



Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences



Workshop

Data-Mining mit Administrativdatensätze der Sozialen Sicherheit
SECO, 17.10.2013

▸ Soziale Arbeit

Programm

Teil 1: Ergebnisse vorstellen (1.25 Std.)

- 1/4h Rahmenbedingungen und Datengrundlagen
- 1h Präsentation der Analysen und Resultate: Ziel, Daten, Aufbereitung, Resultate (Verlaufstypologie, Verlaufsmerkmale, Systemindikatoren, Risikoprofile, Prognose)

Teil 2: Monitoring-Tool (0.75 Std.)

- 1/2h Auswertungsmöglichkeiten des Cube zeigen: 2 Bsp. über Excel, 2 Bsp. über Reporting Services
- 1/4h Fragen

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences



Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Einleitung

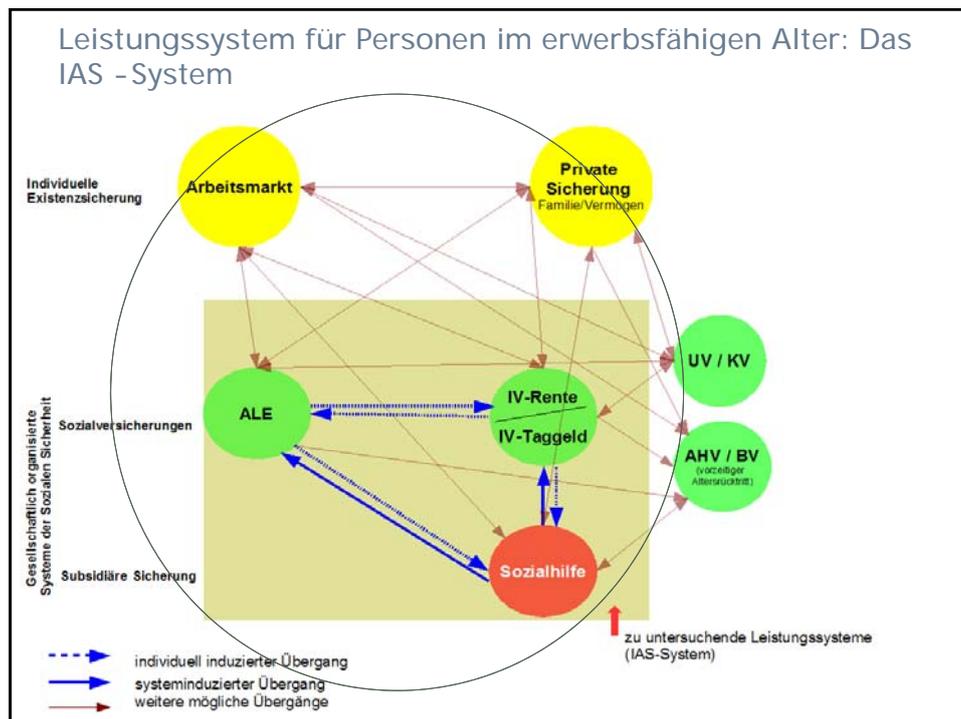
Vorhandene Administrativdaten

- **Arbeitslosenversicherung** :
 - ASAL Daten des Auszahlungssystems
 - → AVAM Daten der RAV (weitere Merkmale, Massnahmen etc.)
- **Invalidenversicherung**
 - Renten, Früherfassung und Massnahmen
- **Sozialhilfe**
 - Erfassung in den Sozialdiensten bzw. Gemeinden
 - → Zusammentragen und Harmonisierung durch BFS

Im Rahmen des Monitorings SHIVALV des BSV werden diese Daten zu einem Datensatz kombiniert, harmonisiert.

Zusätzlich: Daten der **individuellen Konten der AHV** (v.a. Erwerbseinkommen bzw. Angabe der Erwerbstätigkeit je Monat)

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences



Datenmodell, Datenschutz

- Grosse Datenmengen z.B. 7.3 Mio. Datenrecords (Projekt Übergänge)
- Hohe Anforderungen an den Datenschutz: Verknüpfung über anonyme Projekt-ID (Basis SV-Nummer), Aggregationsgrad der Auswertungen
- Geeignetes Datenmodell: Aufbau einer Datenbank mit den Originaldatensätzen, Abfrage von Arbeitsdatensätzen je nach Fragestellung, konsolidierte Aufbereitung in Cube
- hohe Flexibilität des Datenmodells zur Auswertung verschiedener Dimensionen (Soziodemographie, Ort, Zeit)
- Qualitätssicherung und Plausibilisierung der Ergebnisse

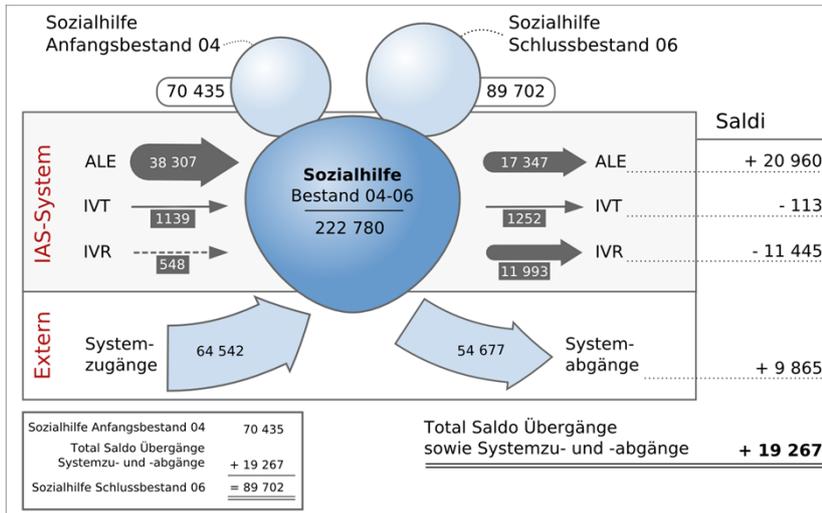
Projekt IV-ALV-SH-Übergänge, Information BSV Plattform 24.3.09

