



Institut für Qualitätssicherung und
Transparenz im Gesundheitswesen

Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL

Stand: 6. April 2017

Impressum

Thema:

Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL

Autoren:

Johannes Hengelbrock, Dr. Michael Höhle

Stand:

6. April 2017

Zitierhinweis:

IQTIG (2017): Ereigniszeitanalyse-Methodik für die Follow-up-Indikatoren nach QSKH-RL.

Stand: 06.04.2017. Berlin: IQTIG. URL: <https://www.iqtig.org/downloads/berichte/2017/>

[IQTIG Ereigniszeitanalyse-Methodik-für-Follow-up-Indikatoren-nach-QSKH-RL_2017-04-06.pdf](https://www.iqtig.org/downloads/berichte/2017/IQTIG_Ereigniszeitanalyse-Methodik-für-Follow-up-Indikatoren-nach-QSKH-RL_2017-04-06.pdf)

Herausgeber:

IQTIG – Institut für Qualitätssicherung
und Transparenz im Gesundheitswesen

Katharina-Heinroth-Ufer 1
10787 Berlin

Telefon: (030) 58 58 26-0
Telefax: (030) 58 58 26-999

info@iqtig.org

<http://www.iqtig.org>

Inhaltsverzeichnis

Abbildungsverzeichnis	4
Abkürzungsverzeichnis	4
1 Zusammenfassung	5
2 Hintergrund	6
2.1 Follow-up-Indikatoren	6
2.2 Datengrundlage	7
2.3 Zensierungen	8
2.4 Erfassungsprobleme	10
2.5 Weitergehende Zensierungsproblematik	10
2.6 Zusammenfassung	11
3 Ereigniszeitanalyse	12
3.1 Notation	12
3.1.1 Survivorfunktion und Hazardrate	12
3.2 Kaplan-Meier-Schätzer für die Survivorfunktion	13
3.2.1 Konfidenzintervalle	15
3.3 Auswertung eines nicht-risikoadjustierten Follow-up-Indikators	16
3.4 Risikoadjustierung	17
3.5 SMR aus multiplikativem Hazardmodell	18
3.5.1 Konfidenzintervalle für das SMR	19
3.6 Auswertung eines risikoadjustierten Follow-up-Indikators	20
3.7 Zusammenfassung	20
4 Berechnung in R	22
5 Diskussion und Ausblick	24
Literatur	25

Abbildungsverzeichnis

Abbildung 1: Exemplarisches Lexis-Diagramm der Follow-up-Daten für drei Ersteingriffe 9
Abbildung 2: Kaplan-Meier-Kurven für alle analysierten Ersteingriffe im EJ 2015 14

Abkürzungsverzeichnis

Abkürzung	Erläuterung
KI	Konfidenzintervall
PSN	Patientenpseudonym(e)
QSKH-RL	Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern
SMR	Standardized Mortality Ratio

1 Zusammenfassung

Die Auswertungsanalytik für die Follow-up-Indikatoren gemäß der Richtlinie über Maßnahmen der Qualitätssicherung in Krankenhäusern (QSKH-RL) wird anhand von statistischen Verfahren aus der Ereigniszeitanalyse formuliert. Für nicht-risikoadjustierte und risikoadjustierte Follow-up-Indikatoren wird jeweils ein Verfahren präsentiert, dessen Anwendung exemplarisch anhand der Daten aus dem Erfassungsjahr (EJ) 2015 für Folgeeingriffe aufgrund von prozedurassoziierten Problemen bei implantierten Herzschrittmachern illustriert wird. Die vorgeschlagene Auswertungsanalytik kann mit gängiger Statistiksoftware berechnet werden und ist flexibel für zukünftige Weiterentwicklungen. In der Darstellung und Interpretation orientiert sich die vorgeschlagenen Methodik für nicht-risikoadjustierte bzw. risikoadjustierte Follow-up-Indikatoren an den ratenbasierten bzw. O/E-Indikatoren, die aus den QS-Verfahren nach QSKH-RL bekannt sind. Die beschriebenen statistischen Verfahren legen somit die Auswertungsanalytik für die Bundesauswertung des EJ 2016 fest.

2 Hintergrund

Gemäß Anlage 3 der QSKH-RL werden für die QS-Verfahren *Herzschrittmacherversorgung*, *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung* ab dem EJ 2015 Follow-up-Daten erhoben, um Daten möglicher Wechseleingriffe mit denen der jeweiligen Erstimplantation innerhalb eines Follow-up-Zeitraums über sogenannte Patientenpseudonyme (PSN) verknüpfen zu können. Ziel des vorliegenden Dokuments ist es, ein Verfahren zur Auswertung der in Anhang 1 zur Anlage 3 der QSKH-RL beschriebenen 6 Follow-up-Indikatoren zu präsentieren. Diese sind:

- *Herzschrittmacherversorgung* (4): Hardwareprobleme (Aggregat und/oder Sonde) als Indikation zum Folgeeingriff innerhalb von 8 Jahren, Infektion oder Aggregatperforation als Indikation zum Folgeeingriff innerhalb eines Jahres, Prozedurassoziierte Probleme (Sonden- bzw. Taschenprobleme) als Indikation zum Folgeeingriff innerhalb eines Jahres, Laufzeit des alten Herzschrittmacher-Aggregats unter 4 Jahren bei Ein- und Zweikammersystemen
- *Hüftendoprothesenversorgung* (1): Hüft-Endoprothesenwechsel bzw. -komponentenwechsel im Verlauf
- *Knieendoprothesenversorgung* (1): Knie-Endoprothesenwechsel bzw. -komponentenwechsel im Verlauf

Dabei beschränkt sich das vorgestellte Auswertungskonzept auf den *ersten Folgeeingriff zu jedem Ersteingriff*. Da Zensierungen und konkurrierende Ereignisse eine wichtige Rolle für die Auswertung spielen, wird das Follow-up-Verfahren im Kontext der statistischen Ereigniszeitanalyse beschrieben. Ziel dieses Dokuments ist dabei die detaillierte und statistisch transparente Darstellung der vorgeschlagenen biometrischen Methodik. Das Dokument richtet sich daher vorwiegend an Leserinnen und Leser mit statistischem Hintergrundwissen; eine laienverständliche Darstellung der statistischen Methodik wird an dieser Stelle nicht angestrebt.

Die hier vorgeschlagene Auswertungsmethodik wird anhand der Daten der Erfassungsjahre 2015 und 2016 entwickelt, wobei zum Zeitpunkt der Entwicklung der Methodik lediglich Daten des EJ 2015 vollständig zur Verfügung standen. Die statistische Methodik kann daher durch die gewonnene Erfahrung in der praktischen Umsetzung in Zukunft Anpassungen unterworfen sein; wir behalten uns daher ausdrücklich vor, die hier beschriebenen Methoden für Auswertungen der in der Anlage 3 der QSKH-RL beschriebenen Indikatoren für nachfolgende Erfassungsjahre zu verändern und anzupassen.

2.1 Follow-up-Indikatoren

Ziel der quantitativen Komponente der Qualitätsförderung für die Follow-up-Indikatoren ist es, zum gegebenen Auswertungszeitpunkt für jeden Leistungserbringer zu bestimmen, ob die Revisionsraten für Ersteingriffe des Leistungserbringers aus einer vorgegebenen Zeitperiode auffällig sind. Ergebnisse, die als **rechnerisch auffällig** eingestuft werden, sind dann von den auf Landesebene beauftragten Stellen **Strukturierten Dialog** weiter qualitativ zu untersuchen. Dies bedeutet, dass ein wichtiges Ergebnis des Follow-up-Verfahrens nicht nur die Auffälligkeitseinstufung ist, sondern auch dass relevante Informationen für den Strukturierten Dialog bereitgestellt werden.

Beim Follow-up-Verfahren zur Qualitätsförderung in der QSKH-RL handelt es sich aus statistischer Sicht um ein sequentielles Monitoring auf Einrichtungsebene: Durchgängig finden neue Ersteingriffe bei Patientinnen und Patienten in den drei QS-Verfahren statt, die dann als Teil des Follow-

up-Verfahrens weiterverfolgt werden. Dabei können beim Follow-up-Verfahren Folgeeingriffe bei einem anderen Leistungserbringer als dem des Ersteingriffs vorgenommen werden. Neu ist, dass solche Fälle in dem Follow-up-Verfahren über ein Patientenpseudonym verknüpft werden können. Die Übermittlung der Daten zu Erst- und Folgeeingriffen erfolgt jährlich im Rahmen der QS-Verfahren *Herzschrittmacherversorgung*, *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung*. Übermittelt wird zu jedem Fall tagesgenaue Information über den Zeitpunkt des Folgeeingriffs sowie vorgegebene patientenseitige Information wie z. B. Alter, Geschlecht und Risikofaktoren, die durch die Basisbögen der QS-Verfahren erhoben werden. Beim Follow-up-Verfahren nach QSKH-RL handelt es sich um ein **passives Follow-up**: Erst wenn ein Folgeeingriff vorgenommen wird, werden neue Information zu dem betroffenen Fall übermittelt. Solange keine neue Informationen zu einem Fall vorliegen, wird davon ausgegangen, dass keine Folgeeingriffe stattgefunden haben und der Fall weiter unter Beobachtung, d. h. unter Risiko für einen Folgeeingriff, steht. Eine Problematik stellen hier **Todesfälle** dar, die nicht im Rahmen des QS-Verfahrens übermittelt werden; dies wird in Abschnitt 2.4 genauer diskutiert. Prinzipiell erweitert sich der Datenpool von Ersteingriffen, die nachverfolgt werden, pro Erfassungsjahr, jedoch werden Fälle nach dem in der QSKH-RL geregelten maximalen Follow-up-Zeitraums nicht weiter verfolgt.

