

Eine vereinfachte Methode zur Erhebung der Legalbewährung

Dipl. Psych. Dr. Harald Meyer
Universität Bamberg
Inst. für Theoretische Psychologie
Markusstr. 2, 96045 Bamberg
Harald.meyer@ppp.uni-bamberg.de

1. Problemstellung

Verkehrssicherheit ist ein hohes Gut. Bei den Anbietern¹ von Maßnahmen zur Wiederherstellung der Verkehrssicherheit nach Verkehrsauffälligkeit (Nachschulung und Verkehrspsychotherapie) besteht zunehmende Einigkeit im Ziel, ihre Maßnahmen strengen Qualitätskontrollen zu unterwerfen. Dabei hat die sog. „Legalbewährung“² als Kontrollinstrument einen besonders hohen Stellenwert. Unabhängig von unterschiedlichen Benennungen meint man damit

- den Prozentsatz π der Gesamtklientel eines Anbieters, der nach Beendigung der Maßnahme (und nachfolgender Wiedererteilung der Fahrerlaubnis) innerhalb eines vorgegebenen Zeitraums Δt noch im Besitz der Fahrerlaubnis ist. Diese Bedeutung hat der Begriff im vorliegenden Text.
- den komplementären Prozentsatz $1 - \pi$ der Personen, die ihre Fahrerlaubnis in diesem Zeitraum wieder verloren haben, also rückfällig geworden sind.

Unterstützend sowohl für die Einigkeit im Ziel als auch für die Bedeutung der Legalbewährung waren sicher: die Vorbildwirkung des ALKOEVA-Projekts der BAST (Nickel, Jakobshagen, Winkler, 1985; Winkler, Jakobshagen und Nickel, 1988), das im Jahre 2000 ebenfalls von der BAST veranstaltete Expertengespräch zur Klärung von

¹ Ich verwende die Maskulinform stellvertretend für Männer und Frauen.

² Auch „Legalbewährungsrate“, „Legalbewährungsquote“, „Rückfallwahrscheinlichkeit“ oder „Rückfallquote“ genannt. Der Einfachheit halber verwende ich den Begriff „Legalbewährung“.

Fragen bezüglich der Ausführungen des § 70 FeV und die internationale Tagung zum Thema „driver improvement“ in Salzburg im Spätjahr 2001³.

Der Einigkeit im Ziel stehen offene Fragen in der Durchführungsmodalitäten gegenüber, besonders in folgenden Punkten:

1. Wie oft sollen Qualitätskontrollen anhand der Legalbewährung durchgeführt werden?
2. Welchen Beobachtungszeitraum Δt soll man wählen?
3. Welches Design soll man wählen?

Zur ersten Frage: Wie bei jeder Dienstleistung muss auch bei den Maßnahmen zur Wiederherstellung der Verkehrssicherheit mit Qualitätsschwankungen gerechnet werden, die in den Personen liegen, die diese Maßnahmen durchführen. Zum einen, weil sich diese Personen verändern, sie werden erfahrener, vielleicht aber auch nachlässiger usw.⁴, zum anderen, weil das Personal der Anbieterorganisationen fluktuiert, erfahrene Mitglieder verlassen die Organisation, unerfahrene treten ein. Die dramatischen Folgen von Qualitätseinbußen für die Verkehrssicherheit auf Deutschlands Straßen verlangen daher, die Qualität der angebotenen Maßnahmen zur Wiederherstellung der Verkehrssicherheit regelmäßig zu überprüfen und nachzuweisen. Es genügt demnach nicht, dass der Anbieter X einmal eine Evaluationsstudie durchgeführt hat und sich für alle Zeiten auf deren Ergebnissen „ausruht“. Diese Überlegung legt zwangsläufig nahe, nach einfachen und kostengünstigen Methoden zur Erhebung der Legalbewährung zu suchen.