6

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Systemübergänge (2009)

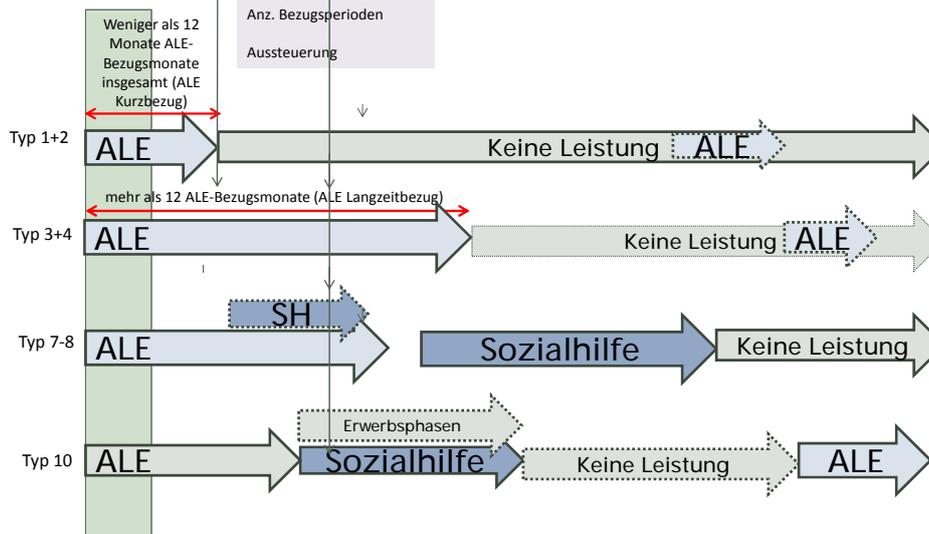
Bilanz der Personenflüsse 2004-2006



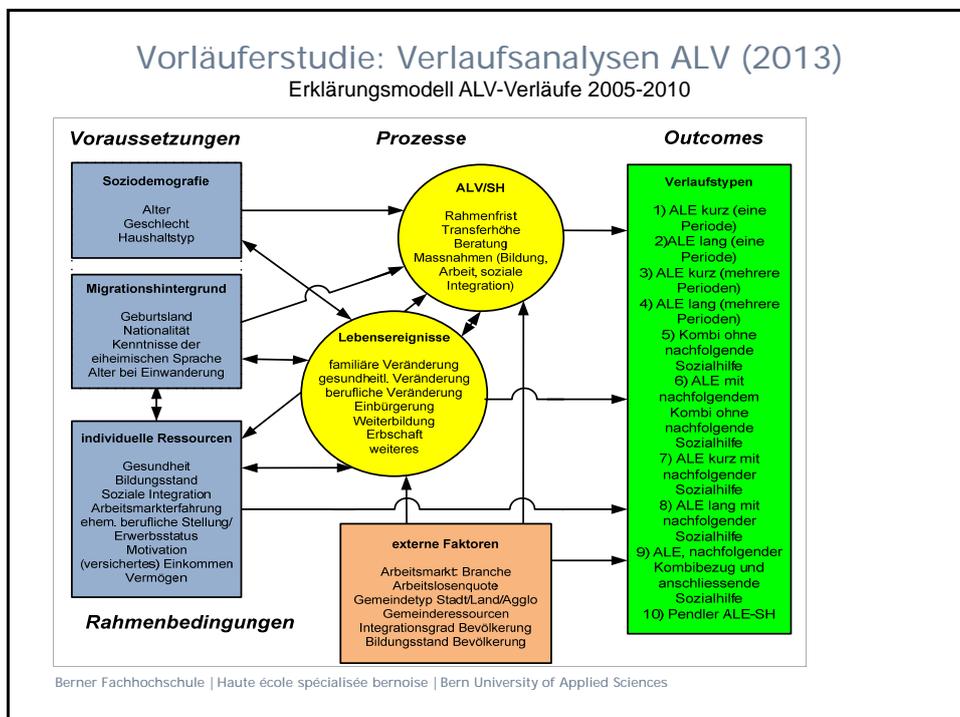
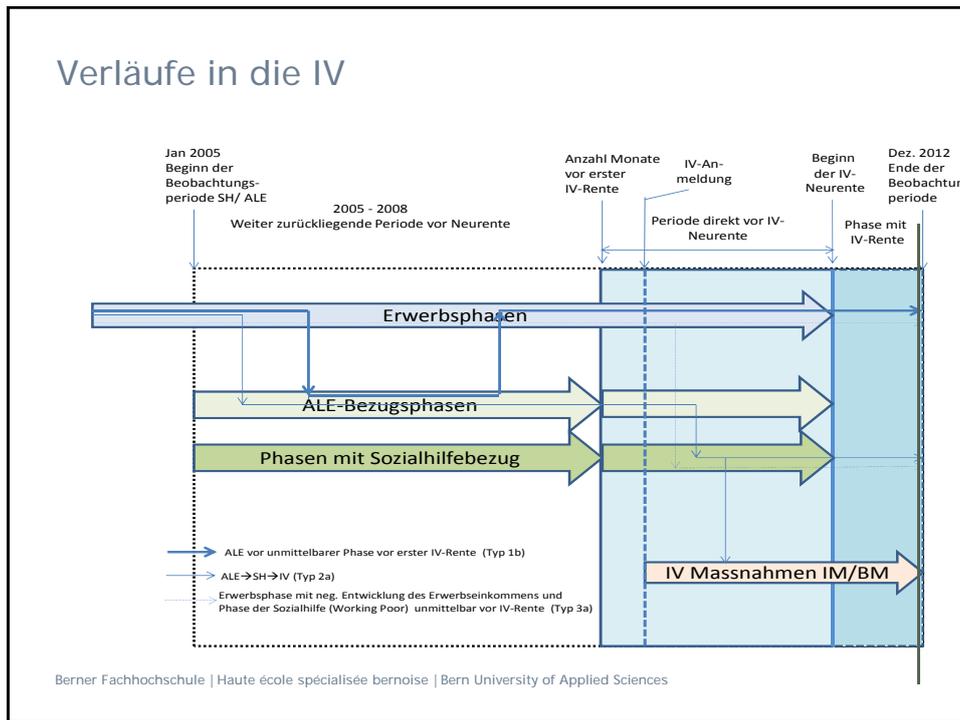
Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences
 Projekt IV-ALV-SH-Übergänge, Information BSV Plattform 24.3.09

