

Biometrische Grundlagen der Tierseuchenbekämpfung

Johannes Hofrichter
26. 5. 2010



Daten, Statistik
und Risikobewertung

Austrian Agency for Health and Food Safety (AGES)



Überblick

- 1 Population - Stichprobe
- 2 Stichprobenumfang
- 3 Sensitivität und Spezifität
- 4 räumlich- zeitlicher Verlauf
- 5 Risikobewertung

Fragestellungen

- Wie viele Tiere sind krank?
- Kommt die Krankheit in Österreich überhaupt vor?
- Gibt es eine zeitlichen Trend?
- Wie ist die räumliche Verteilung?
- Wie hoch ist das Risiko?



Wie viele Tiere sind krank?

Für welche Population soll eine Aussage gemacht werden?

- z.B. *Brucella suis*
 - Mastschweine
 - Hausschwein
 - Wildschwein
 - alle Schweine
 - alle Tiere die für *Brucella suis* empfänglich sind (Schwein, Feldhase, ...)



Wie ist eine Population definiert

- Gesamtheit aller gleichartigen Objekte, die hinsichtlich eines Merkmales untersucht werden
- Beispiele
 - ganze Spezies (z.B. Schweine)
 - charakteristische Teilmenge (z.B. Mastschweine $> 85\text{kg}$)
- mittels Erhebung (Stichprobe) möchte man Aufschlüsse über diese Population machen
- Stichprobe muß repräsentativ sein für die Population!



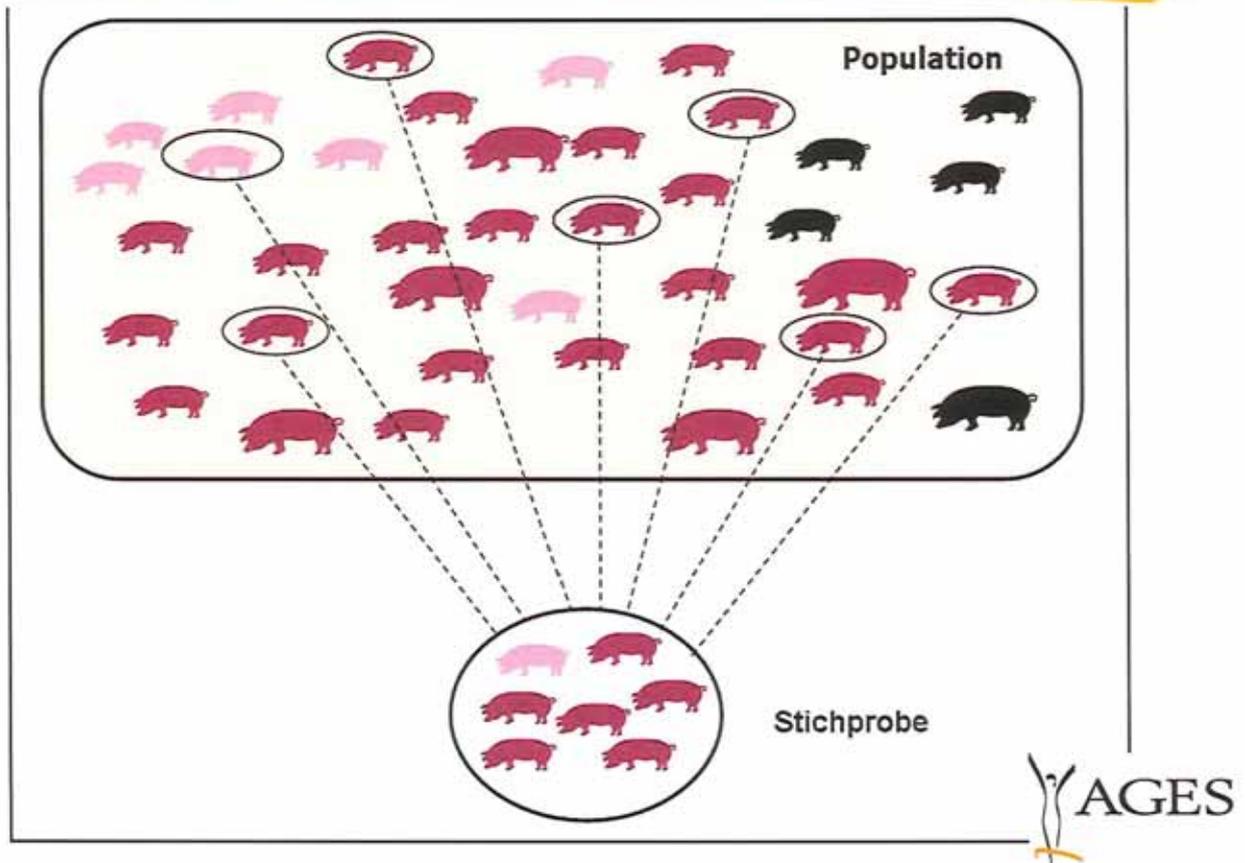
Repräsentative Stichprobe

- Mit Hilfe der Stichprobe sollen Informationen über die Population gesammelt werden
- Eine repräsentative Stichprobe soll ein Abbild der Population sein, sie soll die Heterogenität der Population widerspiegeln.
 - Altersstruktur
 - Geschlecht
 - Rasse
- Repräsentativ muss immer auf eine Population bezogen werden (Tierpopulation, menschliche Population, produzierte Lebensmittel, ...)
- Ziel: Bericht über den Status quo

Wie komme ich zu einer repräsentativen Stichprobe?