Insgesamt unterscheidet sich das vorliegende Entscheidungsproblem von einem sequentiellen Monitoring von Ereignissen (siehe z. B. Hardoon u. a. 2007 sowie Gandy u. a. 2010), weil zwar stetig neue Folgeeingriffe stattfinden, jedoch gemäß QSKH-RL nur einmal jährlich eine Entscheidung bezüglich der rechnerischen Auffälligkeit von Einrichtungen getroffen werden soll. In der Literatur gibt es auch andere Beispiele von Follow-up-Indikatoren (siehe z. B. Burroughs u. a. 2006; Ghenim u. a. 2009); hier werden die Eingriffe jedoch meistens für einen spezifischen Follow-up-Zeitraum (z. B. 30 Tage oder 1 Jahr) komplett nachverfolgt: Nur Fälle aus dem Beobachtungszeitraum, die den kompletten Follow-up-Zeitraum beobachtet wurden, werden analysiert. Bei dieser Vorgehensweise werden Fälle, deren Follow-up-Zeitraum zum Auswertungszeitpunkt noch nicht die gewünschte Dauer erreicht haben, ignoriert. Je nach Länge der Beobachtungsperiode und des Follow-up-Zeitraums kann der Anteil von aktuellen Ersteingriffen, die bei einem solchen Vorgehen zum Auswertungszeitpunkt ignoriert werden, erheblich sein. Handelt es sich z. B. um ein 3-Monats-Follow-up für einen Beobachtungszeitraum von einem Jahr, werden zum Jahresende $3/12 = 25\%$ der Fälle noch nicht über den kompletten Follow-up-Zeitraum beobachtet worden sein und würden somit in einer Analyse zum Jahresende ignoriert. Da aber teilweise sehr lange Follow-up-Zeiträume für die Follow-up-Indikatoren nach QSKH-RL vorgesehen sind, ist es daher relevant, auch Ersteingriffe mit unvollständigen Follow-up-Zeiträumen zu berücksichtigen. Die **Überlebensdauer- bzw. Ereigniszeitanalyse** bietet hierfür einen geeigneten statistischen Analyserahmen. Für eine Einführung siehe z. B. Klein und Moeschberger (2003).

2.2 Datengrundlage

Eingriffe sind ab dem EJ 2015 mit PSN versehen, sodass es möglich ist, Erst- und Folgeeingriffe ab diesem Zeitpunkt miteinander zu verknüpfen. Die Datenerhebung für ein Erfassungsjahr ist so gestaltet, dass alle stationären Eingriffe in den QS-Verfahren im Erfassungsjahr berichtet werden, die vordefinierte Einschlusskriterien erfüllen. Für die QS-Verfahren *Herzschrittmacherversorgung*, *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung* ist der Zeitpunkt der Aufnahme entscheidend: Dieser muss – z. B. für das EJ 2015 – zwischen dem 1. Januar 2015 und dem 31. Dezember 2015 liegen. Die Entlassung (und somit auch die Eingriffsoperation) muss bis spätestens 31. Dezember 2016 erfolgen, damit der Fall für das Erfassungsjahr berücksichtigt wird. Ansonsten

wird der Fall nicht in die Analyse eingeschlossen (auch nicht im Folgejahr). Die genauen Details dazu sind im QS-Filter des jeweiligen QS-Verfahrens festgelegt.

Betrachtet wird also ein Datensatz, welcher für einen bestimmten kalendarischen **Beobachtungszeitraum** (s_{anfang} , s_{ende}) alle relevanten Ereignisse zu den Eingriffen eines Follow-up-Verfahrens enthält. Der aus den Ereignisdaten zusammengeführte Datensatz mit relevanten Ereignissen wird im Folgenden die zu analysierende **Ereignishistorie** genannt. Das PSN entspricht einer pseudonymisierten Version der Krankenversichertennummer der gesetzlichen Krankenkasse, weshalb für privat versicherte Patientinnen und Patienten kein PSN existiert. In der Auswertung können daher nur Patientinnen und Patienten der gesetzlichen Krankenversicherung mit einem gültigen PSN berücksichtigt werden.

Das PSN wird außerdem pro QS-Verfahren (*Herzschrittmacherversorgung*, *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung*) unterschiedlich generiert, weshalb keine QS-Verfahrens-übergreifende Analysen möglich sind; im QS-Verfahren *Herzschrittmacherversorgung* erfolgt jedoch eine Zusammenfassung der Auswertungsmodule 09/1 – 09/3, welche für die Follow-up-Indikatoren daher gemeinsam ausgewertet werden können.

In den Bereichen *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung* gibt es zusätzlich eine **Seitenlokation der Endoprothese** (links bzw. rechts), die bei der Analyse zu berücksichtigen ist. Daher beziehen sich die Analysen der Erst- und Folgeeingriffe hier also auf Hüft- bzw. Knieendoprothesen derselben Seite bei gesetzlich versicherten Patientinnen und Patienten. Hat beispielsweise eine Patientin beidseitig künstliche Hüften bekommen, gehen die Implantationen als zwei Behandlungsfälle in die Analysen ein.

2.3 Zensierungen

In der Terminologie der Ereigniszeitanalyse beschreibt eine **Zensierung** den Abbruch eines Follow-up, bevor das Ereignis eingetreten ist. Zur Illustration verschiedener Zensierungen und Verlaufsdaten der vorliegenden Art, bei denen auch die Kalenderzeit des Ersteingriffs eine Rolle spielt, eignet sich das **Lexis-Diagramm**. Abbildung 1 zeigt ein exemplarisches Lexis-Diagramm auf Basis hypothetischer Follow-up-Daten für die Erfassungsjahre 2015 und 2016:

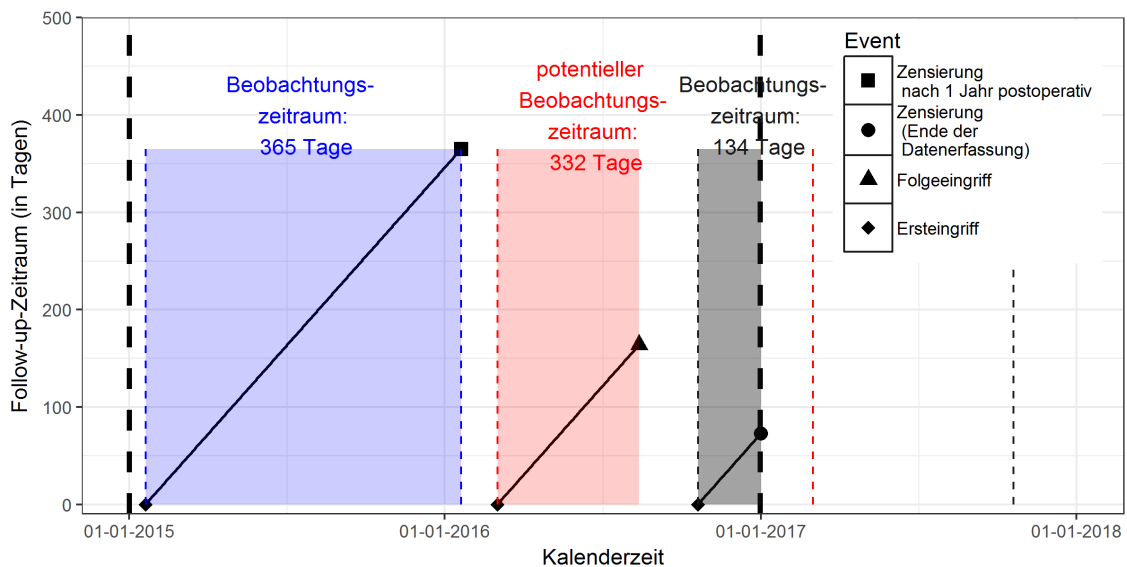


Abbildung 1: Exemplarisches Lexis-Diagramm der Follow-up-Daten für drei Ersteingriffe

Auf der x-Achse ist die Kalenderzeit, auf der y-Achse der Follow-up-Zeitraum seit Ersteingriff aufgetragen. Für den Indikator „Folgeeingriffe auf Grund von prozedurassoziierten Problemen“ ist eine Beobachtungsdauer von einem Jahr definiert, d. h., bei allen Fällen werden nur die ersten 365 Tage nach Datum der Implantation in die Analyse eingeschlossen. Der Fall mit Ersteingriff Anfang 2015 wird daher Anfang 2016 rechtszensiert. Die anderen beiden Fälle hätten dagegen (in den Erfassungsjahren 2015 und 2016) keine 365 Tage beobachtet werden können, da das Ende ihres Beobachtungszeitraums hinter dem Zensierungsdatum am 31. Dezember 2017 liegt. Würden daher nur Fälle mit kompletter Beobachtungsdauer von 365 Tagen in die Analyse eingehen, müssten beide Fälle mit Ersteingriff nach dem 1. Januar 2016 von der Analyse ausgeschlossen werden. Die vorgeschlagenen Verfahren aus der Ereigniszeitanalyse ermöglichen dagegen, die bisher gesammelten Informationen über diese Fälle in die Analyse miteinzubeziehen.