Neubezug ALE

Verläufe von ALE Neubeziehenden



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences



Ziele und Rahmen Dataming-Projekt

- Arbeiten mit grossen gesamtschweizerischen Registerdatenmengen, z.B. für 2010: SHIVALV N=663'583 Datenrecords, IK N=1'005'577, ASAL N=2'392'000
 - strukturierende Verfahren aus Marketing anwenden, induktives Vorgehen, datengetrieben
 - Finden von Konstellationen und Indikatoren: nicht kausale Interpretation, Risikoprognose
- Interdepartamentales Projekt zusammen mit TI
 → Budget 154'500, davon Cash-Beitrag SECO 26'000
 → Software: SPSS Modeler, STATA, SQL

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Zeitplan

- Start: September 2012
- Schulung Modeler-Programm: Januar 2013
- Aufbau Datenbank, exploratives Data Mining: Frühjahr
- 1. workshop an BFH: Juli

10	Modul 3 Aufsetzen Cube für Auswertungen	01.05.2013	30.05.2013						
11	Prototyp Cube für SECO	30.05.2013		★					
12	Modul 4 Gebrauchsanleitung, Fragen SECO	03.06.2013	02.07.2013						
13	Workshop	17.06.2013		★					
14	Modul 2 Exploratives Data Mining 2. Phase	01.07.2013	30.08.2013						
15	Modul 3 Anpassung Cube, Test SECO	01.08.2013	13.09.2013						
16	Definitives Auswertungstool	16.09.2013		★					
17	Modul 4 Schlussbericht	15.08.2013	29.11.2013						
18	Bericht für SECO	31.10.2013		★					
				Jun 2013	Jul 2013	Aug 2013	Sep 2013	Okf 2013	Nov 2013

- Weitere Veröffentlichung: BFH-Bericht, Artikel in Zeitschrift

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Projekt Datamining: Fragestellungen

Fragestellungen:

- 1. Welche Methoden und welches Vorgehen aus dem Bereich des Data Mining lassen sich erfolgreich zur Analyse eines Administrativdatensatzes anwenden?
- 2. Lassen sich mittels dieser komplexen Methoden interpretierbare Klassen (Verlaufsmuster) identifizieren?
- 3. Können solche typische Klassen in multivariaten Regressionsanalysen durch Risikofaktoren erklärt werden?
- 4. Lassen sich diese Methoden im Bereich der Qualitätssicherung der Daten anwenden, um die Qualität der Daten zu evaluieren?
- 5. Lassen sich die als relevant festgestellten Methoden des Data Mining in einem Analyse-Instrument für die Langzeitbeobachtung konsolidieren?

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Konzeptionelle Grundlagen der Datenbearbeitung

- Gesamtschweizerische Daten, Beobachtungszeitraum 2005 bis 2010 (72 Monate bzw. 24 Quartale)
 - Geeignetes Datenmodell mit hoher Flexibilität zur Auswertung verschiedener Dimensionen (Soziodemographie, Ort, Zeit)
 - Nur Leistungsbezüge/Unterbrüche > 1 Monat, Überlappungen von Erwerb und ALV wurden zugeschnitten
- > quartalsweise Aufbereitung erscheint zweckmässig (2v3Mt)
- Definition einer Eintrittskohorte für Verlaufstypen
 - Qualitätssicherung und Plausibilisierung: Harmonisierungen BSV, Erfahrung aus vorangehenden Projekten, workshops

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Ergebnisse

Analysen und Resultate

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Programm

Teil 1: Ergebnisse vorstellen (1.25 Std.)

- ▶ 1/4h Rahmenbedingungen und Datengrundlagen
- ▶ 1h Präsentation der Analysen und Resultate: Ziel, Daten, Aufbereitung, Resultate (Verlaufstypologie, Verlaufsmerkmale, Systemindikatoren, Risikoprofile, Prognose)
- ▶ *Analysen und Resultate*
 - ▶ Datenaufbereitung und Verlaufsindikatoren
 - ▶ Verlaufsmuster
 - ▶ Risikoprofile
 - ▶ Systemindikatoren

Teil 2: Monitoring-Tool (0.75 Std.)

