

## **Berechnung von Konfidenzintervallen für die Population Impact Number (PIN)**

Ralf Bender  
Institut für Qualität und  
Wirtschaftlichkeit im  
Gesundheitswesen (IQWiG)  
Dillenburger Str. 27  
D-51105 Köln  
Ralf.Bender@iqwig.de

Ulrich Grouven  
Institut für Qualität und  
Wirtschaftlichkeit im  
Gesundheitswesen (IQWiG)  
Dillenburger Str. 27  
D-51105 Köln  
Ulrich.Grouven@iqwig.de

### **Zusammenfassung**

Die Zahl "Number Needed to Treat" (NNT) wird vor allem bei der Ergebnisdarstellung randomisierter klinischer Studien häufig verwendet, um die Stärke des Effekts einer neuen Behandlung oder Intervention zu beschreiben. Die Zahl NNT gibt die durchschnittliche Anzahl von Patienten an, die über einen bestimmten Zeitraum die neue Behandlung erhalten müssen, um bei einem Patienten z.B. ein ungünstiges Ereignis zu verhindern. Zur Intervallschätzung von NNTs wurden sowohl Methoden als auch Software für die erforderlichen Berechnungen entwickelt. Zur Anwendung des NNT-Konzepts in der Public-Health-Forschung wurde ein weiteres Maß vorgeschlagen, die so genannte Population Impact Number (PIN). Während NNT durch den Kehrwert der Differenz der Risiken von zwei unterschiedlichen Gruppen (z.B. Interventions- und Kontrollgruppe) gegeben ist, lässt sich PIN als Kehrwert der Differenz zwischen den Risiken in der Population vor und nach einer populationsweiten Intervention darstellen. Auf diese Weise geht in das Effektmaß PIN der Anteil der Personen ein, deren Risiko sich durch die Intervention verändern würde. Dadurch ist PIN gerade zur Darstellung der Effekte populationsweiter Interventionen geeignet. Es fehlen jedoch bislang Methoden zur Berechnung von Konfidenzintervallen für PIN. In diesem Beitrag wird ein SAS/IML<sup>®</sup>-Macro vorgestellt, mit dessen Hilfe Konfidenzintervalle für das Maß PIN berechnet werden können. Das Programm wird beispielhaft auf Routinedaten der Qualitätssicherung zur Darstellung des Populationeffekts einer Mindestmengenregelung für Knieprothesen angewendet. Das SAS/IML<sup>®</sup>-Macro ist ein nützliches Hilfsmittel zur Punkt- und Intervallschätzung der Population Impact Number in der Public-Health-Forschung.

**Schlüsselwörter:** Konfidenzintervalle; Number Needed to Treat (NNT); Population Impact Number (PIN); Public Health; Risikodifferenz.

## **1 Einleitung**

Die Zahl "Number Needed to Treat" (NNT) wird seit einigen Jahren vor allem bei der Ergebnisdarstellung randomisierter klinischer Studien häufig verwendet, um die Stärke des Effekts einer neuen Behandlung oder Intervention zu beschreiben (Bender, 2005). Die Zahl NNT gibt die durchschnittliche Anzahl von Patienten an, die über einen bestimmten Zeitraum die neue Behandlung erhalten müssen, um bei einem Patienten z.B.

ein ungünstiges Ereignis zu verhindern. Das Maß NNT kann in analoger Weise auch den Behandlungseffekt in Bezug auf wünschenswerte Ereignisse beschreiben. Zur Intervallschätzung von NNT wurde auf der Basis der Wilson-Score-Methode (Newcombe, 1998; Bender, 2001) ein SAS/IML<sup>®</sup>-Programm entwickelt (Bender, 2000).

Zur Anwendung des NNT-Konzepts in der Public-Health-Forschung wurde von Heller & Dobson (2000) ein weiteres Maß vorgeschlagen, die so genannte Population Impact Number (PIN). Während NNT durch den Kehrwert der Differenz der Risiken zweier Gruppen (z.B. Interventions- und Kontrollgruppe) gegeben ist, lässt sich PIN als Kehrwert der Differenz zwischen den Risiken in der Population vor und nach einer populationsweiten Intervention darstellen. Auf diese Weise geht in das Effektmaß PIN der Anteil der Personen ein, deren Risiko sich durch die Intervention verändern würde. Dadurch ist PIN gerade zur Darstellung der Effekte populationsweiter Interventionen geeignet. Es fehlen jedoch bislang Methoden zur Berechnung von Konfidenzintervallen für das Maß PIN. Als Ersatz für Konfidenzintervalle wurde lediglich vorgeschlagen Sensitivitätsanalysen durchzuführen (Heller et al., 2002).

