

Ermittlung von Standards für anforderungsgerechte Datenqualität bei Verkehrserhebungen

FA 1.169

Forschungsstelle: IVT Research GmbH, Mannheim

Bearbeiter: Bäumer, M. / Hautzinger, H. /
Kathmann, T. / Schmitz, S. /
Sommer, C. / Wermuth, M.

Auftraggeber: Bundesministerium für Verkehr, Bau und
Stadtentwicklung, Bonn

Abschluss: März 2010

1 Aufgabenstellung

Den bisherigen Richtlinien zu Verkehrserhebungen ist gemeinsam, dass sie – wenn überhaupt – nur sehr wenige Aussagen zur erreichbaren Datenqualität enthalten. Normative Vorgaben und konkrete Handlungsanweisungen, die zu einer Verbesserung der Datenqualität von Erhebungen führen, fehlen in der Regel für die meisten Erhebungsverfahren. Abgesehen von Einzelaspekten, wie beispielsweise den Kernelementen für Haushaltsbefragungen zum Verkehrsverhalten, gibt es keine Qualitätsstandards für die Konzipierung, Durchführung und Auswertung einer Verkehrserhebung.

Ziel der vorliegenden Studie ist es, mit Blick auf verschiedene Datennutzer und Arten der Datenverwendung wissenschaftlich abgesicherte Qualitätsstandards für Verkehrserhebungen zu erarbeiten. Im Kern sollten Hinweise gegeben werden, durch welche konkreten methodischen Ansätze und praktischen Maßnahmen man für die unterschiedlichen Erhebungsverfahren im Verkehrswesen (Zählungen, Messungen, Verhaltensbeobachtungen und Befragungen) die jeweils bestmögliche Datenqualität erreichen kann. Die Ergebnisse dieses Projekts sollen darüber hinaus auch als eine Grundlage für die Fortschreibung der neuen "Empfehlungen für Verkehrserhebungen" (EVE) dienen.

2 Untersuchungsmethodik und -ergebnisse

Im Rahmen des Projekts sollten die nachfolgenden Forschungsfragen untersucht werden:

- Was ist Datenqualität und wie kann der Datenqualitätsbegriff für Verkehrserhebungen konkretisiert werden?
- Was sind aussagekräftige Qualitätsindikatoren zur Messung der Qualität von Daten aus Verkehrserhebungen?
- Wie lassen sich bei Verkehrsdaten Qualitätsstufen unterscheiden und Qualitätsanforderungen formulieren?
- Welche Empfehlungen zur Sicherung der Datenqualität lassen sich für Verkehrserhebungen geben?

Bevor man sich der Beantwortung dieser Fragen zuwendet, sollte man sich vergegenwärtigen, welche Arten von Verkehrserhebungen es gibt.

Neben der Erhebungsform (Zählungen, Messungen, Verhaltensbeobachtungen und Befragungen) lassen sich Verkehrserhebungen nach dem Erhebungsort klassifizieren. Grundsätzlich können Verkehrserhebungen

- im Verkehrssystem und
- außerhalb des Verkehrssystems

stattfinden, wobei die Erhebungen im Verkehrssystem nach den Verkehrsmitteln

- Fuß- und Radverkehr,
- Kfz-Verkehr und
- öffentlicher Verkehr

unterschieden werden können. Verkehrserhebungen außerhalb des Verkehrssystems können darüber hinaus

- am Aktivitätenort (Einkaufs- und Freizeitort, Ort einer privaten Erledigung, Arbeitsplatz, Ausbildungsstätte),
- im Haushalt oder
- im Unternehmen bzw. Betrieb

erfolgen.

Nicht jede Erhebungsform lässt sich mit den klassifizierten Erhebungsorten kombinieren: Verkehrszählungen, Messungen und Verhaltensbeobachtungen finden im Haushalt und Unternehmen keine Anwendung. Die in Bild 1 dargestellte Klassifikation (nach Erhebungsform und Erhebungsort) wurde im Rahmen des Projekts erarbeitet und den Detailbetrachtungen zugrunde gelegt.

2.1. Was ist Datenqualität und wie kann der Datenqualitätsbegriff für Verkehrserhebungen konkretisiert werden?

Ausgehend von einer allgemeinen, an den Ansätzen des Qualitätsmanagements orientierten Definition von Datenqualität lässt sich ein umfassendes Datenqualitätskonzept darstellen, welches im Bereich der amtlichen Statistik auf europäischer Ebene Anwendung findet. Dieses Konzept kann gut auf den Bereich der Verkehrserhebungen übertragen werden und ermöglicht eine strukturierte Betrachtung der Fehlerquellen bei den wichtigsten Erhebungsformen. Dies ist eine Grundvoraussetzung, um Strategien und Maßnahmen zur Vermeidung oder Reduzierung systematischer Fehler zu entwickeln. Im Rahmen dieses Projekts werden zu den wesentlichen Bereichen:

- Stichprobenfehler / Anforderungen an die Präzision,
- Systematischer Fehler / Anforderungen an die Richtigkeit sowie
- Verzerrungen durch Nonresponse

konkrete Beispielrechnungen durchgeführt.

