einzufangen, der über das Fenster hinausreicht; beides verzerrt die Excess Mass. Ein sparsameres Modell ist robuster gegen solche Artefakte, kann jedoch den systematischen Verlauf der Verteilung verfehlen. Eine Übersicht über verschiedene Modelle ist in Abbildung 9 im Anhang dargestellt.

4 Ergebnisse

4.1 Empirischer Nachweis von Punktwerthäufungen/Bunching

Die deskriptive Analyse der empirischen Verteilung der gewichteten Gesamtpunktzahlen (Abbildung 2) deutet auf eine Konzentration von Begutachtungen unmittelbar oberhalb der gesetzlichen Schwellenwerte hin. Dieser visuelle Befund wird durch den Dichtetest nach Cattaneo et al. (2018) formal bestätigt: Für sämtliche Schwellenwerte werden statistisch signifikante Ergebnisse erzielt ($p < 0{,}0001$). Die Ergebnisse weisen somit auf systematische Diskontinuitäten in der Verteilung der Punktwerte an den Schwellenwerten hin und liefern deutliche Hinweise auf das Vorliegen von Fallverschiebungen.

4.2 Identifikation mechanischer systembedingter Effekte – Die Rolle des Begutachtungsinstruments

Um die Rolle des Begutachtungsinstruments für das beobachtete Bunching zu untersuchen, zeigt Abbildung 4 die Verteilung der gewichteten Gesamtpunktzahlen auf Basis der Datensimulation. Für einen unmittelbaren Vergleich stellt Abbildung 5 außerdem die Differenzen zwischen der empirischen und der simulierten Verteilung dar. Dabei zeigt sich ein deutliches Muster: In den real beobachteten Daten liegen die Häufigkeiten unmittelbar vor den Schwellenwerten (gepunktete Linien) systematisch niedriger. Unmittelbar nach den Schwellenwerten liegen die Häufigkeiten hingegen deutlich höher als in der Simulation. Auf Basis der simulierten Daten wäre also kaum Bunching zu erwarten. Die Ergebnisse sprechen somit gegen die Annahme, dass das Bunching primär durch die Konstruktion des Begutachtungsinstruments oder die Gewichtung der Punktwerte verursacht wird.

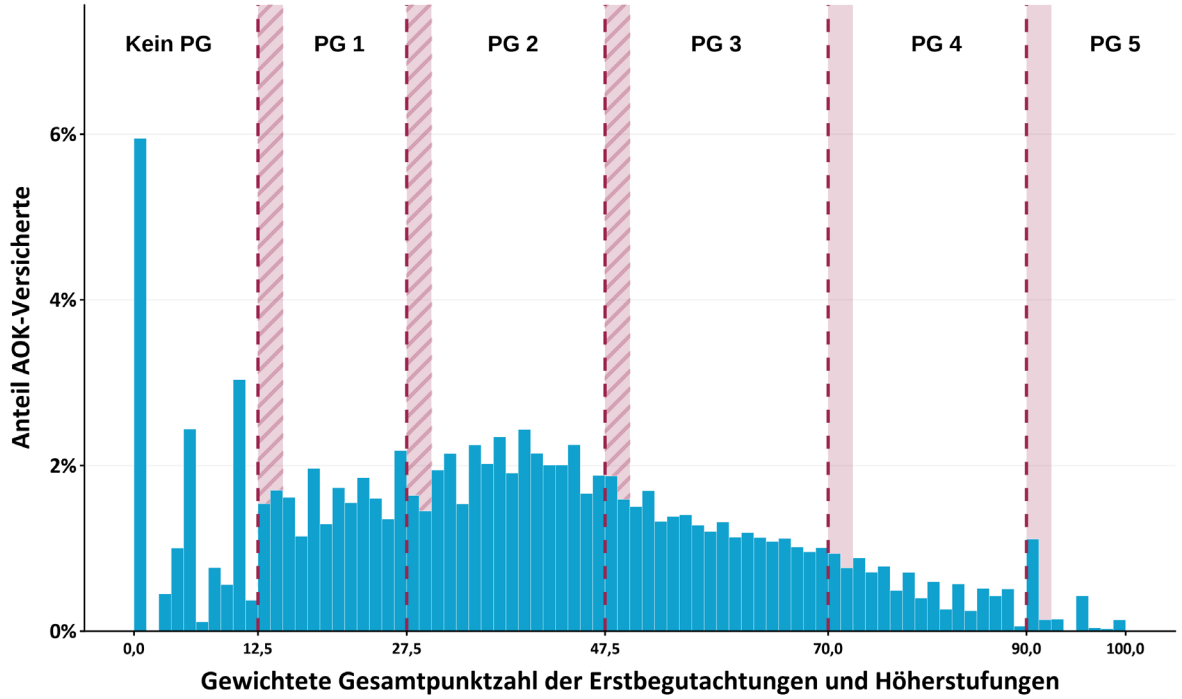
Die Abweichungen zwischen Simulation und empirischer Verteilung konzentrieren sich im Wesentlichen auf einen Bereich von ein bis zwei Modulpunkten um die jeweiligen Schwellenwerte. In diesen Bereichen genügt für die Einstufung in einen höheren Pflegegrad häufig bereits eine geringfügige Veränderung in einem einzelnen Modul.

Welches Modul für eine solche Anpassung genutzt wird, kann zwischen den Begutachtungen variieren. Solange die Anpassungen nicht systematisch über dasselbe Modul oder dieselbe Modulkombination erfolgen, wirken sie sich selbst bei mehreren Millionen Fällen nur begrenzt auf die Korrelationen und Randverteilungen der Modulpunkte aus. Auf aggregierter Ebene bleiben solche Anpassungen daher weitgehend unsichtbar. Die Simulation erzeugt folglich eine kontrafaktische Verteilung, die nur in geringem Maße von anreiz- und verhaltensbedingten Anpassungen beeinflusst wird.

Ein weiterer Faktor, der die Verteilung der Gesamtpunkte beeinflussen kann, ist der Selektionseffekt. Personen, die einen Pflegegrad beantragen, weisen in der Regel bereits gesundheitliche oder funktionale Einschränkungen auf und unterscheiden sich daher von der Gesamtbevölkerung. Dieser sogenannte Selektionseffekt kann die gesamte Punkteverteilung verschieben. Er erklärt jedoch nicht, warum sich Fälle bei ganz bestimmten Punktwerten häufen.

Die Qualität der simulierten Daten wurde anhand der Verteilungen der Modulpunkte, der Korrelationsstruktur sowie ausgewählter Modulkombinationen ähnlich der von Nolting et al. (2026, S. 70 ff.) geprüft. Die Ergebnisse zeigen, dass die Simulation die empirischen Daten in allen betrachteten Aspekten realitätsnah abbildet. Eine ausführlichere Beschreibung ist im Anhang im Abschnitt Qualität der Simulation gegeben.

