Modellsteckbriefe für *QS HSMDEF-HSM-IMPL*

Auswertungsjahr 2025

Stand: 8. Juli 2025

Dieses Dokument enthält Hintergrundinformationen zu den im QS-Verfahren **Herzschrittmacher-Implantation** verwendeten Risikoadjustierungsmodellen, sowie eine Leseanleitung mit Erläuterungen zu den dargestellten Informationen.

Inhalt

IQTIG — Institut für Qualitätssicherung und Transparenz im Gesundheitswesen
Katharina-Heinroth-Ufer 1 10787 Berlin | info@iqtig.org | www.iqtig.org | (030) 58 58 26-0

1 QI 101800: Dosis-Flächen-Produkt

Grundgesamtheit	Alle Patientinnen und Patienten mit implantiertem Einkammer- (VVI, AAI, Leadless Pacemaker) bzw. VDD-System, Zweikammersystem (DDD) oder CRT-System, bei denen eine Durchleuchtung durchgeführt wurde
Zähler	Patientinnen und Patienten mit einem Dosis-Flächen-Produkt - über 900 cGy x cm^2 bei Einkammer- (VVI, AAI, Leadless Pacemaker) oder VDD-System - über 1.700 cGy x cm^2 bei Zweikammersystem (DDD) - über 4.900 cGy x cm^2 bei CRT-System

1.1 Datenbasis und Modellentwicklung

Die Modellschätzung basiert auf der Grundgesamtheit des Erfassungsjahres 2023.

Anzahl Fälle in der Modellschätzung	Davon mit Zählerereignis	Anteil
74.874	3.837	5,12 %

1.1.1 Leistungserbringereffekte

Das Modell wurde unter Berücksichtigung von Leistungserbringereffekten als *zufällige Effekte* geschätzt. Die geschätzte Standardabweichung der Leistungserbringereffekte beträgt $\hat{\tau} = 1{,}188$. Das genaue Vorgehen wird im Begleitdokument *Leistungserbringereffekte bei der Risikoadjustierung* beschrieben, welches u.a. mit den endgültigen Rechenregeln zum Auswertungsjahr 2022 (hier herunterladen) veröffentlicht wurde.

1.1.2 Normative Setzung von Koeffizienten

Das Modell enthält Koeffizienten, die normativ gesetzt wurden: Das Modell für die Risikofaktoren wurde zusammen mit zufälligen Effekten für Leistungserbringer geschätzt. Diese wurden für die Vorhersage der fallbasierten Risiken auf Null gesetzt und die Regressionskonstante entsprechend so angepasst, dass das Modell auf dem Entwicklungsdatensatz gesamt-kalibriert ist (O/E = 1).

1.1.3 Veränderungen zum Vorjahr

Das Vorjahresmodell wurde auf den Schätzdaten neu gefittet und leicht angepasst. Neue Einflussvariablen kamen nicht hinzu.

1.1.4 Erklärung zu „in sample"-Angaben

Mit „in sample" bezeichnete Angaben in diesem Modellsteckbrief basieren auf dem zur Modellentwicklung genutzten Datensatz und auf dem geschätzten Modell vor normativer Setzung von Koeffizienten. Demgegenüber stehen die mit „Ergebnisse der Bundesauswertung" bezeichneten Angaben. Sie basieren auf dem Datensatz der Bundesauswertung des aktuellen Auswertungsjahres

und dem für die Bundesauswertung eingesetzten Modell (s. Abschnitt „Beschreibung des Risikomodells“), also nach normativer Setzung von Koeffizienten. Die geschätzten Leistungserbringereffekte werden für die Schätzung der „in sample“-Risiken berücksichtigt.

1.1.5 Weiterführende Informationen

Detaillierte Informationen über die Datenerhebung und die Berechnung der Qualitätsindikatoren entnehmen Sie bitte den über die Verfahrensübersicht des IQTIG zugänglichen Dokumenten zu Spezifikation und Rechenregeln. Die Bundesauswertungen des IQTIG liefern im Kapitel *Basisauswertung* zudem beschreibende Statistiken zur Grundgesamtheit des QS-Verfahrens.

1.2 Beschreibung des Risikomodells

Risikokoeffizienten aus der logistischen Regression. Die Referenzwahrscheinlichkeit beträgt 7,1 % (Odds: 0,0764).

Risikofaktor	Regressions-koeffizient	Std.-Fehler	Z-Wert	Odds-Ratio (mit 95 %-Vertrauensbereich)
Konstante	-2,571423	0,055902	-49,58	
BMI				
BMI linear bis 32	0,083355	0,005502	15,15	
BMI linear ab 32	0,049370	0,006582	7,50	
BMI unbekannt oder unplausibel	0,171398	0,087269	1,96	1,187 (1,000 – 1,408)

1.2.1 Odds-Ratios

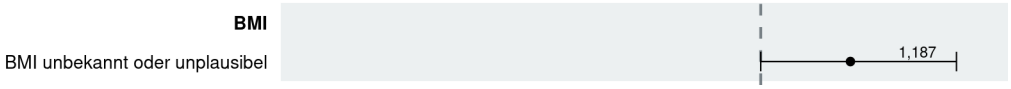


Abbildung 1: Odds-Ratios (grafische Darstellung).