Population - Stichprobe



Einfache Zufallsauswahl

- einfache Zufallsauswahl
- jedes Objekt hat die gleiche Chance gezogen zu werden
- Auswahl z.B. über Zufallszahlengenerator (z.B. in Excel)
- Problem:
 - Nummerierung der Objekte notwendig
 - Erreichbarkeit (technisch und ökonomische Gründe)
 - praktische Auswahl im Bestand

Einfache systematische Auswahl

- Beispiel
 - Auswahl von 10% der Tiere
- systematische Auswahl
 - jedes 10. Tier wird ausgewählt
 - Zufallszahl zwischen 1 und 10 z.B. 4
 - jedes 4., 14., 24., 34., ... Tier wird untersucht
- sehr praktikabel
- Achtung:
 - Systematik der Entnahme darf keine unbekannte Systematik der Grundgesamtheit treffen.



Auswahlmöglichkeiten

Bequemlichkeit	bewusst	zufällig	systematisch



Inferenzstatistik

- Die Übertragung von Ergebnissen aus der Stichprobe auf die zugrundeliegende Population wird in der Statistik Inferenz genannt
- Im Prinzip handelt es sich dabei um eine Schätzung von unbekanntem Größen bzw. Parametern (z.B. Mittelwert oder Anteilswert) der Population mit Hilfe einer Stichprobe
- Die Schätzung wird dabei von mehreren Umständen beeinflusst
 - Variabilität (Variability)
 - Zufälligkeit (Randomness)
 - Messfehler (measurement error)
 - Sensitivität des diagnostischen Test
 - Spezifität
 - ...



Zu beantwortende Fragen

Sicherheit

Mit welcher Sicherheit (Wahrscheinlichkeit) kann man einen (unbekannten) Parameter der Population aus der Stichprobe berechnen?

Genauigkeit

Welches Vertrauen (Genauigkeit) kann in dessen Berechnung gelegt werden?



Beispiel - Parameterschätzung 1

Annahme:

10% der Tiere in der Population sind krank.

Stichprobe von Umfang $n = 150$

- 12 positive Tiere
- geschätzte Prävalenz $p = 0.08$; 8%



Prävalenz

- Relativer Anteil von kranken Tieren in einer Population zu einem bestimmten Zeitpunkt.

Prävalenz

$$P = \frac{\text{Anzahl der kranken Tiere zu einem bestimmten Zeitpunkt}}{\text{Anzahl der Tiere in der Risikopopulation zu einem bestimmten Zeitpunkt}}$$

Beispiel:

- In einem Betrieb mit 150 Schweinen sind 12 positiv auf *Brucella suis* getestet.

$$p = \frac{12}{150} = 0.08 = 8\%$$



Beispiel - Parameterschätzung 2

Ergebnis abhängig von der Stichprobe ($n = 150$):

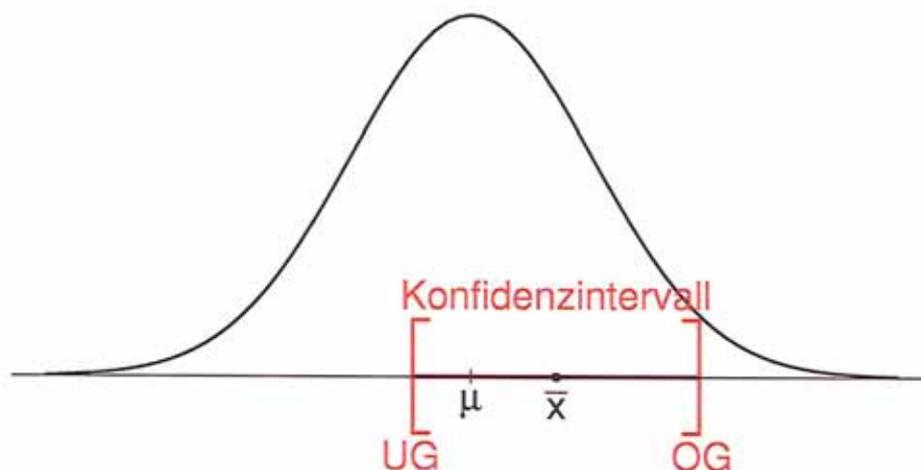
Stichprobe	pos. Tiere	Prävalenz
1	16	10.7%
2	18	12.0%
3	10	6.7%
4	9	6.0%
5	21	14.0%

In wie weit kann ich einem Wert vertrauen?



