

Malgorzata Roos, Ulrich Helfenstein

Abt. Biostatistik, ISPM, Universität Zürich

---

## Methodologisches Editorial

In diesem methodologischen Kommentar geht es um einen Artikel (A)<sup>1</sup>, einen darauf folgenden Beitrag mit Replik (B)<sup>2</sup> und einen weiteren Letter in dieser Ausgabe von *Sozial- und Präventivmedizin* (C)<sup>3</sup>. Im folgenden nehmen wir an, dass der Leser mit der Thematik und den Daten der drei Beiträge in den Grundzügen vertraut ist. Der Klarheit halber geben wir zuerst Mini-zusammenfassungen der Beiträge (A), (B) und (C), so wie wir sie verstanden haben. Anschliessend gehen wir auf einige spezielle Punkte der methodologischen Problematik bei der statistischen Modellbildung ein.

### Mini-Zusammenfassungen der Beiträge (A), (B) und (C)

(A) Das Modell dieses Artikels zur Schätzung der Anzahl Drogenkonsumenten ist kein in der üblichen medizinischen Statistik verwendetes elementares Standardmodell. Es besteht aus einer verallgemeinerten Poissonverteilung, deren Rate eine Mischung von 2 Gamma-Verteilungen ist; oder dazu äquivalent aus einer Mischverteilung aus zwei Negativbinomialverteilungen. Dieses Modell erlaubt eine bessere Anpassung an die Daten als die einfache Poissonverteilung.

Die Schlussfolgerung der Datenanalyse lautet sinngemäss: Positiver zeitlicher Trend der Anzahl der Drogenkonsumenten von 1990 bis 1993.

(B) Die Autoren dieses Beitrages kritisieren einige Punkte der Methodik von (A) und schlagen Verbesserungen vor (verallgemeinerte Poissonverteilungen, deren Rate eine Mischung von 3 Gammaverteilungen ist). Insbesondere geben sie neben den Punktschätzungen eine Schätzung der Ungenauigkeit. Dazu benützen sie eine anspruchsvolle neuere Methode (Markov-chain-Monte-Carlo). Ausserdem wird die Wichtigkeit des Parsimonie-Prinzips betont: Falls zwei Modelle den gleichen "badness-of-fit" ergeben, wird dasjenige Modell vorgezogen, welches weniger unbekannt Parameter enthält (eine statistische Form von Ockham's razor). Mit Hilfe von Informationskriterien wird aus 5 vorgeschlagenen Modellen das am wenigsten schlechte ausgewählt. Aus (B) respektive der Replik von (A) auf (B) folgt sinngemäss eine etwas abgeschwächte Form der Schlussfolgerung von (A): „Anzahl Drogenkonsumenten von 1993 signifikant höher als Anzahl von 1990“.

(C) Der Autor dieses Beitrages ist der Ansicht, dass die „statischen“ Modelle (A) und (B) falsch ge-

wählt seien. Statt dessen schlägt er „dynamische Modelle“ vor. Wir vermuten, dass der Autor die Rohdaten für seine Analysen benützt. Da in einem "Letter" nur wenig Platz zur Verfügung steht, muss er natürlich in den "References" auf weiterführende Literatur zur Erklärung seiner „dynamischen“ Modelle verweisen. Uns ist diese Literatur in der kurzen verfügbaren Zeit nicht zugänglich. Wir müssen uns deshalb mit der sinn gemässen Schlussfolgerung des Autors begnügen: „Kein signifikanter zeitlicher Trend auffindbar“.

### Diskussion

Wie man sieht, führt die Analyse derselben Problematik bei (A, B) zu anderen Schlussfolgerungen als bei (C).

Folgende Punkte erscheinen uns relevant:

#### 1. Vorhersage

(i) Ein Modell mit zeitlich geordneten Daten kann helfen, die gegebenen Daten besser zu verstehen, *ohne* eine Vorhersage auf die folgenden Jahre zu versuchen.

