

Interactive Visual Analysis of Image-Centric Cohort Study Data

P. Klemm¹, S. Oeltze¹, K. Lawonn¹, K. Hegenscheid², H. Völzke², B. Preim¹

Abstract—ToDo: Reviewer-Feedback: The abstract should be formulated more concisely and focus on What? Why? How? What follows? Epidemiological population studies impose information about a set of subjects (a *cohort*) to characterize disease-specific risk factors. Cohort studies comprise of heterogeneous variables (*features*) describing the medical condition as well as demographic and lifestyle factors and more recently, medical image data. We propose an Interactive Visual Analysis (*IVA*) approach that enables epidemiologists to examine both image-based as well as non-image data, e.g., sociodemographic features and attributes derived from the image data. This is achieved by combining brushing and linking enabled coordinated information visualization views and interactive 3D shape renderings with epidemiological data representations such as pivot tables and key figures as association measures. The presented concepts are applied by experts to gather and evaluate hypotheses about the aging process of the lumbar spine. It allows a more flexible comparison between image and non-image data. The new framework enables hypotheses validation and generation by incorporating human pattern recognition as well as data mining methods. Using all reliable information from the image segmentation linked to non-image features aims to unveil *associations* by applying an iterative analysis approach.

Index Terms—Interactive Visual Analysis, Epidemiology, Spine

1 INTRODUCTION

Epidemiology aims at characterizing health and disease by determining risk factors. Clinical problems, such as the selection of diagnostic tools and efficient treatment, are tackled using results of epidemiological research. Also, the introduction of preventive measures in medicine and beyond is based on epidemiological research, where, for example, subgroups with increased risk are identified [12]. Observations made by clinicians in the daily routine are translated into hypotheses for epidemiological research. These are used to determine environmental and lifestyle factors as well as medical examination results that may influence a disease. Potentially useful data variables (*features*) are gathered using structured interviews and clinical examinations. Methods like regression analysis are employed to check the attribute list for statistical soundness.

Longitudinal population-based studies, such as the Study of Health in Pomerania (SHIP) [42], gather as much information as possible about a defined sample of people (a *cohort*). The cohort consists of several thousands of people who are randomly selected to avoid any bias. The subjects are selected without focus on a certain disease. A large size of the cohort is therefore essential to investigate differences between healthy and diseased people. Cohort studies often include medical image data. Segmentation of these data enables an anatomical structure analysis and an analysis of possible correlations with other features. Semi-automatic techniques are more promising but also challenging, since the employed modalities, such as magnetic resonance imaging (MRI) and ultrasound, are subject to inhomogeneity and noise. Analyzing spatial data with respect to other epidemiological factors requires techniques that reach beyond standard statistical methods.

Compiling a list of features for tests of statistical resilience based on experience-driven hypotheses leaves out other features in the data which potentially interact with a disease. This also applies to the chosen landmarks used to quantify medical image data information. The standard workflow lacks methods, which identify correlations that the epidemiologists did not consider. Also, only a small subset of variables can be concurrently analyzed, depending on the statistical measure.

We propose an Interactive Visual Analysis (*IVA*) approach [38] for the combined analysis of image and non-image data. Visual queries and direct feedback of Visual Analytics systems allow for a fast exploration of the data space incorporating many different variables. Intended as an extension to the well-established epidemiological tools it provides a way to rapidly validate hypotheses and to trigger *hypothesis generation* using data mining methods, such as clustering. *Hypothesis generation* gains importance since the number of epidemiological features increases and the focus shifts towards more complex relations involving more than two features.

Our contributions are:

- an Interactive Visual Analysis workflow for cohort study data to allow both, hypothesis-driven analysis and hypothesis generation—based on a characterization of the standard epidemiological workflow,
- visualization techniques, which incorporate both information visualization and 3D rendering of organ shapes as well as combining them with epidemiological graphics and key figures,
- highlighting interesting subject groups and feature associations using shape-based clustering and statistical contingency measures,

We applied our approach to a data set compiled to analyze diseases related to the lumbar spine and aim to determine features, which indicate pathological changes. This data set comprises 127 features and 2 sequences of MRI data from 2,333 patients. We implemented the presented methods using modern web technologies, such as WebGL, D3.js and NodeJS, making them easily accessible for the domain experts. This aims to trigger a fast feedback loop between us and the epidemiologists.