Abbildung 4: Verteilung der simulierten gewichteten Gesamtpunkte

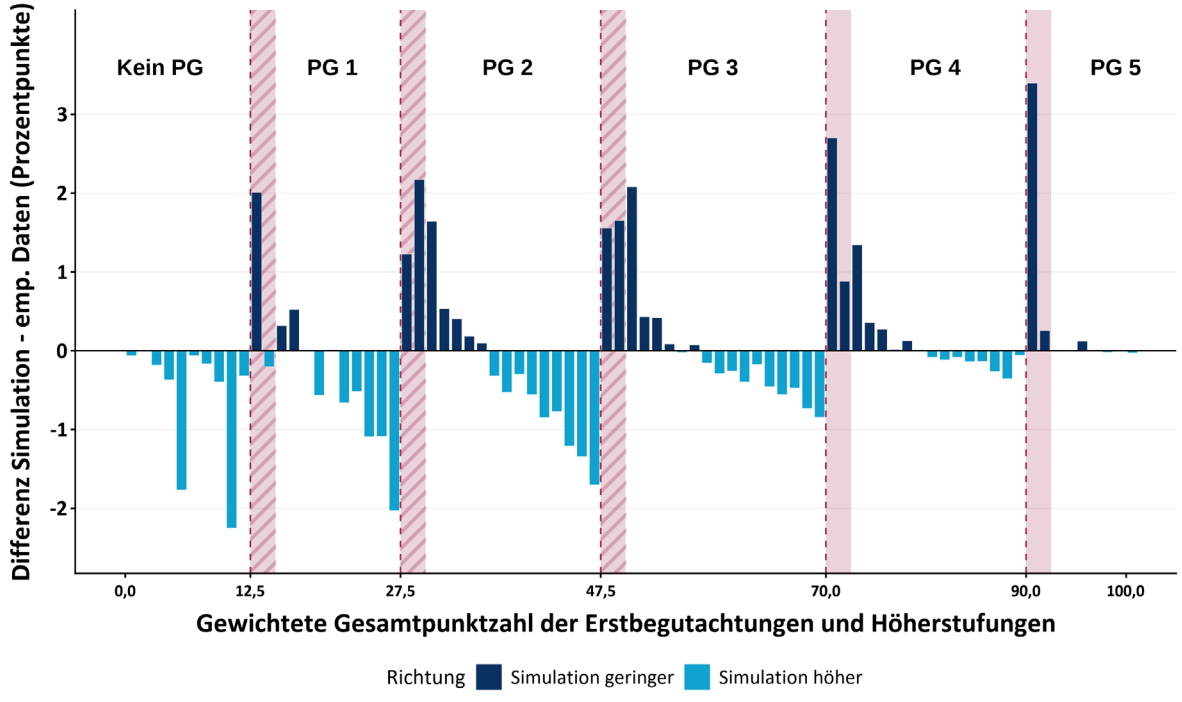


Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen.

Quelle: eigene Berechnung auf Basis von MD-Begutachtungsdaten für AOK-Versicherte 2024

© WIdO 2026

Abbildung 5: Abweichungs-Analyse: Reale Daten vs. Simulation



Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen. Positive/negative Werte bedeuten, dass hier die Simulation weniger/mehr Beobachtungen generiert hat als in den tatsächlichen Daten vorkommen.

Quelle: eigene Berechnung auf Basis von MD-Begutachtungsdaten für AOK-Versicherte 2024

© WIdO 2026

4.3 Quantifizierung des anreiz- und verhaltensbedingten Bunchings

Tabelle 4 fasst die Ergebnisse der drei in Kapitel 3 beschrieben Ansätze zusammen und stellt sie den Ergebnissen einer reinen Ceteris-Paribus-Fortschreibung gegenüber. Hierbei beziehen sich die Berechnungen auf die gesamte SPV. Dafür wurde auf die Zahlen von Nolting et al. (2026) zurückgegriffen. Teil A der Tabelle enthält die statischen Projektionen dieser Fortschreibung sowie die Ergebnisse des simulationsbasierten Ansatzes. Teil B zeigt die Befunde des historischen Vergangenheitsvergleichs mit dem Referenzjahr 2017 und des theoretisch-ökonometrischen Ansatzes auf Basis der Excess-Mass-Schätzung[3].

Ausgewiesen werden jeweils die erwarteten Veränderungen in der Pflegegradzuweisung. Dargestellt sind sowohl die betroffenen absoluten Fallzahlen von Erstbegutachtungen und Höherstufungen als auch deren relativer Anteil an der Gesamtpopulation. Tabelle 4 ermöglicht damit eine Abschätzung, in welchem Umfang die erwarteten niedrigeren Pflegegrade auf Bunching-Effekte zurückzuführen sind. Die Ceteris-Paribus-Fortschreibung ergibt, dass nach der Anpassung der Schwellenwerte etwa 19 Prozent aller Begutachtungen zu einem niedrigeren Pflegegrad führen würden. Wird stattdessen die simulierte Verteilung zugrunde gelegt, reduziert sich dieser Anteil auf etwa 10 Prozent. Eine Modellierung auf Basis der Verteilung des Jahres 2017 kommt zu einem ähnlichen Ergebnis: Hier würden sich in rund 11 Prozent der Begutachtungen die Pflegegrade verändern. Auch der in der ökonomischen Literatur verbreitete Excess-Mass-Ansatz liefert vergleichbare Befunde.

Tabelle 4: Vergleich der durch eine Schwellenwertanpassung verursachten Einstufungsunterschiede unter verschiedenen Berechnungsansätzen von Anreiz- und Verhaltensanpassungseffekten

	Anzahl Änderungen Pflegegrad (*n* = 2.580.165)	Anteil an allen Erstbeg. und Höherstufung	Anzahl Änderungen Pflegegrad (*n* = 2.580.165)	Anteil an allen Erstbeg. und Höherstufung
Teil A	**Ceteris-Paribus-Ansatz**		**Simulationsbasierter Ansatz**	
kein PG anstatt PG1	140.223	5,4 %	91.781	3,6 %
PG1 anstatt PG2	180.098	7,0 %	86.927	3,4 %
PG2 anstatt PG3	172.122	6,7 %	89.815	3,5 %
Änderungen - insgesamt	**492.443**	**19,1 %**	**268.524**	**10,4 %**
Teil B	**Vergangenheitsbasierter Ansatz**		**Ökonometrischer Ansatz**	
kein PG anstatt PG1	75.825	2,9 %	91.730	3,6 %
PG1 anstatt PG2	103.976	4,0 %	116.553	4,5 %
PG2 anstatt PG3	108.199	4,29 %	86.879	3,4 %
Änderungen - insgesamt	**288.001**	**11,2 %**	**295.162**	**11,4 %**

Teil A enthält Berechnungen unter Ceteris-Paribus-Annahme und dem simulationsbasierten Ansatz. Teil B zeigt Ergebnisse basierend auf einem Vergangenheitsvergleich mit dem Jahr 2017 und einem ökonometrischen Ansatz mittels vorgegebener Verteilung (Polynom 20. Grades). Es handelt sich um Hochrechnungen auf die SPV. Dazu wurde auf Daten von Nolting et al. (2026) zurückgegriffen.

Quelle: eigene Darstellung, Datenjahr 2024. Anpassung der Schwellenwerte nur für die Pflegegrade 1 bis 3. © WIdO 2026

[3] Die Modellierung der theoretischen Verteilung im Rahmen der Excess-Mass-Schätzung erfordert restriktive Setzungen bezüglich der funktionalen Form und der Freiheitsgrade. Die hier berichteten Werte basieren auf einer Polynomapproximation 20. Grades (die grafische Anpassungsgüte ist in Abbildung 8 im Anhang dokumentiert).

Unabhängig vom gewählten Ansatz zur Bestimmung der kontrafaktischen Verteilung ergibt sich also ein stabiles Ergebnismuster: Unter Berücksichtigung des Bunchings liegt der Anteil der potenziellen Begutachtungen, die eine niedrigere Einstufung erhalten würden, nur bei 55 Prozent bis 60 Prozent der ursprünglich erwarteten Fallzahl im statischen Modell. Mit anderen Worten: Etwa 40 Prozent bis 45 Prozent der durch eine reine Ceteris-Paribus-Betrachtung prognostizierten Effekte könnten durch Anreiz- und Verhaltensanpassungen kompensiert werden. Für die politische Praxis bedeutet dies, dass die in den Gesetzesentwürfen ausgewiesenen Einsparpotenziale der Schwellenwertreform deutlich niedriger ausfallen könnten.

4.4 Potenzielle finanzielle Auswirkungen

Es liegen bereits verschiedene Schätzungen vor, die das mögliche Einsparpotenzial einer Anpassung der Schwellenwerte für die SPV quantifizieren. Das BMG veranschlagt die mittelfristig erwarteten Finanzwirkungen der Anpassung der Begutachtungssystematik mit 4,2 Milliarden Euro. Auch der vom GKV-Spitzenverband in Auftrag gegebene IGES-Bericht (Nolting et al. 2026) geht von erheblichen Auswirkungen auf die administrative Pflegebedürftigkeit und die Ausgaben der SPV aus. Die Facharbeitsgruppen „Versorgung“ und „Finanzierung“ im Zukunftspakt Pflege (Zukunftspakt Pflege 2025) bezifferten das kurzfristige Einsparpotenzial auf 0,1–0,2 Beitragssatzpunkte (ca. 2–4 Milliarden Euro). Im Frühjahresgutachten des SVR Wirtschaft wird für das Jahr 2040 sogar eine Einsparung von 0,7 Beitragssatzpunkten (ca. 14 Milliarden Euro) projiziert.

