

Ist HIT der Hit?

Licht und Schatten bei der Verwaltungsdatennutzung für die Agrarstatistik

Reiner Seitz

Die amtliche Statistik steht vor der permanenten Herausforderung, die Belastungen der Bürgerinnen und Bürger durch Befragungen zu minimieren und wo immer möglich auf die Nutzung bereits vorhandener Daten zurückzugreifen.

Ein Beispiel für die Nutzung von Verwaltungsdaten ist der Rückgriff auf die Daten des Herkunft- und InformationssystemTier (kurz HIT) für die Viehbestandsbeobachtung. An diesem Beispiel lässt sich zeigen, wie ein System, das ursprünglich für ganz andere Zwecke gedacht war, sich zum Wohl aller Beteiligten für statistische Zwecke nutzen lässt, wenn Verwaltungszweck und Statistikauftrag einigermaßen kompatibel sind. Es lässt sich aber auch zeigen, dass die statistische Nutzung von Verwaltungsdaten an Grenzen stößt, wenn Ziel und Intention des Verfahrens und Auftrag der Statistik zu weit auseinander liegen und nur mit sehr großem Aufwand angenähert werden können.

Im November 2008 wurden in Baden-Württemberg rund 1 Mill. Rinder, darunter 361 000 Kühe, in insgesamt rund 22 000 Stallungen gehalten. Diese Informationen wurden durch Auswertung der Daten im Herkunft- und InformationssystemTier (kurz HIT) gewonnen. Diese Datenbank wurde im Zuge der BSE-Krise aufgebaut und sollte es ermöglichen, den lückenlosen Werde-



Abbildung: Was bin ich? Schlacht- oder Zuchtfärsen? Milch- oder Mutterkuh?
Bildnachweis: ©Miroslaw/www.pixelio.de

gang eines jeden Rindes in Deutschland und den körperlichen Verbleib zu dokumentieren. Es handelt sich um ein zentrales Verfahren, mit dem alle Rinderhalter Deutschlands (und alle die mit Rindern handeln, sie bewegen, schlachten, ...) täglich die Veränderungen melden. Der Meldeweg für ein Tier in HIT beginnt mit der Geburt und endet mit dem Tod. Durch Auswertungen der Meldungen zu einem bestimmten Stichtag lässt sich ein Bild des Rinderbestandes gewinnen. Dies macht man sich in der Agrarstatistik seit Mai 2008 für die Zwecke der Viehbestandsbeobachtung zunutze. Seither kann auf die zweimalige statistische Erhebung der Rinderbestände im Lauf eines Jahres (3. Mai und 3. November) in den landwirtschaftlichen Betrieben verzichtet werden.

Jahrelanger Vorlauf

Bevor es soweit war, mussten jedoch eine Reihe methodischer Fragen geklärt und Vereinbarungen über die technischen Abläufe getroffen werden. Die Nutzung der HIT-Datenbank setzte zudem die Zustimmung der EU-Kommission voraus. Die umfangreichen methodischen Vorarbeiten waren im Wesentlichen darauf zurückzuführen, dass die EU in Dingen der Viehbestandsbeobachtung bis ins Detail vorschreibt, welche Ergebnisse in welcher sachlichen Gliederung und zu welchem Termin geliefert werden müssen. Bei Rindern zählen das Alter, das Geschlecht und die Nutzungsart (Milch, Fleisch) zu den Liefermerkmalen. Während sich die Merkmale Geschlecht und Alter in HIT problemlos ermitteln lassen, enthält HIT jedoch keine originären Informationen zur Nutzungsart der Rinder.

Diese fehlende Information war die zentrale Herausforderung für die Nutzung von HIT. Hier mussten Lösungen gesucht und gefunden werden, die eine Erfüllung der Lieferverpflichtungen ermöglichten, ohne die HIT-Datenbank zusätzlich zu belasten. Es war Grundlagenarbeit erforderlich, die im Wesentlichen vom Statistischen Bundesamt geleistet wurde.¹ In umfangreichen Analysen wurden modelltheoretische Lösungsansätze für zwei Ergebnisgrößen erarbeitet.

Dipl.-Ing. agr. Reiner Seitz ist Leiter des Referats „Landwirtschaftliche Struktur- und Anbauverhältnisse, Wein-, Forst- und Fischereiwirtschaft“ im Statistischen Landesamt Baden-Württemberg.