2 EPIDEMIOLOGICAL BACKGROUND

In this section we describe the epidemiological workflow and associated requirements.

2.1 Epidemiological Workflow

The diversity of epidemiology is reflected in the different experts who work at cohort studies, ranging from specialized doctors to medical computer scientists with focus on biometrics, and statisticians. Epidemiologists follow a workflow mainly driven by statistic tools to validate hypotheses about disease-specific risk factors. Following Thew et al [37], the workflow can be characterized as follows:

• ¹Otto-von-Guericke University Magdeburg, Germany.

• ²Ernst-Moritz-Arndt-University Greifswald, Germany.

Manuscript received 31 March 2013; accepted 1 August 2013; posted online 13 October 2013; mailed on 4 October 2013.

For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

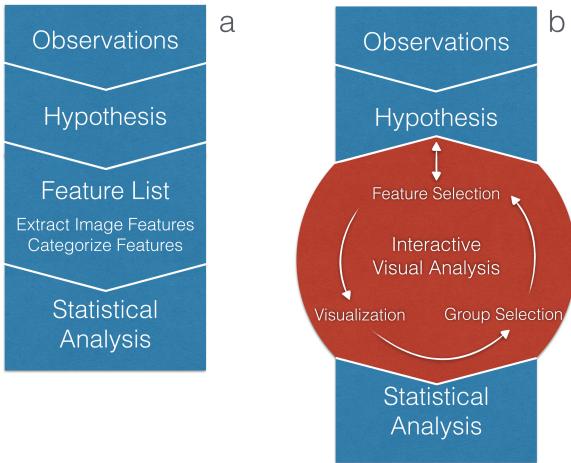


Fig. 1. IVA tools complement parts of the epidemiological workflow instead of replacing them. The appropriate combination of statistical and interactive-driven analysis shows promising potential to unveil the information in the data. The standard epidemiology workflow consists of four steps (a). IVA complements this workflow and is used to elaborate the feature selection (b). The iterative red highlighted part is called the *IVA Loop* and is described in more detail in Figure 2.

1. A hypothesis is derived from observations made by clinicians in their daily routine.
2. A set of features depicting conditions affected by the hypothesis is compiled accordingly.
3. Confounding features are identified and taken into account (for example using stratification).
4. Statistical methods, such as regression analysis, assess the association of selected features with the investigated disease.

The workflow is shown in Figure 1 (a) and serves as orientation for our approach although it does not consider image data. We focus, on the potential of image data and attempt to support *hypothesis generation*.

Reproducibility of results is an epidemiological key requirement. It is difficult to achieve, since many physicians are involved when thousands of test persons are examined and interviewed. Thus, both intra- and inter-observer variability needs to be low for all aspects of a cohort study examination. Longitudinal studies require the acquired features to be comparable for evaluation.

In longitudinal cohort studies, grouping subjects using epidemiological features is essential in order to allow per-group risk determination. Grouping depends on the underlying hypothesis. Age for example is divided into groups (e.g. in 20 year steps) when investigating its influence on a condition. These groups strongly depend on the condition of interest and therefore there is no standard for their categorization.

To determine whether a subject is prone to be affected by a certain disease, *relative risks* are determined including confidence intervals that indicate the certainty of that feature being a risk factor. Statistical correlations are prone to *confounding*, meaning that the association of two features is influenced by a third feature, which needs to be isolated. A famous example for a confounder is the association between shoe size and mortality, where it can be observed that people with larger shoe size have a smaller life expectation than people with small shoe size. The shoe size is actually associated with gender, where women have smaller feet than men but also a longer life expectation. *An extreme form of confounding is called Simpson's Paradox [22], where the confounder reverses the observed effect between two features. A famous example is the Berkeley University gender bias case, where there were significantly less women admitted than man. Analysis later showed that women applied mostly for strongly*

restricted subjects and when only looking at acceptance divided by departments there was actually a small bias toward women.

Statistical tools such as SPSS¹ play a major role for analyzing epidemiological data. Graphic data representation is largely used to present results rather than gaining insight.

2.2 Epidemiological Data

Access to epidemiological data is limited and usually controlled by ethics committees. Large studies, such as the SHIP require written proposals reasoning about each requested features and the expected outcome. Since the number of features and modalities is steadily increasing, the epidemiological community is aware of the lack of specialized analysis methods, making an acceptance more likely. Due to the high sensibility of the data, all identity revealing information of a person, such as name or place of residence is anonymized.