Im Folgenden wird gezeigt, dass auf der Basis der Wilson-Score-Methode auch Konfidenzintervalle für das Effektmaß PIN berechnet werden können. Durch eine Erweiterung des SAS/IML<sup>®</sup>-Programms von Bender (2000) können Punkt- und Intervallschätzungen für NNT und PIN durchgeführt werden. Das Programm wird beispielhaft auf Routinedaten der Qualitätssicherung zur Darstellung des Populationseffekts einer Mindestmengenregelung für Knieprothesen angewendet.

## 2 Effektmaße für binäre Daten

Zur Beschreibung des Unterschieds zwischen zwei Gruppen bei binären Zielvariablen gibt es eine Vielzahl von Effektmaßen. Sehr häufig verwendet werden relative Effektmaße wie das Odds Ratio (OR) oder das relative Risiko (RR) und absolute Effektmaße wie die Differenz zweier Risiken. Wir betrachten hier nur absolute Effektmaße, die sich aus der Differenz von Risiken ableiten.

Wir betrachten hierzu zwei binäre Merkmale, nämlich Exposition ja ( $E=1$ ) und nein ( $E=0$ ) sowie Krankheit ja ( $K=1$ ) und nein ( $K=0$ ). Es soll der Effekt einer Intervention beschrieben werden, die eine Population vor einer Exposition schützt, so dass niemand mehr dieser Exposition ausgesetzt ist. Unter Annahme eines kausalen Zusammenhangs zwischen der Exposition und der Erkrankung, ist es sinnvoll die folgenden Wahrscheinlichkeiten zu betrachten.

- $P(K=1)$  = Wahrscheinlichkeit der Krankheit in der Bevölkerung
- $P(K=1|E=1)$  = Wahrscheinlichkeit der Krankheit unter den Exponierten
- $P(K=1|E=0)$  = Wahrscheinlichkeit der Krankheit unter den Nicht-Exponierten
- $P(E=1)$  = Prävalenz der Exposition in der Bevölkerung

In der Praxis müssen diese Wahrscheinlichkeiten aus einer Studie mit adäquatem Design geschätzt werden. Um den Effekt einer Intervention zu beschreiben wird in der Regel eine prospektive Interventionsstudie benötigt. Wir gehen im Folgenden davon aus,

dass adäquate Daten vorliegen, die sich in Form einer Vierfeldertafel darstellen lassen und konzentrieren uns zunächst auf die theoretischen Eigenschaften der betrachteten Effektmaße.

## **2.1 Die Number Needed to Treat (NNT)**

Die Number Needed to Treat (NNT) ist der Kehrwert der gewöhnlichen Risikodifferenz (RD) zwischen zwei Gruppen. Mit den Bezeichnungen von oben gilt

$$\begin{aligned}RD &= P(K=1|E=1) - P(K=1|E=0) \\NNT &= 1/RD\end{aligned}$$

Im Allgemeinen ist NNT die durchschnittliche Zahl an Personen, die behandelt werden müssen, um bei einem Patienten ein bestimmtes unerwünschtes Ereignis zu verhindern. RD und NNT sind geeignete Maße um den Effekt medizinischer Therapien auf Patienten zu beschreiben. Hier verstehen wir unter "Behandlung" eine Intervention, die eine gesamte Population vor einer schädlichen Exposition schützt, so dass niemand mehr dieser Exposition ausgesetzt ist. In diesem Fall kann die Zahl NNT auch interpretiert werden als die durchschnittliche Zahl exponierter Personen, von denen ein Erkrankungsfall auf die Exposition zurückzuführen ist und durch Schutz vor der Exposition verhindert werden könnte. Die Verwendung von RD und NNT hat hier den Nachteil, dass die Expositionsprävalenz in der Bevölkerung nicht berücksichtigt wird.

