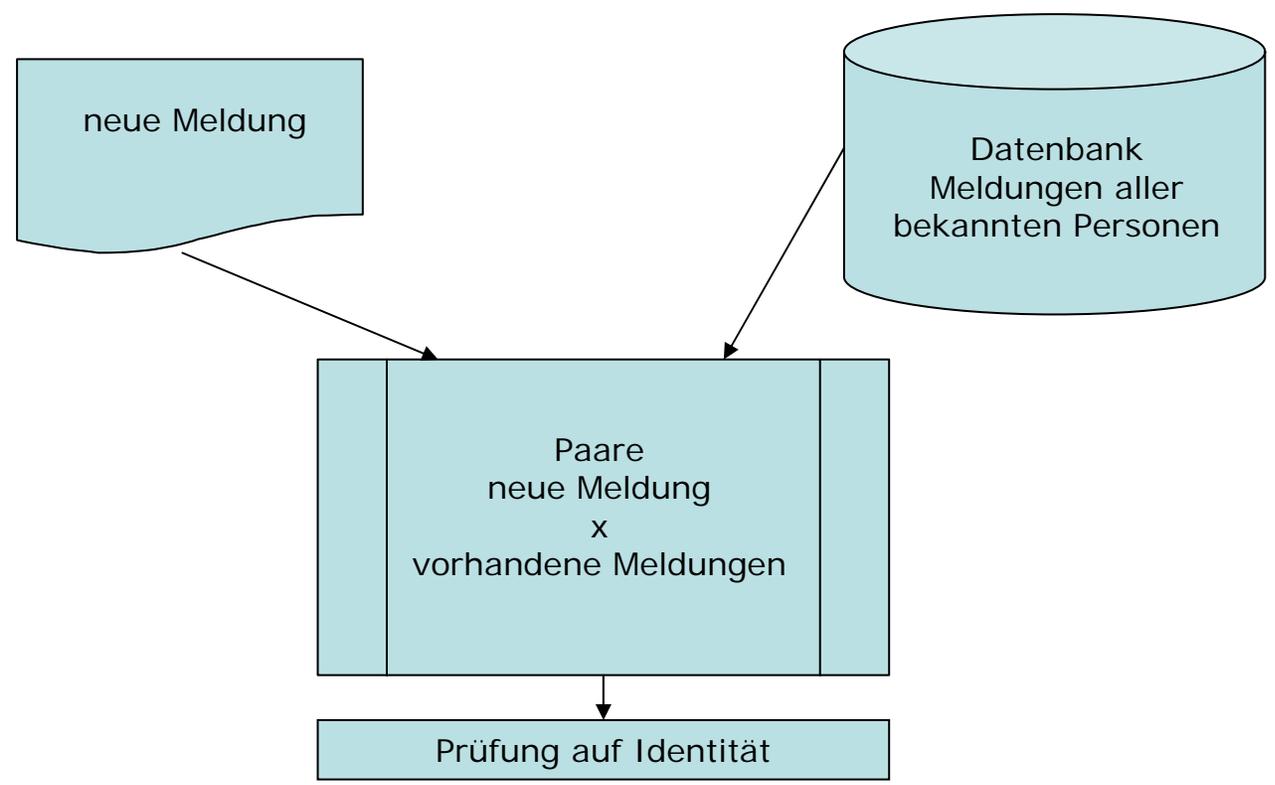


Pseudonymisierung und Record Linkage im EKR NRW

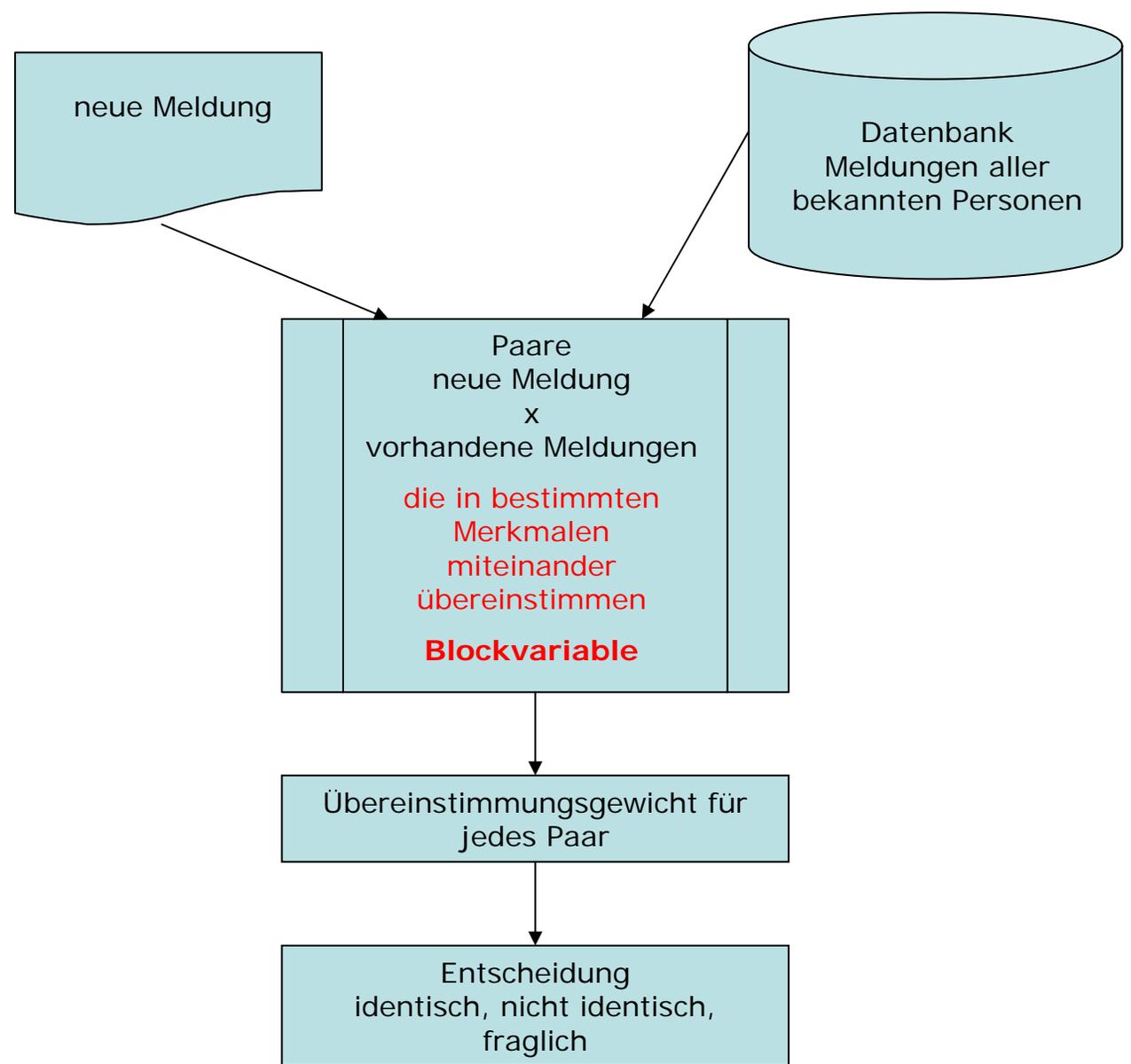
- Modifikationen in Nordrhein-Westfalen
- Record Linkage Verfahren
- Mammographie-Screening

- Meldungen sind nur elektronisch möglich
- Zerlegung, Normierung und Einwegverschlüsselung geschieht in der Meldestelle
- Symmetrische Verschlüsselung geschieht in einem speziellen Pseudonymisierungsdienst
- Java-Applikation in jeder Meldestelle
- Online-Verbindung von jeder Meldestelle

- Verzicht auf Kontrollnummern (BW1, BW2, DDR-Namenscode)
- Zusätzliche Kontrollnummern (Straßennamen, Hausnummern)