Epidemiological data are highly heterogeneous and incomplete. Information about medical history and examinations, genetic conditions, geographical data, questionnaire results and image data yield a complex data space for each subject. For ethical or medical reasons some features cannot be gathered for each subject. The most obvious example are women-specific questions about menstrual status or number of born children. Follow-up examinations or questions about conditions like medications taken after a diagnosed disease also yield features only available for a small amount of subjects.

Indicators for medical conditions as well as questions about a subject's lifestyle are often *dichotomous*—they have two manifestations (*Yes or No*). Dichotomous data can also be derived by aggregating features to yield only two manifestations (e.g. subjects younger or older than 50 years). Medical examinations mostly comprise categorical (e.g. levels of back pain) and continuous values (e.g. age or body size). *Analysis of the data is usually carried out by calculating correlations, which is challenging due to the data type heterogeneity. Parameter correlation can also be associated with confounding, which can not be automatically predicted but rather has to be judged by a domain expert. Sparse populated features are also hard to assess statistically as to few data samples have a higher chance to distort the real underlying distributions.*

2.2.1 Image acquisition

Imaging techniques involving ionizing radiation for the subject are not suitable for ethical reasons. Therefore, MRI is the main method for collecting cohort study imaging data. The image quality is a tradeoff between accuracy and affordability [32]. This often yields image resolutions inferior to those of clinical day-to-day practice, which makes their analysis more challenging. The equipment used to gather medical image data is not updated or changed during a longitudinal study to ensure comparability in and between acquisition cycles.

2.2.2 Image analysis

Take up huge amount of space! Decisions have to be made about *comparison* and *quantification* of image data. Segmentation masks representing the voxels of an anatomical structure would be ideal, since many different key figures, e.g., volume, largest diameter or aspect ratio, can be derived from them. Since reliable and efficient segmentation techniques are not available in general, epidemiologists are forced to measure the data by hand, which is a very tedious work with respect to the number of necessary landmarks and the number of subjects. Information derived by landmarks, such as top and bottom point of a vertebra, are by far not as expressive and versatile as segmentation masks describing its whole shape. They are also prone to a high inter-observer variability and are hard to reproduce. This gains even more importance when analyzing multiple time steps. Morphometric information from landmarks comprises thickness, diameter or length of a structure as well as grey value distribution in an area (used for determining the type of tissue).

¹Product of IBM; <http://ibm.com/software/analytics/spss/>

2.3 The Study of Health in Pomerania (SHIP)

After the pioneering Rotterdam study (started in 1990), several MR imaging study initiatives were initiated. They slightly differ in clinical focus, acquired data and epidemiological research questions. Starting in 1997 with a cohort of 4,308 subjects, the SHIP, located in Northern Germany, aims to characterize health and disease in the widest range possible [42]. Data is collected without focus on a group of diseases. This allows the data set to be queried regarding many diseases and conditions. Subjects were examined in a 5-year time span, continuously adding new parameters including MRI scans in the last iteration [18]. The MRI protocol features a rich number of sequences. A second cohort SHIP-Trend was established in 2008. The protocols for examining the subjects between SHIP and SHIP-Trend remained the same, making them comparable. The overall examination time for each person attending the study is two days.

3 PRIOR AND RELATED WORK

Reviewer 1: The related work section includes merely a set of related tools. I would have liked to see a more comprehensive section that discusses the previous related work in more detail and describes why they are important and relevant to this work and how your work differs or improves upon theirs.

Quality: The work contains many references. But there are potentially some missing; e.g., the framework talks about utilizing multiple coordinated views (although the details of what is actually linked is vague) and therefore cited papers such as Baldonado's work in AVI 2000, or Roberts 2007 on Coordinated multiple views, or Weaver's Cross-filtered Views for multidimensional visual analysis 2010 should be cited. The paper really proposes a framework; however it was unclear how this was implemented into the presented system. Therefore this should be improved.