Für die Daten des EJ 2016, welche in der Bundesauswertung für das EJ 2016 analysiert werden, stellt der 31. Dezember 2016 einen sogenannten **Rechtszensierungszeitpunkt** des Follow-up-Zeitraums dar. Das heißt, dass Ersteingriffe – je nach Kalenderdatum des Ersteingriffs – unterschiedlich lange unter Risiko für Folgeeingriffe standen. Ein Eingriff, der im Januar 2016 vorgenommen wurde, wird also länger nachverfolgt als ein Eingriff, der im Dezember 2016 stattgefunden hat – mit entsprechend größerer Wahrscheinlichkeit eines Folgeeingriffs. Damit alle Fälle innerhalb eines Zeitraums trotz unterschiedlicher Beobachtungsdauer adäquat analysiert werden können, müssen diese Art der Rechtszensierungen bei der Analyse berücksichtigt werden.

Wechseleingriffen, deren Ersteingriff vor 2015 stattgefunden hat, kann kein Ersteingriffszeitpunkt zugeordnet werden. Es ist lediglich bekannt, dass der Eingriff vor dem 1. Januar 2015 stattgefunden haben muss. Diese Situation stellt eine sogenannte **Linkszensierung** dar. In der Analyse könnten prinzipiell jedoch nur linkszensierte Fälle, nicht aber links- (und rechts-) zensierte Fälle ohne beobachteten Folgeeingriff berücksichtigt werden, da letztere nicht in den Datensatz eingehen. Dies würde zu einer Unterschätzung der Survivorfunktion (d. h. zu einer Überschätzung der Anteile an Fällen mit Folgeeingriff) führen. Aus diesem Grund und weil diesen Fällen kein Leistungserbringer des Ersteingriffs zugeordnet werden kann, werden linkszensierte Fälle komplett

von der Analyse ausgeschlossen.

Eine letzte mögliche der Zensierungsform besteht darin, dass die QSKH-RL im Anhang 1 zur Anlage 3 für das QS-Verfahren *Herzschrittmacherversorgung* konkrete Follow-up-Perioden festlegt: bei Hardwareproblemen 8 Jahre, bei einer Infektion 1 Jahr, bei prozedurassoziierten Problemen 1 Jahr, bei einem Aggregatwechsel 4 Jahre. Daher ist ein Fall nach der entsprechenden maximalen Laufzeit des Indikators zu zensieren (= Rechtszensierung), obwohl eventuell Ereignisse nach der maximalen Beobachtungszeit auftreten und in einem übermittelten Datensatz ggf. vorhanden sind.

2.4 Erfassungsprobleme

Fälle mit Patientinnen und Patienten, die noch im Krankenhaus versterben, werden in den Follow-up-Verfahren von der Analyse ausgeschlossen. Todesfälle außerhalb der stationären Behandlung werden im jetzigen Follow-up-Verfahren nach QSKH-RL nicht erhoben. Für lange Follow-up-Zeiträume bei Fällen mit hoher Sterblichkeit ist der Anteil an Fällen mit Patientinnen und Patienten, die in der Follow-up-Periode versterben, vermutlich nicht unerheblich. Die Wahrscheinlichkeit, dass eine männliche Person, die 2015 ihren 80-jährigen Geburtstag erlebt hat, ihren 81. Geburtstag nicht mehr erlebt, liegt bei 6 % (Statistisches Bundesamt 2017). Basierend auf diesen Zahlen und einer Follow-up-Zeit von 8 Jahren wäre die Wahrscheinlichkeit, dass diese Person in 2023 nicht mehr am Leben ist, bereits bei ca. 45 %. Populationsbasierte Kohortenstudien legen zusätzlich nahe, dass die Mortalitätsraten z. B. für Patientinnen und Patienten mit implantiertem Herzschrittmacher in der Zeit nach der Operation noch deutlich höher sind (5,5 % innerhalb der ersten 6 Monate; siehe Kirkfeld u. a. 2014). Da Todesfälle außerhalb des stationären Aufenthaltes nicht Teil des Follow-up-Verfahrens sind, würde fälschlicherweise angenommen werden, dass bereits verstorbene Patientinnen und Patienten weiterhin unter Risiko für einen Folgeeingriff stehen.

Prinzipiell wird somit der Anteil an Folgeeingriffen einer Einrichtung zu positiv (also zu niedrig) eingeschätzt. Gibt es patientenseitige Gründe für Unterschiede in den Sterblichkeitsraten der Leistungserbringer (z. B. durch ältere Patientinnen und Patienten), werden Einrichtungen mit höherer Sterblichkeitsrate dadurch tendenziell bevorteilt. Diese Problematik wird bei Indikatoren mit kurzer Follow-up-Zeit in der Regel seltener auftreten als bei solchen mit langer Follow-up-Zeit.

2.5 Weitergehende Zensierungsproblematik

Für das QS-Verfahren *Hüftendoprothesenversorgung* liegen Ereigniszeitpunkte auf den Tag und die Uhrzeit genau vor, bei den QS-Verfahren *Herzschrittmacherversorgung* und *Knieendoprothesenversorgung* werden Ereigniszeitpunkte dagegen nur mit Tagesgenauigkeit erfasst; es handelt sich bei den QS-Verfahren *Herzschrittmacherversorgung* und *Knieendoprothesenversorgung* also um **doppelt intervallzensierte** und bei dem QS-Verfahren *Hüftendoprothesenversorgung* um **intervallzensierte** Daten (Klein und Moeschberger 2003). Dies führt dazu, dass **Bindungen** bei den Ereigniszeitpunkten auftreten können, d. h., mehrere Fälle können die gleiche beobachtete Lebenszeit (in Tagen) haben. Verglichen mit den langen Follow-up-Zeiten der Indikatoren scheint die Problematik der Intervallzensierung vernachlässigbar und wird im Folgenden daher nicht weiter betrachtet.

In der Analyse wird der Zeitpunkt zwischen der Operation des Ersteingriffs und der Operation des Folgeeingriffs betrachtet. Somit enthalten die Daten für das EJ 2015 auch die sogenannten

Überlieger, d. h. Fälle, die erst im Januar 2016 den Folgeeingriff hatten, aber noch in 2015 stationär aufgenommen wurden. Vereinfachend wird der Folgeeingriff von Überliegern im Januar des Folgejahres in den folgenden Analysen nicht berücksichtigt, wenn dies das letzte Erfassungsjahr ist.

Für die Bundesauswertung für das EJ 2016 wird somit angenommen, dass alle Fälle ab dem 31. Dezember 2016 nicht mehr unter Beobachtung stehen. Die Beobachtungszeit von Überliegern aus dem EJ 2016 geht damit nur bis zum Ende des Jahres in die Analyse des EJ 2016 ein. Die Beobachtungszeit im Januar wird nur in zukünftigen Analysen des EJ 2017 berücksichtigt.

Da es für einen konkreten Eingriff wie z. B. aufgrund von prozedurassoziierten Problemen nur um die Zeit zwischen Ersteintritt und dem Folgeeingriff aufgrund von prozedurassoziierten Problemen geht, wird angenommen, dass alle anderen Eingriffe (wie z. B. Aggregatwechsel oder Hardwareprobleme) eine Rechtszensierung der Follow-up-Zeit darstellen. Prinzipiell sind diese als **Competing-Events** aufzufassen, werden aber der Einfachheit halber als Rechtszensierungen betrachtet. Es wird also angenommen, dass diese konkurrierenden Events unabhängig vom betrachteten Ereignis geschehen.

Ein potentielles Problem bei der Datenerhebung ist, dass Eingriffe fälschlicher Weise als Ersteinträge dokumentiert werden, obwohl bereits ein früherer Eingriff in der Ereignishistorie für das PSN in den Daten vorliegt. Sollte eine solche Situation auftreten, wird in der Datenaufbereitung nur der zeitlich erste Eingriff pro Seite berücksichtigt, ein möglicher weiterer „Ersteintritt“ gilt als Zensierungsereignis.

2.6 Zusammenfassung

Da Zensierungen bei der Analyse der Follow-up-Indikatoren zu berücksichtigen sind, kann die gewöhnliche Auswertungsanalytik nach QSKH-RL für Raten bzw. risikoadjustierte Raten nicht unmittelbar verwendet werden. Daher wird auf Verfahren aus der statistischen **Ereigniszeitanalyse** zurückgegriffen, die im Folgenden detailliert beschrieben wird.

3 Ereigniszeitanalyse

Im Folgenden wird das sequentielle Entscheidungsproblem der Follow-up-Verfahren im Kontext der statistischen Ereigniszeitanalyse dargestellt.

3.1 Notation

Im Folgenden wird die mathematische Notation zur Beschreibung der vorhandenen Daten und der Auswertungsanalytik dargelegt. Der Abschnitt orientiert sich hierbei an der Notation von Klein und Moeschberger (2003). Es wird angenommen, dass Informationen zu Ersteingriffen von insgesamt m Leistungserbringern vorliegen, welche im **Beobachtungszeitraum** ($s_{\text{anfang}}, s_{\text{ende}}$) vorgenommen wurden. Die Information zu diesen Fällen wird über die Landesstellen für Qualitätssicherung und die Vertrauensstelle an das IQTIG übermittelt. Die zu analysierende Ereignishistorie setzt sich also aus den Daten mehrerer Jahre zusammen.

Betrachtet wird der j . Fall beim i . Leistungserbringer, $i = 1, \dots, m$ und $j = 1, \dots, n_i$, in dieser Ereignishistorie.