- ▶ 1/2h Auswertungsmöglichkeiten des Cube zeigen: 2 Bsp. über Excel, 2 Bsp. über Reporting Services
- ▶ 1/4h Fragen

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Datenaufbereitung

Tabelle 5: Codierung der Verlaufstypen

Einfache Codierung (1. Version)					
Zustandskombination: 1: zutreffend 2: nicht zutreffend Egal: 1 oder 2					
Code 1	ALV	IVR	SH	IVT	ER
A0	1	2	2	2	2
A1	1	2	2	2	1
B0	1	1	2	egal	2
B1	1	1	2	egal	1
D0	1	2	2	1	2
D1	1	2	2	1	1
E0	1	1	1	egal	2
E1	1	1	1	egal	1
I0	2	1	2	egal	2
I1	2	1	2	egal	1
C0	1	2	1	2	2
C1	1	2	1	2	1
J0	2	1	1	egal	2
J1	2	1	1	egal	1
U0	2	2	1	1	2
U1	2	2	1	1	1
S1	2	2	1	2	1
S0	2	2	1	2	2
T0	2	2	2	1	2
T1	2	2	2	1	1
X1	2	2	2	2	1
X0	2	2	2	2	2
F0	1	2	1	1	2
F1	1	2	1	1	1

Quelle: Darstellung BFH

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Einfache Codierung (2. Version)						
Zustandskombination						Code 2
ALV						A0
ALV und Erwerb						A1
ALV SH						C0
ALV SH mit Erwerb						C1
IV-Taggeld						T0
IV-Taggeld mit Erwerb						T1
SH kombi IVR und/oder IVT, ev. ALV	U0	J0	E0	F0	>	kombi1
SH kombi IVR und/oder IVT, ev. ALV mit Erwerb	U1	J1	E1	F1	>	kombi1_er
Sozialhilfe						S0
Sozialhilfe mit Erwerb						S1
IV-Rente, ev. IVT						I0
IV-Rente mit Erwerb, ev. mit IVT						I1
Erwerb						X1
Kein Einkommen und keine Leistungen						X0
IV-Rente und/oder IVT mit ALV	B0	D0			>	kombi2
IV-Rente und/oder IVT mit ALV mit Erwerb	B1	D1			>	kombi2_er

te
von
ierung.
rmatoren

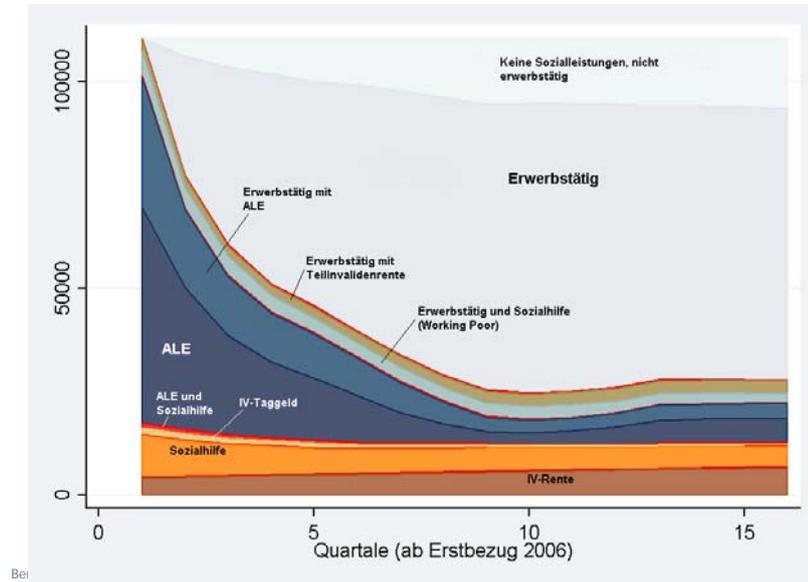
zur
1X1

Eingrenzung der betrachteten Verläufe

- ▶ Eingrenzen der Analysen auf **Neueintritte**. Verlauf startet mit Bezug bei einem der vier Systeme. Betrachteten von **Verläufe über 4 Jahre** (16 Quartale)
- ▶ Eingrenzen auf **Bevölkerung im Erwerbsalter** (keine Personen, die im Beobachtungsfenster das Rentenalter erreichen)
- ▶ Grundlage der Analysen sind Verläufe von Personen, die 2006 zum ersten Mal mit der Sozialen Sicherung in Kontakt gekommen sind:
n=110'508

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Erste visuelle Betrachtung – Zustandsverteilung Eintrittskohorte 2006 – Gesamtsystem



Erwerbsintegrationsindex – Distanz zur Erwerbstätigkeit

–Lässt sich die Komplexität eines Verlaufes gesamtsystemisch einheitlich auf sinnvolle Weise numerisch abbilden?

- Methode der Sequenzdatenanalyse (Biographie-, DNA-Forschung). «Distanz» zwischen Sequenzen wird berechnet.
- ▶ Alle Verläufe werden mit einem idealtypischen Verlauf einer **durchgehend Erwerbstätigkeit** verglichen -> Distanz zur Erwerbstätigkeit

Seq 0: X1X1X1X1X1 X1X1X1 X1X1X1X1X1X1X1X1X1

Seq 1: A0X1X1X1X1X1X1X1X1X1X1X1X1X1X1X1X1X1

Seq 2: IO

- Es resultiert eine Variable, die die Komplexität eines einzelnen Verlaufes in einer Zahl verdichtet.
 - Tiefe Werte = geringe Distanz zur Erwerbstätigkeit
 - Hohe Werte = weit entfernt von der Erwerbstätigkeit

Erwerbsintegrationsindex – Distanz zur Erwerbstätigkeit

–Lässt sich die Komplexität eines Verlaufes gesamtsystemisch einheitlich auf sinnvolle Weise numerisch abbilden?