Vor dem Hintergrund dieser zum Teil unterschiedlichen Einschätzungen werden im Folgenden mögliche Einsparungen für die SPV als Referenzpunkt ermittelt. Diese dienen anschließend als Grundlage, um die Auswirkungen der zuvor identifizierten Bunching-Effekte auf die erwarteten finanziellen Entlastungen zu quantifizieren.

Zur Schätzung der finanziellen Auswirkungen wurden die durchschnittlichen Ausgaben der SPV auf Basis der amtlichen Statistiken (PJ1 und PG2) ermittelt. Die Analyse konzentriert sich auf jene Kostenstellen, die differenziert nach Pflegegraden ausgewiesen werden. Dazu zählen die zentralen Leistungsbereiche der häuslichen Versorgung sowie der teil- und vollstationären Pflege. Tabelle 5 stellt die durchschnittlichen monatlichen Kosten pro Person und Pflegegrad sowie die jeweilige Kostendifferenz zum nächstniedrigeren Grad dar.

Die auf Durchschnittswerten basierende Vorgehensweise ist mit Einschränkungen verbunden. So bleiben insbesondere individuelle Verläufe der Pflegebedürftigkeit und Veränderungen der Leistungsinanspruchnahme im Zeitverlauf unberücksichtigt. Für die Abschätzung der Größenordnung der zu erwartenden finanziellen Effekte ist dieser Ansatz jedoch plausibel.

Tabelle 5: Durchschnittliche monatliche Kosten pro Person nach Pflegegrad und Differenz zum nächstniedrigen Pflegegrad im Jahr 2024

	kein PG	PG 1	PG 2	PG 3	PG 4	PG 5
Durchschnittliche Ausgaben	-	103 €	426 €	851 €	1.417 €	1.832 €
Differenz zum PG zuvor	-	103 €	324 €	425 €	566 €	415 €

Quelle: eigene Berechnungen basierend auf amtlichen Statistiken (PJ1 und PG2)

© WIdO 2026

Tabelle 6 führt die Kosten pro Person mit den von der Reform betroffenen Begutachtungen (Tabelle 4) zusammen und weist das theoretische Einsparpotenzial im Sinne einer Ceteris-Paribus-Fortschreibung aus. Zunächst werden die erwarteten Einsparungen bei einer ausschließlichen Anpassung der Schwellenwerte für die Pflegegrade 1 bis 3 dargestellt. Auf dieser Ebene liegen auch die Erkenntnisse zum Bunching aus der vorliegenden Analyse vor. Ergänzend wird unter Verwendung der Daten von Nolting et al. (2026) das Einsparpotenzial für den Fall ausgewiesen, dass sowohl die Schwellenwerte der Pflegegrade als auch die Modulschwellen angepasst werden. Dabei wird angenommen, dass Bunching-Effekte auf beiden Ebenen in vergleichbarem Umfang auftreten.

Allein durch die Anpassung der Schwellenwerte für die Pflegegrade 1 bis 3 ergibt sich für die SPV in der Ceteris-Paribus Betrachtung ein jährliches Einsparpotenzial von etwa 1,8 Milliarden Euro. Diese Berechnung unterstellt jedoch, dass die betroffenen Pflegebedürftigen volle zwölf Monate Leistungen der SPV beziehen. Berücksichtigt man die innerhalb eines Jahres zu erwartenden Sterbefälle nach Pflegegrad, reduziert sich das jährliche theoretische Einsparvolumen auf ca. 1,5 Milliarden Euro. Für die kombinierte Anpassung der Pflegegrad- und Modulschwellen ergibt sich ein Einsparpotenzial von ca. 3,8 Milliarden Euro, was in etwa einer Verdopplung entspricht. Unter Berücksichtigung der unterjährigen Sterblichkeit reduziert sich dieser Betrag auf ca. 2,9 Milliarden Euro.

Tabelle 6: Übersicht über das theoretische jährliche Einsparpotenzial

Kostenersparnis	allgemeine Schätzung	unter Berücksichtigung der Überlebensrate
Schwellenwertanpassung Pflegegrade 1 bis 3	1,75 Mrd. €	1,49 Mrd. €
Schwellenwertanpassung Pflegegrade 1 bis 3 + Module	3,81 Mrd. €	2,89 Mrd. €

Die Berechnung beruht auf Daten von Nolting et al. (2026) und Tabelle 5, um ein Ergebnis für die gesamte SPV zu erhalten.

Quelle: eigene Berechnungen

© WIdO 2026

Die Werte in Tabelle 6 stellen jedoch lediglich das Einsparpotenzial im Sinne einer Ceteris-Paribus-Betrachtung dar und berücksichtigen keine Anreiz- und Verhaltensanpassungen. Tabelle 7 überträgt die Ergebnisse der Bunching-Analysen auf die erwarteten Finanzwirkungen unter der Annahme, dass die Effekte für die Modulschwellen gleich sind. Dabei zeigt sich, dass sich das Einsparvolumen unter Berücksichtigung von Anreiz- und Verhaltensanpassungen auf etwa die Hälfte des ursprünglich prognostizierten Wertes reduzieren könnte. Für die derzeit diskutierte Anhebung der Pflegegrad- und Modulschwellen entspricht dies einer Verringerung der erwarteten Einsparungen um knapp 2 Milliarden Euro.

Tabelle 7: Einsparpotenziale der SPV unter verschiedenen Annahmen zu Anreiz- und Verhaltensanpassungseffekten

Verhaltensanpassung	Ceteris-Paribus-Ansatz	Simulationsbasierter Ansatz	Vergangenheitsbasierter Ansatz	Ökonometrischer Ansatz
	(1)	(2)	(3)	(4)
Schwellenwertanpassung Pflegegrade 1 bis 3	1,75 Mrd. €	0,91 Mrd. €	1,05 Mrd. €	1,01 Mrd. €
Schwellenwertanpassung Pflegegrade 1 bis 3 + Module	3,81 Mrd. €	1,98 Mrd. €	2,28 Mrd. €	2,20 Mrd. €

Identischer Bunchingeffekt für beide Varianten angenommen.

Quelle: eigene Berechnungen basierend auf den Werten aus Tabelle 4 und Tabelle 5

© WIdO 2026

5 Diskussion

Das SGB XI definiert Pflegebedürftigkeit und übersetzt Unterstützungsbedarfe in Leistungsbeträge, indem es diese in fünf Pflegegrade einteilt und jedem Grad bestimmte Leistungsansprüche zuweist. Charakteristisch für dieses System sind die Sprünge an den Grenzen zwischen zwei Pflegegraden. Während Veränderungen der Punktzahl fernab dieser Schwellenwerte meist ohne Auswirkungen auf den Leistungsanspruch bleiben, können sie in ihrer Nähe erhebliche Folgen haben: Mitunter entscheidet im deutschen System bereits ein einziger Punkt bei nahezu identischen Pflegebedarfen über mehrere hundert Euro zusätzlicher Leistungen pro Monat.

Die Anpassung der Schwellenwerte bei der Pflegebegutachtung bietet daher – auf den ersten Blick – ein erhebliches Potenzial, die Ausgaben der SPV zu senken. Eine Ceteris-Paribus-Betrachtung zeigt allein für die Anpassung der Schwellenwerte der Pflegegrade 1 bis 3 ein jährliches Einsparvolumen von knapp 1,8 Milliarden Euro. Übertragen auf die Anpassung der Schwellenwerte auch auf Modulebene ergibt sich ein Einsparpotenzial von 3 bis 4 Milliarden Euro pro Jahr. Diese Größenordnung deckt sich mit dem Finanztableau des PNOG-Entwurfs. Demnach weist die Anpassung der Begutachtungssystematik das größte Einsparpotenzial aller Vorschläge auf.