- Sei T_{ij} die Anzahl an Tagen zwischen dem Operationsdatum des Folgeeingriffs und dem Operationsdatum des Ersteingriffs. Wenn ein solcher Folgeeingriff gar nicht vorkommt, wird $T_{ij} = \infty$ definiert.
- Sei C_{ij} der entsprechende Rechtszensierungszeitpunkt des Follow-up. Meistens ist dies der Abstand zwischen dem Operationszeitpunkt des Ersteingriffs und dem letztmöglichen Erfassungszeitpunkt des Beobachtungszeitraums, z. B. ist es für den zum Zeitpunkt der Entwicklung der statistischen Methodik vorliegenden Datenpool aus dem EJ 2015 der 31. Dezember 2015. Ein Competing-Event, wie z. B. ein Aggregatwechsel bei der Beobachtung von prozedurassoziierten Problemen, würde die Beobachtungsdauer jedoch früher abbrechen.

Beobachtet wird nun die **Überlebenszeit**

$$T_{ij}^* = \min(T_{ij}, C_{ij}),$$

wobei im konkreten Fall der Follow-up-Indikatoren mit Überlebenszeit das „Überleben“ des Implantats bzw. Herzschrittmachers und nicht der Patientin oder des Patienten gemeint ist. Für obige Zufallsvariable wird des Weiteren ein Zensierungsindikator

$$\delta_{ij} = I(T_{ij} \leq C_{ij})$$

definiert, welcher angibt, ob die Überlebenszeit T_{ij}^* ein Wechseleingriff für den Fall im Beobachtungszeitraum (1) oder eine Zensierungszeit angibt (0). Die genaue Dauer bis zum Wechseleingriff wird also nur beobachtet, falls das Ereignis vor oder genau zum Zeitpunkt der Zensierung eintritt.

3.1.1 Survivorfunktion und Hazardrate

Es wird im Folgenden angenommen, dass die Überlebenszeiten bzw. Ausfallzeiten T_{ij} unabhängige und identisch verteilte Zufallsvariablen aus einer Verteilung mit Träger $[0, \infty)$ sind, deren Verteilungsfunktion $F(t)$ ist. Sei

$$S(t) = P(T > t) = 1 - F(t)$$

die sogenannte **Survivorfunktion**, welche den Anteil an Fällen angibt, die bis zum Zeitpunkt t noch kein Ereignis, d. h. Folgeeingriff, hatten. Durch die Annahme der identisch verteilten Zufallsvariable werden an dieser Stelle also keine patientenspezifischen Risikofaktoren berücksichtigt. Auch wird die Verteilungsfunktion als homogen über die Fälle sowie den Beobachtungszeitraum gesehen (keine Periodeneffekte).

Anstelle der Survivorfunktion wird in der Überlebenszeitanalyse oft auch mit der **Hazardrate** (auf Deutsch auch **Ausfallrate** genannt) der Überlebenszeiten gearbeitet. Diese kann als momentanes Risiko für den Tod bzw. einen Ausfall aufgefasst werden. Da die Hazardrate für die Überlebensdauer T oft von der Dauer von T abhängt, wird sie als Funktion der Überlebensdauer angegeben, d. h. sie wird als $\lambda(t) \geq 0$ geschrieben. Für einen kleinen Zeitraum Δt gilt dann, dass die Wahrscheinlichkeit für einen Folgeeingriff im Zeitraum $[t, t + \Delta t)$ bei einem Ersteingriff, der bis kurz vor t noch keinen Folgeeingriff hatte, approximativ $\Delta t \cdot \lambda(t)$ ist. Zusätzlich zur Hazardrate definiert man auch die **kumulierte Hazardrate** als $\Lambda(t) = \int_0^t \lambda(u) du$. Man kann zeigen, dass folgender Zusammenhang zwischen Survivorfunktion und der Hazardrate bzw. der kumulierten Hazardrate besteht: $S(t) = \exp(-\Lambda(t))$ (Klein und Moeschberger 2003). Diese Größen werden im folgenden zur Definition der Follow-up-Indikatoren verwendet. Wir wenden uns zuerst der Schätzung der Survivorfunktion zu.

3.2 Kaplan-Meier-Schätzer für die Survivorfunktion

Ein nicht-parametrischer Schätzer für die Survivorfunktion von rechts zensierten Ereignisdaten unter der Annahme, dass die Zensierungen nicht-informativ sind, ist der **Kaplan-Meier-Schätzer**, welcher auf folgende Art definiert ist (Klein und Moeschberger 2003, Abschnitt 4.2):

$$\hat{S}(t) = \begin{cases} 1 & \text{falls } t < t_1, \\ \prod_{t_k \leq t} \left(1 - \frac{d_k}{y_k}\right) & \text{falls } t_1 \leq t \end{cases}$$

Im obigen sind $t_1 < t_2 < \dots < t_D$ die D eindeutigen Ereigniszeitpunkte in der zu analysierenden Ereignishistorie. Für einen spezifischen Ereigniszeitpunkt t_k ist y_k die Anzahl an Fällen (hier also Ersteingriffe), welche zum Zeitpunkt t_k unter Risiko stehen, d. h. Fälle mit einer Beobachtungszeit $\geq t_k$, welche Fälle mit Ereignis zum Zeitpunkt t_k oder später sind. Des Weiteren ist d_k die Anzahl an Ereignissen zum Zeitpunkt t_k , es gilt $d_k \leq y_k$. Für ein Zahlenbeispiel der Kaplan-Meier-Schätzung siehe z. B. Beispiel 4.1 in Klein und Moeschberger (2003).

Der Kaplan-Meier-Schätzer ist eindeutig definiert für alle $t \leq t_{\max}$, wobei t_{\max} die längste Studienzeit (Zensierung oder Ereignis) ist. Falls diese Zeit einem Ereignis entspricht, ist $\hat{S}(t_{\max}) = 0$, entspricht die Zeit jedoch einer Zensierung, ist $\hat{S}(t_{\max})$ nicht definiert. Klein und Moeschberger (2003, S. 99-100) enthält einige Vorschläge, wie der Schätzer für Zeitpunkte mit $t > t_D$ trotzdem fortgeschrieben werden kann. Eine konservative (d.h nach oben verzerrte) Konvention ist anzunehmen, dass $\hat{S}(t) = \hat{S}(t_{\max})$, also die Survivorfunktion wird einfach fortgeschrieben, indem angenommen wird, dass für längere Lebenszeiten keine weiteren Ereignisse auftreten und weitere Zensierungen erst zum Zeitpunkt unendlich geschehen. Dieser Konvention wird hier gefolgt.

Der Kaplan-Meier-Schätzer für verschiedene Zeitpunkte kann grafisch beispielsweise wie in Abbildung 2 dargestellt werden – in diesem Fall für alle Fälle auf Bundesebene und einen exemplarischen Leistungserbringer anhand der Daten aus dem EJ 2015:

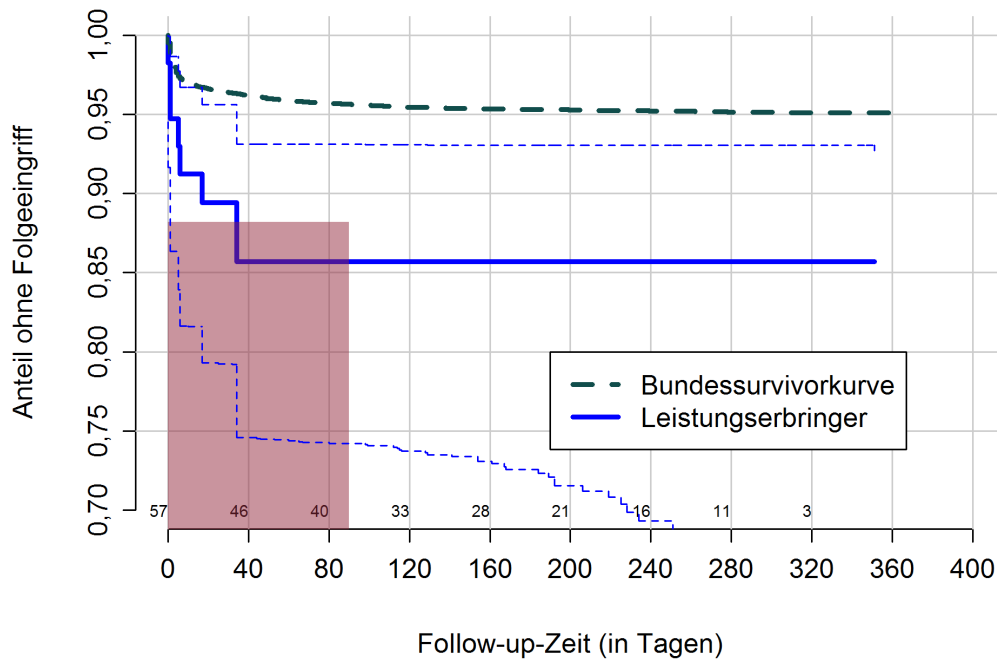


Abbildung 2: Kaplan-Meier-Kurven für alle analysierten Ersteingriffe im EJ 2015 (= Bundessurvivorkurve) sowie für einen exemplarischen Leistungserbringer

Direkt über der x-Achse dargestellt ist die jeweilige Anzahl an Ersteingriffen, die zu den dargestellten Zeitpunkten noch unter Beobachtung, d. h. unter Risiko für einen Folgeeingriff, stehen. Daraus wird ersichtlich, dass für den hier anonymisierten Leistungserbringer im Beobachtungszeitraum ursprünglich 57 Ersteingriffe unter Risiko für einen Wechsel standen. Anhand der Sprünge in der Kurve sieht man, dass es bei insgesamt 8 der betrachteten Ersteingriffe zu einem Wechsel kam. Zum Zeitpunkt 40 Tage nach Ersteingriff, zu welchem bereits alle 8 Folgeeingriffe stattgefunden haben, stehen nur noch 46 Ersteingriffe unter Beobachtung (aufgrund von 8 Folgeeingriffen und 3 Zensurierungen, die vor diesem Zeitpunkt stattgefunden haben).