Die zweite Frage bedarf ebenfalls der Klärung, weil sich nachweisbar die Legalbewährung als Funktion von Δt (vgl. z. B. Jacobshagen, 1996, Abb. 1, S. 258) ändert. Der Funktionstyp ist jedoch unbekannt. Daher ist eine möglicherweise beabsichtigte Gegenüberstellung der Ergebnisse verschiedener Anbieter oder die Festlegung eines Normwertes, an dem die Qualität der Maßnahmen der verschiedenen Anbieter gemessen werden kann, nur möglich, wenn sich die empirisch errechneten Legalbe-

³ Nicht zuletzt wurde diese Einigkeit auch auf der Regensburger Fachtagung deutlich, deren Beiträge dieser Sammelband wiedergibt.

⁴ Die regelmäßige Supervision ist hier eine Kontrollmethode, die Erhebung der Legalbewährung die zweite, objektivere.

währungs-Ergebnisse auf ein einheitliches Beobachtungsintervall Δt beziehen. Es zeichnet sich – vermutlich wiederum dem Vorbild des ALKOEVA-Projekts folgend – die Festlegung ab: $\Delta t = 3$ Jahre (nach Wiedererteilung der Fahrerlaubnis).

Bei der dritten Frage geht es um die Entscheidung: Kontrollgruppendesign, ja oder nein. Die Argumente beim oben erwähnten Expertengespräch der BAST vermitteln den Eindruck, als sei das Kontrollgruppendesign „mehrheitsfähig“. Gegen diese Alternative sprechen allerdings zwei gewichtige Argumente: 1. Aufwand und Kosten, die mit einem Kontrollgruppendesign verbunden sind und 2. die Sicherheitsansprüche der Gesellschaft, die sich nicht durch den Vergleich Experimental- versus Kontrollgruppe befriedigen lassen, sondern allein durch den Nachweis, dass das Ergebnis der Maßnahme (Nachschulung, Therapie) die allgemeinen Erwartungen an die Legalbewährung der Maßnahmeempfänger erfüllt.

Zum ersten Gegenargument, dem Aufwand und den Kosten. Eine der Experimentalgruppe (also der Personen mit Therapie oder Nachschulung.) vergleichbare Kontrollgruppe muss dieser in ergebnisrelevanten Variablen entsprechen. Ein Anbieter, der seine Maßnahme auf diese Art evaluieren will, muss also a) die (hoffentlich) ergebnisrelevanten Variablen bei der EG erheben und dann b) eine Möglichkeit finden, die dazu passende KG zu rekrutieren. Das ist de facto nur über das Verkehrszentralregister (VZR) des KBA in Flensburg möglich. Zum Mehraufwand, der durch die Kontrollgruppenbildung entsteht, kommen auch noch Mehrkosten. Denn die weitaus kostengünstigere Möglichkeit, die benötigte Information über die Legalbewährung durch Abfrage im zentralen Fahrerlaubnisregister (ZFER) zu erhalten, entfällt hier. Das Kontrollgruppendesign scheidet daher für viele Anbieter schon aus Kostengründen aus.

Zum zweiten Gegenargument, den Sicherheitsansprüchen der Gesellschaft: Wie bei der ersten Frage nach dem Beobachtungszeitraum dürfte auch bei der Bevorzugung des Kontrollgruppendesigns das Vorbild der ALKOEVA-Studie die Gedanken bewegen. Zusätzlich kann es sein, dass Psychologen aufgrund ihrer Ausbildung bei dem Gedanken an eine empirische Studie die Kontrollgruppe „automatisch“ mitdenken. Die statistischen Methoden der Qualitätsprüfung, die in der Industrie Gang und Gäbe sind und die ohne Kontrollgruppe auskommen, gehören nicht zum Kanon der psy-