Konfidenzintervall

$$P(UG \leq \mu \leq OG) = 1 - \alpha$$



Genauigkeit vs Sicherheit

- $[UG, OG]$... Konfidenzintervall
 - je kleiner das Intervall, desto größer ist die Genauigkeit der Schätzung
- Genauigkeit
 - Länge des Intervalls: $OG - UG$
- Sicherheit
 - Wahrscheinlichkeit $1 - \alpha$, dass der Parameter im Intervall liegt
 - α ... Irrtumswahrscheinlichkeit



Konfidenzintervall - Beispiel 2

Untersuchung von 150 Schweinen auf *Brucella suis* (fiktives Beispiel!!)

- $m = 12$ positiv
- Anteilsschätzer: $12/150 = 0.08$
- 95% Konfidenzintervall: [4.43% – 13.45%]
- 99% Konfidenzintervall: [3.40% – 15.26%]



Konfidenzintervall - Beispiel 2

Was heißt das für die Ergebnisse aus den 5 versch. Stichproben?

Stichprobe	pos. Tiere	Prävalenz	95% KIV
1	16	10.7%	[6.42% – 16.49%]
2	18	12.0%	[7.41% – 18.16%]
3	10	6.7%	[3.41% – 11.78%]
4	9	6.0%	[3.02% – 10.91%]
5	21	14.0%	[9.09% – 20.49%]

Wie kann die Größe (Genauigkeit) des Intervalls beeinflusst werden?



Größe eines Konfidenzintervalls

- Sicherheit
 - Höhere Sicherheit → größeres KIV
- Anteilswert (Prävalenz)
 - je näher bei 0 oder 1 → kleineres KIV
 - worst case: 50%
- Stichprobenumfang
 - größerer Stichprobenumfang → kleineres KIV



Stichprobenumfang

Unterschiedliche Fragestellungen:

- Bestimmung eines Parameters mit einer vorgegebenen Genauigkeit
 - Prävalenz TBC im Rotwild
 - mittleres Gewicht bei Absetzferkeln
- Erkennen von einem Unterschied
 - Rasse, Inland/Ausland, ...
- Nachweis bezüglich Freedom from Disease
 - IBR/IPV bei Rindern



Datenanforderung für die Berechnung

- Sicherheit $1 - \alpha$
- Genauigkeit bzw. Unterschied den ich erkennen will!
- Streuung des Merkmals bzw. geschätzte Prävalenz
- Power $1 - \beta$ (üblicherweise 50% oder 80%)
- Größe der Population → Endlichkeitskorrektur



Stichprobenumfang - Prävalenz

Bestimmung der Prävalenz

- Unterschiedl. Sicherheit und Prävalenz (siehe Tabelle)
- Genauigkeit: $\pm 2\%$
- Power: 50%
- Größe der Population 100.000

Prävalenz	Sicherheit		
	90%	95%	99%
10%	609	865	1493
20%	1083	1537	2654
30%	1421	2017	3484
50%	1691	2401	4147



Größe des Stichprobenumfangs

- mit steigender Genauigkeit steigt der Stichprobenumfang
- mit steigender Sicherheit steigt der Stichprobenumfang
- mit steigender Power steigt der Stichprobenumfang
- je größer die Variabilität des interessierenden Merkmals, desto größer der Stichprobenumfang
- wenn der berechnete Stichprobenumfang n mehr als 10% der Größe N der Population beträgt, kann dieser mit der Endlichkeitskorrektur verringert werden

$$n_e = \frac{1}{\frac{1}{n} + \frac{1}{N}}$$



Stichprobenumfang - Prävalenz - Endlichkeitskorr.

Bestimmung der Prävalenz bei endlicher Population

- Sicherheit: 95%
- Genauigkeit: $\pm 2\%$
- Power: 50%
- Prävalenz und Größe der Population siehe Tabelle

Prävalenz	Größe der Population		
	∞	5000	2000
10%	865	738	604
20%	1537	1176	870
30%	2017	1438	1005
50%	2401	1623	1092



Freedom from Disease

Fragestellung:

Wie viele Tiere/Betriebe in einem Betrieb/Land muß ich untersuchen, um mit einer vorgegebenen Sicherheit sagen zu können, dass dieser/s Betrieb/Land frei von dieser Krankheit ist?

Was muß ich wissen?