Bei 2.000 neuen Meldungen pro Tag
und 1.000.000 bekannten Meldungen
→
mehr als 2.000.000.000 Prüfungen pro Tag



Bei der Berechnung des Übereinstimmungsgewichtes werden berücksichtigt:

Matchvariable:

Name, Vorname, Geburtsname,
Geburtstag, Geburtsmonat, Geburtsjahr,
Geschlecht, Wohnort

In die Berechnung des Übereinstimmungsgewichtes gehen ein:

Wahrscheinlichkeit (m), dass identische Merkmalsausprägungen vorliegen,
wenn es sich um identische Personen handelt

Wahrscheinlichkeit (u), dass identische Merkmalsausprägungen vorliegen,
wenn es sich um verschiedene Personen handelt

Identität bzw. Nicht-Identität der Merkmalsausprägungen

Die Wahrscheinlichkeiten m können aus den vorliegenden Meldungen geschätzt werden
(identische Ausprägungen bei identischen Personen):

KRNL - empirische Ermittlung der Wahrscheinlichkeiten W_m

Ansicht

ende **Wm ermitteln** speichern

ID	Name	Meldungen	falsche	richtige	Wm	Datum
1	NAME	1.285.361	8.070	1.277.291	0,9937216	28.04.2008
2	VORNAME	1.362.730	16.230	1.346.500	0,98809008	28.04.2008
3	GEBURTSNAME	65.037	724	64.313	0,98886787	28.04.2008
4	GEBURTSTAG	1.254.257	4.260	1.249.997	0,99660356	28.04.2008
5	GEBURTSMONAT	1.254.387	2.891	1.251.496	0,99769528	28.04.2008
6	GEBURTSJAHR	1.254.387	1.902	1.252.485	0,99848372	28.04.2008
7	GESCHLECHT	1.254.387	61	1.254.326	0,99995137	28.04.2008
8	GEMEINDEKENNZIFFER	1.250.909	9.558	1.241.351	0,99235915	28.04.2008

Die Wahrscheinlichkeiten u können aus den vorliegenden Meldungen geschätzt werden
(identische Ausprägungen bei unterschiedlichen Personen):

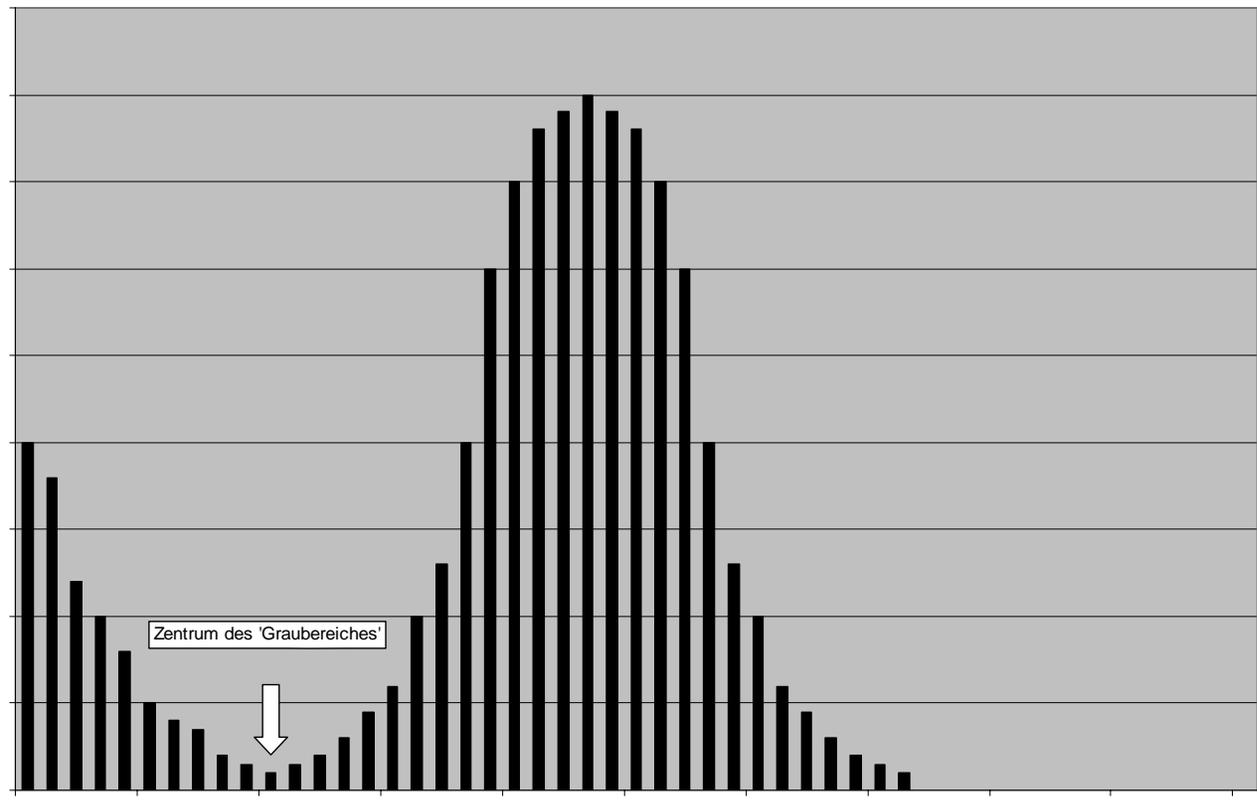
KRNW - Häufigkeiten der Matchvariablen ermitteln