1.2.2 Einfluss stetiger Variablen

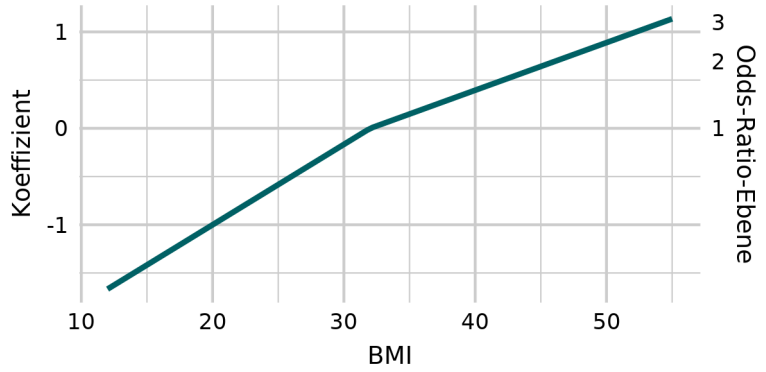


Abbildung 2: Einfluss der stetigen Variable BMI.

1.2.3 Verteilung der Risiken (Ergebnisse der Bundesauswertung)

Die folgende Grafik und Tabelle zeigt die bundesweite Verteilung der geschätzten Risiken für das Auswertungsjahr 2025

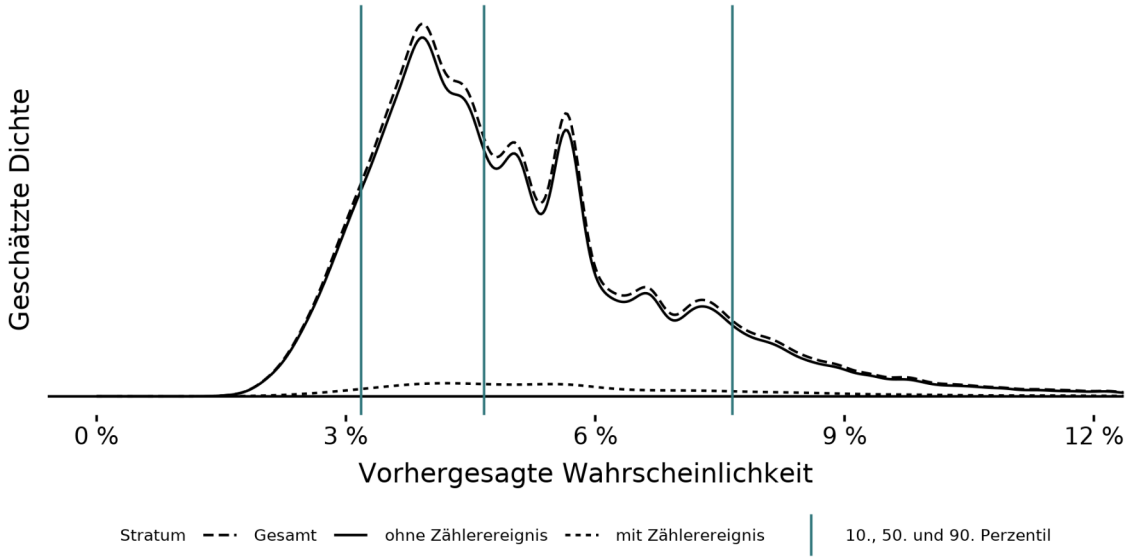


Abbildung 3: Dichtediagramm zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

Statistiken zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

		Geschätzte Risiken	
Ereignis	**Anzahl Fälle**	**Mittelwert**	**Median**
mit Zählerereignis	3.875	5,84 %	5,36 %
ohne Zählerereignis	72.830	5,08 %	4,63 %
Gesamt	**76.705**	**5,12 %**	**4,66 %**

1.3 Eigenschaften des geschätzten Modells

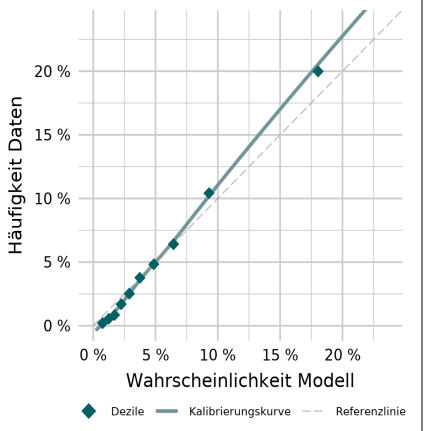
1.3.1 Kennzahlen zur Vorhersagegüte

	AUC	Brier-Score	Nagelkerkes Pseudo-R^2
in sample	0,800	0,045	0,176

1.3.2 Kalibrierung (in sample)

Kalibrierungstabelle nach Risiko-Dezilen (in sample).

	Dezil	Erwartet	Beobachtet	Statistik	Kalibrierungsdiagramm
1	[0,00195 ... 0,0102]	0,75 %	0,21 %	29,17	
2	(0,0102 ... 0,0147]	1,25 %	0,55 %	29,67	
3	(0,0147 ... 0,0197]	1,71 %	0,84 %	33,69	
4	(0,0197 ... 0,0254]	2,25 %	1,70 %	10,42	
5	(0,0254 ... 0,0329]	2,90 %	2,52 %	3,74	
6	(0,0329 ... 0,0426]	3,75 %	3,77 %	0,01	
7	(0,0426 ... 0,0555]	4,86 %	4,84 %	0,01	
8	(0,0555 ... 0,0756]	6,45 %	6,41 %	0,02	
9	(0,0756 ... 0,115]	9,29 %	10,42 %	11,34	
10	(0,115 ... 0,704]	18,04 %	19,99 %	19,33	



In den Spalten 'Erwartet' und 'Beobachtet' sind die jeweiligen Mittelwerte dargestellt. Ein Aufsummieren der Spalte 'Statistik' ergibt die Teststatistik nach Hosmer und Lemeshow zur Modellkalibrierung.