¹ Walther, Matthias: Nutzung von Verwaltungsdaten für die Agrarstatistik, in: *Wirtschaft und Statistik* 9/2003, S. 849–857 und Walther, Matthias: Verwaltungsdatennutzung für die Viehbestandsbeobachtung, in: *Wirtschaft und Statistik* 8/2004, S. 845–853.

Modelltheoretische Bestimmung von Nutzungskategorien

Schlachtfärsen: Für den statistischen Nachweis der Schlachtfärsen (weibliche Rinder zum Schlachten, ohne Kühe) wird ein Quotenverfahren verwendet. Die Zahl der Schlachtfärsen wird als Anteil an der Zahl der weiblichen Rinder insgesamt bestimmt. Diese Quotierung wird einmal für Rinder zwischen 1 und 2 Jahren und ein weiteres Mal für die 2 Jahre und älteren Rinder durchgeführt. Die Schlachtfärsenquote wird durch Auszählung der in der HIT-Datenbank gemeldeten Schlachtungen der letzten 12 Monate in Relation zum Gesamtbestand vor 12 Monaten ermittelt. Mit diesem Verfahren findet eine reale Größe Eingang in das Schätzverfahren. Allerdings wird damit im aktuellen Bestand die Entwicklung aus der Vergangenheit fortgeschrieben. Veränderungen im Entscheidungsverhalten der Tierhalter werden somit nicht abgebildet.

Milchkühe: Dieses Merkmal ist von erheblicher agrarstruktureller Bedeutung, weshalb an die modelltheoretische Lösung besondere Anforderungen zu stellen sind. In HIT sind Informationen zu Alter, Geschlecht und Abkalbestatus² verfügbar. Damit lässt sich ohne Weiteres die Gesamtzahl der Kühe bestimmen. Eine Aufteilung auf Milchkühe, Ammen- und Mutterkühe sowie auf Schlachtkühe ist jedoch nicht möglich. In HIT sind jedoch zwei weitere Informationen verfügbar, die für eine Ableitung der Nutzungsart verwendet werden können. Das sind die Rinderrasse und die Produktionsrichtung der Haltung. Die Rinderrasse ist insofern hilfreich, als bestimmte Rinderrassen nahezu eindeutig einer bestimmten Nutzung der Tiere zugeordnet werden können. Dieses Merkmal reicht für eine sichere Zuordnung jedoch nicht aus – insbesondere nicht in Bundesländern, in denen die sogenannten Zweinutzungsrinderrassen stark vertreten sind. Dazu zählt Baden-Württemberg mit den hier zahlreich vertretenen Rassen Fleckvieh und Braunvieh.

In ergänzenden Untersuchungen wurde erkannt, dass eine zufriedenstellende Qualität der Zuordnung der Kühe zu einer der Unterkategorien dann erreicht wird, wenn zusätzlich für die Rinderhaltung die sogenannte Produktionsrichtung bekannt ist. Mit der Produktionsrichtung wird der betriebliche Schwerpunkt (Milch, Mast, etc.) beschrieben. Dieses Merkmal war in HIT von Anfang an angelegt, stand aber nicht in allen Fällen belastbar und flächendeckend zur Verfügung. Erst nach einer Aktualisierung der Angabe zur Produktionsrichtung war der Weg frei für eine Nutzung von HIT für die Zwecke der Viehbestandsbeobachtung.

Konsolidierung der Daten im Zeitablauf

Ein weiteres Charakteristikum der HIT-Datenbank ist ihre dynamische Veränderung durch die täglichen Meldungen der Beteiligten. Doch nicht jede Meldung erfolgt in der Praxis taggenau, wodurch sich eine Zeitverschiebung zwischen realem Vorgang und Meldung in der Datenbank ergibt. Untersuchungen haben gezeigt, dass ein Zeitraum von mindestens 4 Wochen abzuwarten ist, bevor sich ein weitgehend konsolidierter Stichtagsbestand abbilden lässt.