Ansicht

ende start

ID	Name	Anzahl	Klassen	Wu-Datum
1	NAME	1.285.359	118.930	28.04.2008
2	VORNAME	1.362.728	14.877	28.04.2008
3	GEBURTSNAME	65.037	24.953	28.04.2008
4	GEBURTSTAG	1.254.255	3.761	28.04.2008
5	GEBURTSMONAT	1.254.387	12	28.04.2008
6	GEBURTSJAHR	1.254.387	124	28.04.2008
7	GESCHLECHT	1.254.387	3	28.04.2008
8	GEMEINDEKENNZIFFER	1.250.909	3.502	28.04.2008

Häufigkeitsverteilung des Übereinstimmungsgewichts



nicht identisch	fraglich	identisch
-----------------	----------	-----------

KRNW - Manuelle Nachbearbeitung des Record Linkage																					
Ansicht Extra																					
ende		Personengruppen einlesen					K	<	>	>	save	save & weiter	zurückstellen	drucken	27 von 287 [aktuelle Gruppe : 501]				ändern		
P	H	ID	Typ	N	VN	GN	J	GT	GM	GJ	S	STR	HNR	PLZ	Ort	Diag	ICD	Topo	Morph	Text	
1	D	446.535	PBF	A B A B			A	3	1944	M	A B A	57234	Wilnsdorf	03/2007	C61	C6191	82113			Makroskopie: 1. Prostata und Samenblase: 60 g	
1	D	446.535	IMO	A B A B			A	3	1944	M	A B A	57234	Wilnsdorf	01/2007	C61	C619	85503			Prostata-Ca. 2007: pT3a pN0(0/22) M0, G3, aAz	
1	D	446.535	IMO	A B A B			A	3	1944	M	A B A	57234	Wilnsdorf	01/2007	C61	C619	85503			Prostata-Ca. 2007: pT3a pN0(0/22) M0, G3, aAz	
1	D	446.535	IMO	A B A B			A	3	1944	M	A B A	57234	Wilnsdorf	01/2007	C61	C619	85503			Prostata-Ca. 2007: pT3a pN0(0/22) M0, G3, aAz	
1	D	446.535	IMO	A B A B			A	3	1944	M	A B A	57234	Wilnsdorf	01/2007	C61	C619	85503			Prostata-Ca. 2007: pT3a pN0(0/22) M0, LO, VO,	
2	D	502.204	PBF	C D AC D			A	3	1944	M	C D B	46147	Oberhausen	07/2007	C71.9	C719	94013			Befundbericht: Vorbefunde: N/07/1263 Klinische	
2	D	502.204	IMD	C D AC D C D			A	3	1944	M	C D B	46147	Oberhausen	07/2007	C71.9	C719	80003			Bösartige Neubildung: Gehirn, nicht näher beze	
3	D	515.033	PBF	E F AE F			A	3	1944	M	E F C	47623	Kevelaer	08/2007	D09.0	C679	81302			Makroskopisch: Ein 1,2 x 0,6 x 0,6 cm großes g	
4	D	627.422	IMO	G H A G			A	3	1944	M	GH I D	46397	Bocholt	06/2008	D09.0	C679	81302			Urotheliales Harnblasen-Ca.	
4	D	627.422	PBF	G H A G			A	3	1944	M	GH I D	46397	Bocholt	06/2008	D09.0	C672	81302			*** Begutachtung *** a+b I. a) Wir erhielte	
	N	2.768	IMO	I J A G			A	3	1944	M	J F E	45897	Gelsenkirchen	02/2008	C18.2	C182	81403			ED 01/08 Colon ascendens Adeno-Ca G2, pT3 pN2	