2 QI 51191: Sterblichkeit im Krankenhaus

Grundgesamtheit	Alle Patientinnen und Patienten
Zähler	Verstorbene Patientinnen und Patienten

2.1 Datenbasis und Modellentwicklung

Die Modellschätzung basiert auf der Grundgesamtheit des Erfassungsjahres 2023.

Anzahl Fälle in der Modellschätzung	Davon mit Zählerereignis	Anteil
75.305	1.101	1,46 %

2.1.1 Leistungserbringereffekte

Das Modell wurde unter Berücksichtigung von Leistungserbringereffekten als *zufällige Effekte* geschätzt. Die geschätzte Standardabweichung der Leistungserbringereffekte beträgt $\hat{\tau} = 0{,}591$. Das genaue Vorgehen wird im Begleitdokument *Leistungserbringereffekte bei der Risikoadjustierung* beschrieben, welches u.a. mit den endgültigen Rechenregeln zum Auswertungsjahr 2022 (hier herunterladen) veröffentlicht wurde.

2.1.2 Normative Setzung von Koeffizienten

Das Modell enthält Koeffizienten, die normativ gesetzt wurden: Das Modell für die Risikofaktoren wurde zusammen mit zufälligen Effekten für Leistungserbringer geschätzt. Diese wurden für die Vorhersage der fallbasierten Risiken auf Null gesetzt und die Regressionskonstante entsprechend so angepasst, dass das Modell auf dem Entwicklungsdatensatz gesamt-kalibriert ist (O/E = 1).

2.1.3 Veränderungen zum Vorjahr

Das Vorjahresmodell wurde auf den Schätzdaten neu gefittet und leicht angepasst. Neue Einflussvariablen kamen nicht hinzu.

2.1.4 Erklärung zu „in sample"-Angaben

Mit „in sample" bezeichnete Angaben in diesem Modellsteckbrief basieren auf dem zur Modellentwicklung genutzten Datensatz und auf dem geschätzten Modell vor normativer Setzung von Koeffizienten. Demgegenüber stehen die mit „Ergebnisse der Bundesauswertung" bezeichneten Angaben. Sie basieren auf dem Datensatz der Bundesauswertung des aktuellen Auswertungsjahres und dem für die Bundesauswertung eingesetzten Modell (s. Abschnitt „Beschreibung des Risikomodells"), also nach normativer Setzung von Koeffizienten. Die geschätzten Leistungserbringereffekte werden für die Schätzung der „in sample"-Risiken berücksichtigt.

2.1.5 Weiterführende Informationen

Detaillierte Informationen über die Datenerhebung und die Berechnung der Qualitätsindikatoren entnehmen Sie bitte den über die Verfahrensübersicht des IQTIG zugänglichen Dokumenten zu Spezifikation und Rechenregeln. Die Bundesauswertungen des IQTIG liefern im Kapitel *Basisauswertung* zudem beschreibende Statistiken zur Grundgesamtheit des QS-Verfahrens.

2.2 Beschreibung des Risikomodells

Risikokoeffizienten aus der logistischen Regression. Die Referenzwahrscheinlichkeit beträgt 0,25 % (Odds: 0,0025).

Risikofaktor	Regressions-koeffizient	Std.-Fehler	Z-Wert	Odds-Ratio (mit 95 %-Vertrauensbereich)
Konstante	-5,986611	0,108733	-54,99	
Alter (linear zwischen 70 und 95 Jahren)	0,034645	0,004719	7,34	
ASA-Klassifikation				
ASA-Klassifikation 3	1,153509	0,095723	12,05	3,169 (2,627 – 3,823)
ASA-Klassifikation 4	2,428055	0,110429	21,99	11,337 (9,130 – 14,076)
ASA-Klassifikation 5	3,819037	0,231474	16,50	45,560 (28,944 – 71,716)
Ätiologie: infarktbedingt	0,628663	0,177897	3,53	1,875 (1,323 – 2,657)
AV-Block				
AV-Block I. oder II. Grades	-0,594305	0,104662	-5,68	0,552 (0,450 – 0,678)
AV-Block III. Grades	0,151269	0,068165	2,22	1,163 (1,018 – 1,330)
Nierenfunktion				
Nierenfunktion: Kreatinin > 1,5 mg/dl bis <= 2,5 mg/dl	0,872433	0,074899	11,65	2,393 (2,066 – 2,771)
Nierenfunktion: Kreatinin > 2,5 mg/dl, nicht dialysepflichtig	1,573443	0,111310	14,14	4,823 (3,878 – 5,999)
Nierenfunktion: Kreatinin > 2,5 mg/dl, dialysepflichtig	1,853454	0,118129	15,69	6,382 (5,063 – 8,044)

2.2.1 Odds-Ratios

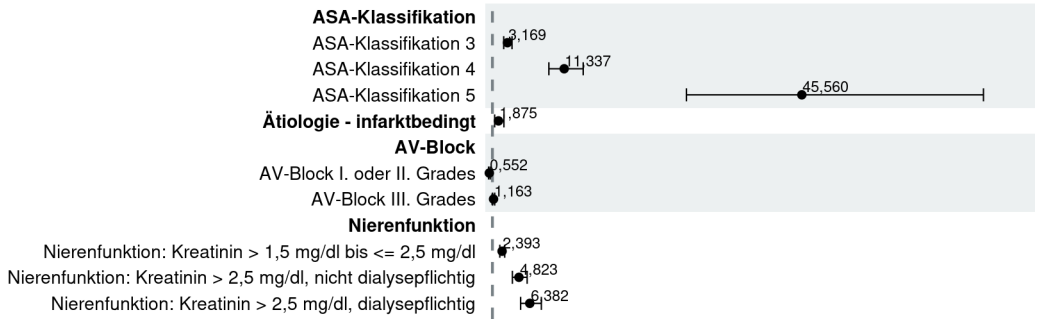


Abbildung 4: Odds-Ratios (grafische Darstellung).