Die vorgesehene Anpassung der Schwellenwerte entspricht den Empfehlungen des Expertenbeirats aus dem Jahr 2013. In der Gesetzesbegründung wird sie ausdrücklich als „Rückkehr“ zu diesen Werten bezeichnet mit dem Ziel, die Entwicklung der Pflegeprävalenzen abzuschwächen. Dabei ist zu berücksichtigen, dass die Erweiterung des Pflegebedürftigkeitsbegriffs und die damit verbundene Einbeziehung eines größeren Personenkreises ausdrückliche Ziele der damaligen Reform waren. Mit der nun vorgesehenen „Rückkehr“ ist die implizite Annahme verbunden, dass der stärker als erwartete Anstieg der Pflegeprävalenz seit 2017 wesentlich auf die seinerzeitige Abweichung von den vorgeschlagenen Werten zurückzuführen ist. Die Tragfähigkeit dieser Annahme hängt jedoch auch davon ab, auf welcher wissenschaftlichen Grundlage die Schwellenwerte ursprünglich festgelegt wurden. Diese waren nicht das Ergebnis einer streng wissenschaftlichen Analyse, sondern eine pragmatische, auf Plausibilität ausgerichtete Annäherung (Rothgang 2026). Darüber hinaus vernachlässigt die Diskussion der vorgeschlagenen Reform zentrale wissenschaftliche Erkenntnisse zur Anreizsetzung durch Schwellenwerte.

Diese Anreize spiegeln sich in der empirischen Verteilung der Gesamtpunkte in Form von auffälligen Häufungen von Bewertungen direkt oberhalb der Schwellenwerte wider und unterstreichen die Limitationen einer Ceteris-Paribus-Betrachtung. Solche Häufungen sind in der internationalen Literatur als „Bunching“ bekannt. Dieses Phänomen wurde im Kontext der Pflegeversicherung bereits in anderen Ländern wissenschaftlich dokumentiert. So ist für Japan nachgewiesen, dass Begutachtende die Bewertungspunkte gezielt über die nächste Stufengrenze heben (Ishihara und Takahashi 2022). Für Spanien sind dieselben Sprünge in der Verteilung der Punktwerte dokumentiert (Hernández-Pizarro et al. 2020), für Frankreich die Häufung an der Zugangsschwelle zur Pflegeleistung (Roy 2025). Die ökonomische Theorie erwartet genau das: Wo Leistungen an Schwellen springen (sogenannte „Notches“), entstehen extreme Anreize, die große Effekte erzeugen können (Kleven und Waseem 2013; Kleven 2016).

Die vorliegende Studie zeigt mittels einer Simulation zunächst, dass das in Deutschland beobachtete Bunching nicht mechanisch-systembedingt durch die Berechnungslogik der gewichteten Gesamtpunktzahl verursacht wird. Die tatsächliche Anzahl an Personen direkt über den Schwellenwerten ist im Vergleich zur erwarteten Anzahl bis zu doppelt so groß.

Welche Gründe das Bunching verursachen, muss tiefergehend analysiert werden. Die vorliegenden Ergebnisse dieses Beitrags legen, unter Berücksichtigung der internationalen Literatur und Evidenz, nahe, dass anreiz- und verhaltensbezogene Faktoren hier eine wichtige Rolle spielen. In weitergehenden Analysen ist zu untersuchen, welche Wirkmechanismen hier greifen.

Alternative Bewertungsmodelle

Wenn die Schwellenwerte selbst die Ursache des Bunchings sind, liegt die Frage nahe, ob eine andere Ausgestaltung oder ein grundlegend anderes Bewertungsmodell diese Effekte vermeiden könnte. Ein internationaler Vergleich der Begutachtungsmodelle von OECD-Ländern zeigt, dass die verfügbaren Varianten jeweils eigene Zielkonflikte mit sich bringen (Llena-Nozal et al. 2025).

Innerhalb des bestehenden, stufenförmigen Modells ist die Anzahl der Stufen eine wichtige Stellschraube. Der Vorteil einer feineren Abstufung besteht darin, dass mit der geringeren Höhe der einzelnen Leistungssprünge auch der punktuelle Anreiz abnimmt, eine bestimmte Schwelle zu überschreiten. Zugleich wächst jedoch die Zahl der Schwellen, bei denen strategische Erwägungen überhaupt eine Rolle spielen können: Aus wenigen punktuellen Anpassungen wird ein breiter gestreuter Aufwärtsdruck. Zudem würden kleine Bewertungsunterschiede, die bei einem ordinal angelegten Instrument wie dem NBA unvermeidlich mit Messunschärfe behaftet sind, häufiger leistungsrelevant. Eine feinere Abstufung ist dementsprechend ein Kompromiss zwischen Anreizdämpfung und Steuerbarkeit. Der Extremfall eines vollständig linearen Systems ohne Stufen ist in der Praxis nicht bekannt (Llena-Nozal et al. 2025).

Eine mögliche Alternative verzichtet auf eine einheitliche landesweite Punktebewertung und bemisst die Leistung über eine individuelle Bedarfsfeststellung. In Schweden und Dänemark entscheiden kommunale Stellen im Einzelfall über Art und Umfang der Leistungen, ohne dass eine landesweite Einteilung den Anspruch festlegt (Llena-Nozal et al. 2025). Der Vorzug dieses Ansatzes liegt in der stärkeren Orientierung am konkreten Einzelfall. An die Stelle der Anreiz- und Verhaltenseffekte an den Schwellen tritt jedoch der Ermessensspielraum der bedarfsfeststellenden Kommune. Zudem ist eine Bedarfsbemessung im Einzelfall wahrscheinlich verwaltungsaufwendiger, weniger transparent und anfälliger für regionale Unterschiede, da sich Einzelfallentscheidungen nur begrenzt standardisieren lassen. Die Niederlande zentralisierten ihre zuvor regional organisierte Begutachtung, um solche Unterschiede zu verringern (Llena-Nozal et al. 2025).

Praxis- und Politikimplikationen

Aus den Befunden und dem internationalen Vergleich lassen sich zwei Implikationen ableiten: Erstens sollten die im PNOG-Entwurf ausgewiesenen Einsparpotenziale mit Vorsicht bewertet werden, da sie weitgehend auf statischen Betrachtungen beruhen. Die vorliegenden Ergebnisse legen nahe, dass Anreiz- und Verhaltensanpassungen einen Teil (bis zu 50 Prozent) der erwarteten Einsparungen kompensieren könnten. Wenn ein erheblicher Anteil der Begutachtungen in unmittelbarer Nähe der Schwellenwerte erfolgt, ist zu erwarten, dass sich entsprechende Häufungen bei einer Verlagerung der Grenzen zumindest teilweise an die neuen Schwellen verschieben. Diese Ergebnisse sind jedoch nicht so zu verstehen, dass die Schwellenwerte stärker angehoben werden sollten, um die Sparziele dennoch zu erreichen. Wenn der steigende Zugang zum Leistungssystem als Problem identifiziert wird, wäre eine Evaluation und anschließende Reform des Begutachtungssystems selbst nachhaltiger als eine pauschale Verschiebung. Inhaltlich wäre dabei zu klären, wie das Instrument angepasst werden muss, um die begrenzten finanziellen Ressourcen zielgenau zu verteilen. Eine pauschale Anhebung der Schwellenwerte spart zwar Geld, ist aber nicht zwangsläufig mit den geringsten Einbußen an Versorgungssicherheit und -qualität verbunden. Mögliche Verschlechterungen in diesen Bereichen können ihrerseits Folgekosten für das Pflege- und Gesundheitssystem verursachen (Serrano-Alarcón et al. 2022). Für eine belastbare Einschätzung sind daher unter anderem eine Analyse der Pflegebedürftigen nahe der Schwellenwerte sowie der Effektivität von Pflegeleistungen in Deutschland erforderlich.