- *Leistungsbezüge sind unterschiedlich weit weg von der Erwerbstätigkeit*
- Die empirisch beobachtete Möglichkeit ausgehend von bestimmten Leistungsbezügen in den Zustand «Erwerbstätig» zu gelangen, wird mit modelliert: ALV ist etwa «näher» als IV.

$Status_{t_0}$	$Pr(Erwerbstätig)_{t_0+1}$
ALV	21.09
ALV und Erwerb	37.88
ALV und Sozialhilfe	5.40
ALV, Sozialhilfe und Erwerb	12.00
IV-Rente	0.01
IV-Rente und Erwerb	0.39
Sozialhilfe	2.09
Sozialhilfe und Erwerb	15.03
Keine Sozialleistungen, nicht erwerbst.	10.78

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

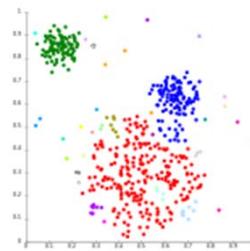
Ergebnisse

Verlaufsmuster

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Verlaufstypologie – Datengestützte Verlaufstypologie über eine 2-step Clusteranalyse

- Mittels **Clusteranalyse** ist es möglich, Beobachtungen so zu gruppieren, dass die Beobachtungen innerhalb eines Clusters möglichst ähnlich und die Unterschiede zwischen den Cluster möglichst gross sind.



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

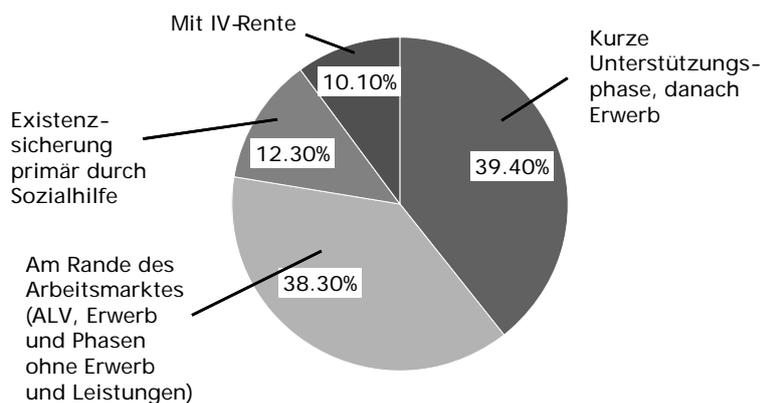
Verlaufstypologie – Datengestützte Verlaufstypologie über eine 2-step Clusteranalyse

- *22 Inputvariablen zur Charakterisierung von Verläufen*: Zustand am Anfang, nach einem Jahr und nach 4 Jahren (nominal), Dauer in Zuständen (16), Anzahl Zustandswechsel, Anzahl unterschiedlicher Zustände und Distanz zur Erwerbstätigkeit (metrisch)

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Verlaufskluster

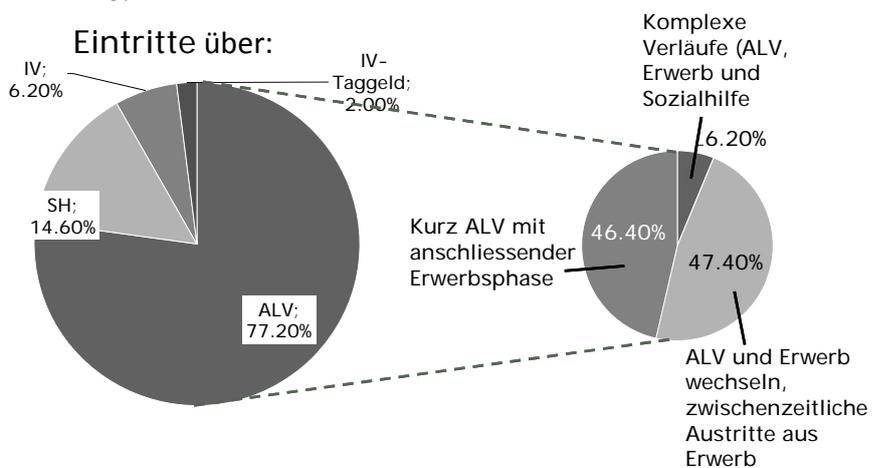
Vier Grundtypen von Verläufen in der gesamtsystemischen Sicht



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Verlaufskluster

Drei Grundtypen von Verläufen mit Fokus auf die ALV



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

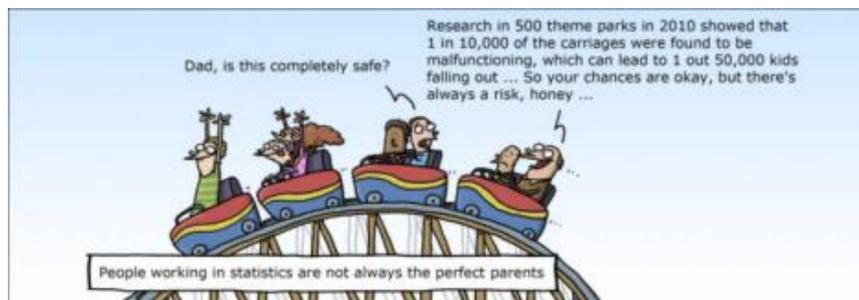
Ergebnisse

Risikoprofile

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Risikoprofile – Grundlegende Überlegungen

Was heisst Risiko?