Reviewer 2 - I would suggest to also cite the original Generalized Pairs Plot and not only the adaptation as GPLOMS [21]: J. W. Emerson and W. A. Green. *gpairs: The Generalized Pairs Plot, 2012. R package version 1.1.*

- A related process model for Visual Analytics that focuses on hypotheses and models: T. Lammarsch, W. Aigner, A. Bertone, S. Miksch, and A. Rind, Towards a Concept how the Structure of Time can Support the Visual Analytics Process, in Proceedings of International Workshop on Visual Analytics (EuroVA 2011) in conjunction with EuroVis 2011, Goslar, Germany, 2011, pp. 912.
- VISITORS is related work in visualization of medical variables of cohorts: D. Klimov, Y. Shahar, and M. Taieb-Maimon, Intelligent visualization and exploration of time-oriented data of multiple patients, Artificial Intelligence in Medicine, vol. 49, no. 1, pp. 1131, May 2010.
- another, forthcoming related paper: P. Angelelli, S. Oeltze, C. Turkay, J. Haasz, E. Hodneland, A. Lundervold, A. Lundervold, B. Preim, and H. Hauser, Interactive Visual Analysis of Heterogeneous Cohort Study Data, IEEE Computer Graphics and Applications, vol. Early Access Online, 2014.

Reviewer 3 What the prior work section needs at the end is a paragraph stating how their work differs from the work they summarise i.e. it uses more non-image data than Turkay and uses different visualisations, and more importantly differentiation from Steenwijk et al. very similar work (saying they use a different meaning for the word cohort does not seem justification enough, given the large overlap), only then can we be sure there is a novel contribution here. Given the number of features of epidemiological data sets and their different manifestations, the strength of different visualization techniques needs to be combined [4, 25]. The principal component analysis (PCA) and similar techniques are able to reduce the dimensions by extracting most expressive components, but make the influence of each variable hard to convey.

Turkay and colleagues [40] focus on *hypothesis generation* using parallel assessment of multiple data features makes their work

closest to ours albeit our emphasis is on processing models derived from medical image data segmentation and variables with categorical manifestations. Their methods aim to amplify a hypothesis generation process for analyzing data of a Norwegian aging study. Statistical measures of continuous features, such as mean, standard deviation, skewness, or inter-quartile range, are used to create *dimension plots*. These transform dimensions into data points and make them comparable with respect to the derived descriptive measures, making them comparable. The method is strongly dependent on the descriptive measures of the epidemiological factors.

Hypotheses based on observations of changes in these plots may impose *overfitting* to the data because the measures highlight only parts of the statistical changes. Our approach sticks more to the information extracted from the segmented image data and variable associations with non-image epidemiological factors.

Visualizing Image and Non-Image Data. Gresh et al. [16] proposed WEAVE, one of the first systems which concurrently analyzed medical image and non-image data using linked views. Blaas et al. [2] presented a similar system which analyzed medical image data and variables derived from them using views from the feature and physical space. They incorporated data mining methods such as dividing the data space by using a k-nearest neighbor technique and the PCA. Steenwijk et al. [36] employ a relational database to organize the data for visualizing subject data using linked views such as parallel coordinates, scatterplots and time plots. Zhang et al. [44] provide a web-based system for analyzing subject groups with linked views and batch-processing capabilities for categorizing new subject entries into the data set. Their understanding of a cohort differs from the understanding of the term in an epidemiological context by denoting every parameter-divided subject group as individual cohort.

Visualizing Heterogenous Non-Image Data. Generalized Pairs Plots (GPLOMS) are an information visualization technique comparing heterogenous features pairwise using a plot-matrix grouped by type [21]. They are useful to gain an overview over numerous variables and their distributions. Histograms, bar charts, scatterplots and heat maps are used to visualize variable combinations with regard to their type. Dai et al. [10] explored risk factors by incorporating choropleth maps of epidemiological features (e.g., mortality rates in a region) with parallel coordinates, bar charts and scatterplots with integrated regression lines. Their findings yielded a *Concept Map*, which linked cancer-related associations via graph edges. Chui et al. [7] visualized associations in time-dependent epidemiological data using time-series plots highlighting risk factor differences in age and gender.

Commercial Data Visualization. Commercial systems such as Tableau² or Spotfire³ provide a rich user interface that enables a Visual Analytics approach. With little effort, linked views can be created, but the data processing possibilities, such as derivation of new variables or the 3D rendering capabilities, are very limited. These systems with their origin in business intelligence applications do not support epidemiological workflows well, in particular workflows that also consider medical image data.