KRNW - Manuelle Nachbearbeitung des Record Linkage																						
Ansicht: Extra																						
ende		Personengruppen einlesen					K	<	>	>	save	save & weiter	zurückstellen	drucken	106 von 287 [aktuelle Gruppe : 932]				ändern			
P	H	ID	Typ	N	VM	GN	I	GT	GM	GJ	S	STR	HNR	PLZ	Ort	Diag	ICD	Topo	Morph	Text		
1	D	9.327	UEM	A B A B				A	12	1938	W	AB C A		45657	Recklinghausen	04/1986	C50.9	C509	85203	Mamma-Carcinom li. (T4,N2).Car-cino/8 cm gr.,ge		
2	D	102.109	IMO	C D A C				A	12	1938	W	D C B		45772	Marl	12/1996	C50.4	C504	85003	Mamma-Ca. re.		
3	D	199.837	IMO	E F A C				B	12	1938	W	E FC C		48231	Warendorf	02/2002	C16.3	C163	81403	Adeno-Ca. des Magenantrums (intestinaler Typ n		
3	D	199.837	PBF	E F A C				B	12	1938	W			48231	Warendorf	02/2002	C16.3	C163	81443			
4	D	359.817	IMO	G H A B				B	12	1938	W	G C DE		58313	Herdecke	02/2006	C18.0	C180	81403	Lokalisation: Zökum, einschl. Ileozökalklappe,6		
4	D	359.817	PBF	G H A B				B	12	1938	W	G C DE		58313	Herdecke	02/2006	C18.0	C180	84803	Klinischer Befund: Zoekum-Carcinom Morphologis		
5	D	366.311	IMO	I J A B				C	12	1938	W	H I F		44329	Dortmund	03/2006	D05.1	C509	85002	Lokalisation: Mamma o.n.A.,Histologie: Dukta		
5	D	366.311	IMO	I J A B				C	12	1938	W	H I F		44329	Dortmund	03/2006	D05.1	C509	85002	Lokalisation: Mamma o.n.A.,Histologie: Dukta		
5	D	366.311	IMO	I J A B				C	12	1938	W	H I F		44329	Dortmund	03/2006	D05.1	C509	85002	Lokalisation: Mamma o.n.A.,Histologie: Dukta		
6	D	386.198	PBF	K L A D				B	12	1938	W			40880	Ratingen	05/2006	C20	C209	81403	I: Makroskopisch ein eröffnetes, 17 cm langes		
6	D	386.198	IMO	K L A D				B	12	1938	W	J G		40880	Ratingen	05/2006	C20	C2092	81403	Diagnose: Rektum-Ca 8 cm ab ano, ED 05/06Neb		
6	D	386.198	IMO	K L A D				B	12	1938	W	J G		40880	Ratingen	05/2006	C20	C2092	81403	Diagnose: Rektum-Ca 8 cm ab ano, ED 05/06Neb		
6	D	386.198	IMO	K L A D				B	12	1938	W	J G		40880	Ratingen	05/2006	C20	C2092	81403	Diagnose: Rektum-Ca 8 cm ab ano, ED 05/06Neb		
7	D	390.951	PBF	M N A C				A	12	1938	W	K C H		33428	Harsewinkel	06/2006	C44.3	C4432	80903	10495/2006		
8	D	422.346	PBF	O P A B				D	12	1938	W	L M I		40211	Düsseldorf	12/2006	C34.9	C349	80003	Makroskopisch 3 bis vierstecknadelkopfgroße Ge		
8	D	422.346	PBF	O P A B				D	12	1938	W	L M I		40219	Düsseldorf	02/2007	C34.9	C349	82463	I: (LK Lig. pulmonale inferius N 2). II: Separ		
8	D	422.346	IMO	O P A B				D	12	1938	W	L M I		40219	Düsseldorf	01/2007	C34.3	C343	82463	Diagnose: Neuroendokrines Karzinom re. Lungenu		
8	D	422.346	IMO	O P A B				D	12	1938	W	L M I		40219	Düsseldorf	01/2007	C34.3	C343	82463	Diagnose: Neuroendokrines Karzinom re. Lungenu		
