

Data security conform evaluation of the Mammography Screening



A. Thiel¹, S. v. Gehlen²,
I. Urbschat³, J. Kieschke³, P. Jensch^{1,2}



www.mammographie-screening.org



¹ Carl von Ossietzky Universität, Oldenburg,

² OFFIS, Oldenburg;

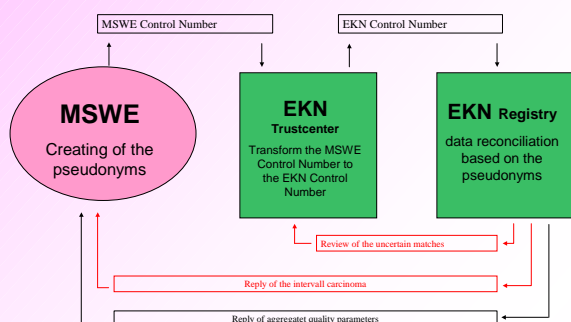
³ Epidemiologisches Krebsregister Niedersachsen (EKN), Oldenburg, Hannover



Purpose

For the quality assurance in the mammography screening project in the Weser-Ems region (MSWE) the interval cancer rate, the breast cancer incidence rate and the cancer mortality rate are very important parameters and can only be identified by the epidemiological cancer registry in the Lower Saxony (EKN). In the initial phase of MSWE the comparison between screening data and the data of the EKN was planned. For data security reasons the control number system of the EKN has to be used.

Dataflow



Methods

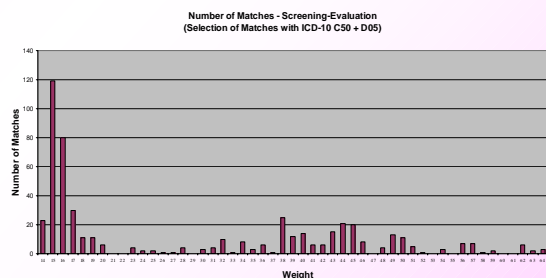
For every woman (aged between 50 and 69) attending the screening (12.913 women, ~65% of the target population) a hash code will be generated and encrypted according to the control number system of the EKN. This system standardises the way of writing the attributes (forename, name, birthdate and further more) and added phonetic codes. Due to security reasons this pseudonym is created in the MSWE.

The alignment with the EKN database took place over a probabilistic record linkage procedure. It was examined whether the samples of two data records speaks rather for or against the affiliation of the data records to a person. The computation of the weights took place on the basis of the characteristics of the attributes. For example: a match of the birthdate has a heigher weight as a match on the phonetic code of the forename. These single weights are summarized. Due to empirical values adjusting below an alignment weight of 14 no person agreement was accepted; for a weight above 36 an automatic unification took place. For the range between further clarification took place interactively.

Results

The comparison of the 12.913 screend women with the aprox. 280.000 records in the EKN took about 4 ½ h. Between 0,03 - 1% of all screened women were uncertain matches, which require a further clarification (4 - 119 women). Messages with a match weight under 20 practically never belong together according to experience (115 cases). Only 4 cases (0,03 %) exhibited a weight between 20 and 25, for these a mapping based on the pseudonyms is not possible.

Linkage Distribution



Linkage Result

Matches with ICD-10 C50+D50 (weight > 14)	523
Chronological plausibility given	401
Safe match given (weight > 36 or interactiv identified)	244
Uncertain Matches	4 - 119
- weight between 20 and 25	4
- weight under 20	115
Interactive identified mismatches	38
Intervall carcinoma found	6

Conclusions

First effects of the mammography screening are promptly evaluable by alignment with population-based cancer registers. With the use of the control number system of the epidemiological cancer registers the alignment on pseudonyms under observance of data security-legal regulations is feasible. The sensitivity record linkage procedures should be evaluated at a larger screening population, whereby also the effect of a lowering of the error tolerance is to be examined for the decrease from uncertain matches.



kindly supported by

Deutsche Krebshilfe