Visualizing Shape Variance. Comparing tissue between many subjects requires shape variance visualizations. Caban et al. [6] investigated the suitability of variance visualizations of shape distribution models and concluded that users favor spherical glyph representations over deformation grids and likelihood volumes. The distribution of shapes in a space derived from a PCA is plotted by Busking et al. [5] in a 2D-projected plane of the space. Interpolated views can be created by the user in a separate view as well as comparisons in a contour view. Interpolation is carried out by mesh morphing. The distance to the mean shape is color-coded. We incorporate the idea of combining 3D shape rendering with information visualization techniques. Hermann et al. [20] identify local deformation changes by investigating shape-related differences on rodent mandibles. User-specified regions

²Owned by Tableau Software; <http://tableausoftware.com>

³Owned by TIBCO; <http://spotfire.tibco.com>

of interests are mapped to associated anatomic covariation using tensor visualization. This method enables rapid hypotheses validation and was able to reproduce textbook knowledge.

SHIP Data Analysis. Klemm et al. [24] visualized lumbar spine variabilities based on a semi-automatic shape detection algorithm of 490 participants of the SHIP-2 cohort. Hierarchical agglomerative clustering divided the population into shape-related groups. As proof of concept, a relation between the size of the segmented shape and measured size of the subjects was shown. This work focuses on incorporating these derived data as new features of the overall data set, making it possible to include it into the hypothesis validation and generation process. When applying clustering techniques to the non-image data it was found that *k*-Prototypes and DBSCAN are appropriate in the epidemiological context, but is strongly dependent on the chosen variables and distance measure [23]. Niemann et al. [28] presented an interactive data mining tool for the assessment of risk factors of hepatic steatosis, the fatty liver disease. Association rules created by data mining methods can be analyzed interactively with their tool and highlight potentially overlooked features.

Interactive Visual Analysis The strength of the *IVA* approach described in the next section is its versatility with respect to the application field [25]. Oeltze et al. [29] combined a linked view representation of results from a statistical analysis with feature localizations of the tissue perfusion. This was focused on individual patients and not on cohort study data.

While we take similar steps when analyzing the data, such as employing statistical tests, our data are not associated to a geographical or temporal context, which affects the techniques described in the following section.

4 IMAGE CENTRIC COHORT STUDY DATA IN INTERACTIVE VISUAL ANALYSIS CONTEXT

We described the epidemiological workflow followed by domain experts where reproducibility and statistical integrity are important aspects (recall Subsection 2.1). Figure 1 (a) shows this workflow as a consecutive series of steps. Introducing the *IVA* principle to the epidemiological domain aims to compensate the weaknesses of the existing workflow, rather than replacing it. In the current state, the workflow treats the data like a black box. Statistical tests on features associated to a hypothesis yield a value for deciding whether the data supports the hypothesis or not. Features not included in the analysis may potentially support the chosen hypothesis by discriminating the population in the expected way, but are not highlighted in any way. This becomes even more important when the workflow is adapted to the analysis of the medical image data, where domain experts annotate landmarks tediously to derive measures, such as diameters. This leaves out the majority of information in the image data by abstracting it to single values. It is easily possible that omitted information would heavily influence the result. Considering more complex parts of the data would make those results more trustworthy and also could identify possible anatomical confounders—an epidemiological research result in itself.

IVA tries to illuminate the black box by making the domain experts part of an iterative feature selection process (see Fig. 1 (b)). It also aims to project back into the hypothesis formulation step to amplify hypothesis generation. This has to be handled with care, since *overfitting* of expectations to the data is an imminent danger [40].

Domain and Range Variables. In the *IVA* context, data are characterized by a combination of independent variables, such as space and/or time, and dependent variables, like temperature or pressure. Two kinds of views are employed to inspect the data:

- *physical views* [30], such as direct volume rendering, show information in the context of the spatio-temporal observation space [29], while
- *attribute views*, such as scatter plots and parallel coordinates, show relationships between multiple data attributes.

Transferred to epidemiological data, the residential area of cohort subjects could be interpreted as *space*, the different assessment cycles of a longitudinal study as *time*, and the image and non-image data as *dependent variables*. However, our current work neglects geographical and temporal aspects. Instead we employ a more abstract model and consider the subjects as living in a joint image space where each of them is represented by a particular segmented organ or structure. For instance, the lumbar spine is segmented over all subjects and all lumbar spines are co-registered spanning a joint space. Then, two types of dependent variables exist:

1. the socio-demographic data and medical examination results and
2. variables derived from the segmented structures, e.g., spinal curvature or misalignment of the vertebrae.

An alternative of the image space would be the shape space generated by extracting the major modes of variation from all segmentation results [5]. Based on our abstract model, the three analysis patterns of *IVA* can be employed.