9	D	460.582	PBF	Q R A C				A	12	1938	W	N I J		45257	Essen	03/2007	C34.3	C343	80703	Befundbericht: Klinische Angaben: Plattene		
9	D	460.582	PBF	Q R A C				A	12	1938	W	N I J		45257	Essen	10/2007	C34.9	C349	80703	Befundbericht: Vorbefunde: E/07/7184, E/07		
9	D	460.582	IMD	Q R A C Q R				A	12	1938	W	N I J		45257	Essen	06/2007	C34.1	C341	80003	Bösartige Neubildung: Oberlappen (-Bronchus)		
9	D	460.582	IMD	Q R A C Q R				A	12	1938	W	N I J		45257	Essen	11/2007	C80	C809	80003	Bösartige Neubildung ohne Angabe der Lokalisat		
10	D	471.205	PBF	S T A B				B	12	1937	W	O P K		44143	Dortmund	06/2007	C50.9	C509	84803	Material: Mastastanzen li. - Screening - Patho		
10	D	471.205	IMO	S T A B				B	12	1937	W	O P K		44143	Dortmund	05/2007	C50.9	C509	85003	05/07 Mamma-Carcinom links pT2 pN0 M0 G2 L0 V0		
10	D	471.205	IMO	S T A B				B	12	1937	W	O P K		44143	Dortmund	05/2007	C50.9	C509	85003	05/07 Mamma-Carcinom links pT2 pN0 M0 G2 L0 V0		
10	D	471.205	PBF	S T A B				B	12	1937	W					06/2008	C50.9	C509	80103	invasives Ca anderer Spezieller Typ, Grading=1		
11	D	484.538	IMD	U V A B				B	4	1938	W	Q	L	44359	Dortmund	05/2007	C44.6	C4461	80913	Superfizielles Basaliom (Rumpfhautbasaliom) li.		
12	D	546.747	PBF	W X A C				A	12	1938	W	R C M		42853	Remscheid	12/2007	C17.2	C172	82463	*** Mikroskopie *** *** ERGÄNZENDE DIAGNOSE *		
12	D	546.747	PBF	W X A C				A	12	1938	W	R C M		42853	Remscheid	12/2008	C44.5	C4453	80903	Klinische Verdachtsdiagnose: BCC C44.9 Die		
13	D	549.713	PBF	Y Z A C				A	12	1938	W	S T F		44532	Lünen	11/2007	C45.0	C384	90503	Befundbericht: Klinische Angaben: Zytologi		
13	D	549.713	IMD	Y Z A C Y Z				A	12	1938	W	S T F		44532	Lünen	11/2007	C45.0	C384	90503	Mesotheliom der Pleura		
14	D	592.254	PBF	a b A C				B	12	1938	W	U	N	56640	Iserlohn	02/2008	C44.6	C4463	80913	Hautexcidat: Oberarm li. Klinisch: Basaliom		
15	D	608.704	PBF	c d E F				A	12	1938	W	V WC C		40229	Düsseldorf	03/2008	C34.0	C3402	80713	Makroskopisch I: Vier stecknadelkopfgroße		
16	D	621.552	PBF	e f A B				A	12	1937	W	X Y O		40476	Düsseldorf	05/2008	C20	C209	84803	Eingesandt wurden: I. Eine 2,3 cm lange und m		
16	D	621.552	IMO	e f A B				A	12	1937	W	X C O		40476	Düsseldorf	05/2008	C20	C2091	81403	Diagnose: Fortgeschrittenes muzinöses A		
	N	3.560	IMD	I J A B				A	12	1938	W			40549	Düsseldorf	11/2008	C44.3	C4431	80973	BCC vom nodulären und nodulozystischen Typ der		
	N	3.697	PBF	g h A B				B	12	1938	W	Z a H		44329	Dortmund	11/2008	C50.9	C509	90203	Material: Mastatumor links #10807852# Pathol		