(ii) Ein Modell mit zeitlich geordneten Daten kann zum Ziel haben,

*Vorhersagen* über zukünftige Ereignisse zu machen: Das Modell als "prediction machine".

Bei allen drei Beiträgen war das Ziel (i) von Interesse.

Bei der Modellbildung von (A) war es ausserdem implizit von Interesse, das Ziel (ii), die Vorhersage, mindestens qualitativ zu verfolgen ((A), S. 105). Eine Zunahme in den Jahren 1990 bis 1993 könnte darauf hindeuten, dass es z. B. Sinn macht, für die folgenden Jahre die vermehrten nötigen finanziellen und personellen Ressourcen bereitzuhalten.

Da spätere Daten ebenfalls verfügbar wurden (Replik von (A) auf (B)) und eine Abnahme der Konsumenten beobachtet wurde, könnte man vermuten, dass das Modell als Vorhersagehilfe versagt hat. Man könnte die Daten jedoch auch so interpretieren, dass ein möglicherweise unbekannter „Mechanismus“ eine Abweichung bewirkt hat.

Für (C) ist die Beobachtung einer späteren Abnahme nicht überraschend. Das Modell (C) hat ja keinen „deterministischen Trend“ behauptet. Das Modell als qualitative Vorhersagemaschine hat „funktioniert“.

## 2. Abhängigkeiten, Modellunsicherheit, Vertrauensbereiche

Falls „Zeitreihendaten“ vorliegen, kann nicht ohne weiteres angenommen werden, dass die beobachteten Anzahlen in aufeinanderfolgenden Jahren unabhängig sind. Im vorliegenden Datensatz erscheint uns eine solche Abhängigkeit nicht unplausibel, da z. B. ein Drogenabhängiger im Jahr 1991 auch als Drogenabhängiger im Jahr 1992 auftauchen kann.

Abhängigkeit könnte auch innerhalb eines Jahres zwischen verschiedenen Personen wesentlich sein. Bekannte haben die Tendenz "Cluster" zu bilden. Anzeigen, die bei einem der polizeilichen Einsätze erfolgten, können abhängig

sein. Diese versteckten möglichen Abhängigkeiten, könnten bewirken, dass die tatsächlichen Vertrauensbereiche breiter sind als in (B) angegeben.

Die in (B) erwähnte Modellunsicherheit (Auswahl aus 5 Modellen) führt ebenfalls zur Vergrößerung der Ungenauigkeit der Schätzungen. Wenn diese zusätzlichen Effekte in Betracht gezogen werden könnten, wäre es möglich, dass sich der Unterschied zwischen 1990 und 1993 auch „durch Zufall“ erklären lässt.

## 3. Verteilungen der Anzeigeraten

Es wurde in den Arbeiten (A) und (B) vorgeschlagen, dass sich die Drogenkonsumenten aus 2 resp. 3 Untergruppen zusammensetzen, in welchen die Anzeigeraten Gamma-verteilt sind.

Es ist zu beachten, dass die Form der Gamma-Verteilungen durch den "Shape-parameter"  $r$  bestimmt wird: Für  $r > 1$  ist die Dichte unimodal und hat den Wert 0 für Anzeigerate = 0. Für  $r \leq 1$  ergeben sich links-schiefe Verteilungen mit dem Spezialfall der Exponentialverteilung für  $r = 1$ .

Eine aufgrund des Vorwissens bestimmte Fixierung der beiden Shape-parameter in jeder der beiden Untergruppen a priori auf  $r = 4$  bedeutet, dass insbesondere Exponentialverteilungen ausgeschlossen werden. Eine Begründung für diese spezielle Wahl der Shape-parameter wäre angebracht. Graphische Darstellungen der geschätzten Verteilungen könnten für die Fachleute hilfreich sein, um zu beurteilen, ob die Verteilungen der Anzeigeraten in den Untergruppen plausibel sind oder nicht.