2.2.2 Einfluss stetiger Variablen

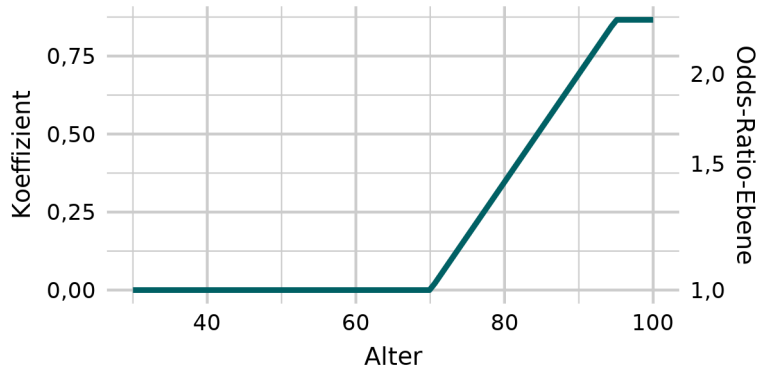


Abbildung 5: Einfluss der stetigen Variable Alter.

2.2.3 Verteilung der Risiken (Ergebnisse der Bundesauswertung)

Die folgende Grafik und Tabelle zeigt die bundesweite Verteilung der geschätzten Risiken für das Auswertungsjahr 2025

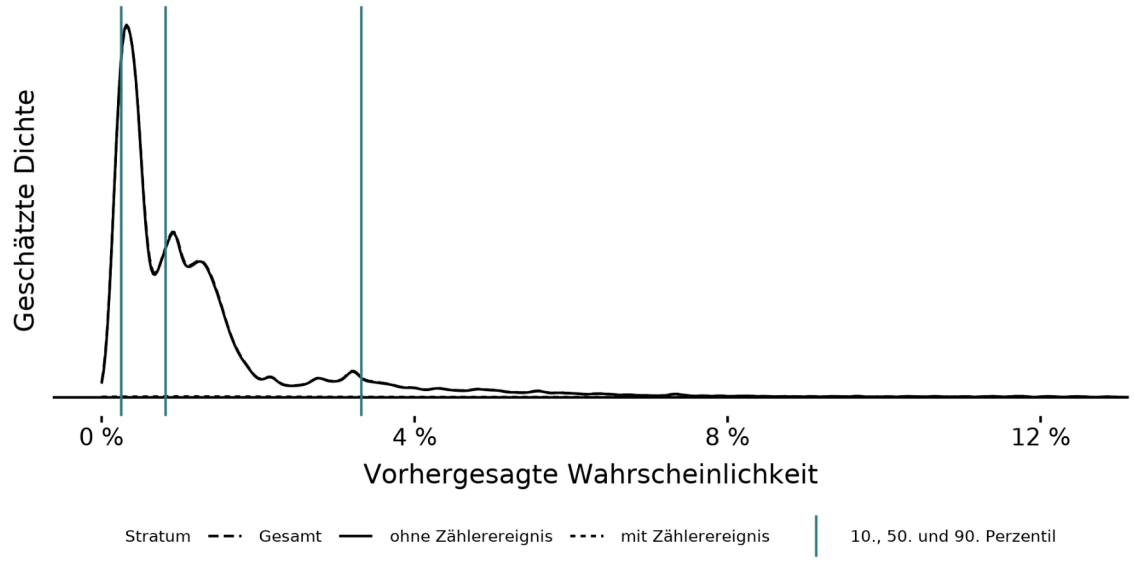


Abbildung 6: Dichtediagramm zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

Statistiken zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

		Geschätzte Risiken	
Ereignis	**Anzahl Fälle**	**Mittelwert**	**Median**
mit Zählerereignis	1.009	5,19 %	2,54 %
ohne Zählerereignis	76.213	1,45 %	0,81 %
Gesamt	**77.222**	**1,49 %**	**0,82 %**

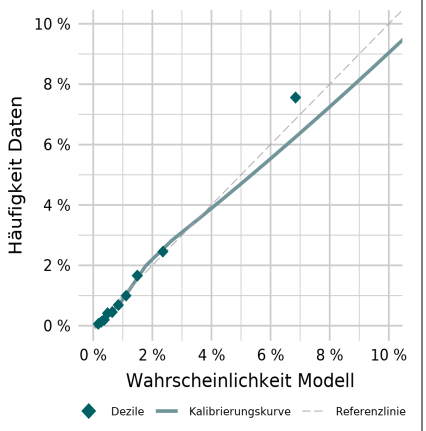
2.3 Eigenschaften des geschätzten Modells

2.3.1 Kennzahlen zur Vorhersagegüte

	AUC	Brier-Score	Nagelkerkes Pseudo-R^2
in sample	0,834	0,014	0,183

2.3.2 Kalibrierung (in sample)

Kalibrierungstabelle nach Risiko-Dezilen (in sample).