Local Investigation: refers to the inspection of dependent variables with respect to certain subsets of the image or shape space. For instance, the epidemiologist selects several lumbar spines with a common characteristic in the image or shape space and wishes to inspect the associated dependent variables in an attribute view [20]. The selection step requires dedicated interaction techniques for defining a subset. Alternatively, derived shape-related variables opposed in an attribute view or automatic techniques for shape clustering may be employed [24]. Clustering algorithms can be used to investigate associations between shape groups and other non-image variables. Analysis of outliers can indicate segmentation errors or a group of subjects sharing a pathology.

Feature Localization: refers to the search for structures in the image or shape space with a defined characteristic. The epidemiologist may be interested in all female subjects with lower back pain and wishes to see the corresponding spines in a physical 3D view.

Multivariate Analysis: refers to an investigation of multi-variate properties of the dependent data by specifying a feature in one attribute view and at the same time analyzing the value distribution with respect to other variables in other attribute views. Epidemiologists may define a feature in a scatter plot of the body mass index (BMI) and age to inspect the result in a histogram of body height. These associations may also be summarized using pivot tables which are widely used in epidemiology.

4.1 Data Preprocessing

A number of transformations are carried out in the data preprocessing stage.

Non-Image Data. Data obtained using questionnaires or medical tests are often stored using statistical packages such as SPSS, which have a proprietary data format with limited export capabilities. Exporting the data in the respective tool to a CSV file and then converting it to file types that are easily manageable, such as JSON or XML, makes it readable for all modern programming languages. A data dictionary stores information about each manifestation of a feature. Detailed description of data variables, its meaning as well as unit of measurement are stored as a lookup table. Missing data are denoted using error codes indicating their cause ranging from ethical to medical and personal issues (recall Subsection 2.2).

Image Data. Information about anatomical structures, such as diameter or volumes, is extracted from the image data. This is either done manually by experts setting landmarks (sometimes supported by algorithms connecting the landmarks such as graph cuts [15]) or by a (semi-)automatic detection, registration and segmentation. These algorithms have to deal with a large inter-subject variability of the anatomical structure and need to create reproducible results [32]. In principle, model-based approaches are effective for detection [33] and segmentation [13, 14]. If a segmentation yields only binary masks

separating the structures, algorithms such as *Growing and Adaptive Shapes* can be applied creating a surface grid where each point is comparable throughout the population [11]. Grey value comparison is usually used to measure the quantity of fat, water, and—application-specific—the iron content (liver) or the distribution of grey and white brain tissue.

Morphometric features are derived to allow for statistical comparison of the tissue, which incorporates mostly positions, diameters, volumes and relative distances and alignment to other structures.

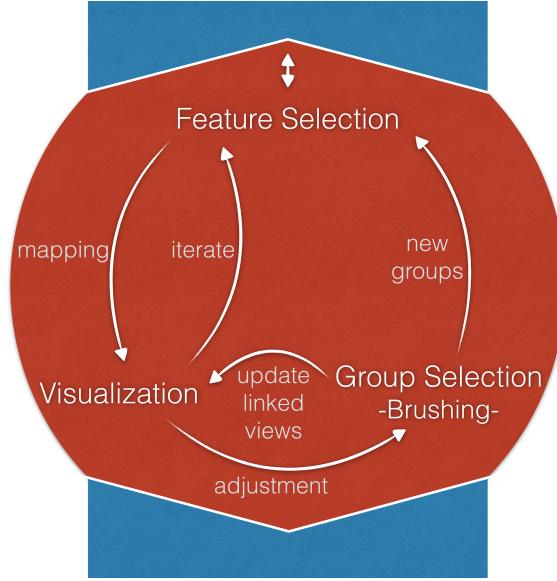


Fig. 2. Detailed *IVA Loop* as extension from Figure 1. Usually starting with a selection of a feature of interest (user-driven or via data mining techniques), the data are mapped using a visualization technique appropriate for the selected data types. The data are visualized in the range and domain space, which can then be brushed, yielding new groups, to be investigated using further features. Note that adjacent steps are directly connected via feedback loops allowing for an iterative refinement and giving as much freedom to the user as possible.