2.2 Was sind aussagekräftige Qualitätsindikatoren zur Messung der Qualität von Daten aus Verkehrserhebungen?

Aufbauend auf dem Datenqualitätskonzept von Eurostat lassen sich Indikatoren für Verkehrserhebungen ableiten, wobei es im Hinblick auf die Relevanz der einzelnen Indikatoren teilweise Abweichungen zum Bereich der amtlichen Statistik gibt. Zur Beurteilung der Qualität von Verkehrserhebungen sind in erster Linie die Indikatoren, welche die Genauigkeit beschreiben, relevant. Auf Basis verschiedener Literaturquellen wurden Indikatoren der Datenqualität für Verkehrserhebungen definiert. Diese sind in Abbildung 2 zusammenfassend dargestellt.

Erhebungsform		Datenerhebung					
		im Verkehrssystem			außerhalb des Verkehrssystems		
		Fuß- /Rad- verkehr	Kfz- Verkehr	Öffentlicher Verkehr	am Aktivitäts- ort (Einkauf, Freizeit...)	im Haushalt	im Betrieb/ Unter- nehmen
Beobachtung	Verkehrszählung	Fußgänger Radfahrer	Kfz Kfz-Insassen Ruhender Verkehr	Fahrgäste im Verkehrsmittel/ an Haltestellen ÖV-Fahrzeuge	Besucher	—	—
	Messung	Geschwindigkeit, Zeitlücken, Raumlücken, Verweildauer			Besucher	—	—
	Verhaltensbeobachtung	Konflikte, Verhaltensweisen, Bewegungsmuster			Besucher	—	—
Befragung	Befragung im Verkehrswesen	Fußgänger Radfahrer	Kfz-Führer Kfz-Insassen (Kordon/Parkplatz)	Fahrgäste im Verkehrsmittel/ an Haltestellen Umsteiger	Besucher	Personen/ Haushalts- mitglieder	Beschäftigte

Bild 1: Klassifikation von Verkehrserhebungen

	Qualitätsaspekt	Qualitätsindikator
Relevanz	Kundenorientierung bei Aufgabenstellung und Untersuchungsziel	<ul style="list-style-type: none"> Erfahrungen/Referenzen/ Zertifizierung des Erhebungsinstitutes
	Wissenschaftlichkeit bei Erhebungskonzept, Durchführung und Auswertung	<ul style="list-style-type: none"> Vorliegen von Qualitätsverpflichtungen, Geheimhaltungsvereinbarungen, etc.
Genauigkeit	Stichprobenfehler	<ul style="list-style-type: none"> Standardfehler, Variationskoeffizient
	Abdeckungsfehler	<ul style="list-style-type: none"> Abdeckung der Zielgrundgesamtheit durch die Auswahlgrundgesamtheit (Anteilswert) Eigenschaften der nicht erfassten Einheiten der Zielgrundgesamtheit
	Nonresponse-Fehler	<ul style="list-style-type: none"> Ausschöpfungsquote/Erfassungsquote Umgang mit den Nonresponse-Fällen
	Messfehler	<ul style="list-style-type: none"> Anteil der nicht erfassten Merkmale (Item-Nonresponse) Umgang mit den Einheiten, bei denen einzelne Merkmale nicht erfasst wurden Eigenschaften der Einheiten, bei denen einzelne Merkmale nicht erfasst wurden Abweichungen zwischen Erhebungsergebnissen und Ergebnissen vorheriger und vergleichbarer Erhebungen, die als valide eingestuft werden können (falls vorhanden)
Aktualität und Pünktlichkeit	Termintreue	<ul style="list-style-type: none"> Differenzen zwischen angekündigtem Termin und tatsächlicher Präsentation der Ergebnisse der Verkehrserhebung
	Aktualität	<ul style="list-style-type: none"> Zeitspanne zwischen Ende der Feldarbeit und Vorlage der Ergebnisse
Zugänglichkeit und Klarheit	Dokumentation	<ul style="list-style-type: none"> Metadaten-Dokumentation (Beschreibung des Verfahrens): Vollständigkeit definierter Metadaten

Bild 2: Qualitätsindikatoren

2.3 Wie lassen sich bei Verkehrsdaten Qualitätsstufen unterscheiden und Qualitätsanforderungen formulieren?

Diese Frage lässt sich nicht pauschal beantworten, da sich keine allgemeingültigen Abgrenzungen von Qualitätsstufen für die einzelnen Indikatoren über alle Erhebungsformen formulieren lassen. Deshalb wurde diese Fragestellung im Rahmen dieses Projekts immer im Zusammenhang mit dem Vorgehen zur Sicherstellung der Qualität betrachtet, also mit der Frage.