- Prävalenz (Studie, Literatur, Expertenmeinung, ...)
- Größe der Population (zB Anzahl der Tiere im Betrieb)
- Sicherheit
- Sensitivität und Spezifität des diag. Test



"Genauigkeit" diagnostischer Tests

- Reproduzierbarkeit
 - Übereinstimmendes Ergebnis bei wiederholter Anwendung des Tests am selben Individuum
- Validität
 - Fähigkeit, ein wahres Maß für die getestete Eigenschaft zu liefern

Validität hat zwei wesentliche Aspekte:

- Sensitivität
 - Fähigkeit, **positives** Ergebnis zu liefern, wenn das Individuum die Krankheit hat
- Spezifität
 - Fähigkeit, **negatives** Ergebnis zu liefern, wenn das Individuum die Krankheit **nicht** hat



Vierfeldertafel eines Tests

		Gesundheitszustand		Summe
		erkrankt	nicht erkrankt	
Test- ergebnis	positiv	a	b	a+b
	negativ	c	d	c+d
Summe		a+c	b+d	a+b+c+d = n

$$\text{Sensitivität: } a/(a+c)$$

$$\text{Spezifität: } d/(b+d)$$



Vierfeldertafel eines Tests - Beispiel

		Gesundheitszustand		Summe
		erkrankt	nicht erkrankt	
Test- ergebnis	positiv	80	63	143
	negativ	20	837	857
Summe		100	900	1000

$$\text{Sensitivität} = 80/100 = 0.8 = 80\%$$

$$\text{Spezifität} = 837/900 = 0.93 = 93.0\%$$



Freedom from Disease - Beispiel

- Prävalenz und Größe der Population (N) siehe Tabelle
- Sicherheit: 95%
- Sensitivität: 90% und Spezifität 99.9%

N	Prävalenz		
	20%	10%	5%
100	15	28	49
500	15	31	95
2000	15	32	100
∞	16	32	102



Vierfeldertafel eines Tests

		Gesundheitszustand		Summe
		erkrankt	nicht erkrankt	
Test- ergebnis	positiv	a	b	a+b
	negativ	c	d	c+d
Summe		a+c	b+d	a+b+c+d = n

$$\text{Sensitivität: } a/(a+c)$$

$$\text{Spezifität: } d/(b+d)$$

$$\text{Wahre Prävalenz: } (a+c)/n$$

$$\text{Scheinbare Prävalenz: } (a+b)/n$$



Vierfeldertafel eines Tests - Beispiel

		Gesundheitszustand		Summe
		erkrankt	nicht erkrankt	
Test- ergebnis	positiv	80	63	143
	negativ	20	837	857
Summe		100	900	1000

$$\text{Sensitivität} = 80/100 = 0.8 = 80\%$$

$$\text{Spezifität} = 837/900 = 0.93 = 93.0\%$$

$$\text{Wahre Prävalenz} = 100/1000 = 0.1 = 10\%$$

$$\text{Scheinbare Prävalenz} = 143/1000 = 0.143 = 14.3\%$$



Schätzung der wahren Prävalenz

- notwendige Voraussetzung:
 - Sensitivität: S_e
 - Spezifität: S_p
 - geschätzte (scheinbare) Prävalenz: \hat{p}

Punktschätzer für die wahre Prävalenz: π

$$\hat{\pi} = \frac{\hat{p} + S_p - 1}{S_e + S_p - 1}$$



Wahre Prävalenz Beispiel

Punktschätzer:

$$\hat{\pi} = \frac{\hat{p} + S_p - 1}{S_e + S_p - 1} = \frac{\frac{143}{1000} + \frac{837}{900} - 1}{\frac{80}{100} + \frac{837}{900} - 1} = 0.1$$

Konfidenzintervall (Genauigkeit a):

$$a = 1.96 * \sqrt{\frac{0.1(1 - 0.1)}{1000(0.8 + 0.93)^2}} = 0.011$$

Konfidenzintervall: [0.089; 0.111] bzw. [8.9%; 11.1%]



Vorhersagewerte

		Gesundheitszustand		Summe
		erkrankt	nicht erkrankt	
Test- ergebnis	positiv	a	b	a+b
	negativ	c	d	c+d
Summe		a+c	b+d	a+b+c+d = n

- positiver Vorhersagewert: $a/(a + b)$
 - Wahrscheinlichkeit, dass ein Test positives Tier wirklich positiv ist.
 - hängt von der Prävalenz, der Sensitivität und der Spezifität ab.
- negativer Vorhersagewert: $d/(c + d)$



Kleiner Test

- Die Prävalenz einer Erkrankung liegt bei 1/1000. Die Sensitivität des Testverfahrens liegt bei 99%, die Spezifität bei 98%.
- Wie hoch ist die Wahrscheinlichkeit, dass ein positiv getestetes Tier wirklich positiv ist?
 - 99%
 - 98%
 - ca. 95%
 - ca. 50%
 - ca. 5%
 - 2%
 - 1%