## 4. Klarheit der Darstellung

In regulären Artikeln wie (A), die anspruchsvolle statistische Methoden beschreiben, sollten alle Schritte so klar und vollständig dargestellt werden, dass sie vom

statistisch gebildeten Leser nachvollziehbar sind. Dazu gehören auch scheinbar belanglose Details wie das korrekte Zitieren. Wenn z. B. auf ein mathematisches Buch verwiesen wird, sollten auch die betreffenden Seiten angegeben werden, wo die interessierenden Formeln oder die Herleitung einer Formel gefunden werden können. Alle unnötigen Formeln oder Herleitungen sollten jedoch weglassen werden, um dem Leser das Verständnis nicht noch zusätzlich zu erschweren.

## 5. Daten, Qualität und Interpretation

(C) greift das wichtige Problem der Datenqualität auf. (C) weist insbesondere darauf hin, dass die Daten nicht zwangsläufig einen Anstieg des Drogenkonsums bedeuten. Ein Anstieg könnte auch durch vermehrte Aktivität der Polizei vorgetäuscht werden. Da wir Laien auf dem Gebiet der Drogenforschung sind, können wir zu diesem wichtigen Thema nichts beitragen.

## Schlussfolgerung

– Wir konnten uns insbesondere aus Gründen des Zeitdrucks nicht über alle Details Klarheit verschaffen. Wegen der oben aufgeführten zusätzlichen Unsicherheitsfaktoren würden wir jedoch zur Vorsicht warnen bezüglich der Schlussfolgerung von (A) respektive Replik von (A) auf (B): „signifikant ansteigender Trend“.

– Die Modelle von (A, B) und (C) müssen sich gegenseitig nicht unbedingt ausschliessen. Aus unserer Sicht repräsentieren sie verschiedene Zugänge zu denselben komplexen Fragen. Dass verschiedene Modelle mit verschiedenen Interpretationen nebeneinander bestehen können, hätte gut auch innerhalb von (B) allein auftreten können: Wenn nämlich die Informationskriterien AIC (oder BIC)

von zwei Modellen gleich sind, sich aber in dem “badness-of-fit Term” und in der Anzahl Parameter unterscheiden.

– Wir haben gesehen, dass die mathematisch-statistischen Anforderungen an den modernen Epidemiologen auf diesem Gebiet äusserst anspruchsvoll geworden sind. Falls er sich nicht der Mühe unterzieht, die angewandten statistischen Methoden in ihren Stärken und Schwächen gründlich zu studieren, ist es fraglich, ob er z.B. bei Drogenproblemen wissenschaftlich fundiert mitdiskutieren kann.

– Es sollten freiwillige, unabhängige und qualifizierte Reviewer ausgewählt werden, welche einen

mathematisch orientierten Artikel nicht zur Publikation vorschlagen, bevor die Autoren die wesentlichen Unklarheiten bereinigt und einen auch didaktisch hilfreichen Beitrag bereitgestellt haben.

---

#### Literaturverzeichnis

- 1 *Knolle H.* Zunehmender Heroin- und Kokainkonsum in der Schweiz seit 1990: Anwendung eines verallgemeinerten Poisson-Ansatzes auf Anzeigedaten. *Soz Präventivmed* 1997; 42: 105–113.
- 2 *Zwahlen M, Neuenschwander BE.* Zunehmender Heroin- und Kokainkonsum in der Schweiz seit 1990: Anwendung eines verallgemeiner-

ten Poisson-Ansatzes auf Anzeigedaten. Ein “Letter to the editor” mit Replik des Autors in: *Soz Präventivmed* 1997; 42:321–323.

- 3 *Estermann J.* Doch kein zunehmender Heroin- und Kokainkonsum in der Schweiz? *Soz Präventivmed* 1998; 43:173–174.

---

#### Korrespondenzadresse

Dr. M. Roos  
 Dr. U. Helfenstein  
 Institut für Sozial- und Präventivmedizin der Universität Zürich  
 Abteilung Biostatistik  
 Sumatrastr. 30  
 CH-8006 Zürich  
 Fax +41 16344986