2.4 Welche Empfehlungen zur Sicherung der Datenqualität lassen sich für Verkehrserhebungen formulieren?

Der Titel des Forschungsprojekts "Ermittlung von Standards für anforderungsgerechte Datenqualität bei Verkehrserhebungen" suggeriert, dass Standards im Sinne von "festen Kochrezepten" erzeugt werden können. Die Erhebungspraxis zeigt jedoch, dass es so etwas aufgrund der Vielfalt der Fragestellungen, unterschiedlichen Anforderungen und finanziellen Möglichkeiten nicht geben kann.

Es können zwar zahlreiche Vorgaben hinsichtlich der Datenqualität gemacht werden, vergleichsweise häufig werden allerdings von Auftraggebern in Ausschreibungen für Verkehrserhebungen konkrete Werte zum Stichprobenumfang und damit indirekt zum Stichprobenfehler vorgegeben. Zu den systematischen Fehlern werden in der Regel aber keine oder nur wenige Vorgaben festgelegt. Dies hängt vor allem damit zusammen, dass

- detaillierte Kenntnisse über systematische Fehler und deren Zusammenhänge mit den wesentlichen Aspekten einer Erhebung häufig fehlen,
- unklar ist, in welcher Größenordnung notwendige Vorgaben zur Datenqualität liegen sollen,
- häufig die Bereitschaft oder Möglichkeit, finanzielle Mittel aufzubringen, um eine hohe Datenqualität zu erreichen, nicht vorhanden ist.

Die Qualität von Erhebungsergebnissen steht immer im Spannungsfeld der finanziellen und personellen Ressourcen und des Erhebungsumfangs und -aufwands. Die Konzeption der Erhebung resultiert demnach aus einem Abwägungsprozess zwischen der gewünschten Datenqualität, dem Erhebungsumfang und den vorhandenen finanziellen und personellen Ressourcen. Dabei ist es die Aufgabe für die Verantwortlichen der Erhebung, die in Abhängigkeit der Aufgabenstellung optimale Mischung zwischen diesen drei Elementen zu finden. In der Regel werden die finanziellen Mittel durch den Auftraggeber vorgegeben, sodass das verantwortliche Erhebungsinstitut nur noch die Prioritäten zwischen Quantität und Qualität setzen kann. Methoden und Maßnahmen zur Qualitätssicherung binden im Allgemeinen Ressourcen, die auf der anderen Seite zu Einsparungen an der Quantität, d. h. beim Stichprobenumfang und Erhebungsinhalt je Teilnehmer führen müssen. Andererseits erhöht eine Reduktion des Stichprobenumfangs den Stichprobenfehler, sodass im Extremfall die Repräsentativität der Erhebung, d. h. die Verallgemeinerungsfähigkeit der Erhebungsergebnisse nicht mehr gegeben ist.

Die Ziele und Vorgaben zur Qualität sollten daher immer im Zusammenhang mit den zur Verfügung stehenden finanziellen Mitteln und dem geplanten Erhebungsumfang gesehen werden. Je nach Aufgabenstellung sind bei der Erhebungskonzeption entsprechende Prioritäten zu setzen. Das optimale Erhebungskonzept muss damit nicht das Konzept sein, welches zur höchsten Datenqualität führt, sondern dasjenige, welches unter den gegebenen Rahmenbedingungen, insbesondere den finanziellen Restriktionen, die Untersuchungsziele am besten erfüllt.

3 Folgerungen für die Praxis

Während des gesamten Erhebungsprozesses kann die Qualität der Erhebungsergebnisse durch unterschiedliche Verfahren bzw. Maßnahmen gesichert werden. Neben spezifischen Handlungsempfehlungen für die einzelnen Erhebungsformen gelten übergreifend folgende Grundsätze zur Qualitätssicherung (die Reihenfolge entspricht dem Erhebungsablauf):