## **2.2 Die Population Impact Number (PIN)**

Ein NNT-ähnliches Maß mit Populationsbezug wurde von Heller & Dobson (2000) vorgeschlagen und als Population Impact Number (PIN) bezeichnet. PIN lässt sich darstellen als Kehrwert der Differenz des Erkrankungsrisikos in der Bevölkerung und des Erkrankungsrisikos unter den nicht exponierten Personen, d.h. es gilt

$$\begin{aligned}PRD &= P(K=1) - P(K=1|E=0) \\PIN &= 1/PRD\end{aligned}$$

In Ermangelung einer allgemein üblichen Bezeichnung für diese Risikodifferenz verwenden wir hier den Begriff "populationsbezogene Risikodifferenz" (PRD), da dies gerade der Zähler des in der Epidemiologie üblichen populationsbezogenen attributablen Risikos (PAR) ist.

PIN kann interpretiert werden als die durchschnittliche Zahl von Personen aus der Gesamtbevölkerung, von denen ein Erkrankungsfall auf die Exposition zurückzuführen ist. Im Gegensatz zu NNT ist PIN abhängig von der Expositionsprävalenz und damit geeignet zur Darstellung des Effekts der Exposition in der Bevölkerung (Heller et al., 2002). Analog kann PIN verwendet werden zur Darstellung des Effekts einer populationsweiten Intervention, die die Bevölkerung vor einer schädlichen Exposition schützt. PIN ist dann die durchschnittliche Zahl von Personen aus der Gesamtbevölkerung, von denen

eine Person durch die Intervention bezüglich der betrachteten Zielvariable einen Nutzen hat.

### 3 Berechnung von Konfidenzintervallen

Ein allgemeiner Ansatz zur Berechnung von Konfidenzintervallen für NNT ist das Vertauschen und Invertieren der entsprechenden Konfidenzgrenzen von RD (Bender, 2001). Hierbei muss die ungewöhnliche Skala von NNT beachtet werden, die dazu führt, dass bei einem nicht signifikanten Effekt der Konfidenzbereich für NNT aus zwei Halbinsintervallen besteht. Während die übliche Wald-Methode häufig inadäquate Ergebnisse liefert, stellt die Wilson-Score-Methode für viele praxisrelevante Situationen eine adäquate Methode zur Intervallschätzung von RD dar (Newcombe, 1998), die auch zur Intervallschätzung von NNT geeignet ist (Bender, 2001). Zur Durchführung der entsprechenden Berechnungen wurde das SAS/IML<sup>®</sup>-Programm "nnt\_ci.sas" entwickelt (Bender, 2000).

Da zwischen PRD und RD die Beziehung

$$\text{PRD} = P(E=1) \times \text{RD}$$

gilt, hängt unter der Annahme einer festen Expositionsprävalenz die Schätzung von PRD nur noch von RD ab. In diesem Fall lassen sich somit Konfidenzintervalle für PRD aus den Konfidenzgrenzen von RD durch Multiplikation mit  $P(E=1)$  ableiten. Konfidenzintervalle für PIN erhält man dann wiederum durch das Vertauschen und Invertieren der entsprechenden Konfidenzgrenzen von PRD. Wie bei NNT muss bei der Intervallschätzung von PIN die ungewöhnliche Skala von PIN beachtet werden.

Zur Durchführung der notwendigen Berechnungen wurde das SAS/IML<sup>®</sup>-Programm "nnt\_ci.sas" (Bender, 2000) erweitert. Mit Hilfe des neuen SAS/IML<sup>®</sup>-Macros "pin\_nnt\_ci.sas" lassen sich nun Punkt- und Intervallschätzungen von NNT und PIN durchführen. Beim Aufruf des Macros durch

```
%PIN (e1, n1, e2, n2);
```

ist es hierbei erforderlich für  $e_i$  ( $i=1,2$ ) die Zahl der beobachteten Ereignisse und für  $n_i$  ( $i=1,2$ ) die Stichprobenumfänge der beiden Gruppen einzusetzen. Günstigerweise wählt man für Gruppe 1 diejenige mit dem höheren und für Gruppe 2 diejenige mit dem niedrigeren Risiko aus. Als Output erhält man neben deskriptiven Angaben Punkt- und Intervallschätzungen für RD, NNT, PRD und PIN zu den Konfidenzniveaus 95% und 99%. Das Programm ist auf der CD des Tagungsbands verfügbar und außerdem im Internet erhältlich unter <http://www.rbsd.de/softw.html/>.