chologischen Methodenausbildung. Nun war aber unter den damaligen Gegebenheiten die Intention der ALKOEVA-Projektnehmer, nachzuweisen, dass die Schulung auffälliger Kraftfahrer deutlich effektiver ist als nichts zu tun, sinnvoll und notwendig. Es ging darum, in einem frühen Stadium die Existenzberechtigung des Schulungskonzepts empirisch zu belegen. Diese Phase des Beweisen-Müssens ist aber vorbei. Heutzutage geht es vielmehr darum, sich einer weit schwierigeren Herausforderung zu stellen, dem Sicherheitsanspruch der Gesellschaft. Dem sprichwörtlichen „Mann auf der Straße“ genügt es nämlich nicht, dass Alkoholtäter nach einer Maßnahme signifikant seltener rückfällig werden als (vergleichbare) Personen, die diese Maßnahme nicht durchlaufen haben. Er erwartet eine viel höhere Verbindlichkeit als ihm diese Auskunft geben kann: Er möchte am liebsten die Zusicherung, dass ihm dieser Alkoholtäter niemals betrunken ins Auto fährt oder gar sein Kind tötet. Und wenn absolute Gewissheit nicht möglich ist, so erwartet er doch, dass das Gefährdungspotenzial, das von diesem Menschen ausgeht, die Grenze des Tolerierbaren nicht überschreitet - Kontrollgruppe hin, Kontrollgruppe her. Er wäre wohl kaum zufrieden gestellt, wüsste er, dass im Schnitt von 100 Personen, die eine Nachschulung durchlaufen, 20 im Laufe der nächsten 3 Jahre wieder rückfällig werden, wie die einschlägigen empirischen Studien belegen. Dass das prozentuell signifikant weniger ist, als wenn keine derartige Maßnahme erfolgt wäre, dürfte ihn wenig berühren. Ganz in diesem Sinne äußern sich z.B. Sohn und Meyer-Gramcko (1998): Ziel einer verkehrstherapeutischen Maßnahme sei das *Vermeiden* weiterer Verkehrsauffälligkeiten; sie sprechen nicht von *Verringern*. Zum Nachweis, dass eine Maßnahme diesem Sicherheitsanspruch genügt, ist eine Kontrollgruppe überflüssig. Es genügt ein Design, wie es bei der statistischen Qualitätskontrolle Gang und Gäbe ist, wenn vorgegebene Qualitätsstandards zu erfüllen sind: durch den statistischen Nachweis, dass ein Produkt oder eine Dienstleistung vereinbarten Standards genügt. Zwar existiert bisher keine explizit festgelegte normative Größe „Legalbewährung nach einer Maßnahme“⁵. Aber das ist eher ein Versäumnis der Politik und Gesetzgebung, und es ist keine Frage, dass diese normative Größe über dem Wert von 80 Prozent liegen müsste, um in der Bevölkerung akzeptiert zu werden.

⁵ Ich nenne sie im Folgenden π^* .

Aus diesen Gründen ist die hier vorgeschlagene Methode eine Methode ohne Kontrollgruppe. Zum Abschluss der Vorüberlegungen ist noch eine vierte Frage zu beantworten:

4. Muss es eine Totalerhebung sein oder genügt eine Stichprobe?

Die jahrzehntelangen praktischen Erfahrungen mit der statistischen Qualitätskontrolle - auch in Produktionsbereichen mit extrem hohem Sicherheitsanspruch - zeigen: die Stichprobe genügt. Vor allem, wenn man eine Prüfung regelmäßig wiederholen will, was der Qualitätsgedanke eigentlich vorschreibt, können durch das Stichprobenverfahren Aufwand und Kosten im erträglichen Rahmen gehalten werden. Dieses Argument mag in einem geförderten wissenschaftlichen Forschungsprojekt weniger Gewicht haben. Um so mehr müssen selbstständige niedergelassene Verkehrstherapeuten oder Träger von Schulungsmaßnahmen darauf achten.

Diese Überlegungen führten zu der Methode, die ich im folgenden Abschnitt zur Diskussion stelle.

2. Methode

2.1. Voraussetzungen und Bezeichnungen

Die Methode geht von einer normativen Setzung π^* der Legalbewährung aus, bezogen auf einen festgelegten Beobachtungszeitraum Δt . Die Festlegung von Δt dürfte kein Problem sein. Wer π^* den konkreten Wert zuweist, ist offen. Diese Wertzuweisung können z.B. einvernehmlich die Anbieter der Maßnahmen vornehmen. Auch kann jeder Anbieter für sich selbst den konkreten Wert von π^* bestimmen, gewissermaßen als vermutete Forderung der Gesellschaft. Am meisten würde es überzeugen, legten die gewählten Repräsentanten der Gesellschaft diesen Normwert fest. Wahrscheinlicher aber wird er sich gewissermaßen als Durchschnittswert aus einschlägigen Publikationen im Laufe der Zeit von selbst definieren, durch die Fachpresse „zu rechtgerückt“ werden und irgendwann eine stabile Größe sein. Kurzum, es ist realistisch, auch diesen Wert als gegeben anzunehmen.