In der Diskussion weitgehend unberücksichtigt bleiben zudem die Auswirkungen der erforderlichen Übergangsregelung. Da rechtzeitig eingereichte Anträge noch nach den bisherigen Schwellenwerten bewertet werden, entsteht ein starker Anreiz, Anträge vorzuziehen. Entsprechende Hinweise liegen bereits vor (vgl. Sozialverband VdK Deutschland 2026). Leistungen würden dadurch früher beantragt und der ohnehin stark belastete Medizinische Dienst zusätzlich beansprucht. Da die Leistungen rückwirkend ab dem Monat der Antragstellung gewährt werden, könnten die Ausgaben kurzfristig sogar steigen.

Zweitens spricht der internationale Vergleich dafür, den Fokus auf die Weiterentwicklung des bestehenden Systems zu legen, solange keine belastbare Evidenz für die Überlegenheit eines anderen Modells vorliegt. In dieselbe Richtung weist die IGES-Evaluation (Nolting et al. 2026), indem sie die grundlegende Frage aufwirft, welche Funktion das Begutachtungsinstrument künftig erfüllen soll: Soll es primär den Zugang zu Leistungen abgrenzen oder zugleich Grundlage für Prävention, Versorgungsplanung und Steuerung dienen? Internationale Erfahrungen können für diese funktionale Neubestimmung wichtige Orientierungspunkte bieten, da zugangsbegrenzende und steuerungsorientierte Ansätze in vielen Ländern eingeführt wurden (Llena-Nozal et al. 2025). Verknüpft mit dieser funktionalen Ausrichtung, rücken auch prozedurale Aspekte des bestehenden Systems in den Fokus, die im internationalen Vergleich oft unterschiedlich gehandhabt werden. Hierzu zählen etwa die rechtliche Ausgestaltung von Karenzzeiten nach der Antragstellung, die Systematik und Befristung von Wiederholungsbegutachtungen sowie die Organisation und Beauftragung des Begutachtungsprozesses selbst (vgl. § 18 und § 18a SGB XI). Eine Evaluation dieser verwandten Regelungen erlaubt es, das Begutachtungsinstrument nicht isoliert, sondern als Teil eines optimierten Gesamtverfahrens weiterzuentwickeln.

Das PNOG greift den Bedarf an einer solchen Weiterentwicklung grundsätzlich auf, indem es in § 18f SGB XI (über die bereits enthaltenen Bestimmungen zur Weiterentwicklung in § 18e hinaus) eine kontinuierliche und institutionalisierte Evaluation der Begutachtung vorsieht. Offen bleibt jedoch, wann daraus konkrete Vorschläge für dessen Weiterentwicklung hervorgehen und in welchem Umfang die Empfehlungen des vorgesehenen Beirats tatsächlich in Anpassungen des Instruments überführt werden.

6 Fazit

Die vorliegende Studie untersucht erstmals systematisch das Auftreten von Bunching an den Schwellenwerten der deutschen Pflegebegutachtung und deren Bedeutung für Reformüberlegungen. Die empirischen Analysen haben gezeigt, dass es (1) Häufungen von Pflegebegutachtungspunktzahlen unmittelbar oberhalb der Schwellenwerte in Deutschland gibt. (2) Diese lassen sich nicht allein durch das Begutachtungsinstrument erklären, sondern sind (3) bis zu 50 Prozent durch Anreiz- und Verhaltenseffekte verursacht.

Für die Reformdebatte bedeutet dies, dass die Einsparpotenziale der Schwellenwertanpassungen bisher überschätzt werden. Unter Berücksichtigung von Anreiz- und Verhaltenseffekten dürften die tatsächlichen Einsparungen nur etwa halb so hoch ausfallen. Mittelfristig wird das realisierbare Einsparpotenzial damit um mehrere Milliarden Euro niedriger ausfallen als erwartet. Kurzfristig sind aufgrund von Vorzieheffekten sogar Mehrausgaben möglich, da Anträge vor Inkrafttreten neuer Schwellenwerte gebündelt gestellt und Leistungen rückwirkend ab dem Monat der Antragstellung gewährt werden.

Im Einklang mit der internationalen Evidenz verdeutlichen unsere Ergebnisse die strukturellen Limitationen von Schwellenwertanpassungen in gestuften Systemen. Die Befunde legen nahe, dass eine Verschiebung von Anspruchsschwellen die Anreiz- und Verhaltensreaktionen nicht beseitigt, sondern überwiegend verlagert.

Die Ergebnisse sprechen weder für eine noch stärkere Erhöhung der Schwellenwerte noch gegen jede Anpassung von Zugangskriterien. Sie zeigen jedoch, dass Schwellenwertpolitik allein ein begrenztes Steuerungsinstrument ist. Wenn der Leistungszugang als zu weitreichend, zu wenig zielgenau oder fiskalisch nicht tragfähig bewertet wird, muss die Reformfrage breiter gestellt werden: Nicht nur die Höhe einzelner Schwellen, sondern das Begutachtungssystem und dessen versorgungspolitische Zielstellung insgesamt sind zu überprüfen.

Anhang

Qualität der Simulation

Die Validierung der Simulationsqualität erfolgte über die Analyse spezifischer Modulkombinationen ähnlich der von Nolting et al. (2026, S. 70 ff.) vorgenommenen Betrachtung. Rein theoretisch existieren je nach Punktwert zwischen 2 und 82 verschiedene Kombinationen, die es ermöglichen, genau diesen Punktwert zu erreichen. Empirisch bleiben viele dieser Kombinationen aufgrund der Korrelationen der Module allerdings unbesetzt. Außerdem gibt es einige Kombinationen, die nur sehr schwach besetzt sind. Das Ziel der Simulation ist es nicht, die Daten bis ins letzte Detail nachzubilden, sondern eine realistische Population von Pflegebedürftigen zu simulieren. Unter Berücksichtigung der Relevanz von Kombinationen, gemessen anhand der Fallzahlen, kann die effektive Anzahl berechnet werden. Diese liegt aufgrund der Gewichtung deutlich unterhalb der tatsächlichen Anzahl.

Um die Simulation zu validieren, wurden die tatsächlichen Besetzungszahlen der Kombinationen mit den simulierten verglichen. Dabei wurde einerseits betrachtet, wie viele Kombinationen pro Punktwert vorkommen, und ob die Rangfolge der Kombinationen bezüglich ihrer Häufigkeit übereinstimmt. Wenn Simulation und Realität bezüglich der erzielten Modulpunktkombinationen übereinstimmen, ist dies ein Indiz für eine passgenaue Simulation. Insgesamt zeigt sich, dass das empirische Muster durch die Simulation präzise abgebildet werden kann:

Anzahl der Kombinationen

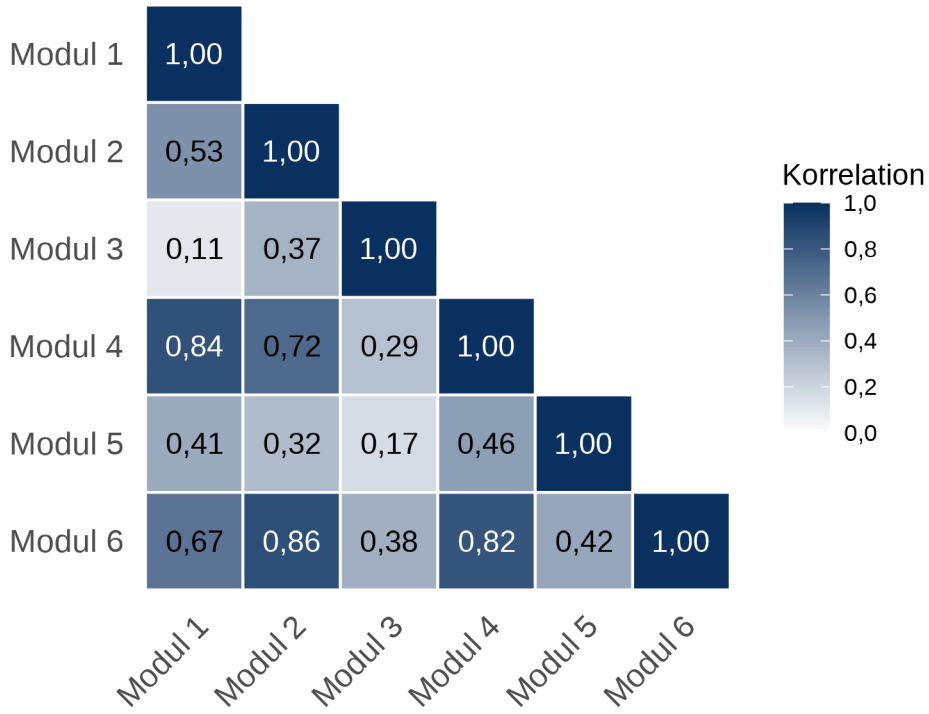
- **Ungewichtet**: Die Anzahl der tatsächlich belegten Merkmalskombinationen wird in der Tendenz unterschätzt. Es gibt nur geringe Abweichungen in den Variationskoeffizienten. Zudem wurde der mittlere absolute Fehler in den Standardabweichungen (MAE) bestimmt. Diese Kennzahl beschreibt den mittleren absoluten Modellfehler, ausgedrückt in Einheiten der empirischen Standardabweichung der Originaldaten. Obwohl es sich hierbei formal nicht um ein Maß für Mittelwertdifferenzen (wie Cohens d) handelt, kann die Höhe dennoch analog interpretiert werden: Da die Abweichung an der natürlichen Streuung der empirischen Verteilung normiert wird, indizieren Werte unter 0,2 bis 0,5 in Anlehnung an klassische Effektstärkebenchmarks (Cohen, 1988) einen statistisch und praktisch kleinen Modellfehler. Der MAE für die ungewichtete Anzahl an Kombinationen liegt bei 0,36.
- **Gewichtet:** Unter Berücksichtigung der Fallzahlen pro Kombination, also mit einer Gewichtung mit den Häufigkeiten jeder Kombination, nähern sich die Variationskoeffizienten weiter an. Zudem sinkt der MAE auf 0,27.

Die Ergebnisse zeigen, dass die Simulation die reale relative Streuung der Daten gut widerspiegelt und typische Begutachtungsbewertungen generieren kann. Gleichzeitig ist eine Limitation der Simulation, dass seltene Begutachtungsbewertungen nicht umfangreich dargestellt werden können. Dies ist für die Beantwortung der Fragestellung nicht problematisch, da die Detailvielfalt seltener Ausreißer-Kombinationen nicht entscheidend für das Bunching ist.

Rangfolge der Kombinationen

Die Rangfolge der häufigsten Kombinationen ist konsistent. Ein Abgleich der 15 häufigsten Kombinationen (für die Einstufungen in keinen Pflegegrad sowie in die Pflegegrade 1 bis 3) zeigt ein**e** weitgehende Übereinstimmung zwischen den simulierten Daten, den vorliegenden AOK-Daten und den Ergebnissen von Nolting et al. (2026).

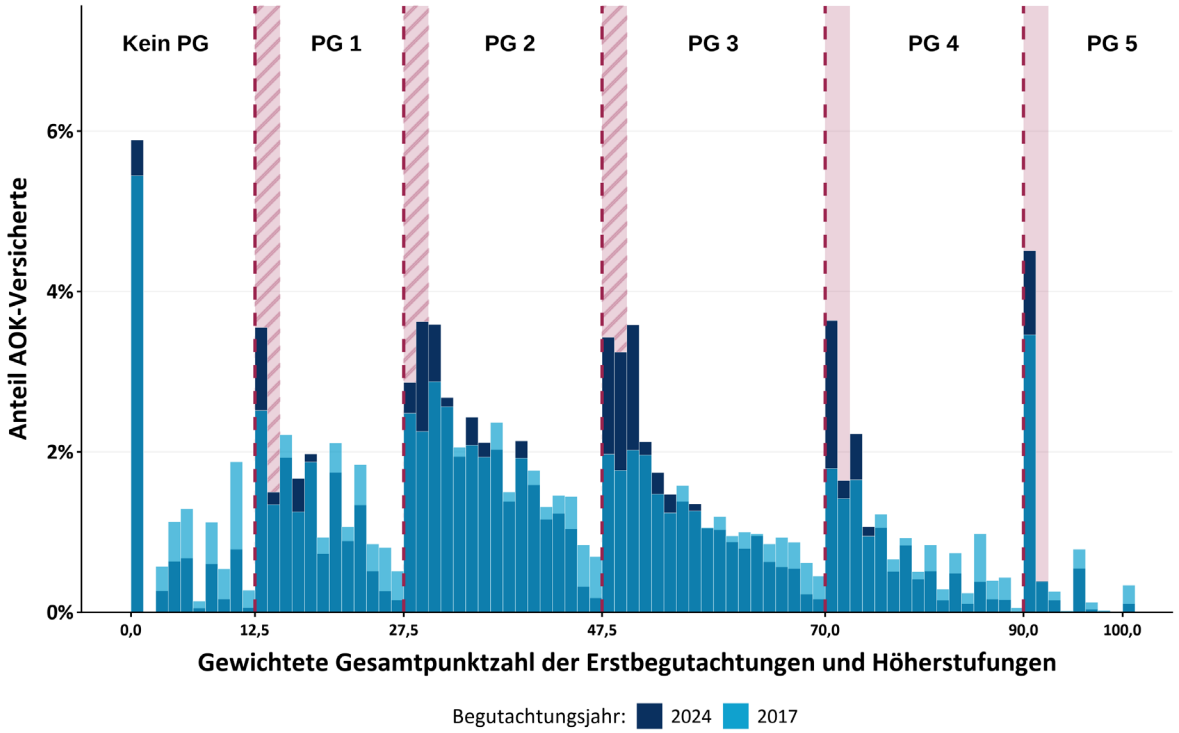
Abbildung 6: Korrelation der Modulpunkte



Quelle: MD-Begutachtungsdaten von AOK-Versicherten aus dem Jahr 2024

© WIdO 2026

Abbildung 7: Verteilung der Gesamtpunktzahl 2017 vs. 2024

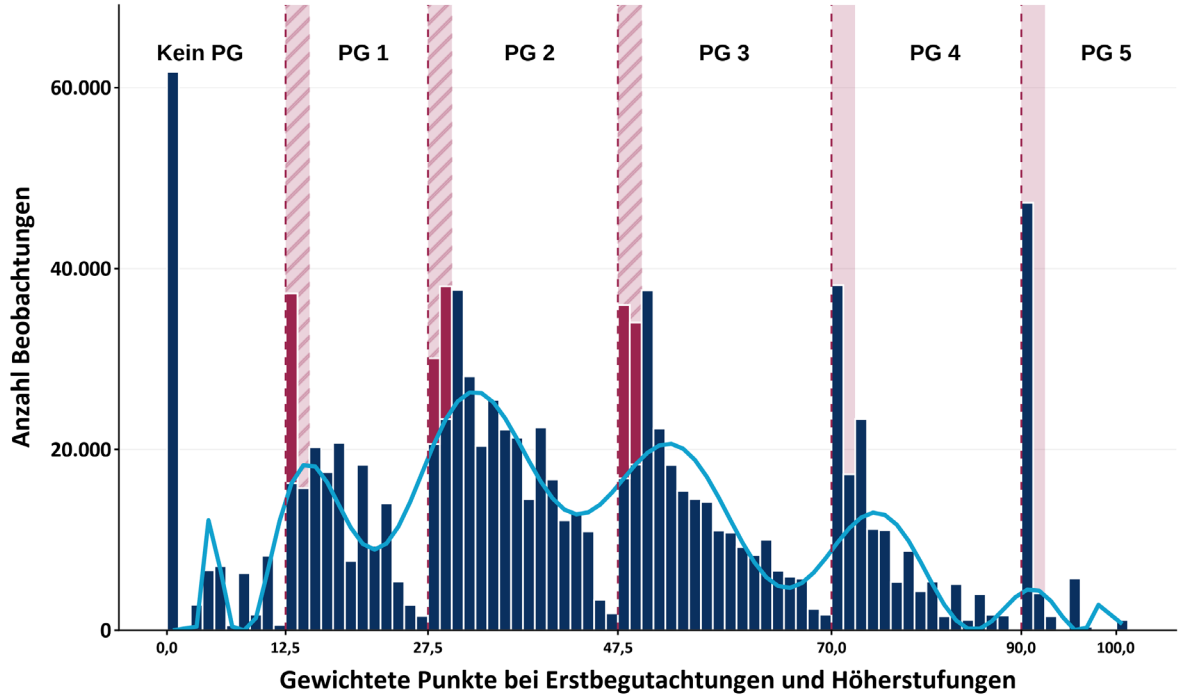


Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen.