Bei HIT handelt es sich um eine zentrale Datenbank für ganz Deutschland. Die Nutzung hat damit ganz zwangsläufig zur Folge, dass für alle Bundesländer ein einheitliches Verfahren etabliert wird. Die Nutzung von HIT fiel auch in die Phase der verstärkten Bemühungen um zentralisierte Aufbereitungen der amtlichen Statistik, bei der die Aufbereitung der Daten einer Statistik in nur einem Bundesland für alle Bundesländer erfolgt. Die zentrale Aufbereitung von HIT ist daher auch ein Baustein des Masterplans zu Optimierung der föderalen Statistik. Im Fall von HIT ist die zentrale Aufbereitung in Schleswig-Holstein/Hamburg angesiedelt.

Sekundärstatistik in zentraler Produktion – effektiver geht's nicht

Nachdem die modelltheoretischen Lösungsansätze gefunden und die technischen Voraussetzungen für eine zentrale Produktion von HIT geschaffen waren, geht die Nutzung von HIT jetzt schnell und effektiv vonstatten. Der erwartete Erfolg bei der Nutzung von Verwaltungsdaten stellt sich ein: Es ist keine Befragung der Betriebe erforderlich, die Zuarbeit in den statistischen Landesämtern beschränkt sich auf die qualitative Prüfung der Ergebnisse und die gegebenenfalls erforderliche Einarbeitung der Geheimhaltung. Für Bund und Länder stehen bereits gut 14 Tage nach Abruf der Datenbank bundesweit aktuelle, vergleichbare und hochinformativere Ergebnisse zur Verfügung. Eleganter und effektiver können statistische Ergebnisse kaum gewonnen werden: Verwaltungsdatenutzung at its best, wie man so schön in Neudeutsch sagt.

Aber ...

Die Nutzung von HIT im Rahmen der halbjährlichen Viehbestandserhebung ist die „Schokoladenseite“, weil die in HIT vorhandenen Datenlücken durch modelltheoretische Lösungsansätze in angemessener Weise geschlossen

² Der Abkalbestatus gibt an, ob ein Rind bereits ein oder mehrere Kälber hat.

werden können. Die modellhafte Ableitung basiert allerdings auf einer ganzen Reihe von Annahmen, die im statistischen Durchschnitt zwar durchaus zutreffend sind, im Einzelfall aber auch daneben liegen können. Und das führt zur Schattenseite der HIT-Nutzung. Die HIT-Datenbank soll nämlich nicht nur für die Viehbestandsbeobachtung, sondern auch für die Strukturhebung in der Landwirtschaft genutzt werden. In der Strukturbeobachtung geht es nicht mehr nur um die summarische Auswertung von HIT, vielmehr sollen die Angaben aus HIT für jeden Einzelfall mit den konventionell erhobenen übrigen Angaben zusammengeführt werden. Durch die Verknüpfung der Daten entsteht ein umfassendes Strukturbild der landwirtschaftlichen Betriebe. Dabei können allerdings die vorhandenen methodischen Brüche und definitorischen Abweichungen sichtbar werden.

... Bruchlinien werden sichtbar

So werden die Einheiten in HIT und Agrarstatistik unterschiedlich definiert. In HIT sind Rinderhaltungen an ihren Standorten nach veterinärmedizinischen Kriterien erfasst, in der Statistik zählen landwirtschaftliche Betriebe als technisch-wirtschaftliche Einheiten (mit ggf. mehreren Produktionsstätten). Auch die modelltheoretisch und statistisch ermittelten Milchkühe können in Wirklichkeit Ammen- und Mutterkühe (und umgekehrt) sein, und bei genauem Hinsehen finden sich Einheiten, bei denen ein länderübergreifender Austausch notwendig ist, weil der Rinderbestand in dem einen Land und der zugehörige Betrieb in einem anderen Land liegt. Damit kein falsches Bild entsteht: Die Annahmen, die der Nutzung für die Viehbestandsbeobachtung zugrunde liegen, gelten vielfach auch für die Einzelfallnutzung. In der ganz großen Mehrheit der Fälle trifft die modellhaft getroffene Annahme zu, dass HIT-Haltung und landwirtschaftliche Rinderhaltung identisch sind oder dass die „Modell-“Milchkühe auch in Wirklichkeit Milch geben.