Der Referenzwert zur Bestimmung der rechnerischen Auffälligkeit wird hier beispielhaft zum Zeitpunkt 90 Tage nach Ersteingriff berechnet und beträgt 88,21 %. Da die Survivorkurve nur abfällt, aber nie ansteigt, wird das Ergebnis eines Leistungserbringer als rechnerisch auffällig eingestuft, wenn sein Anteil von Fällen ohne Folgeeingriff zum Zeitpunkt 90 Tage oder vorher unter diesen Referenzwert fällt (als rot markierter Bereich dargestellt). Die Berechnung der für den Leistungserbringer dargestellten Konfidenzintervalle sowie des Referenzwerts wird im Folgenden beschrieben.

3.2.1 Konfidenzintervalle

Konfidenzintervalle für $S(t)$ können auf verschiedene Arten bestimmt werden, z. B. als punktweise Konfidenzintervalle oder als gemeinsames Konfidenzband. Für punktweise Konfidenzintervalle eignet sich besonders die in Fay und Brittain (2016) beschriebene Methode, die auch bei kleinen Fallzahlen gute Überdeckungswahrscheinlichkeiten garantiert. Hierbei handelt es sich um exakte Konfidenzintervalle anhand von mid-p-Werten. Falls es bei den analysierten Daten keine Zensierungen gibt, entsprechen diese Intervalle der Methodik für nicht-risikoadjustierte Raten, die im Abschlussbericht des IQTIG zu planungsrelevanten Qualitätsindikatoren beschrieben ist (IQTIG 2016). In der statistischen Software R (R Core Team 2016) können diese Intervalle mittels dem `bpcp`-Paket (Fay und Brittain 2016) berechnet werden. Untenstehend eine kurze Erläuterung des Verfahrens; Genaueres ist in Fay und Brittain (2016) nachzulesen.

Seien wie oben die eindeutigen Überlebenszeiten t_1, \dots, t_D , d. h. Bindungen sind bei den Überlebenszeiten möglich. Dies wird erweitert notationell um $t_0 = 0$ und $t_{D+1} = \infty$. Dadurch wird die positive Zeitachse in $D + 1$ Intervalle $(0, t_1], (t_1, t_2], \dots, (t_D, \infty)$, unterteilt. Mit d_i und y_i bezeichnen wir, wie bereits zuvor, für jedes Intervall die Zahl an Ereignissen zum Zeitpunkt t_i bzw. die Anzahl an Fällen unter Risiko zu Beginn des i . Intervalls. Falls Ereignis und Zensierung im gleichen Intervall geschieht, wird angenommen, dass die Zensierungen nach den Ereignissen geschehen.

Zwei Vektoren an Zufallsvariablen \mathbf{a}^- und \mathbf{b}^- der Länge $D + 2$ werden nun auf folgende Art festgelegt: Seien $a_0^- = 1$ und $b_0^- = 0$. Für $i = 1, \dots, D + 1$:

- Falls $d_i = 0$, dann ist $a_i^- = 1$ und $b_i^- = 0$.
- Falls $d_i > 0$, dann ist $a_i^- = y_i - d_i + 1$ und $b_i^- = d_i$.

Eine **Beta-Produkt-Variable** mit Parametervektoren $\mathbf{a} = (a_0, \dots, a_{D+1})'$ und $\mathbf{b} = (b_0, \dots, b_{D+1})'$ wird auf folgende Art definiert:

$$\text{BP}(\mathbf{a}, \mathbf{b}) = \prod_{i=0}^{D+1} B(a_i, b_i),$$

wobei $B(a, b)$ eine betaverteilte Zufallsvariable mit den Parametern $a > 0$ und $b > 0$ angibt. In diesem Zuge seien $B(1, 0) \equiv 1$ und $B(0, 1) \equiv 0$. Des Weiteren sei

$$W^-(t_i) = \prod_{h=0}^i B(a_h^-, b_h^-),$$

d. h. eine Beta-Produkt-Variable basierend auf den oben definierten Parametervektoren bis zum i . Index. Somit ist $W^-(t_i) \sim \text{BP}(\mathbf{a}_i^-, \mathbf{b}_i^-)$ mit $\mathbf{a}_i^- = (a_0^-, \dots, a_i^-)'$ und $\mathbf{b}_i^- = (b_0^-, \dots, b_i^-)'$. Hierauf aufbauend wird

$$\begin{aligned} W^+(t_i) &= W^-(t_i)B(y_{i+1}, 1) \\ &= \text{BP}((a_0^-, \dots, a_i^-, y_{i+1}), (b_0^-, \dots, b_i^-, 1)') \\ &= \text{BP}(\mathbf{a}_i^+, \mathbf{b}_i^+), \end{aligned}$$

mit $\mathbf{a}_i^+ = (a_0^-, \dots, a_i^-, y_{i+1})'$ und $\mathbf{b}_i^+ = (b_0^-, \dots, b_i^-, 1)'$ definiert. Sei nun für $t_{i-1} \leq t < t_i$

$$W^*(t) = UW^-(t_{i-1}) + (1 - U)W^+(t_i),$$

wobei $U \sim B(\frac{1}{2})$ eine Bernoulli-verteilte Zufallsvariable mit Wahrscheinlichkeit 1/2 ist. Ein zweiseitiges $(1 - \alpha) \cdot 100\%$ exaktes mid-p-Konfidenzintervall für $\bar{S}(t) = P(T \geq t)$ mit $t_{i-1} \leq t < t_i$ ist nun definiert als

$$\left(q \left\{ \frac{\alpha}{2}, W^*(t) \right\}, q \left\{ 1 - \frac{\alpha}{2}, W^*(t) \right\} \right),$$

wobei z. B. $q \{ \alpha/2, W^*(t) \}$ das $\alpha/2$ -Quantil von $W^*(t)$ angibt. Dieses Quantil könnte einfach, aber für hohe Präzision nur sehr rechenaufwendig, über Monte-Carlo-Simulation bestimmt werden. Alternativ können die beiden Beta-Produktvariablen W^- und W^+ wie in Fay und Brittain (2016) jeweils durch Beta-Variablen mit übereinstimmenden 1. und 2. Momenten approximiert werden. Die Quantile der entsprechenden Beta-Verteilungen können dann kombiniert werden, um das gewünschte Quantil zu erhalten. Dafür ist die Quantilsfunktion der Betaverteilung `qbeta` sowie eine numerische Nullstellensuche mittels `uniroot` nötig. Siehe Fay und Brittain (2016) bzw. R-Quellcode der Funktion `bpcp::bpcpMidp.mm` für Details zur Bestimmung der notwendigen Quantile.

3.3 Auswertung eines nicht-risikoadjustierten Follow-up-Indikators

Auf Basis der oben dargestellten Methodik kann ein nicht-risikoadjustierter Follow-up-Indikator wie folgt ausgewertet werden: Sei τ ein gewählter Vergleichszeitpunkt für einen nicht-risikoadjustierten Follow-up-Indikator. Dies könnte z. B. 30 oder 90 Tage (oder später) nach dem Ersteingriff sein. Die Auffälligkeitseinstufung eines Leistungserbringers Einrichtung erfolgt auf folgende Art:

1. Pro Leistungserbringer $i = 1, \dots, m$ wird $\hat{S}_i(\tau)$ anhand der Daten des Beobachtungszeitraums ($s_{\text{anfang}}, s_{\text{ende}}$) bestimmt.
2. Berechne das empirische untere x%-Perzentil der Werte in $\left\{ \hat{S}_i(\tau) \right\}_{i=1}^m$. Analog der Perzentilbestimmung bei Indikatoren bei QS-Verfahren nach QSKH-RL sollen bei der Berechnung nur die Ergebnisse von Leistungserbringern mit einer Mindestanzahl an Ersteingriffen im Beobachtungszeitraum eingehen. Der erhaltene Wert ρ_τ funktioniert im Folgenden als **Referenzwert** für den Follow-up-Indikator. Für Follow-up-Indikatoren, bei denen die Auffälligkeitseinstufung nicht perzentilbasiert, sondern auf Basis eines festen Referenzwertes stattfindet, entspricht ρ_τ diesem Referenzwert.
3. Ein Leistungserbringer wird als **rechnerisch auffällig** eingestuft, wenn

$$\hat{S}_i(\tau) < \rho_\tau.$$

Dabei ist zu beachten, dass jegliche statistische Unsicherheit bei der Schätzung der Ergebnisse der Leistungserbringer bei dieser Einstufung gemäß der Richtlinie nicht berücksichtigt wird. Trotzdem sollten, wie für die jetzigen Indikatoren bei QS-Verfahren nach QSKH-RL, entsprechende Konfidenzintervalle der Schätzung mit angegeben werden. Das Ergebnis eines Leistungserbringers für prozedurassoziierte Probleme im QS-Verfahren *Herzschrittmacherversorgung* kann somit tabellarisch auf folgende Weise dargestellt werden:

LE	Erst- eingriffe	Vergleichs- zeitpunkt	Ereignisse	Anteil ohne Ereignis	95%-KI	Referenzwert
X	57	90	8	85,71 %	74,21 % – 93,10 %	88,21 %

Dabei bezeichnet LE das Leistungserbringerpseudonym und Anteil ohne Ereignis den über den Kaplan-Meier-Schätzer berechneten Anteil an Fällen ohne Folgeeingriff zum Vergleichszeitpunkt mit entsprechendem 95%-Konfidenzintervall und Referenzwert. Die Werte sind wie folgt zu interpretieren: Zum Vergleichszeitpunkt 90 Tage nach Ersteingriff hatten unter Berücksichtigung der unterschiedlichen Beobachtungszeiten im Beobachtungszeitraum 2015 bis 2016 von allen Fällen des Leistungserbringers X 85,71 % keinen Folgeeingriff. Da dieser Anteil unter dem Referenzwert liegt, wird das Ergebnis dieses Leistungserbringers als rechnerisch auffällig eingestuft.