Zeitliche Analysen

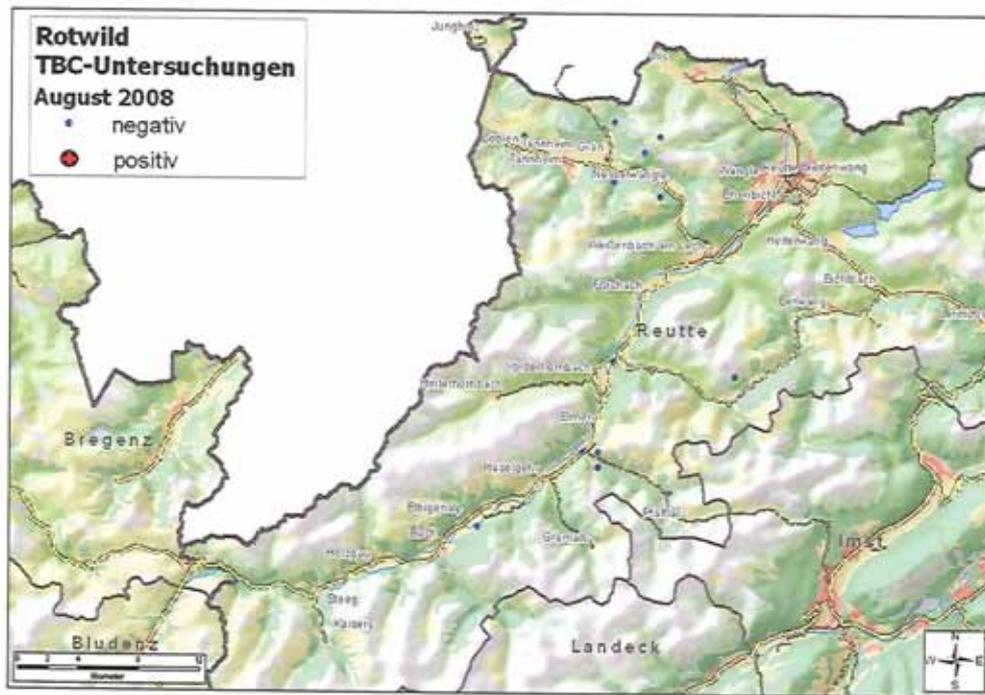
- Trendanalyse
 - Rückgang bzw. Anstieg über die Jahre
 - Welchen Effekt hat eine Impfung?
 - Saisonale Schwankungen
- Erkennen eines Ausbruches
 - Ab wie vielen Fällen kann ich von einem Ausbruch sprechen?
 - Kritische Zahl von Erkrankungen



Räumliche Analysen

- regionale Unterschiede
- Ausbreitung
 - Richtung
 - Transportwege
- effiziente Abdeckung eines Stichprobenplanes





Vorraussetzungen für räuml. zeitl Analysen

Valide Daten:

- Identifikation (LFBIS, OM-Nummer, Probennummer, ...)
- Datum (Abschuß, Auffinden, Probenziehung, ...)
- geographische Lage (Koordinaten, GemNr., ...)
- Größe der Population
- ...



Definition Risikobewertung (OIE)

- release assessment
 - Verteilung der Cul.obs. in Österreich
- exposure assessment
 - Exposition von Wiederkäuern gegenüber Cul.obs.
 - Tierdichte von Wiederkäuern
- consequence assessment
 - Handelsrestriktionen
 - Leistungseinbußen
 - Tod der Tiere
- risk estimation
 - Wahrscheinlichkeit, dass Wiederkäuer infiziert werden
 - Abschätzen der Eintrittswahrscheinlichkeit und Bewerten der Konsequenzen



Ziel

Ziel ist eine Risikokarte für das Auftreten der Blauzungkrankheit

- Vektoren: Culicoides obsoletus (Cul.obs.)
 - Modellierung in Abhängigkeit von Wetterdaten und GIS-Daten
- Tierdichte
 - Rinder, Schafe, Ziegen, Wild