### 4 Beispiel: Mindestmengen für Knieprothesen

Auf dem Hintergrund möglicher Assoziationen zwischen Behandlungsmenge und Ergebnisqualität wurden in der Vergangenheit für verschiedene Leistungen des Gesundheitswesens Mindestmengen (MM) je Krankenhaus festgelegt, die pro Jahr erfüllt sein müssen, damit diese Leistung angeboten werden darf (Schröder et al., 2007). So wurde

z.B. im Jahr 2005 für die primäre Knieprothesenimplantation vom Gemeinsamen Bundesausschuss (G-BA) eine Mindestmenge von 50 Implantationen pro Krankenhaus und Jahr festgelegt. Im Rahmen eines Auftrags des G-BA an das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) wurden zur Untersuchung von Schwellenwerten bezüglich Mindestmengen für Knieprothesen Daten der externen Qualitätssicherung der Bundesgeschäftsstelle Qualitätssicherung gGmbH (BQS) ausgewertet (IQWiG, 2005). Hierbei wurden verschiedene regressionsbasierte Ansätze zur Ermittlung von Schwellen- und Benchmarkwerten angewendet (Bender & Grouven, 2006; Grouven et al., 2008). In dieser Arbeit konzentrieren wir uns jedoch auf die Darstellung des populationsbezogenen Effekts einer Mindestmengenregelung mit Hilfe von PIN. Vergleichend wird zusätzlich auch NNT betrachtet. Da es sich bei den verwendeten Daten um Routinedaten handelt, die nicht im Rahmen einer prospektiven Interventionsstudie erhoben wurden, haben die Ergebnisse bezüglich einer Mindestmengenregelung nur hypothesengenerierenden Charakter.

Eine Mindestmenge unterteilt letztlich die Gesamtheit aller Krankenhäuser in solche mit niedriger und hoher Fallzahl. Betrachtet man als Indikator der Ergebnisqualität die postoperative Infektion und geht man von einem monoton sinkenden Infektionsrisiko mit steigender Patientenzahl aus, so kann man die Unterbringung in einem Krankenhaus mit niedriger Fallzahl als Exposition ansehen. In dieser Situation würden die Ergebnisqualität (gemessen am sinkenden Infektionsrisiko) und auch die Prävalenz der Exposition mit steigender Mindestmenge immer weiter zunehmen.

Als Beispiel betrachten wir hier die Ergebnisse für den Ergebnisqualitätsindikator Infektion bezüglich des vom IQWiG geschätzten Benchmarkwerts  $MM=116$  sowie bezüglich des vom G-BA im Jahr 2005 festgelegten Werts  $MM=50$  (IQWiG, 2005).

**Tabelle 1:** Anzahl von Infektionen und Gesamtanzahl von Patienten mit Knieprothese in Krankenhäusern mit niedriger und hoher Fallzahl bei Verwendung der Mindestmengen  $MM=116$  und  $MM=50$

Mindestmenge	Patienten aus Krankenhäusern mit niedriger Fallzahl		Patienten aus Krankenhäusern mit hoher Fallzahl	
	Infektionen	Gesamtanzahl	Infektionen	Gesamtanzahl
MM = 116	365	31361	505	79188
MM = 50	135	8251	735	102098

Mit den Daten aus Tabelle 1 erhält man mit Hilfe des SAS/IML<sup>®</sup>-Macros "pin\_nnt\_ci.sas" die folgenden Ergebnisse:

MM=116: NNT = 187 (95%-Konfidenzintervall: 149 – 246)  
 PIN = 664 (95%-Konfidenzintervall: 529 – 872)  
 MM=50: NNT = 109 (95%-Konfidenzintervall: 82 – 152)  
 PIN = 1460 (95%-Konfidenzintervall: 1098 – 2031)

Die Ergebnisse bezüglich NNT bedeuten, dass unter der Annahme einer kausalen Beziehung zwischen Fallzahl und Infektionsrisiko bei Verwendung der Grenze  $MM=116$  unter 187 Patienten aus Krankenhäusern mit niedriger Fallzahl eine Infektion mehr zu erwarten ist als bei Krankenhäusern mit hoher Fallzahl. Bei Verwendung der Grenze  $MM=50$  wären es eine Infektion mehr unter 109 Patienten. Diese Ergebnisse verdeutlichen den Nachteil des Effektmaßes NNT. Der Effekt gemessen als NNT ist bei  $MM=50$  größer als bei  $MM=116$ , obwohl durch  $MM=50$  eine geringere Ergebnisqualität erreicht wird als durch  $MM=116$ .