Quelle: MD-Begutachtungsdaten von AOK-Versicherten aus den Jahren 2017 und 2024

© WIdO 2026

Abbildung 8: Modellierung der Anreiz- und Verhaltenseffekte durch eine theoretische Verteilung

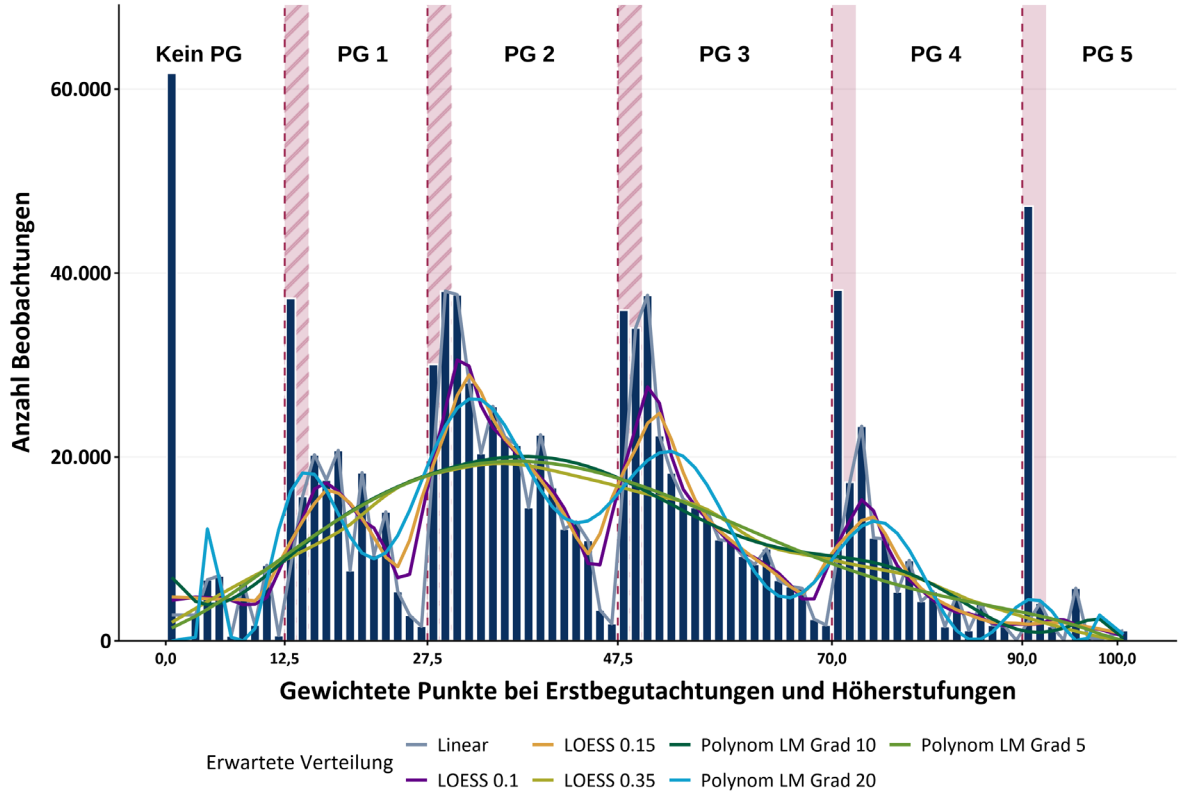


Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen. Schwellenwerte aus Fit ausgeschlossen; markiert: positive Überschussmasse an ausgewählten Schwellen im Fenster [Schwelle, Schwelle + 2,5 Punkte]. Zur Schätzung wurde ein Polynom LM Grad 20 verwendet.

Quelle: eigene Berechnungen und MD-Begutachtungsdaten von AOK-Versicherten aus dem Jahr 2024

© WIdO 2026

Abbildung 9: Vergleich der erwarteten Verteilungen

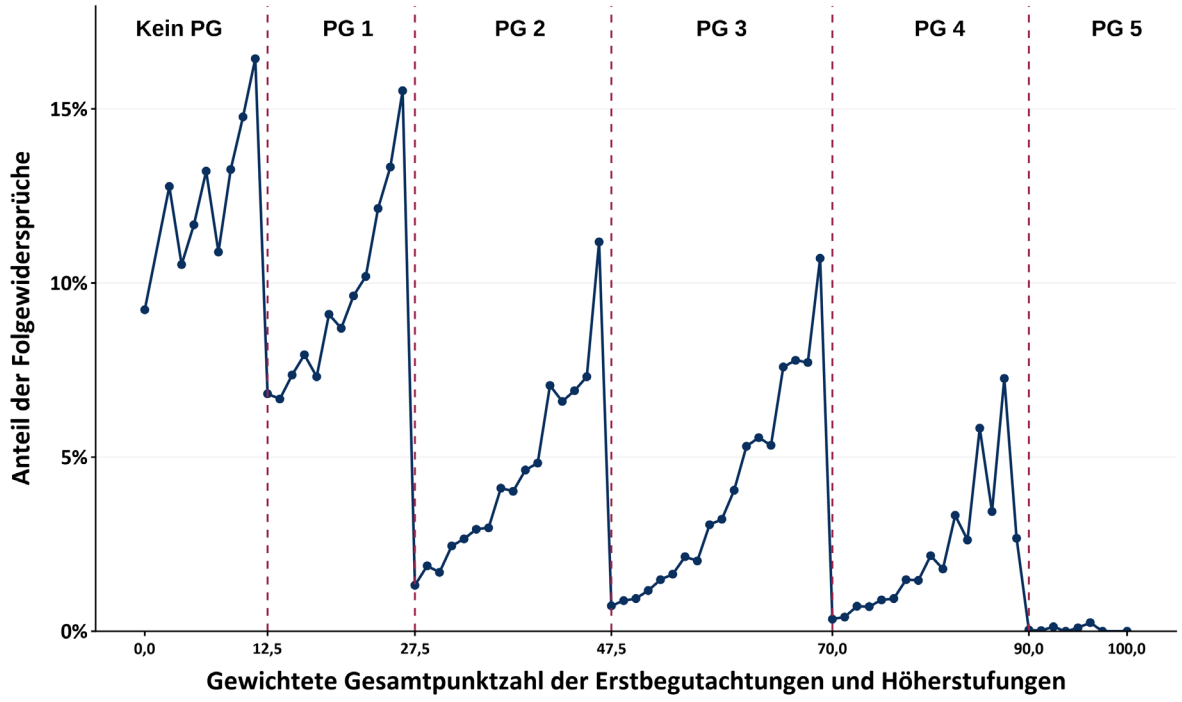


Die vertikalen roten Bereiche markieren einen Bereich von 2,5 Punkten um die bestehenden Schwellenwerte. Rot schraffiert sind die vom Reformvorschlag betroffenen Schwellenwertverschiebungen. Die Schwellenwerte wurden bei der Schätzung der Verteilungen ausgeschlossen.