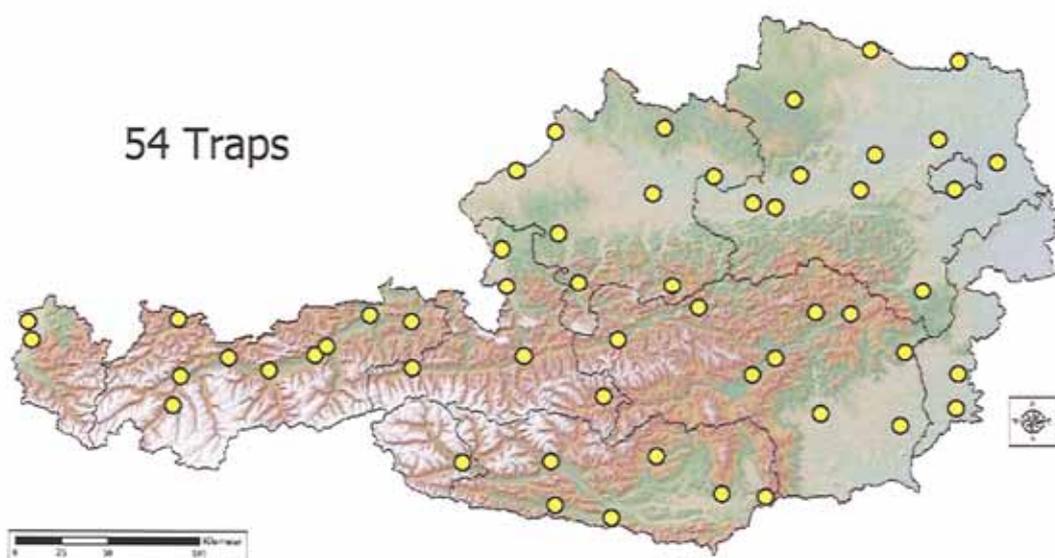


Release Assessment

- Auftreten von Cul. obs. in Österreich
- Datensammlung im Rahmen des EU-Monitoring
- 54 repräsentative Mückenfallen in ganz Österreich
- Beobachtungszeitraum: 31. März bis 30. September 2008 (27 Wochen)
- wöchentlich geleert und "gezählt"