Es bezeichne also:

- π^* den normativ festgelegten Wert der Legalbewährung
- π die (unbekannte) Legalbewährung der Gesamtklientel eines Anbieters (bezogen auf Δt)
- N den Gesamtumfang der Klientel eines Anbieters (während Δt)
- P das Stichprobenergebnis, die empirisch ermittelte Legalbewährung
- n den (noch zu bestimmenden) Umfang der Stichprobe
- σ_P den Standardschätzfehler von P (bei Stichproben der Größe n)

2.2. Qualitätsvereinbarungen

Es entspricht der Vorgehensweise der statistischen Qualitätskontrolle und auch dem Qualitätsgedanken in der Verkehrstherapie und Nachschulung, wenn folgendes festgelegt wird:

- 1) π darf π^* nicht unterschreiten. Andernfalls muß der Anbieter Qualitätsverbesserungsmaßnahmen einleiten, und zwar so lange, bis er die Norm π^* wieder erreicht hat.

Da wir keine Totalerhebung machen, also π nicht erheben, sondern lediglich die Statistik P , legen wir eine kritische Abweichung ε fest:

- 2) Eine Normabweichung $P - \pi^* \leq -\varepsilon$ bedeutet die Aufforderung zum Eingreifen. Der Anbieter muss qualitätsverbessernde Maßnahmen setzen.

Nun hängt die Abweichung $P - \pi^*$ vom Stichprobenumfang ab: Bei geringem Stichprobenumfang sind zufallsbedingte Normabweichungen (falsche Alarme) häufiger als bei großem Umfang. Andererseits soll der Stichprobenumfang aus Kostengründen nicht den Umfang der Grundgesamtheit annehmen. Wir legen daher fest:

- 3) Entspricht die Maßnahme den Anforderungen, ist also $\pi \geq \pi^*$, dann müssen mindestens $1-\alpha$ Prozent aller Stichprobenergebnisse P oberhalb der Marke $\pi^* - \varepsilon$ liegen.

2.3. Bestimmung des Stichprobenumfangs

Qualitätsvereinbarung 3) enthält auch die Festlegung des erforderlichen Stichprobenumfangs. Denn es gilt

$$z_{\alpha} = \frac{-\varepsilon}{\sigma_p}$$

und der Standardschätzfehler σ_p hängt bekanntlich vom Stichprobenumfang n ab. Allerdings tritt hier ein Problem auf, das uns hindert, die üblichen Formeln anzuwenden: Die Grundgesamtheit ist endlich und die Stichprobe kann im Vergleich zu N groß ausfallen. Daher ist es hier unzulässig, wie sonst üblich das Modell des „Ziehens ohne Zurücklegen“ durch das Modell des „Ziehens mit Zurücklegen“ zu ersetzen. Im vorliegenden Fall verändert jeder Zug aus der Urne „Gesamtklientel“ den Zustand der Urne wahrnehmbar. Die Berücksichtigung dieses Sachverhalts liefert

$$z_{\alpha} = \frac{-\varepsilon}{\sigma_p} = \frac{-\varepsilon}{\sqrt{\frac{\pi^*(1-\pi^*)}{n} \cdot \frac{N-n}{N-1}}} \quad (1)$$

Wir quadrieren beide Seiten von Gl (1) und stellen um:

$$\frac{\pi^*(1-\pi^*)}{n} \cdot \frac{N-n}{N-1} = \frac{\varepsilon^2}{z^2} \Leftrightarrow \frac{N-n}{n} = \frac{(N-1) \varepsilon^2}{z^2 \pi^*(1-\pi^*)}$$