Quelle: eigene Berechnungen und MD-Begutachtungsdaten von AOK-Versicherten aus dem Jahr 2024

© WIdO 2026

Abbildung 10: Widerspruchsquote nach gewichteter Gesamtpunktzahl im Jahr 2024



Quelle: MD-Begutachtungsdaten von AOK-Versicherten aus dem Jahr 2024

© WIdO 2026

Abbildungsverzeichnis

Abbildung 1: Bewertungslogik nach dem Neuen Begutachtungsassessment (NBA) 6

Abbildung 2: Verteilung der gewichteten Punkte bei Erstbegutachtungen und Höherstufungen von AOK-Versicherten 9

Abbildung 3: Prozentualer Anteil der Pflegebedürftigen, die maximal 2,5 Punkte über dem Schwellenwert liegen, gemessen an allen Pflegebedürftigen innerhalb des jeweiligen Pflegegrades 10

Abbildung 4: Verteilung der simulierten gewichteten Gesamtpunkte 15

Abbildung 5: Abweichungs-Analyse: Reale Daten vs. Simulation 15

Abbildung 6: Korrelation der Modulpunkte 24

Abbildung 7: Verteilung der Gesamtpunktzahl 2017 vs. 2024 24

Abbildung 8: Modellierung der Anreiz- und Verhaltenseffekte durch eine theoretische Verteilung 25

Abbildung 9: Vergleich der erwarteten Verteilungen 26

Abbildung 10: Widerspruchsquote nach gewichteter Gesamtpunktzahl im Jahr 2024 26

Tabellenverzeichnis

Tabelle 1: Überblick über rohe und gewichtete Punkte 7

Tabelle 2: Übersicht über die Schwellenwerte der gewichteten Gesamtpunktzahl 8

Tabelle 3: Übersicht über mögliche Einstufungsänderungen 8

Tabelle 4: Vergleich der durch eine Schwellenwertanpassung verursachten Einstufungsunterschiede unter verschiedenen Berechnungsansätzen von Anreiz- und Verhaltensanpassungseffekten 16

Tabelle 5: Durchschnittliche monatliche Kosten pro Person nach Pflegegrad und Differenz zum nächstniedrigen Pflegegrad im Jahr 2024 17

Tabelle 6: Übersicht über das theoretische jährliche Einsparpotenzial 18

Tabelle 7: Einsparpotenziale der SPV unter verschiedenen Annahmen zu Anreiz- und Verhaltensanpassungseffekten 18

Literaturverzeichnis

BMG (2013) Bericht des Expertenbeirats zur konkreten Ausgestaltung des neuen Pflegebedürftigkeitsbegriffs.

BMG (2026) Entwurf eines Gesetzes zur Neuordnung der Pflegeversicherung (Pflegeneuordnungsgesetz – PNOG) vom 05.06.26.

Cattaneo MD, Jansson M, Ma X (2018) Manipulation Testing Based on Density Discontinuity. The Stata Journal: Promoting communications on statistics and Stata 18 (1):234-261. doi:10.1177/1536867x1801800115

Chetty R, Friedman JN, Olsen T, Pistaferri L (2011) Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. The Quarterly Journal of Economics 126 (2):749-804. doi:10.1093/qje/qjr013

Dee TS, Dobbie W, Jacob BA, Rockoff J (2019) The Causes and Consequences of Test Score Manipulation: Evidence from the New York Regents Examinations. American Economic Journal: Applied Economics 11 (3):382–423. doi:10.1257/app.20170520

Diamond R, Persson P (2016) The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests. National Bureau of Economic Research Working Paper Series No. 22207. doi:10.3386/w22207

Garcia-Gomez P, Koning P, O'Donnell O, Riumalló-Herl C (2025) Selective exercise of discretion in disability insurance awards. Journal of Policy Analysis and Management 44 (3):816-835. doi:https://doi.org/10.1002/pam.22560

GKV-Spitzenverband (2026) SPV-Kennzahlen 2024. https://www.gkv-spitzenverband.de/gkv_spitzenverband/presse/zahlen_und_grafiken/spv_kennzahlen/spv_kennzahlen.jsp

Hernández-Pizarro HM, Nicodemo C, López Casasnovas G (2020) Discontinuous system of allowances: The response of prosocial health-care professionals. Journal of Public Economics 190:104248. doi:10.1016/j.jpubeco.2020.104248

Ishihara T, Takahashi M (2022) Economic Consequences of Manipulation of Social Insurance Benefits. doi:10.2139/ssrn.3784394

Jürchott K, Seidinger T, Messerle R, Behrendt S (2026) Pflegeeintritt und Erstbegutachtung bei betagten Menschen im Zeitverlauf: Ein empirischer Beitrag auf Basis von Daten der Sozialen Pflegeversicherung und der Medizinischen Dienste. WIdO e-Paper 8. doi:10.4126/FRL01-006527188

Kleven HJ (2016) Bunching. Annual Review of Economics 8:435-464. doi:10.1146/annurev-economics-080315-015234

Kleven HJ, Waseem M (2013) Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. The Quarterly Journal of Economics 128 (2):669-723. doi:10.1093/qje/qjt004

Llena-Nozal A, Araki S, Killmeier K (2025) Needs assessment and eligibility criteria in long-term care: How access is managed across OECD countries. OECD Health Working Papers 181. doi:10.1787/461811c4-en

McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics 142 (2):698-714. doi:10.1016/j.jeconom.2007.05.005

Nolting H-D, Roll P, Wolff JK (2026) Entwicklung der Pflegeprävalenzen und Weiterentwicklungsbedarf des Begutachtungsinstruments: Erklärungsfaktoren für die Entwicklung der Pflegeprävalenzen seit 2017 sowie Bestandsaufnahme zu möglichen Weiterentwicklungsbedarfen des Begutachtungsinstruments. Abschlussbericht für den GKV-Spitzenverband

Rothgang H (2026) „Umverteilung pervers“: Gesundheitsökonom Rothgang kritisiert Warken-Reform. https://www.altenheim.net/umverteilung-pervers-gesundheitsoekonom-rothgang-kritisiert-warken-reform/

Rothgang H, Burfeindt C, Czwikla J, Müller R (2025) BARMER Pflegereport 2025: Pflegeverläufe bei Akutereignissen und Erkrankungen mit progredientem Verlauf. Schriftenreihe zur Gesundheitsanalyse; Band 48

Roy D (2025) Strategic Coding in the Assessment of Long-Term Care Needs: Evidence From France. Health Economics 34 (6):1035-1063. doi:10.1002/hec.4950

Saez E (2010) Do Taxpayers Bunch at Kink Points? American Economic Journal: Economic Policy 2 (3):180-212. doi:10.1257/pol.2.3.180

Serrano-Alarcón M, Hernández-Pizarro H, López-Casasnovas G, Nicodemo C (2022) Effects of long-term care benefits on healthcare utilization in Catalonia. Journal of Health Economics 84:102645. doi:10.1016/j.jhealeco.2022.102645

Sozialverband VdK Deutschland (2026) Verschlechterungen durch Pflege-Reform: VdK ruft Mitglieder zum Handeln auf. https://www.vdk.de/presse/pressemitteilung/verschlechterungen-durch-pflege-reform-vdk-ruft-mitglieder-zum-handeln-auf/

SVR Wirtschaft (2026) Frühjahrsgutachten 2026. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung,

Zukunftspakt Pflege (2025) Fachliche Eckpunkte für eine nachhaltige Struktur- und Finanzierungsreform in der Pflegeversicherung.

Hinweis zur KI-Nutzung: Für die Erstellung dieses Beitrags wurden von den Autoren KI-Instrumente zur Unterstützung bei der Recherche und Formulierung in Anspruch genommen. Die Rechercheergebnisse und Formulierungsvorschläge sowie deren Verwendung wurden durch die Autoren persönlich sorgfältig geprüft, ggf. bearbeitet und werden von diesen verantwortet, um die Richtigkeit und Relevanz sicherzustellen.