Since technology is advancing and storage capacities are growing, it is possible to process larger amounts of data. Meanwhile, there is the desire to generate useful knowledge from this aggregated data, which is not found explicitly or can be identified at once in the data. PaDaWaN, a scalable, digitized, double-pseudonymized, daily updated Data Warehouse provides the opportunity to investigate specific problems.

By selecting and defining the study cohort “incisional hernia or primary ventral hernia” it is possible to validate the Data-Warehouse using information from the EuraHS database (n=510). For that purpose risk factors and intra- and postoperative complications are generated from SAP and textual diagnostic findings. Congruence between the information extracted from the different data sources (regarding the exactly same patients) was calculated using the kappa coefficient (IBM SPSS Statistics 23).

Potential discrepancies between patients’ records and medical reports can be found. For no item a kappa greater than 0.8 was found (good correlation: 0.6-0.8; very good correlation: 0.81-1.0). For Bleeding the kappa was 0.67, for ascites 0.49 (medium correlation), for surgical site infection 0.34 (weak correlation) and for intraoperative complications 0.24 (weak correlation). Early results, using PaDaWaN’s screening tools for specific requests of established risk factors and complications, show accordance with published data.

By establishing Data Warehousing as a clinical research platform, data can be structured and generated faster in the future. The dynamic daily automated data update enables clinical staff to validate and evaluate treatment concepts and results more easily. Furthermore recommendations for future medical reports can be given in order to improve information extraction of Data Warehousing. The results also show, that there is a great discrepancy between data regarding one and the same patient, depending on the intention of data collection. The correlation between administrative data (generated for purposes of health care insurance) and clinical data (generated for communication between doctors) needs to be improved (mean kappa: 0.4).