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

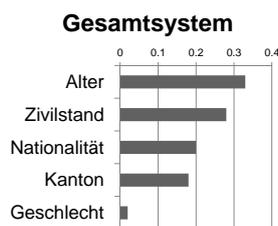
Risikoprofile – Grundlegende Überlegungen

Was heisst Risiko?

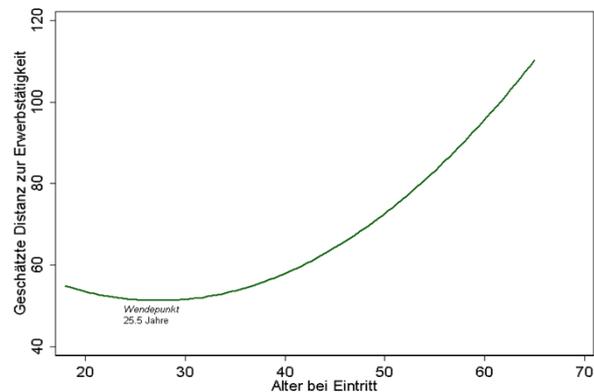
- ▶ In diesem Kontext wird Risiko in Bezug zur Erwerbstätigkeit und zur Komplexität von Verläufen gesetzt.
- ▶ *Wichtig:* Es geht um Risiko von Personen, die mit dem System der Sozialen Sicherheit in Kontakt gekommen sind (nicht um Risiko der Gesamtbevölkerung)
 - ▶ Risikoprofile werden über eine Verbindung von Personenmerkmalen mit «Risikoindikatoren» erstellt
 - ▶ Risikoindikatoren: (a) Distanz zur Erwerbstätigkeit und (b) Verlaufstypen
 - ▶ Personenmerkmale: Merkmale zur Beschreibung der Situation bei Eintritt

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Risikoprofile – Risikofaktoren Gesamtsystem

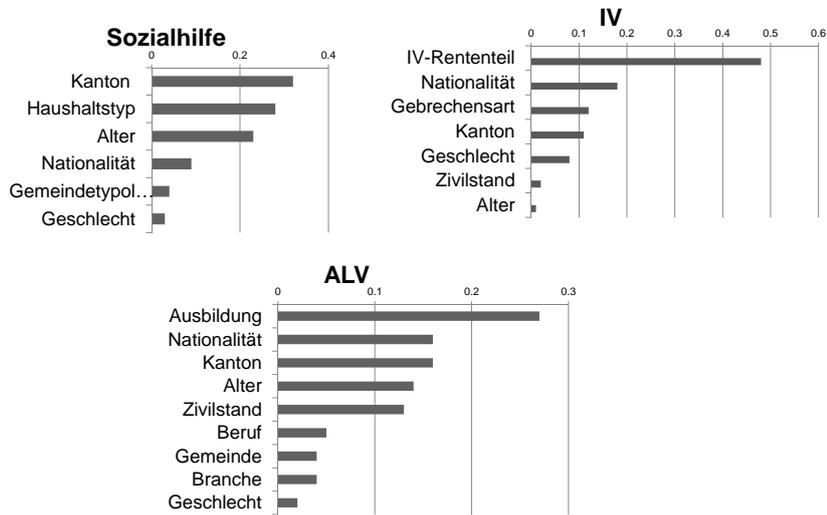


Risiko ist mit 25.5 Jahren am geringste, ab 40 steigt die vorhergesagte Distanz zur Erwerbstätigkeit deutlich an



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

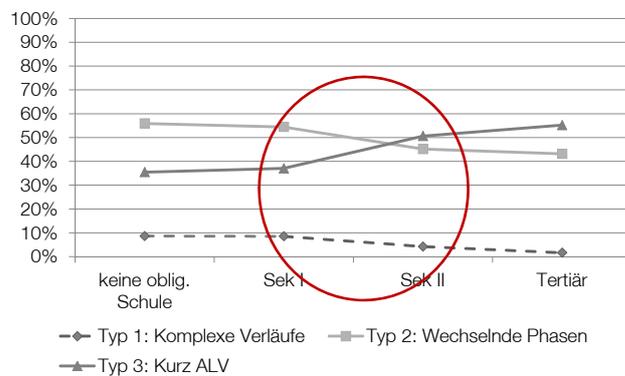
Risikoprofile – Bedeutsamkeit von Prädiktoren nach Eintrittssystemen



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Risikoprofile mit Fokus auf die ALV

Wahrscheinlichkeit der Zugehörigkeit zu den drei Verlaufstypen nach Bildungsunterschieden



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Risikoprofile mit Fokus auf die ALV

Drei Risikogruppen - Wahrscheinlichkeit der Zugehörigkeit zu den drei Verlaufstypen