Nur in einem kleinen Teil der Fälle treffen diese Annahmen eben nicht zu. Leider weiß man a priori nicht, welcher Teil dies ist. Dies hat zur Folge, dass in der gesamten Bearbeitungskette mit großer Aufmerksamkeit gearbeitet werden muss. Dies gilt umso mehr, als an die Qualität der Ergebnisse von Agrarstrukturhebungen besondere Anforderungen zu stellen sind. Schließlich werden die Ergebnisse am Ende bis auf Gemeindeebene publiziert oder im Rahmen der Forschungsdatenzentren einer externen Analyse auf der Grundlage anonymisierter einzelbetrieblicher Datensätze zugänglich gemacht. Die einzelbetriebliche Nutzung

der HIT-Datensätze erfordert daher ganz andere Verfahren und Abläufe, um die notwendige Qualität zu schaffen.

Abweichende Einheitendefinition

Das beginnt bei der Zusammenführung von HIT-Haltung und Betrieb. Der einfachste Weg besteht darin, in den Betrieben die HIT-Nummer(n) zu erheben. Dieses Vorgehen führt allerdings zu der Frage, warum nicht gleich die Rinderbestände erhoben werden. Bei den Rindern handelt es sich immerhin um leicht abzählbare Einheiten, deren zahlenmäßiger Bestand jedem gut informierten Landwirt praktisch ständig präsent ist. Seine HIT-Nummer(n) wird er wohl in jedem Fall nachschlagen müssen. Es kommt hinzu, dass auch auf diesem Weg nicht alle HIT-Nummern gemeldet werden und ein Nachforschen über die Zusammenhänge in den übrigen Fällen trotzdem notwendig ist. Dabei ist die Erhebung der HIT-Nummern in vielen Fällen gar nicht notwendig, weil ein eindeutiger Zusammenhang zwischen HIT-Nummer und einer anderen Verwaltungsnummer, der sogenannten Unternehmensnummer, besteht und über diese der Zusammenhang zwischen den Einheiten hergestellt werden kann.

In Baden-Württemberg wird daher auf die flächendeckende Erhebung der HIT-Nummern verzichtet. Vielmehr werden die verfügbaren Quellen ausgeschöpft und die dann verbleibenden Unklarheiten durch direkte und gezielte Rückfrage bei den Landwirten aufgeklärt. Dieses Verfahren bietet die maximale Entlastung für die Landwirte im Land, hat aber entsprechenden internen Aufwand im Statistischen Landesamt zur Folge.

Mit dem geschilderten Vorgehen lässt sich die formale Einheitenzuordnung in den Griff bekommen. Weitaus diffiziler ist es nach derzeitigem Stand, deutlichen Bestandsabweichungen und fehlerhaften Nutzungszuordnungen auf die Spur zu kommen. Die bisherigen Analysen auf Einzelfallebene zeigen, dass es doch einige Fälle gibt, in denen die Bestandsangaben aus HIT und vorangegangener Strukturhebung erheblich differieren sowie Fälle, in denen die automatisierte Entscheidung über Milchkühe oder Ammen- und Mutterkühe offensichtlich nicht zutreffend ist. Diese Fehler sind für das Gesamtergebnis ohne größere Bedeutung, weil sich die Fehlzuordnungen weitgehend gegenseitig ausgleichen. Bei näherer regionaler oder inhaltlicher Analyse, dem eigentlichen Zweck von Strukturhebungen, werden diese Fehler jedoch sichtbar. Hier hilft nur die aufmerksame einzelbetriebliche Plausibilisierung der Angaben

weiter. Die Nutzung von HIT im Rahmen einer Strukturhebung erfordert damit deutlich mehr Aufwand: zum einen für die Zuordnung der Einheiten, zum anderen für die Verifizierung der zugeordneten Bestände.

Belastbarkeit der Daten überschritten?

Im Zusammenhang mit der Nutzung von HIT für die Strukturbeobachtung zeigt sich, dass eine Verwaltungsquelle auch überstrapaziert werden kann. Wenn eine Verwaltungsquelle Daten nicht enthält, dann müssen diese Datenlücken geschlossen werden. Für die Zwecke der halbjährlichen Viehbestandsbeobachtung hat man Verfahren entwickelt, um diese Lücken zu schließen. Die vorhandenen Unschärfen im Verfahren sind im Hinblick auf das angestrebte Ergebnis vertretbar. Die in diesem Zusammenhang vorgenommene Ertüchtigung des Merkmals Produktionsrichtung in HIT weist allerdings bereits in eine problematische Richtung: Ist es

wirklich eine Entlastung, wenn nun im Verwaltungsverfahren Merkmale von den Betrieben gepflegt werden müssen, um das Verwaltungsverfahren statistisch nutzbar zu machen?