	Dezil	Erwartet	Beobachtet	Statistik	Kalibrierungsdiagramm
1	[0,000494 ... 0,00232]	0,17 %	0,07 %	5,04	
2	(0,00232 ... 0,00326]	0,28 %	0,13 %	5,79	
3	(0,00326 ... 0,00426]	0,37 %	0,20 %	6,19	
4	(0,00426 ... 0,0056]	0,49 %	0,41 %	0,92	
5	(0,0056 ... 0,0074]	0,65 %	0,45 %	4,46	
6	(0,0074 ... 0,00969]	0,85 %	0,69 %	2,27	
7	(0,00969 ... 0,0127]	1,11 %	1,00 %	0,89	
8	(0,0127 ... 0,018]	1,49 %	1,66 %	1,41	
9	(0,018 ... 0,0314]	2,36 %	2,46 %	0,31	
10	(0,0314 ... 0,685]	6,85 %	7,56 %	5,97	

In den Spalten 'Erwartet' und 'Beobachtet' sind die jeweiligen Mittelwerte dargestellt. Ein Aufsummieren der Spalte 'Statistik' ergibt die Teststatistik nach Hosmer und Lemeshow zur Modellkalibrierung.

3 QI 52311: Peri- bzw. postoperative Komplikationen während des stationären Aufenthalts

Grundgesamtheit	Alle Patientinnen und Patienten
Zähler	Patientinnen und Patienten mit Sondendislokation oder -dysfunktion

3.1 Datenbasis und Modellentwicklung

Die Modellschätzung basiert auf der Grundgesamtheit des Erfassungsjahres 2023.

Anzahl Fälle in der Modellschätzung	Davon mit Zählerereignis	Anteil
75.305	1.041	1,38 %

3.1.1 Leistungserbringereffekte

Das Modell wurde unter Berücksichtigung von Leistungserbringereffekten als *zufällige Effekte* geschätzt. Die geschätzte Standardabweichung der Leistungserbringereffekte beträgt $\hat{\tau}$ = 0,935. Das genaue Vorgehen wird im Begleitdokument *Leistungserbringereffekte bei der Risikoadjustierung* beschrieben, welches u.a. mit den endgültigen Rechenregeln zum Auswertungsjahr 2022 (hier herunterladen) veröffentlicht wurde.

3.1.2 Normative Setzung von Koeffizienten

Das Modell enthält Koeffizienten, die normativ gesetzt wurden: Das Modell für die Risikofaktoren wurde zusammen mit zufälligen Effekten für Leistungserbringer geschätzt. Diese wurden für die Vorhersage der fallbasierten Risiken auf Null gesetzt und die Regressionskonstante entsprechend so angepasst, dass das Modell auf dem Entwicklungsdatensatz gesamt-kalibriert ist (O/E = 1).

3.1.3 Veränderungen zum Vorjahr

Das Vorjahresmodell wurde auf den Schätzdaten neu gefittet und leicht angepasst. Eingriffsart Systemumstellung wurde aus dem Modell entfernt. Neue Einflussvariablen kamen nicht hinzu.

3.1.4 Erklärung zu „in sample"-Angaben

Mit „in sample" bezeichnete Angaben in diesem Modellsteckbrief basieren auf dem zur Modellentwicklung genutzten Datensatz und auf dem geschätzten Modell vor normativer Setzung von Koeffizienten. Demgegenüber stehen die mit „Ergebnisse der Bundesauswertung" bezeichneten Angaben. Sie basieren auf dem Datensatz der Bundesauswertung des aktuellen Auswertungsjahres und dem für die Bundesauswertung eingesetzten Modell (s. Abschnitt „Beschreibung des Risikomodells"), also nach normativer Setzung von Koeffizienten. Die geschätzten Leistungserbringereffekte werden für die Schätzung der „in sample"-Risiken berücksichtigt.

3.1.5 Weiterführende Informationen

Detaillierte Informationen über die Datenerhebung und die Berechnung der Qualitätsindikatoren entnehmen Sie bitte den über die Verfahrensübersicht des IQTIG zugänglichen Dokumenten zu Spezifikation und Rechenregeln. Die Bundesauswertungen des IQTIG liefern im Kapitel *Basisauswertung* zudem beschreibende Statistiken zur Grundgesamtheit des QS-Verfahrens.

3.2 Beschreibung des Risikomodells

Risikokoeffizienten aus der logistischen Regression. Die Referenzwahrscheinlichkeit beträgt 0,46 % (Odds: 0,0046).

Risikofaktor	Regressions-koeffizient	Std.-Fehler	Z-Wert	Odds-Ratio (mit 95 %-Vertrauensbereich)
Konstante	-5,372197	0,159313	-34,51	
Geschlecht männlich	0,126982	0,065102	1,95	1,135 (0,999 – 1,290)
BMI				
BMI linear	0,020531	0,006094	3,37	
BMI unbekannt oder unplausibel	0,079764	0,159017	0,50	1,083 (0,793 – 1,479)
ASA-Klassifikation 3, 4 oder 5	0,223012	0,073337	3,04	1,250 (1,083 – 1,443)
Herzinsuffizienz NYHA IV	0,119305	0,299973	0,40	1,127 (0,626 – 2,028)
Vorhofflimmern	-0,061597	0,074949	-0,82	0,940 (0,812 – 1,089)
LVEF				
LVEF linear bis 65	0,006206	0,005054	1,23	
LVEF unbekannt	0,388181	0,163545	2,37	1,474 (1,070 – 2,031)
Systemart				
Systemart Zweikammersystem	1,039956	0,134790	7,72	2,829 (2,172 – 3,685)
Systemart CRT	0,514041	0,208301	2,47	1,672 (1,112 – 2,515)

3.2.1 Odds-Ratios

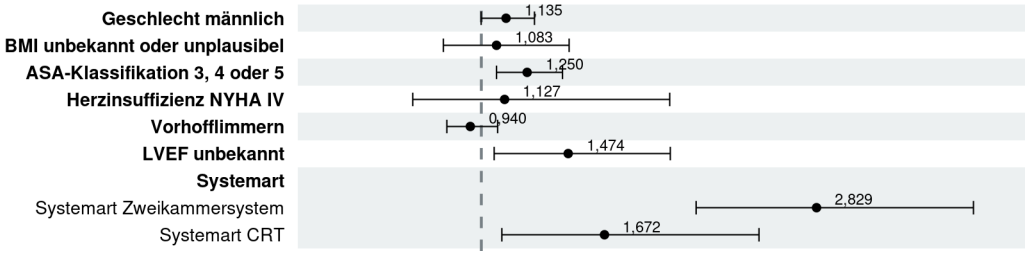


Abbildung 7: Odds-Ratios (grafische Darstellung).