	Typ 1: Komplex	Typ 2: Wechselhaft	Typ 3: Erwerb
kein obligatorischer Schulabschluss aus EU8, geschieden/verwitwet/getrennt, 46 - 65 Jahre	24.7%	62.4%	12.9%
Abschluss auf Sekundarstufe I, aus EU8, zwischen 18 und 25 Jahren	13.4%	56.9%	29.6%
zwischen 45 und 65 Jahre, Schweizer-/In, ledig	8.6%	50.6%	40.8%
Profil von Durchschnittsbeziehenden lediger Mann, zwischen 26 und 45 Jahren, Schweizer, Gastgewerbe, mit Sekundarstufen II Abschluss.	7.6%	37.6%	54.7%
Abschluss auf Sekundarstufe II, unter 23 Jahre	4.1%	38.1%	57.8%

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Ergebnisse

Systemindikatoren

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Systemindikatoren – Beurteilungen von strukturellen Veränderungen

Erweiterung des Blickfeldes mittels Verlaufsperspektive

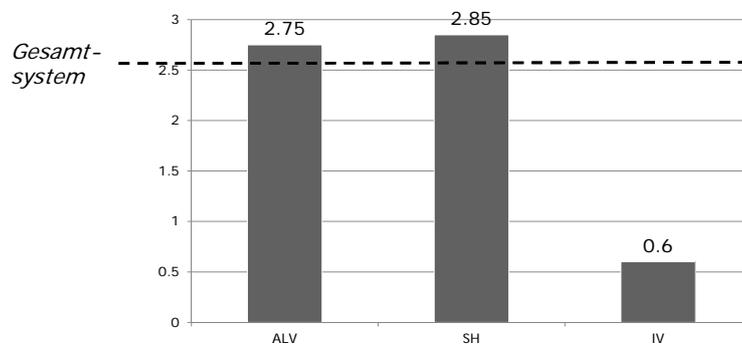
- ▶ Monitoring zur Beobachtung von Entwicklungen und Veränderungen im Kontext der Systeme der Sozialen Sicherheit
- ▶ Bisher durchgeführte regelmässige Beobachtungen sind meist *Querschnittbetrachtungen*, z. Bsp. «Die Lage auf der Arbeitsmarkt»
- ▶ Welche Möglichkeiten und Chancen bietet eine *Verlaufsperspektive* hinsichtlich der Entwicklung von Indikatoren auf der Systemebene?

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Systemindikatoren aus Verlaufsanalyse

Beurteilungen von strukturellen Veränderungen mittels Mobilitätsindikatoren

- ▶ Anzahl Wechsel der Quelle der finanziellen Unterstützung -> Index zur Nachhaltigkeit

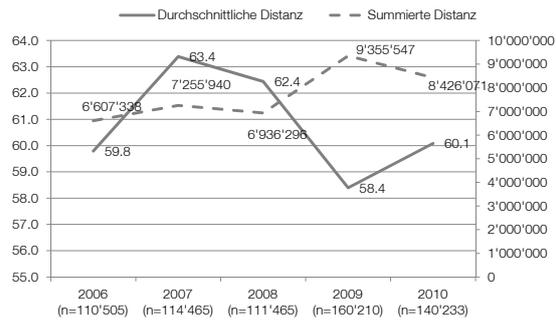


Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Systemindikatoren aus Verlaufsanalyse

Beurteilungen von strukturellen Veränderungen mittels Mobilitätsindikatoren

- ▶ Erwerbsintegrationsindex als Indikator der Belastung der Sozialen Sicherheit



- ▶ Kombiniert Information zu
- ▶ (1) Zahl der Unterstützten nach System
- ▶ (2) «Struktur» der Eintrittskohorten

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences



Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

Monitoring-tool

- ▶ Fachbereich Soziale Arbeit

Programm

Teil 1: Ergebnisse vorstellen (1.25 Std.)

- ▶ 1/4h Rahmenbedingungen und Datengrundlagen
- ▶ 1h Präsentation der Analysen und Resultate: Ziel, Daten, Aufbereitung, Resultate (Verlaufstypologie, Verlaufsmerkmale, Systemindikatoren, Risikoprofile, Prognose)

Teil 2: Monitoring-Tool (0.75 Std.)

- ▶ 1/2h Auswertungsmöglichkeiten des Cube zeigen: 2 Bsp. über Excel, 2 Bsp. über Reporting Services
- ▶ 1/4h Fragen

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Monitoring-tool

- ▶ Aufbau Cube
- ▶ Auswertungsmöglichkeiten ansehen
- ▶ 2 Bsp. über Excel: Abrufen von Daten aus dem Cube
- ▶ 2 Bsp. über SQL-Server Reporting Services

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

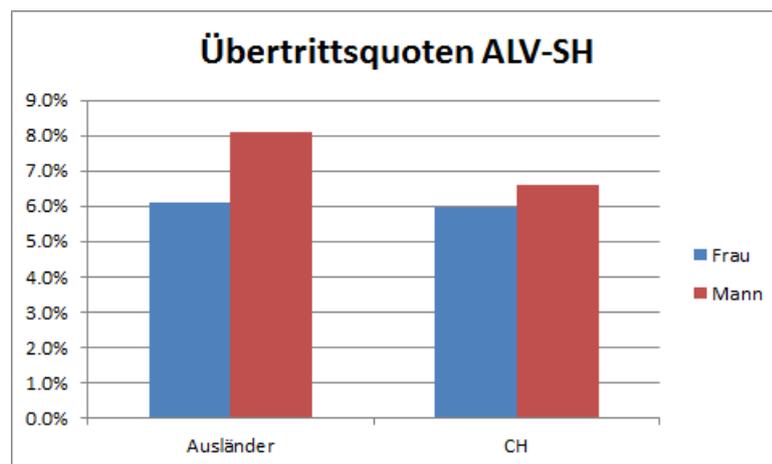
Monitoring-tool

Was ist ein Cube?