Standorte der Mückenfallen

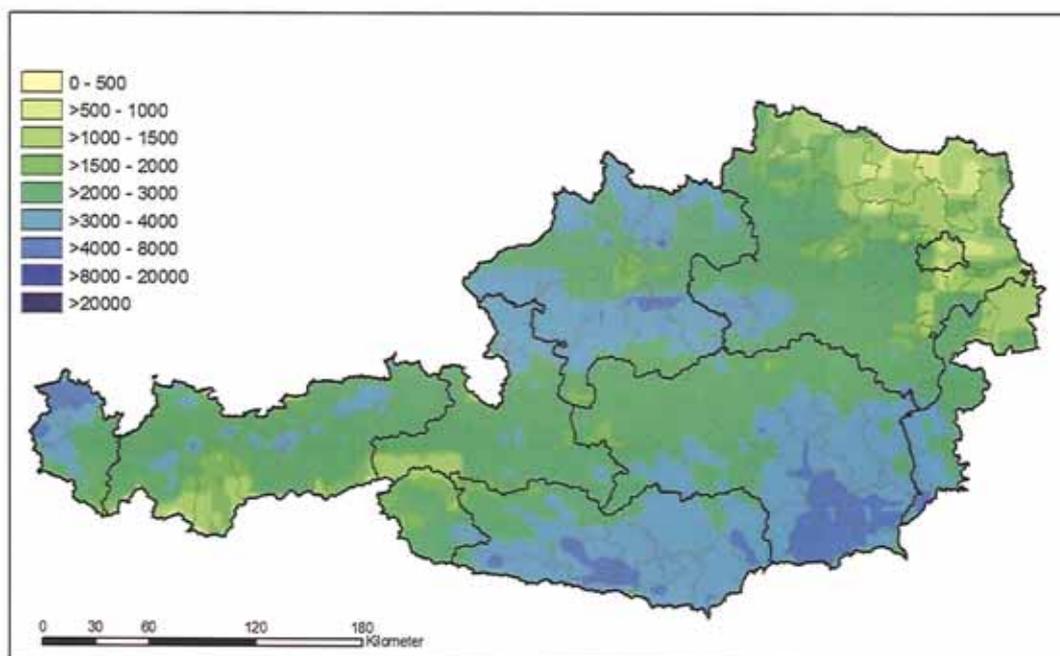


Erklärende Variablen

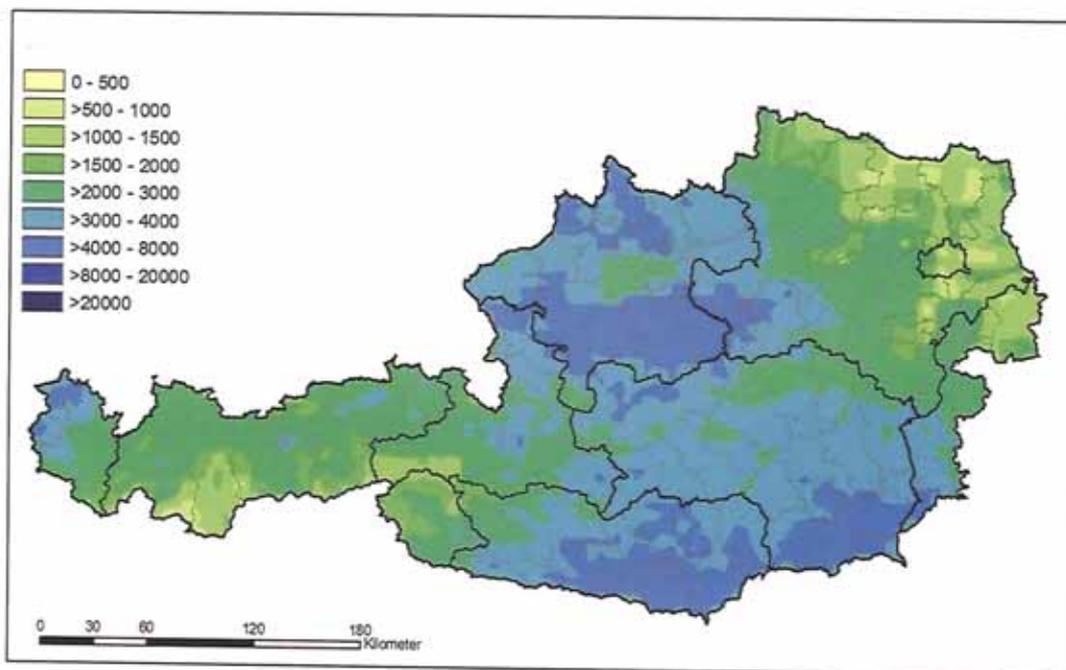
- Daten von 44 Wetterstationen
 - Temperatur (min, max, average)
 - Niederschlag
 - rel. Luftfeuchtigkeit
 - Windgeschwindigkeit
- Daten aus Geographischen Informations Systems (GIS)
 - Landnutzung (Wald, Grass, etc.)
 - Vegetationsindex (NVDI: Quelle NASA)
 - Index für das Vorkommen von Gewässer
 - Seehöhe
 - Tierdichte (Rinder, Schafe, Ziegen, Schweine, Geflügel)



Vorhersage für Woche 19



Vorhersage für Woche 21



Definition Risikobewertung (OIE)

- release assessment ✓
 - Verteilung der Cul.obs. in Österreich
- exposure assessment
 - Exposition von Wiederkäuern gegenüber Cul.obs.
 - Tierdichte von Wiederkäuern
- consequence assessment
 - Handelsrestriktionen
 - Leistungseinbußen
 - Tod der Tiere
- risk estimation
 - Wahrscheinlichkeit, dass Wiederkäuer infiziert werden
 - Abschätzen der Eintrittswahrscheinlichkeit und Bewerten der Konsequenzen

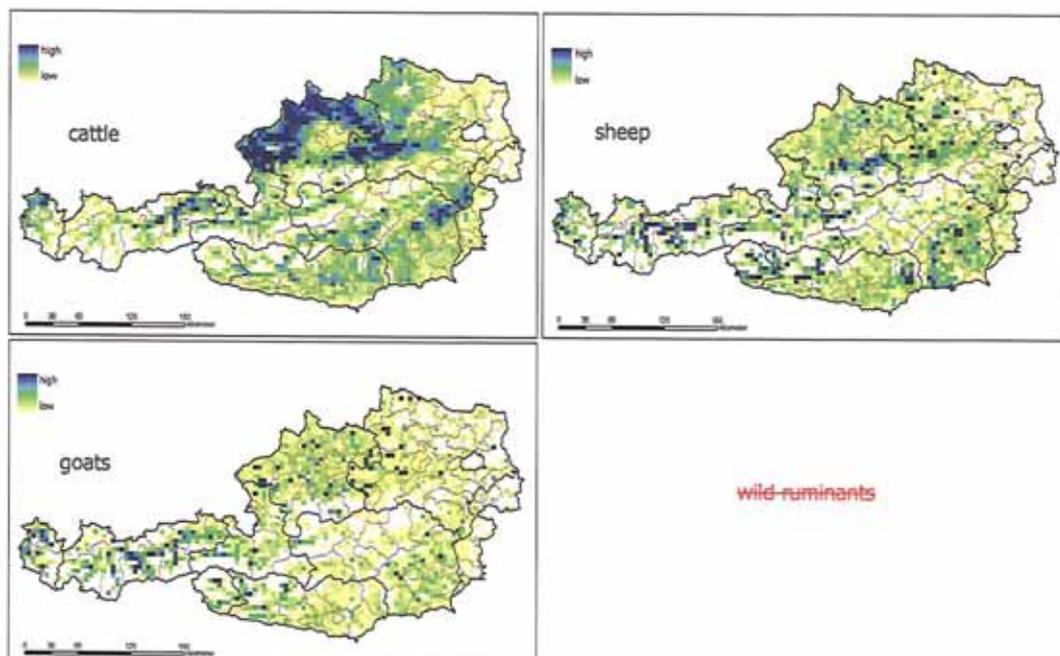


Definition Risikobewertung (OIE)