- eindeutige Definition der Untersuchungsziele, der Aufgabenstellung und des gewünschten Qualitätsstandards,
- eindeutige Festlegung der Rahmenbedingungen (Erhebungszeitraum, finanzielle und personelle Restriktionen),
- Überprüfung des Erhebungsinstruments mithilfe eines Pretests (bei erstmalig durchgeführten Erhebungen und allen Selbstausfüllfragebögen), Durchführung von Tests der Funktionalität bei technischen Geräten,
- sorgfältige Auswahl von geeignetem Erhebungspersonal (z. B. Prüfung der Eignung über Vorstellungsgespräche),
- projektbezogene Einweisung und Schulung des Erhebungspersonals,
- stichprobenartige Überwachung der Feldarbeit (bei Erhebungspersonal vor Ort),
- Vorhalten und Einsatz von "Springern" (Personal, das bei Ausfall des Stammpersonals kurzfristig einspringt),
- Aufbereitung der Erhebungsdaten auf Basis überprüfbarer Regeln (Plausibilitätsprüfung),
- bei langen Erhebungszeiträumen frühe Datenerfassung und -aufbereitung, sodass eine Rückkopplung zur laufenden Erhebung möglich ist sowie
- Dokumentation des Erhebungsablaufs.

Folgendes Schema zur Dokumentation von Verkehrserhebungen ("Metadaten") wird vorgeschlagen:

1. Allgemeine Angaben / Basisangaben / Quellenangaben (z. B. Auftraggeber, Bearbeiter, finanzielle Institution)
2. Gegenstand und Ziele der Erhebung
3. Grundgesamtheit der Erhebung.
(z. B. "Personen im Alter ab x Jahren mit deutscher Staatsangehörigkeit der Gemeinde Y zum 31.12.2008")
4. Zeitliche und räumliche Abdeckung (Dauer der Feldarbeit, geografischer Erhebungsraum)
5. Untersuchungs-/Erhebungseinheiten (z. B. Personen, Kfz)
6. Erhebungsmerkmale (Charakteristika der Untersuchungseinheiten)
7. Angestrebter Genauigkeitsgrad
 - 7.1 Vorgaben zum Stichprobenfehler
 - 7.2 Vorgaben zur statistischen Sicherheit
 - 7.3 Benötigter Stichprobenumfang
8. Messverfahren zur Ermittlung der Merkmalsausprägungen
 - 8.1 Erhebungstechnik
 - 8.2 Eingesetzte Erhebungsinstrumente

- 8.3 Gegebenenfalls Anzahl des eingesetzten Erhebungspersonals
- 8.4 Gegebenenfalls die Art der Incentives
- 8.5 Gegebenenfalls die Methoden der Qualifizierung des Erhebungspersonals
9. Auswahlgesamtheit und Auswahlrahmen (Verzeichnis der Auswahlinheiten)
10. Auswahlverfahren
 - 10.1 Stufengliederung
 - 10.2 Schichtung
 - 10.3 Ziehungsmodell
11. Pretest
 - 11.1 Planung und Durchführung
 - 11.2 Ergebnisse und Konsequenzen
12. Organisation und Ablauf der Erhebung (Feldphase)
 - 12.1 Feldbericht
 - 12.2 Darstellung des Antwortverhaltens
 - 12.3 Gegebenenfalls Nonresponse-Studie
13. Datenaufbereitung und Datenanalyse
 - 13.1 Erfassung, Prüfung, Bereinigung und Korrektur der Daten
 - 13.2 Gewichts- und Hochrechnungsfaktoren
 - 13.3 Hochrechnung / Tabellenprogramm
14. Informationen zur Beurteilung der Qualität der Daten
 - 14.1 Standardfehler
 - 14.2 Abdeckungsfehler
 - 14.3 Nonresponse-Fehler
 - 14.4 Messfehler
15. Ergänzende Materialien (z. B. ein Muster der eingesetzten Erhebungsmaterialien)

Im vorliegenden Projekt wurden mögliche Fehlerquellen bei Verkehrserhebungen im Detail aufgezeigt und Hinweise für deren Vermeidung gegeben. Die Fehlerfortpflanzung bei der Weiterverwendung der Erhebungsdaten in Modellen wurde hingegen nicht betrachtet, da dies nicht Teil der Aufgabenstellung war und man dann auch andere, nachgelagerte Fragestellungen hätte berücksichtigen müssen. Hinsichtlich der Auswirkungen auf die Qualität von Verkehrsmodellen besteht also weiterer Untersuchungsbedarf, da Grundlagendaten aus Verkehrserhebungen sehr häufig als Eingangswerte in Verkehrsmodelle aller Art genutzt werden.

Als zentrales Fazit lässt sich festhalten, dass die Datenqualität bei Verkehrserhebungen immer nur so gut ist, wie das schwächste Glied im statistischen Produktionsprozess. Deshalb sollte immer versucht werden, die richtige Balance zwischen allen relevanten Aspekten zu erreichen. Hierzu kann die Berücksichtigung des oben dargestellten Schemas zur Dokumentation von Verkehrserhebungen im Sinne eines Projekthandbuchs ein wichtiger Baustein sein.