Statt den Anteil an Fällen *ohne* Ereignis darzustellen, ist alternativ auch die Darstellung des Anteils an Fällen *mit* Ereignis möglich, welcher sich als 1 - Anteil an Fällen ohne Ereignis berechnet:

LE	Erst- eingriffe	Vergleichs- zeitpunkt	Ereignisse	Anteil mit Ereignis	95%-KI	Referenzwert
X	57	90	8	14,29 %	6,90 % – 25,79 %	11,79 %

Der Konvention aus der statistischen Ereigniszeitanalyse folgend wird in diesem Dokument und in den Auswertungen immer der Anteil an Fällen ohne Ereignis dargestellt; die hier exemplarisch aufgeführte Darstellung des Anteils an Fällen mit Ereignis ist aber ebenso korrekt.

3.4 Risikoadjustierung

Die Auswertung eines Follow-up-Indikators auf Basis des Kaplan-Meier-Schätzers erfolgte im obigen Abschnitt ohne Risikoadjustierung. Risikoadjustierung kann jedoch auch für Follow-up-Indikatoren von Bedeutung sein: Frühere Analysen z. B. für die QS-Verfahren *Hüftendoprothesenversorgung* und *Knieendoprothesenversorgung*, siehe z. B. AQUA (2015), haben gezeigt, dass patientenseitige Faktoren wie Geschlecht, Alter sowie Diabetes die Zeit bis zum Folgeeingriff beeinflussen. Auch wird die Anzahl an Sonden als Risikofaktor für prozedurassoziierte Probleme erwähnt (AQUA 2015).

In der Onkologie oder der Analyse von klinischen Studien ist das klassische Vorgehen zur Risikoadjustierung von Survivorkurven die **A-posteriori-Schichtung**, welche eng mit der direkten Standardisierung verwandt ist. Die kann entweder durch die gewichtete Kombination mehrerer strata- und einrichtungsspezifischen Survivorkurven oder über gewichtete Kaplan-Meier-Schätzer (Amato 1988) geschehen. Speziell sind hier **adjustierte Überlebensdauerkurven** mit **inverse probability weights** (Cole und Hernan 2004; Xie und Liu 2005) beliebt. Das Problem dieser Ansätze ist jedoch, dass es bei kleinen Fallzahlen oft keine Beobachtung in einem der vorgegebenen Strata gibt. Diese Situation kommt bei kleinen Einrichtungen sogar bei Strata, welche durch Kovariablen wie Geschlecht oder Altersklassen gebildet werden, vor. Auch der übliche Vorgang, Strata zu kombinieren, ist für essentielle Risikofaktoren wie z. B. Geschlecht nicht möglich. Somit kann eine Analyse mittels direkter Standardisierung nur über zusätzliche, teilweise ad-hoc getroffene Annahmen, wie z. B. die Annahme einer Priori Survivorkurve pro Strata, durchgeführt werden.

Im Folgenden wird daher eine zusätzliche Indikatorart vorgeschlagen, bei der eine Risikoadjustierung durch die Berücksichtigung stratifizierter Kaplan-Meier-Kurven möglich ist und die sich an

die direkte Standardisierung anlehnt. Dieser Indikator kann wie ein aus den QS-Verfahren nach QSKH-RL bekannter O/E-Indikator interpretiert und dargestellt werden.

3.5 SMR aus multiplikativem Hazardmodell

Der Indikator auf Basis des Kaplan-Meier-Schätzers ist auf einen konkreten Vergleichzeitpunkt τ festgelegt. Im Gegensatz dazu nimmt der im Folgenden vorgeschlagene Indikator den Vergleich auf den Daten des kompletten Beobachtungszeitraums vor. Die Bewertungsart ist ähnlich dem Vergleich von Hazardraten mittels dem Einstichproben-log-Rank-Test. Im konkreten Modell wird angenommen, dass der Einrichtungseffekt im Rahmen eines multiplikativen Hazardmodells definiert werden kann:

$$\lambda_{ij}(t) = \theta_i \lambda_j^*(t), \quad (1)$$

wobei $\theta_i > 0$ der einrichtungsspezifische Effekt ist, der als zeitlich konstant angenommen wird. Des Weiteren ist $\lambda_{ij}(t)$ die Hazardrate des j . Patienten der i . Einrichtung und $\lambda_j^*(t)$ ist die bundeseinheitliche Hazardrate eines Falles mit den gleichen Charakteristika wie der ij . Fall. Vereinfachend und um die Unabhängigkeit von dem Leistungserbringer bei $\lambda_j^*(t)$ hervorzuheben, wird das i notationell hier nicht angeben. Diese multiplikative Betrachtungsweise geht auf Breslow (1975) zurück. Hier ist auch die Log-Likelihood-Funktion für θ_i im Falle von rechtszensierten Daten definiert:

$$l(\theta_i) = o_i \log(\theta_i) + \theta_i \sum_{j=1}^{n_i} \log(S_{*,ij}(t_{ij})).$$

Im obigen gibt $o_i = \sum_{j=1}^{n_i} \delta_{ij}$ die Anzahl an tatsächlichen Ereignissen für die Einrichtung im Beobachtungszeitraum an und $S_{*,ij}(t)$ ist die entsprechende Survivorfunktion in der Standardpopulation für einen Fall mit den gleichen Charakteristika wie der Fall ij – diese würden beispielsweise über strata-spezifische Kaplan-Meier-Kurven, Cox-Modelle mit Breslow-Schätzern für die kumulierte baseline-hazard-Funktion oder Überlebensstafeln bestimmt werden. Der Maximum-Likelihood-Schätzer für θ_i ist dann

$$\hat{\theta}_i = \frac{o_i}{-\sum_{j=1}^{n_i} \log(S_{*,ij}(t_{ij}))} = \frac{o_i}{e_i}.$$

Diese Größe kann als eine survival-adaptierte **Standardized Mortality Ratio (SMR)** interpretiert werden, weil die beobachtete Anzahl an Ereignissen der Einrichtung (o_i) mit der erwarteten Anzahl an Ereignissen ($e_i = -\sum_{j=1}^{n_i} \log(S_{*,ij}(t_{ij}))$) verglichen wird. Letzteres wird anhand von strataspezifischen Survivalfunktionen ermittelt, welche für die Standardpopulation (hier gewählt als alle Fälle im Beobachtungszeitraum) berechnet wurden. Der Punktschätzer kann daher als aus den QS-Verfahren nach QSKH-RL bekannter O/E-Indikator interpretiert und dargestellt werden. Das Verfahren ermöglicht somit auch direkt eine Risikoadjustierung, indem z. B. geschichtete Survivorkurven auf dem Bundesdatenpool berechnet werden. Falls es keine Stratifizierung gibt, resultiert aus (1), dass $S_i(t) = S_*(t)^{\theta_i}$ gilt.

Anstelle des direkten Maximum-Likelihood-Schätzers kann die Schätzung auch im Rahmen eines Poisson-Regressionsmodells mit Log-link-Funktion formuliert werden:

$$\delta_{ij} \sim \text{Po}(\mu_{ij}),$$

$$\log(\mu_{ij}) = \underbrace{\log(-\log(S_{*,ij}(t_{ij})))}_{\text{offset}} + \beta_{0,i}.$$

Somit ist $\hat{\theta}_i = \exp(\hat{\beta}_{0,i})$. Diese Schätzung ist auch als die **Subject-Years-Methode** bekannt (Berry 1983) und kann mit gängiger GLM-Software in jedem Statistikpaket für jeden Leistungserbringer berechnet werden.

Ein Vorteil des SMR-Ansatzes ist, dass der Vergleich für die komplette Dauer vorgenommen wird. Eventuell kann es jedoch schwer sein, im Strukturierten Dialog anhand eines hohen $\hat{\theta}_i$ -Wertes einen konkreten Zeitbezug herzustellen. Dies ist ggf. einfacher mit dem festgelegten Vergleichszeitpunkt τ oder durch die Visualisierung der einrichtungsspezifischen Survivalkurve herzustellen.

Ein Nachteil des obigen Ansatzes ist jedoch, dass die Veränderung in der Hazardrate nicht unbedingt multiplikativ bzw. auch nicht für alle Risiken gleich sein muss. Zum Beispiel könnte die Veränderung zeitlich abhängig sein, welches z. B. durch eine Qualitätsverbesserung nach dem Strukturierten Dialog der Fall wäre. Zukünftige Erweiterungen des Modell sind hier möglich, können aber bei der zum Zeitpunkt der Entwicklung der Methodik vorliegenden Datenlage nur schwer beurteilt werden.