3.2.2 Einfluss stetiger Variablen

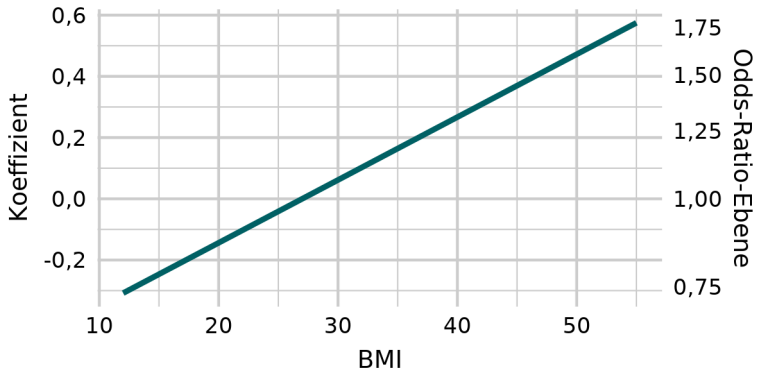


Abbildung 8: Einfluss der stetigen Variable BMI.

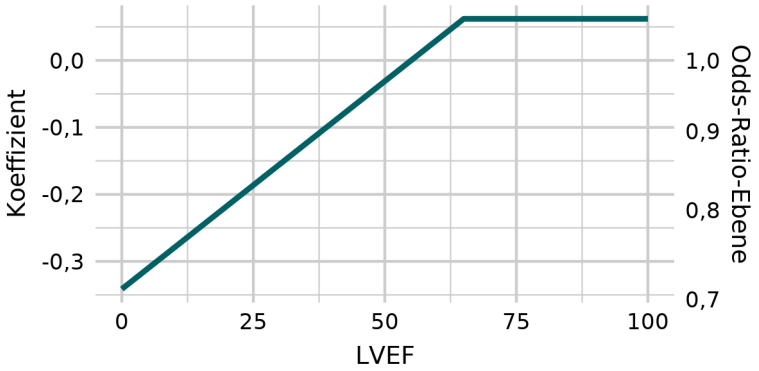


Abbildung 9: Einfluss der stetigen Variable LVEF.

3.2.3 Verteilung der Risiken (Ergebnisse der Bundesauswertung)

Die folgende Grafik und Tabelle zeigt die bundesweite Verteilung der geschätzten Risiken für das Auswertungsjahr 2025

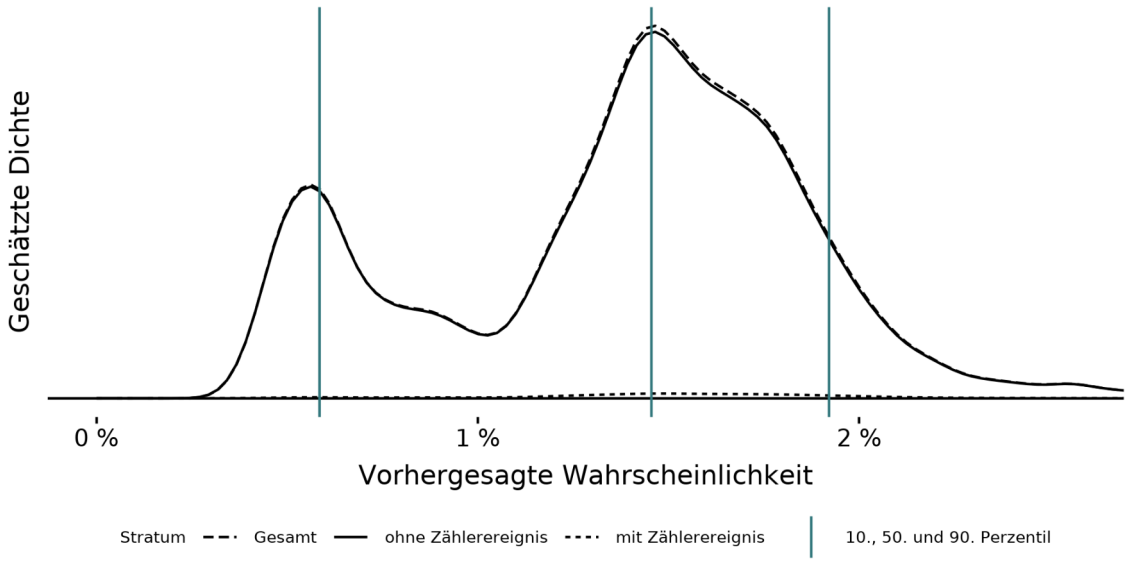


Abbildung 10: Dichtediagramm zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

Statistiken zur Verteilung der Risiken (Ergebnisse der Bundesauswertung).