Im Rahmen der Strukturhebung steigert sich der interne Aufwand im Statistischen Landesamt erheblich, wenn die Verwaltungsdaten noch in adäquater Qualität genutzt werden sollen. Der Aufwand steigt in einem Maße, dass sich die Frage stellt, ob es nicht einfacher wäre, die Merkmale direkt zu erheben. Allerdings ist zu berücksichtigen, dass der Fragebogen für die nächste Strukturhebung – die Landwirtschaftszählung 2010 – mit bis zu 19 Seiten bereits sehr umfangreich ist. Deshalb kommt es gerade unter diesen Umständen darauf an, trotz des Aufwands im Statistischen Landesamt die Belastung der Landwirte in Grenzen zu halten. ■

Weitere Auskünfte erteilt
Reiner Seitz, Telefon 0711/641-26 60,
Reiner.Seitz@stala.bwl.de

kurz notiert ...

Ob für Park, Vorgarten oder Obstwiese: Bäume und Sträucher aus dem Land

In Baden-Württemberg werden auf einer Fläche von fast 2 000 Hektar (ha) Bäume und Sträucher für verschiedenste Verwendungszwecke angezogen. Nach den Ergebnissen der Baumschulerhebung 2008 erfolgt die Anzucht in 327 Baumschulen (mit mindestens 30 Ar Baumschulfläche).

Auf annähernd der Hälfte der Flächen (ca. 950 ha) stehen Ziersträucher und -gehölze für Straßen, Parks, Gärten sowie Landschaftsgehölze. Dazu gehören zum Beispiel Bäume wie Ahorn und Linde, die an Straßen und in Parks gepflanzt werden (ca. 276 ha), Formgehölze wie zum Beispiel Buchs in Kugel- oder Kegelform auf ca. 19 ha, aber auch Bodendecker (33,5 ha), Rhododendren (3,8 ha) sowie Schling- und Kletterpflanzen (2,4 ha). Pflanzfähige Ware an sonstigen Nadelgehölzen wie Kiefer und Thuja wächst auf 285,5 ha. Ebenfalls auf einer größeren Fläche wird das breite Sortiment an sonstigen jungen Ziersträuchern produziert. Hierzu sind Schneeball, Flieder und Forsythien zu rechnen, aber auch typische Heckenpflanzen wie Liguster und Hainbuche, die zusammen auf fast 328 ha angezogen werden. Weit verbreitet in Baden-Württemberg als Land der Sonderkulturen sind Obstgehölze. Auf ungefähr einem Zehntel der Baumschulflächen (188 ha) wachsen junge veredelte Obst- und Rosensträucher bis zum verkaufsfertigen Sta-

dium heran. Der Schwerpunkt liegt hier bei Obstgehölzen (157 ha) wie Apfel-, Birn- und Zwetschgenbäumen.

Deutlicher Flächenrückgang bei Sommergerste

Ende April ist die Anbaufläche von Sommergerste gegenüber dem Vorjahr um 14 % auf 74 400 Hektar (ha) zurückgegangen. Als Gründe werden die hervorragenden Aussaatbedingungen im Herbst 2008 für die konkurrierenden Winterungen und die ungünstige Erlössituation in der Verwendung als Braugerste angeführt. Beim Hafer setzte sich der langjährige Abwärtstrend mit einem Minus von 13 % auf knapp 25 000 ha verstärkt fort: Die wichtigsten Wintergetreidearten Winterweizen (325 000 ha) und Wintergerste (105 500 ha) konnten ihre herausragende Stellung behaupten.

Gewinner im Wettbewerb um die Ackerfläche sind mit Triticale (+ 6 %; 23 400 ha) und Silomais (+ 9 %; 96 700 ha) zwei Fruchtarten, die ursprünglich ausschließlich als Futtermittel dienten, heute aber zunehmend als Substrat in der Biogasproduktion und damit der Energiegewinnung Verwendung finden. Der Anbau von Winterraps wurde auch ohne obligatorische Flächenstilllegung deutlich ausgedehnt (+ 13 %; 80 400 ha), während bei Körnermais (- 2 %; 76 000 ha) leichte Einschränkungen zu verzeichnen sind. ■