Wir lösen den Bruch auf der linken Seite auf, addieren 1 auf beiden Seiten und bilden rechts den gemeinsamen Nenner:

$$\frac{N}{n} - 1 = \frac{(N-1) \varepsilon^2}{z^2 \pi^*(1-\pi^*)} \Leftrightarrow \frac{N}{n} = \frac{(N-1) \varepsilon^2 + z^2 \pi^*(1-\pi^*)}{z^2 \pi^*(1-\pi^*)}$$

Der Kehrwert dieses Quotienten ist die gesuchte Lösung, der Stichprobenumfang, bezogen auf die Gesamtklientel:

$$\frac{n}{N} = \frac{z^2 \pi^* (1 - \pi^*)}{(N - 1) \varepsilon^2 + z^2 \pi^* (1 - \pi^*)} \quad (2)$$

Das folgende Beispiel zeigt die Anwendung der Methode.

3. Anwendungsbeispiel

Die Qualitätsvereinbarungen in diesem Beispiel sind relativ streng, sie lauten:

- $\pi^* = 0.9$
- die maximal tolerierbare Abweichung für P ist $\varepsilon = 0.05$
- Mindestens $1 - \alpha = 99\%$ aller Stichprobenergebnisse P müssen oberhalb der Marke $\pi^* - \varepsilon$ liegen.

Spalte 3 der folgenden Tabelle zeigt die notwendigen Stichprobenumfänge in Abhängigkeit vom Umfang der Gesamtklientel. Offensichtlich ist der Profit dieser Methode umso höher, je größer der Umfang N der Gesamtklientel ist.

N der Gesamtklientel	Stichprobenumfang in % nach Gl (2) errechnet und gerundet	Stichprobenumfang n
100	66%	66 Personen
150	57%	85 Personen
200	49%	99 Personen
500	28%	141 Personen

4. Abschließende Anmerkungen zum Verhältnis von Aufwand und Ertrag bei der vorgeschlagenen Methode

Die Fehlalarmrate ist bei dieser Methode gering. Sie beträgt allgemein α Prozent. Bei den Qualitätsvereinbarungen im vorliegenden Beispiel beträgt sie also ein Prozent. Das schützt Anbieter vor der Gefahr übertriebener Qualitätsanstrengungen.

Andererseits kommt aber auch das Sicherheitsbedürfnis der Gesellschaft bei dieser Methode nicht zu kurz: Unterschreitet z. B. die tatsächliche Legalbewährung die

Normvorgabe genau um den Betrag ε , ist also $\pi = \pi^* - \varepsilon$, erfolgt bei der nächsten Erhebung der Legalbewährung mit 50-prozentiger Wahrscheinlichkeit die Aufforderung zum Eingreifen. Diese Wahrscheinlichkeit wächst rapide an, wenn π die Vorgabe π^* noch weiter unterschreitet: Ist z.B. $\pi = .80$ und $\pi^* = 0.9$, erfolgt der Hinweis auf die Normabweichung bereits mit einer Wahrscheinlichkeit von 95 Prozent, wenn die Gesamtklientel im Beobachtungszeitraum lediglich 100 Fälle umfasst. Beträgt sie aber 150 oder mehr Fälle, erfolgt der Hinweis mit absoluter Sicherheit.

Literatur

Jacobshagen, W. 1996. ALKOEVA und kein Ende? Blutalkohol, 33, 257 – 266.

Nickel, W-R., Jacobshagen, W., Winkler, W. 1985. Die Wirksamkeit von Kursen für wiederholt alkoholauffällige Kraftfahrer. Zeitschrift für Verkehrssicherheit, 31, 67 – 74.

Sohn, J-M., Meyer-Gramcko, F. 1998. Evaluation der Verkehrstherapie. Zeitschrift für Verkehrssicherheit, 44, 170 – 173.

Winkler, W., Jacobshagen, W., Nickel, W-R. 1988. Wirksamkeit von Kursen für wiederholt alkoholauffällige Kraftfahrer. Unfall- und Sicherheitsforschung Straßenverkehr, 64.