Dagegen beschreibt PIN durch Berücksichtigung der Expositionsprävalenz den populationsbezogenen Effekt, der bei kleinerer Mindestmenge geringer wird. Bei Verwendung der Grenze  $MM=116$  ist unter 664 Patienten mit Knieprothese eine Infektion aufgrund der Behandlung in einem Krankenhaus mit niedriger Fallzahl zu erwarten. Von durchschnittlich 664 Patienten mit Knieprothese würde also ein Patient von der Mindestmengenregelung  $MM=116$  profitieren. Bei Verwendung von  $MM=50$  wäre dies erwartungsgemäß ein Patient von 1460 Patienten, also ein deutlich geringerer Effekt als bei  $MM=116$ .

Mit keiner der beiden Mindestmengen erreicht man jedoch eine ausreichende klinische Relevanz, die vorab durch  $PRD=0.4\%$  bzw.  $PIN=250$  festgelegt wurde (IQWiG, 2005). Zudem ergaben weitere Berechnungen zum Qualitätsindikator Beweglichkeit, dass sich insgesamt kein einheitlicher Grenzwert bezüglich der Fallzahl pro Krankenhaus finden lässt, der klar Häuser mit hoher und niedriger Qualität voneinander trennt (IQWiG, 2005).

## 5 Schlussfolgerung

Die populationsbezogene Risikodifferenz (PRD) und deren Kehrwert, die Population Impact Number (PIN) stellen geeignete Maße dar, um den Effekt populationsweiter Interventionen bezüglich binärer Zielvariablen zu beschreiben. Im Gegensatz zur üblichen Risikodifferenz (RD) und deren Kehrwert Number Needed to Treat (NNT) wird die Prävalenz der Exposition in der Bevölkerung bei PRD und PIN berücksichtigt. Mit Hilfe des SAS/IML<sup>®</sup>-Macros "pin\_nnt\_ci.sas" lassen sich Punkt- und Intervallschätzungen für Risikodifferenzen, NNT und PIN durchführen.

## **Literatur**

- [1] R. Bender (2000): Berechnung von Konfidenzintervallen für die Zahl "Number Needed to Treat" (NNT). In: R.H. Bödeker, M. Hollenhorst (Hrsg.): Proceedings der 4. Konferenz für SAS-Anwender in Forschung und Entwicklung, S. 1-7. AG Statistik des Inst. für Med. Informatik, Gießen.
- [2] R. Bender (2001): Calculating confidence intervals for the number needed to treat. *Control. Clin. Trials* 22, 102-110.
- [3] R. Bender, R. (2005): Number needed to treat (NNT). In: P. Armitage, T. Colton (Eds.): *Encyclopedia of Biostatistics*, pp. 3752-3761. Wiley, Chichester.
- [4] R. Bender, U. Grouven (2006): Möglichkeiten und Grenzen statistischer Regressionsmodelle zur Berechnung von Schwellenwerten für Mindestmengen. *Z. ärztl. Fortbild. Qual. Gesundh. wes.* 100, 93-98.
- [5] U. Grouven, H. Küchenhoff, P. Schröder, R. Bender (2008): Flexible regression models are useful tools to calculate and assess threshold values in the context of minimum provider volumes. *J. Clin. Epidemiol.* 61 (in press).
- [6] R.F. Heller, A.J. Dobson (2000): Disease impact number and population impact number: Population perspectives to measures of risk and benefit. *BMJ* 321, 950-952.
- [7] R.F. Heller, A.J. Dobson, J. Attia, J. Page (2002): Impact numbers: Measures of risk factor impact on the whole population from case-control and cohort studies. *J. Epidemiol. Community Health* 56, 606-610.
- [8] IQWiG (2005): Entwicklung und Anwendung von Modellen zur Berechnung von Schwellenwerten bei Mindestmengen für die Knie-Totalendoprothese. Abschlussbericht B05/01a. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Köln.
- [9] R.G. Newcombe (1998): Interval estimation for the difference between independent proportions: Comparison of eleven methods. *Stat. Med.* 17, 873-890.
- [10] P. Schröder, U. Grouven, R. Bender (2007): Können Mindestmengen für Knieprothesen anhand von Routinedaten errechnet werden? Ergebnisse einer Schwellenwertanalyse mit Daten der externen stationären Qualitätssicherung. *Orthopäde* 36, 570-576.