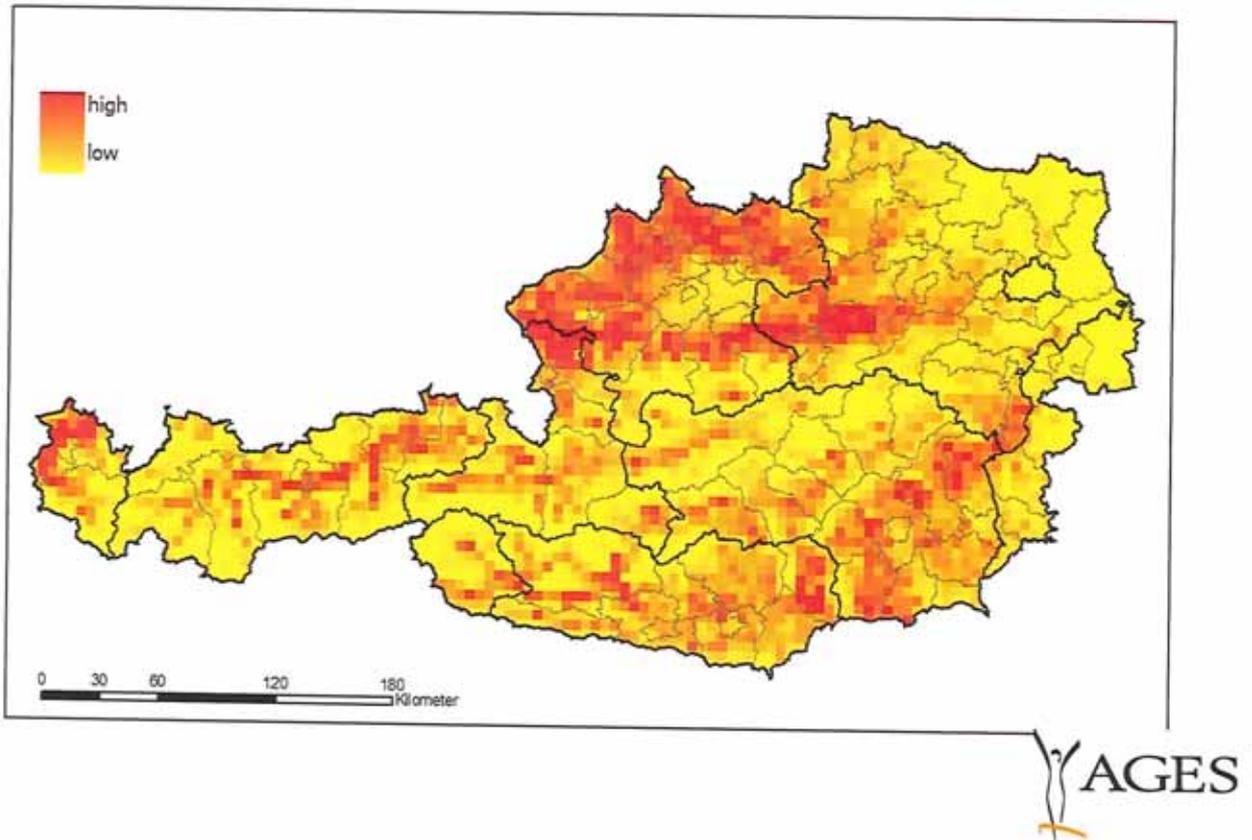
- release assessment ✓
 - Verteilung der Cul.obs. in Österreich
- exposure assessment
 - Exposition von Wiederkäuern gegenüber Cul.obs.
 - Tierdichte von Wiederkäuern
- consequence assessment
 - Handelsrestriktionen
 - Leistungseinbußen
 - Tod der Tiere
- risk estimation
 - Wahrscheinlichkeit, dass Wiederkäuer infiziert werden
 - Abschätzen der Eintrittswahrscheinlichkeit und Bewerten der Konsequenzen



Tierdichte für Wiederkäuer in Österreich



Expositionskarte für Blauzungenkrankheit



Noch zu tun

- Validierung des räumlichen Modelles für Cul.obs. mit aktuellen Daten
- consequence assessment
 - Handelsrestriktionen
 - Leistungseinbußen
 - Tod der Tiere
- risk estimation
 - Wahrscheinlichkeit, dass Wiederkäuer infiziert werden
 - Abschätzen der Eintrittswahrscheinlichkeit und Bewerten der Konsequenzen

Danksagung: Michael Schwarz für die Aufbereitung der GIS-Daten und die schönen Karten

Fazit

statistische Methoden

- Abschätzung der wahren Prävalenz
- Berechnung des Stichprobenumfanges
- räumlich - zeitlich Analysen
- quantitative Risikobewertung

Vorraussetzung: valide Daten!!

- LFBIS- Nummer
- OM - Nummer
- Datum der *Probennahme*
- Koordinaten der *Probennahme*
- ...