		Geschätzte Risiken	
Ereignis	**Anzahl Fälle**	**Mittelwert**	**Median**
mit Zählerereignis	1.007	1,52 %	1,54 %
ohne Zählerereignis	76.215	1,38 %	1,45 %
Gesamt	**77.222**	**1,38 %**	**1,46 %**

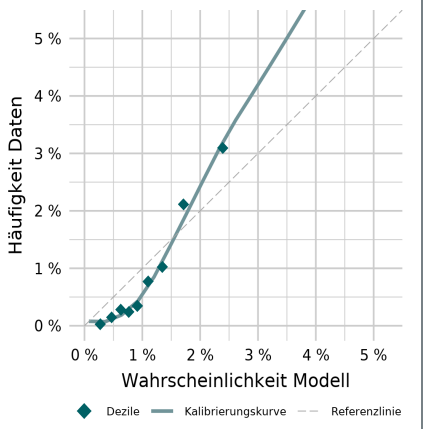
3.3 Eigenschaften des geschätzten Modells

3.3.1 Kennzahlen zur Vorhersagegüte

	AUC	Brier-Score	Nagelkerkes Pseudo-R^2
in sample	0,817	0,013	0,121

3.3.2 Kalibrierung (in sample)

Kalibrierungstabelle nach Risiko-Dezilen (in sample).

	Dezil	Erwartet	Beobachtet	Statistik	Kalibrierungsdiagramm
1	[0,000754 ... 0,00376]	0,27 %	0,03 %	16,80	
2	(0,00376 ... 0,00555]	0,47 %	0,15 %	16,85	
3	(0,00555 ... 0,00699]	0,63 %	0,28 %	14,75	
4	(0,00699 ... 0,00834]	0,77 %	0,24 %	27,52	
5	(0,00834 ... 0,00998]	0,91 %	0,35 %	26,78	
6	(0,00998 ... 0,0121]	1,10 %	0,77 %	7,53	
7	(0,0121 ... 0,015]	1,35 %	1,02 %	5,94	
8	(0,015 ... 0,0197]	1,71 %	2,11 %	7,13	
9	(0,0197 ... 0,0297]	2,39 %	3,09 %	15,95	
10	(0,0297 ... 0,224]	4,22 %	5,79 %	45,59	

In den Spalten 'Erwartet' und 'Beobachtet' sind die jeweiligen Mittelwerte dargestellt. Ein Aufsummieren der Spalte 'Statistik' ergibt die Teststatistik nach Hosmer und Lemeshow zur Modellkalibrierung.

4 Leseanleitung zu den Modellsteckbriefen

Zu einigen Risikoadjustierungsmodellen veröffentlicht das IQTIG ergänzend zu den Angaben in der QIDB sogenannte Modellsteckbriefe. Sie enthalten detaillierte Informationen über die Entstehung der Modelle sowie ihre statistischen Eigenschaften. Für eine Einführung in die Risikoadjustierung von Qualitätsindikatoren siehe das entsprechende Kapitel in der Leseanleitung zur Bundesauswertung.

Die Modellsteckbriefe gliedern sich in drei Abschnitte: „Datenbasis und Modellentwicklung", „Risikomodell des Qualitätsindikators" und „Eigenschaften des geschätzten Modells".

4.1 Datenbasis und Modellentwicklung

Dieser Abschnitt enthält allgemeine Informationen über das Zustandekommen des Modells, die zugrundeliegenden Daten und eventuelle Veränderungen zum Vorjahr.

4.2 Beschreibung des Risikomodells

Dieser Abschnitt enthält eine Beschreibung des für die Berechnung des Qualitätsindikators / der Qualitätsindikatoren in der Bundesauswertung genutzten Modells. Neben der bereits in der QIDB enthaltenen Tabelle der Koeffizienten der Risikofaktoren werden die Einflüsse der diskreten und stetigen Risikofaktoren graphisch dargestellt.

Schließlich wird die empirische Verteilung der Risiken mittels Dichteschätzer dargestellt. Dies vermittelt einen Eindruck darüber, wie sich die Grundgesamtheit des Modells in Hochrisiko- und Niedrigrisikofälle verteilt. Die Verteilung wird stratifiziert für die Fälle mit und ohne Zählerereignisse dargestellt (grob- bzw. feingestrichelte Linie). Die Darstellung ist so gewählt, dass die Fläche unter der grobgestrichelten, feingestrichelten bzw. durchgezogenen Linie proportional ist zur Anzahl an Fällen mit bzw. ohne Zählerereignissen bzw. zur Gesamtanzahl an Fällen. Vertikale Linien markieren das 10., 50. (Median) und 90. Perzentil der Gesamtverteilung.

Zusätzlich zur grafischen Darstellung werden hier Maße für die Lage der geschätzten Risiken tabellarisch berichtet, ebenfalls stratifiziert für die Fälle mit und ohne Zählerereignisse. Dies vermittelt, wie stark sich die Modellvorhersagen im Mittel und im Median zwischen Fällen mit und ohne Zählerereignis unterscheiden. Die Größe dieses Unterschieds wird auch oft als Maß für die Diskriminationsfähigkeit eines Modells interpretiert. Dazu muss allerdings auch die Gesamtprävalenz des Zählerereignisses berücksichtigt werden.

4.3 Eigenschaften des geschätzten Modells

In diesem Abschnitt werden statistische Eigenschaften des Risikoadjustierungsmodells dargestellt. Die in diesem Abschnitt dargestellten Modelleigenschaften beschreiben das Verhältnis des Modells mit der empirischen Grundgesamtheit des Modells. Falls das für die Bundesauswertung genutzte Risikoadjustierungsmodell normativ gesetzte Koeffizienten enthält, so werden die ent-

sprechenden Setzungen in diesem Abschnitt nicht berücksichtigt (siehe dazu auch den Unterabschnitt „Erklärung zu , in sample'-Angaben" im Modellsteckbrief). Dies ist darin begründet, dass die normative Setzung datenunabhängig geschieht.