3.5.1 Konfidenzintervalle für das SMR

Konfidenzintervalle für θ_i sind am einfachsten als likelihood-basierte Konfidenzintervalle zu berechnen. Die Intervalle können auch zur Berechnung von einseitigen Tests der Art $H_0: \theta_i = 1$ vs. $H_1: \theta_i > 1$ verwendet werden. Für einen vorgegebenen Wert θ_0 in der Nullhypothese (meistens $\theta_0 = 1$ wie oben) würde diese Untersuchung dann auf der Likelihood-Quotienten-Statistik

$$T_i(\theta_0) = 2\{l_i(\hat{\theta}_i) - l_i(\theta_0)\} = 2\{o_i \log(\hat{\theta}_i/\theta_0) - (\hat{\theta}_i - \theta_0)e_i\}$$

geschehen. Weil asymptotisch $T_i \sim \chi^2(1)$ gilt, können über die Dualität zwischen Test und Konfidenzintervall entsprechende Konfidenzintervalle bestimmt werden (Lachin 2010). Zum Beispiel kann ein 95%-Konfidenzintervall für θ_i als die beiden Lösungen der Gleichung $T_i(\theta_0) = \chi_{1-0.05}^2(1) = 3,84$ numerisch über eine Nullstellensuche bestimmt werden. Hier gibt $\chi_q^2(1)$ das q -Quantil der χ^2 -Verteilung mit einem Freiheitsgrad an.

Alternativ kann der Poisson-GLM-Ansatz auch verwendet werden, um likelihood-basierte Konfidenzintervalle für θ zu erhalten. Berry (1983, S. 182) kommentiert, dass Score-Intervalle ggf. eine bessere Überdeckungswahrscheinlichkeit bei kleinen Fallzahlen haben. Diese Intervalle stehen jedoch nicht in gängiger Statistiksoftware zur Verfügung. Exakte Poisson-Verfahren sind nicht direkt anwendbar, weil die Poisson-Verteilung nur eine äquivalente Likelihood und nicht das korrekte Verteilungsmodell darstellt. Somit stellen die likelihood-basierten Konfidenzintervalle einen guten Kompromiss zwischen Umsetzbarkeit und guter Überdeckungswahrscheinlichkeit dar. Für Leistungserbringer ohne beobachtete Ereignisse, $o_i = \sum_{j=1}^{n_i} \delta_{ij} = 0$, wird die obere Grenze des Konfidenzintervalls unter der Annahme fixer e_i dabei als $\frac{n_i}{e_i}$ angegeben.

3.6 Auswertung eines risikoadjustierten Follow-up-Indikators

Auf Basis der oben dargestellten Methodik kann ein risikoadjustierter Follow-up-Indikator wie folgt ausgewertet werden: Sei $(s_{\text{anfang}}, s_{\text{ende}})$ der gewählte **Beobachtungszeitraum** für einen risikoadjustierten Follow-up-Indikator. Dies könnte z. B. der Zeitraum vom Ersteingriff bis 365 Tage nach Datum des Ersteingriffs sein. Die Auffälligkeitseinstufung eines Ergebnisses eines Leistungserbringers Einrichtung geschieht wie folgt:

1. Pro Leistungserbringer $i = 1, \dots, m$ wird $\widehat{\text{SMR}}_i$ anhand der Daten des Beobachtungszeitraums $(s_{\text{anfang}}, s_{\text{ende}})$ bestimmt.
2. Berechne das empirische obere x%-Perzentil der Werte in $\left\{ \widehat{\text{SMR}}_i \right\}_{i=1}^m$. Analog der Perzentilbestimmung bei Indikatoren in QS-Verfahren nach QSKH-RL sollen bei der Berechnung nur die Ergebnisse von Leistungserbringern mit einer Mindestanzahl an Ersteingriffen im Beobachtungszeitraum eingehen. Der erhaltene Wert ρ funktioniert im Folgenden als **Referenzwert** für den Follow-up-Indikator. Für Follow-up-Indikatoren, bei denen die Auffälligkeitseinstufung nicht perzentilbasiert, sondern auf Basis eines festen Referenzwertes stattfindet, entspricht ρ diesem Referenzwert.
3. Das Ergebnis eines Leistungserbringers wird als **rechnerisch auffällig** eingestuft, wenn

$$\widehat{\text{SMR}}_i > \rho.$$

Das Ergebnis für einen Leistungserbringer für prozedurassoziierte Probleme im QS-Verfahren *Herzschrittmacherversorgung* kann tabellarisch auf folgende Weise dargestellt werden (wobei der Punktschätzer für das SMR Zwecks Konsistenz mit anderen Indikatoren in den QS-Verfahren nach QSKH-RL hier O / E genannt wird):

LE	n	o	e	o / e	95%-KI	Referenzwert
X	57	8	2,48	3,23	1,48 – 6,01	3,12

Für den O/E-Indikator sind die 8 Folgeeingriffe bei diesem Leistungserbringer mit der erwarteten Zahl von 2,48 Folgeeingriffen zu vergleichen. Das SMR nach Breslow (1975) ist für diesen Leistungserbringer also $8 / 2,48 = 3,23$ (95%-KI 1,48 – 6,01). D. h., die Hazardrate für ein prozedurassoziiertes Problem ist 3,23-mal höher bei diesem Leistungserbringer im Vergleich zur bundesweiten Hazardrate. Da dieser Wert über dem Referenzwert von 3,12 liegt, wird das Ergebnis dieses Leistungserbringers als rechnerisch auffällig eingestuft.

Die in der Tabelle dargestellten Werte sind zusammenfassend wie folgt zu interpretieren: Im Beobachtungszeitraum 2015 bis 2016 hatten im Vergleichszeitraum 8 Patientinnen und Patienten des Leistungserbringers X einen Folgeeingriff, verglichen mit dem risikoadjustierten erwarteten Wert von 2,48, was einem O/E-Wert von 3,23 entspricht. Da dieser Wert über dem Referenzwert liegt, wird das Ergebnis dieses Leistungserbringers als rechnerisch auffällig eingestuft.

3.7 Zusammenfassung

Der Kaplan-Meier-Schätzer ermöglicht eine nicht-parametrische Schätzung des Anteils an Fällen ohne Folgeeingriff zu einem Zeitpunkt τ , zu welcher nach der von Fay und Brittain (2016) beschriebenen Methode exakte Konfidenzintervalle berechnet werden können. Diese Methode garantiert

auch bei kleinen Fallzahlen gute Überdeckungswahrscheinlichkeiten. Der Kaplan-Meier-Schätzer bietet sich daher vor allem für nicht-risikoadjustierte Follow-up-Indikatoren an, wobei prinzipiell auch die Berechnung von stratifizierten Kaplan-Meier-Kurven als Methode zur Risikoadjustierung möglich ist.

Das Auswertungsverfahren auf Basis des multiplikativen Hazardmodells beruht im Gegensatz zum Kaplan-Meier-Schätzer auf einer Modellannahme (zeitlich konstante multiplikative Veränderung der Hazardrate), bietet aber eine gute Möglichkeit der Risikoadjustierung über die Berechnung stratifizierter Kaplan-Meier-Kurven; diese könnte für zukünftige Auswertungen ggf. z. B. mittels Cox-Modell erweitert werden. Die Methodik des multiplikativen Hazardmodells bietet sich daher vor allem für risikoadjustierte Follow-up-Indikatoren an.

Im Rahmen der QS-Verfahren nach QSKH-RL werden daher nicht-risikoadjustierte Follow-up-Indikatoren auf Basis des vorgestellten Kaplan-Meier-Schätzers ausgewertet, risikoadjustierte Follow-up-Indikatoren auf Basis des vorgestellten multiplikativen Hazardmodells.

4 Berechnung in R

Im Folgenden wird die Berechnung der Indikatoren beispielhaft für Folgeeingriffe aufgrund von prozedurassoziierten Problemen bei implantierten Herzschrittmachersystemen mit der statistischen Software R illustriert.¹ Sei `fu` ein `data.frame` mit der Ereignishistorie für alle Fälle im Beobachtungszeitraum und mit `lid_IND` als Faktor-Variable, die die Pseudonyme der Leistungserbringer des Implantationseingriffs enthält und `HSM_System` als Faktorvariable mit den Ausprägungen `Einkammersystem`, `Zweikammersystem` und `CRT_System`. Letztere wird hier für die Risikoadjustierung verwendet. Der nicht-risikoadjustierte Indikator für den Anteil ohne Folgeeingriff kann auf folgende Art mit R bestimmt werden:

```
##Nur Folgeeingriffe im betrachteten Krankenhaus 'KH'
fu_KH <- dplyr::filter(fu, lid_IND==KH)
##Kaplan-Meier Berechnung der Survivor-Kurve
S_bp_kh <- with(fu_KH, bpcp::bpcp(SurvivalZeit_PAP, Event, midp = TRUE, alpha=.05))
##Anteil ohne Folgeeingriff zum Zeitpunkt `tau`
bpcp::StCI(S_bp_kh, tstar=tau, afterMax = 'continue')
```

##	Vergleichszeitpunkt	Anteil ohne Ereignis	95%KI u. Grenze	95%KI o. Grenze
## 1	90	0,85712648	0,74211495	0,93100267

Der risikoadjustierte O/E-Indikator und die zugehörigen Konfidenzintervalle können in R auf verschiedene Arten berechnet werden: manuell oder über die Verwendung von GLM-Software. Hier exemplarisch die GLM-Lösung, welche in R standardmäßig Likelihood-Quotienten basierte Konfidenzintervalle berechnet. Dabei ist `S_star_df` ein `data.frame`, welcher die Anteile an Fällen ohne Folgeeingriffe über alle Leistungserbringer getrennt nach den Ausprägungen der Risikoadjustierungsvariablen (und somit S^* für die drei Gruppen `Einkammersystem`, `Zweikammersystem` und `CRT_System`) für den Zeitraum (s_{anfang} , s_{ende}) enthält:

```
S_star_df <- read.csv2('SurvivorFunktion.HSM_PAP.csv')
##Aufbau des data.frame S_star_df:
head(S_star_df)
```