Zuordnungsfehler

Synonymfehler:

Für eine neue Meldung, die sich in Wahrheit auf eine bekannte Person bezieht, wird fälschlicherweise eine neue Person erzeugt.

Homonymfehler:

Eine neue Meldung, die sich in Wahrheit auf eine bisher unbekannte Person bezieht, wird fälschlicherweise einer bekannten Person zugeordnet.

Zuordnungsfehler

Untersuchung zu Meldungen des Jahres 1998

Datenbank: 101.880 Personen mit 174.815 Meldungen

Im Laufe des Jahres 1998 wurden 27.262 neue Meldungen verarbeitet:

Manuelle Zuordnung unter Verwendung aller Klartexte

Halbautomatische Zuordnung unter Verwendung der
Kontrollnummern
mit manueller Nachbearbeitung

Ergebnis:

Homonymfehlerrate: 0,36 %

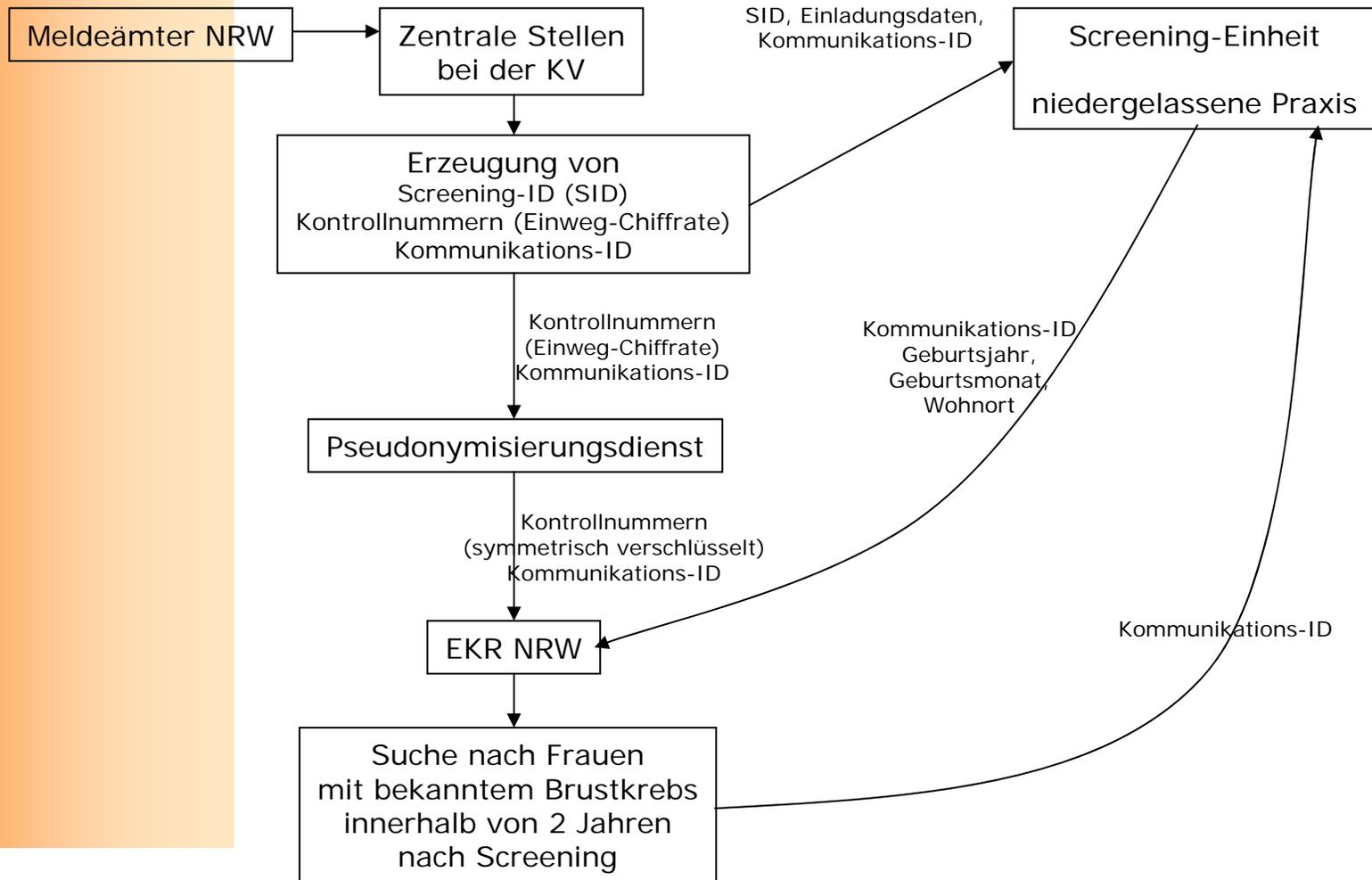
Synonymfehlerrate: 1,81 %

www.krebsregister.nrw.de

volker.krieg@krebsregister.nrw.de

Mammographie-Screening

- Jede Frau im Alter von 50 bis unter 70 Jahren kann alle zwei Jahre eine Mammographie zur Frühdiagnose einer bösartigen Neubildung der Brustdrüse wahrnehmen
- Ein wichtiger Qualitätsindikator zur Einschätzung der Effektivität des Mammographie-Screening-Programmes ist die Häufigkeit der so genannten „Intervall-Karzinome“