- ▶ Die aufbereiteten Personen- und Verlaufsdaten werden in einem kleinen Data Warehouse mit multidimensionalem Datenmodell abgebildet.
- ▶ Das Data Warehouse ist eine relationale SQL Server 2008 R2 Datenbank.
- ▶ Auf Basis des Datawarehouse wird eine SQL Server 2008 R2 Analysis Services OLAP Datenbank mit einem Cube erstellt.
- ▶ OLAP steht für On-Line Analytical Processing, was eine Form von auf Datenanalyse ausgerichtete Datenbanken und Abfragen bezeichnet.

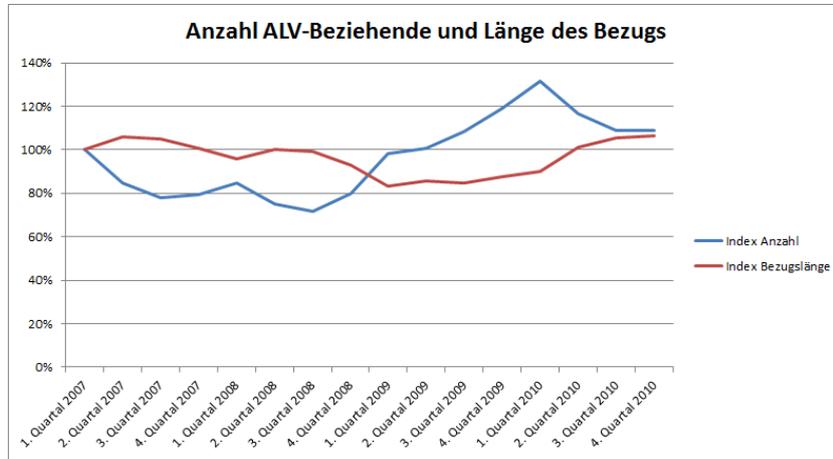
Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Monitoring-tool



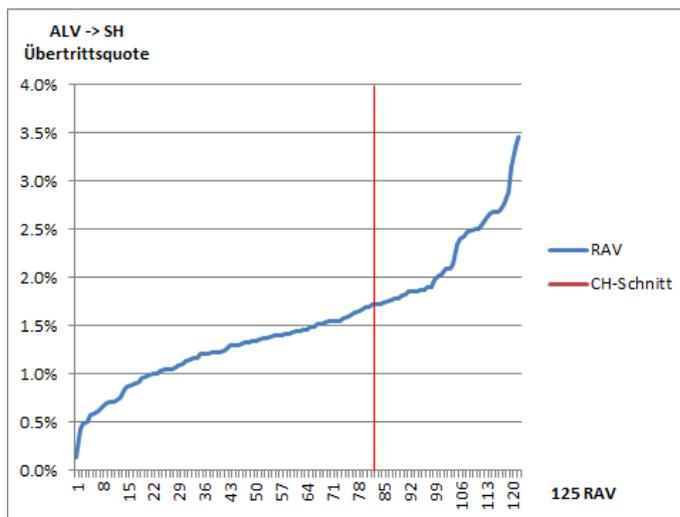
Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Monitoring-tool



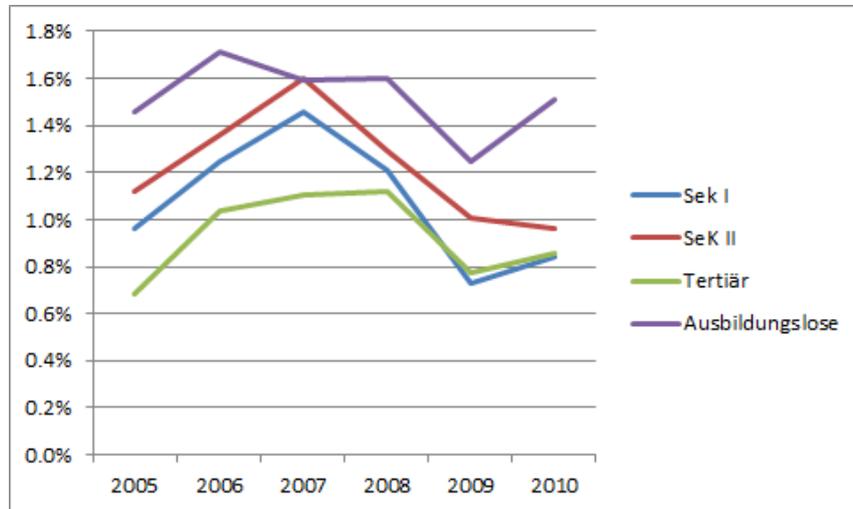
Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Monitoring-tool



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Monitoring-tool



Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences

Fragen

Fragestellungen der BFH :

- ▶ Wie beurteilen Sie die Auswertungen zum Data Mining bzw. zu einzelnen Indikatoren?
- ▶ Inwiefern scheint Ihnen das interaktive Abfrage-tool geeignet für ein Monitoring der Risiken in der ALV?

Mögliche weitere Verwendung:

- ▶ Erstellen von Abfragen aus dem Cube auf Bestellung SECO
- ▶ Nutzung der bestehenden Konzepte und des Abfragetools für Synergien mit weiteren Projekten im Bereich ALV (nachhaltige Erwerbsintegration, RAV-Evaluation)
- ▶ Erstellen einer gemeinsamen Plattform SECO-BFH zur Auswertung

Berner Fachhochschule | Haute école spécialisée bernoise | Bern University of Applied Sciences