##	SurvivalZeit_PAP	CRT_System	Einkammersystem	Zweikammersystem
## 1	0	0,99364459	0,99481068	0,99598637
## 2	1	0,99142001	0,99162728	0,98824048
## 3	2	0,99018352	0,98941274	0,98239237
## 4	3	0,98795449	0,98772629	0,97857299
## 5	4	0,98522670	0,98652576	0,97500633
## 6	5	0,98373732	0,98505686	0,97290371

```
##Berechnung des Strata-spezifischen Kaplan-Meier-Schätzers Sstar
#@' @param hsm_system System des implantierten Herschrittmachers
#@' @param Tstar Beobachtungszeit
#@' @param data data.frame mit stratifizierten Anteilen an Patienten ohne Folgeeingriff
Sstar_func <- function(Tstar, hsm_system, data=S_star_df){
  stopifnot(is.character(hsm_system))
  sstar <- stepfun(data$SurvivalZeit_PAP, c(data[1, hsm_system], data[, hsm_system]))
  return(sstar(Tstar))
}

fu_KH <- fu_KH %>%
  dplyr::rowwise() %>%
  dplyr::mutate(Sstar = Sstar_func(SurvivalZeit_PAP, HSM_System)
  )
```

¹R-Version: 3.3.3 (2017-03-06); BPCP-Paket-Version: 1.3.4

```
##Berechnung des SMR mit GLM-Software
m <- glm(Event ~ offset(log(-log(Sstar))) + 1, family=poisson, data=fu_KH, epsilon=1e-40)

##o, e, Punktschätzer sowie 95% KI
data.frame(o=sum(fu_KH$Event),e=sum(-log(fu_KH$Sstar)), "o/e"=as.numeric(exp(coef(m))),
'95%-KI untere Grenze'=as.numeric(exp(confint(m, level= 0.95))[1]),
'95%-KI obere Grenze' =if_else(sum(fu_KH$Event)==0, nobs(m)/sum(-log(fu_KH$Sstar)),
as.numeric(exp(confint(m, level= 0.95))[2])))

## o e o/e 95%-KI untere Grenze 95%-KI obere Grenze
## 1 8 1,4526708 5,5070976 2,5166372 10,249981
```

5 Diskussion und Ausblick

Mit der Einführung der Follow-up-Indikatoren kommen im Rahmen der QS-Verfahren nach QSKH-RL erstmals Auswertungsverfahren aus der Ereigniszeitanalyse zur Anwendung. Das vorliegende Dokument bietet einen statistischen Rahmen zur rechnerischen Einstufung von Auffälligkeiten anhand von zwei Indikatortypen für die erhobenen Follow-up-Daten: dem Anteil ohne Folgeeingriff zu einem vorgegeben Zeitpunkt τ und dem O/E-Indikator, welcher die tatsächliche Anzahl an Folgeeingriffen in Relation zu der erwarteten Anzahl an Folgeeingriffen setzt. Der erste Indikatortyp ist einfacher zu interpretieren, wobei sich der zweite Indikatortyp einfacher risikoadjustieren lässt. Somit stellen die beiden Indikatortypen die Pendanten zu den nicht-risikoadjustierten bzw. risikoadjustierten Anteilsindikatoren in den QS-Verfahren nach QSKH-RL dar. Die Anwendung beider hier vorgeschlagener Indikatortypen im Rahmen der Auswertung der Follow-up-Indikatoren nach der QSKH-RL wurde in einer Simulationsstudie getestet, welche sich an den Daten des EJ 2015 orientiert und daraus für längere Beobachtungszeiträume extrapoliert. Die Methoden aus der Ereigniszeitanalyse liefern dabei in den analysierten Szenarien eine zuverlässigere Auffälligkeitseinstufung von Leistungserbringerergebnissen als Methoden, die zensierte Beobachtungen nicht in der Analyse berücksichtigen.

Ziel der hier vorgeschlagenen Methodik ist es, eine Analysestrategie für die erstmalige Auswertung der Follow-up-Daten aus den Erfassungsjahren 2015 und 2016 zu bieten. Für zukünftige Auswertungen stellt sich die Frage, wie mit Erst- und Folgeeingriffen, die bereits in vorherigen Auswertungen berücksichtigt wurden, umzugehen ist. Hier bietet die vorgeschlagene Methodik eine große Flexibilität: Bereits analysierte Ersteingriffe könnten beispielsweise so lange in die Analyse eingeschlossen werden, bis die vom Indikator vorgegebene maximale Beobachtungszeit überschritten ist. Dies bedeutet, dass ein Ersteingriff (mit potentielltem Folgeeingriff) mehrfach in die Analysen eingehen würde. Eine andere, mit dem Strukturierten Dialog möglicherweise besser zu vereinbarende Variante ist, in der Analyse eines Erfassungsjahres nur die Beobachtungszeit in diesem einen Erfassungsjahr zu berücksichtigen; Ersteingriffe, die vor diesem Erfassungsjahr stattgefunden haben, würden dann als **linkszensierte** (und ggf. rechtszensierte) **Beobachtungen** in die Analyse eingehen. Dadurch wäre ausgeschlossen, dass ein Folgeeingriff in die Analyse mehrerer Erfassungsjahre eingeht.

Im Rahmen der Bundesauswertung für das EJ 2016 werden nicht-risikoadjustierte Follow-up-Indikatoren auf Basis des vorgestellten Kaplan-Meier-Schätzers ausgewertet, risikoadjustierte Follow-up-Indikatoren auf Basis des vorgestellten multiplikativen Hazardmodells. Für zukünftige Analysen der Follow-up-Indikatoren werden Vor- und Nachteile dieser Auswertungsstrategie anhand von Simulationsstudien und den dann vorliegenden Daten umfassend evaluiert. Eine abschließende Bewertung möglicher Anpassungen und Weiterentwicklungen der Auswertungsmethodik für zukünftige Auswertungen wird jedoch erst mit einem Vergleich auf Basis von Follow-up-Daten aus mehreren Erfassungsjahren möglich sein. Da zum Stand der Entwicklung der Methoden nur die Follow-up-Daten des EJ 2015 vollständig vorlagen, wurde die statistische Auswertungsmethodik so konzipiert, dass sie einen geeigneten Rahmen für die erstmalige Auswertung der Follow-up-Daten aus den Erfassungsjahren 2015 und 2016 bietet und gleichzeitig verschiedene Strategien für zukünftige Auswertungen zulässt.

Literatur

- Amato, D. A. (1988). A generalized Kaplan-Meier estimator for heterogeneous populations. In: *Communications in Statistics – Theory and Methods* 17.1, S. 263–286.
- AQUA (2015). Sonderexport 2015 (Erfassungsjahr 2014) Sonderbericht QSKH-Follow-up. AQUA – Institut für angewandte Qualitätsförderung und Forschung im Gesundheitswesen.
- Berry, G. (1983). The Analysis of Mortality by the Subject-Years Method. In: *Biometrics* 39.1, S. 173–184.
- Breslow, N. E. (1975). Analysis of Survival Data under the Proportional Hazards Model. In: *International Statistical Review* 43.1, S. 45–57.
- Burroughs, Andrew K u. a. (2006). 3-month and 12-month mortality after first liver transplant in adults in Europe: predictive models for outcome. In: *The Lancet* 367.9506, S. 225–232.
- Cole, S. R. und M. A. Hernan (2004). Adjusted survival curves with inverse probability weights. In: *Comput Methods Programs Biomed* 75.1, S. 45–49.
- Fay, M. P. und E. H. Brittain (2016). Finite sample pointwise confidence intervals for a survival distribution with right-censored data. In: *Statistics in Medicine* 35.16, S. 2726–2740.
- Gandy, A. u. a. (2010). Risk-adjusted monitoring of time to event. In: *Biometrika* 97.2, S. 375–388.
- Ghenim, Rabeh u. a. (2009). One-Year Follow-Up of Nonrandomized Comparison between Coronary Artery Bypass Grafting Surgery and Drug-Eluting Stent for the Treatment of Unprotected Left Main Coronary Artery Disease in Elderly Patients (Aged ≥ 75 Years). In: *Journal of Interventional Cardiology* 22.6, S. 520–526.
- Hardoon, Sarah L., James D. Lewsey und Jan H. P. van der Meulen (2007). Continuous monitoring of long-term outcomes with application to hip prostheses. In: *Statistics in Medicine* 26.28, S. 5081–5099.
- IQTIG (2016). Planungsrelevante Qualitätsindikatoren. Abschlussbericht zur Auswahl und Umsetzung. IQTIG – Institut für Qualitätssicherung und Transparenz im Gesundheitswesen.
- Kirkfeld, R.E. u. a. (2014). Complications after cardiac implantable electronic device implantations: an analysis of a complete, nationwide cohort in Denmark. In: *European Heart Journal* 35.18, S. 1186–1194.
- Klein, J.P. und M.L. Moeschberger (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd. *Statistics for Biology and Health*. Springer New York.
- Lachin, John M. (2010). *Biostatistical Methods – The Assessment of Relative Risks*. 2nd. John Wiley & Sons, Inc.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.
- Statistisches Bundesamt (2017). Sterbefälle, Lebenserwartung. <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/Bevoelkerung/Sterbefaelle/Sterbefaelle.html>. Zugegriffen: 10.02.2017.
- Xie, J. und C. Liu (2005). Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. In: *Statistics in Medicine* 24.20, S. 3089–3110.