- Wie oft kommt es vor, dass eine Frau am Mammographie-Screening teilnimmt, eine negative Diagnose erhält und im Zeitraum bis zur nächsten Untersuchung eine bösartige Neubildung der Brustdrüse entwickelt?
- Die Ermittlung der Häufigkeit von Intervall-Karzinomen ist nur in den Krebsregistern möglich



Übereinstimmungsgewicht

Ermittlung des Übereinstimmungsgewichts

Das Übereinstimmungsgewicht ist die Summe aller Übereinstimmungsgewichte der Matchvariablen

$$UEG = \sum_{i=1}^n G_i$$

mit

n = Anzahl der Matchvariablen

G_i = Übereinstimmungsgewicht für jede Matchvariable

Ermittlung des Übereinstimmungsgewichts

Sind die Ausprägungen der beiden Matchvariablen identisch, so berechnet sich das Übereinstimmungsgewicht nach

$$G_+ = \frac{\ln\left(\frac{m}{u}\right)}{\ln(2)}$$

Wenn die beiden Ausprägungen identisch sind, gilt immer **$u < m$**

Deshalb ist G_+ immer **positiv**.

Ermittlung des Übereinstimmungsgewichts

Sind die Ausprägungen der beiden Matchvariablen nicht identisch, so berechnet sich das Übereinstimmungsgewicht nach

$$G_{-} = \frac{\ln\left(\frac{(1-m)}{(1-u)}\right)}{\ln(2)}$$

Wenn die beiden Ausprägungen nicht identisch sind, gilt immer

$$1-u > 1-m$$

Deshalb ist G_{-} immer **negativ**.