Die statistischen Eigenschaften lassen sich dabei grundsätzlich *in sample* und *out of sample* berechnen. *In sample* bedeutet, dass die jeweilige Eigenschaft auf Grundlage des Datensatzes berechnet wird, der auch für die Schätzung des Modells selbst genutzt wurde. *Out of sample* bedeutet, dass die jeweilige Modelleigenschaft auf einem anderen Datensatz (z. B. den Daten aus einem anderen Erfassungsjahr) berechnet wurde. Jeder Modellsteckbrief enthält mindestens Informationen zu den *in-sample*-Eigenschaften des Modells. Darüber hinaus sind die Modelleigenschaften unter Umständen auch für einen oder mehrere *out-of-sample*-Datensätze dargestellt – ebenfalls vor normativen Setzungen von Koeffizienten. Unterschiede zwischen *in-sample*- und *out-of-sample*-Eigenschaften können auf Überanpassung hindeuten.

4.3.1 Kennzahlen zur Vorhersagegüte

Bei den drei dargestellten Kennzahlen AUC, Brier-Score und Nagelkerkes Pseudo-R^2 handelt es sich um weitverbreitete und wichtige Maßzahlen für klinische Vorhersagemodelle. Für die genaue Definition wird z.B. auf das Buch *Clinical Prediction Models* von Ewout Steyerberg (2. Auflage, Springer 2019) verwiesen.

Die **AUC** (*area under the ROC curve*, auch *c-statistic* bzw. Konkordanzstatistik) ist ein Maß für die Diskriminationsfähigkeit des Modells. Der Wert liegt zwischen 0.5 und 1. Der Wert 0.5 tritt nur dann auf, wenn sämtlichen Fällen das gleiche Risiko zugewiesen wird.

Der **Brier-Score** beschreibt, wie gut das Modell das Eintreten von Zählerereignissen vorhersagen kann. Der Wert liegt in der Regel zwischen 0 und 0.25. Der Wert 0.25 tritt nur dann auf, wenn sämtlichen Fällen das Risiko 0.5 zugewiesen wird.

Nagelkerkes Pseudo-R^2 ist ein Maß für den Anteil an der Variabilität oder Unsicherheit über das Zählerereignis, den das Modell erklärt. Der Wert liegt zwischen 0 und 1. Der Wert 0 tritt nur dann auf, wenn sämtlichen Fällen das gleiche Risiko zugewiesen wird.

Es handelt sich bei diesen Kennzahlen nicht um Gütekriterien für Risikoadjustierungsmodelle: Perfekte Werte (also ein AUC von 1, ein Brier-Score von 0 und ein Pseudo-R^2 von 1) sind bei Risikoadjustierungsmodellen nicht möglich und auch nicht wünschenswert, da solche perfekten Werte nur dann auftreten können, wenn sich aus den Risikofaktoren sicher vorhersagen lässt, ob ein interessierende Zählerereignis auftritt oder nicht. In der externen Qualitätssicherung werden jedoch Ereignisse betrachtet, bei denen die Leistungserbringer einen starken Einfluss haben. „Schlechte" Werte der Kennzahlen deuten also auf einen schwachen Einfluss der berücksichtigten Risikofaktoren auf das Zählerereignis hin. Solche Werte treten daher unter anderem dann auf, wenn es, wie beispielsweise bei vielen Prozessindikatoren, nur wenige patientenseitige Risikofaktoren gibt, die berücksichtigt werden sollen.

4.3.2 Kalibrierung

Die Kalibrierung beschreibt, inwiefern die vorhergesagten Risiken mit beobachteten Häufigkeiten in den Daten zusammen passen: Unter allen Fällen mit einem Risiko von x % wird erwartet, dass der Anteil von Fällen mit beobachtetem Zählerereignis in etwa x % beträgt.

Die Kalibrierung wird gelegentlich mit dem Hosmer-Lemeshow-Test überprüft. Dieser Test ist jedoch umstritten: Bei großen Fallzahlen kann der Test sehr kleine p-Werte liefern, auch wenn die Verletzung der Kalibrierung nicht relevant erscheint. Zudem wird in der wissenschaftlichen Literatur diskutiert, welche Anzahl an Freiheitsgraden man für die Verteilung der Teststatistik annehmen sollte. Aussagekräftiger sind die Kalibrierungstabelle sowie das Kalibrierungsdiagramm, aus denen ersichtlich wird, in welchen Bereichen die Kalibrierung verletzt wird. In der Kalibrierungstabelle werden die Fälle nach ihrem vorhergesagten Risiko in zehn gleich große Gruppen, sogenannte Dezile, eingeteilt und geprüft, wie weit innerhalb dieser Gruppen die vorhergesagte und beobachtete Anzahl an Zählerereignissen auseinanderliegen. Diese Gruppierung liegt auch dem Hosmer-Lemeshow-Test zugrunde. Für jede Gruppe ist daher der Beitrag zur Hosmer-Lemeshow-Statistik dargestellt. Im Kalibrierungsdiagramm ist die Kalibrierung ohne Diskretisierung mit Hilfe eines Glätters dargestellt (siehe z.B. *Clinical Prediction Models* von Ewout Steyerberg, 2. Auflage, Springer 2019). Ergänzt wird die Darstellung durch die Werte für die Dezile aus der Kalibrierungstabelle, die als Punkte eingetragen sind.