4.2 Analysis Workflow

Our proposed *IVA* workflow consists of three major steps as illustrated in Figure 2: Feature selection, visualization and brushing. A hypothesis-driven analysis usually starts with the selection of features, or shape groups derived from a shape-based clustering. *Hypothesis generation* with focus on image data starts with a shape-based clustering. [ToDo: Multivariate analysis using the adjacency matrix may have also be explained here, depending on the Application chapter.](#) The feature is mapped using an automatically chosen visualization appropriate for its data type (described in detail in the following section). The visualization techniques have to combine both image-and non-image data in order to set domain and range data in relation to each other. In our system, the visualization can either be brushed or new features can be added to the analysis. Brushing methods are subdivided using the previously described *IVA* patterns. Brushed regions are treated like features, as they divide the subject space just like categorical features. Selecting features also triggers a *multivariate analysis* using contingency values (described in the following section) to highlight features associated with selected features. A sample workflow using interaction and visualization techniques described in the next section can be seen in Figure 3.

5 SYSTEM DESIGN AND VISUALIZATION TECHNIQUES

The suitability of an interaction and visualization technique for epidemiological data depends on its ability to intuitively compare multiple data features at once while highlighting new associations. The

methods have to use domain language and reflect the routines that epidemiologists take in order to be incorporated into their research. Visual evaluations of data are therefore as important as methods allowing for numerical data analysis. In the following sections we present the different parts of our proposed *IVA* system for image-based cohort study analysis.

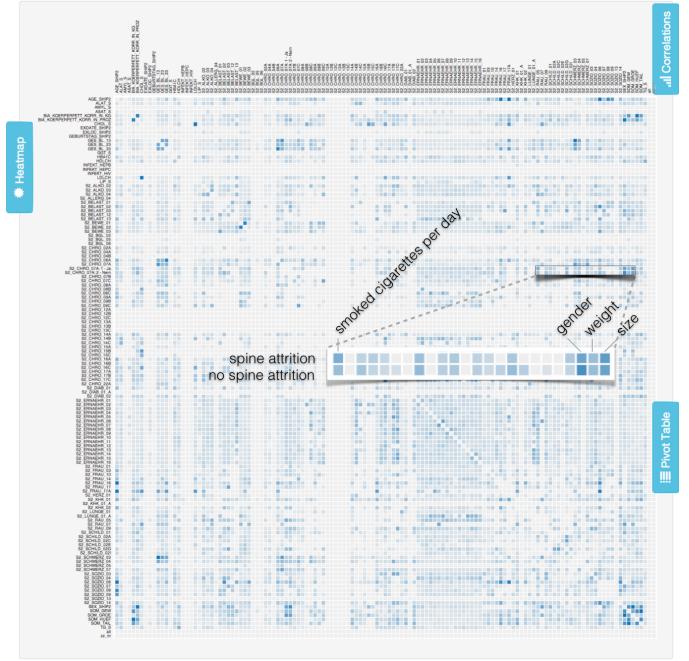


Fig. 4. Adjacency matrix of 129 features (127 data set variables, 2 cluster results) with a grand total of 16,641 combinations. Similarity is calculated using the Cramér's V contingency value. Color brightness encodes association strength. Mouse-over an entry highlights the feature names for better readability. The enlarged excerpt shows associations for shape clusters of subjects with and without diagnosed spine attrition, which show associations between gender, weight, body height and smoking behavior.

5.1 ToDo Rename.System Structure and Design

Early it became clear that we have to rely a lot on online communication with our clinical partners due to their busy schedule and our large spatial distance towards each other. Hence we build our system using modern web technologies, described in detail in Section 5.2. By running the prototypes on server machines, software exchange became as easy as sharing a website link, giving us the opportunity to include the clinical experts easily in the development process. Incorporating the *IVA* workflow for image-centric cohort study data requires *overview visualizations* as well as *multivariate visualizations*, which bring image-derived information in context to non-image features.

The focus on web technologies is not without tradeoffs. Classical UI elements, such as the menu bar or custom right click menus are technically possible, but not common in this domain. In favor of a clean layout, we designed the system without such components. Since the previously described *IVA* workflow allows for many different ways to analyze the data, we tried to make the interface as minimalistic and discrete as possible, treating the resulting space as *canvas* for the data. We divide the workspace into four major parts, as illustrated in Figure 3 and 4.

- The *sidebar*, which contains all epidemiological features. As cluster results group features like categorical features and are part of the sidebar as well (Fig. 3 (a)).
- The *canvas* holding all visualizations. Elements can be added, arranged, resized and removed freely (Fig. 3 (b)).

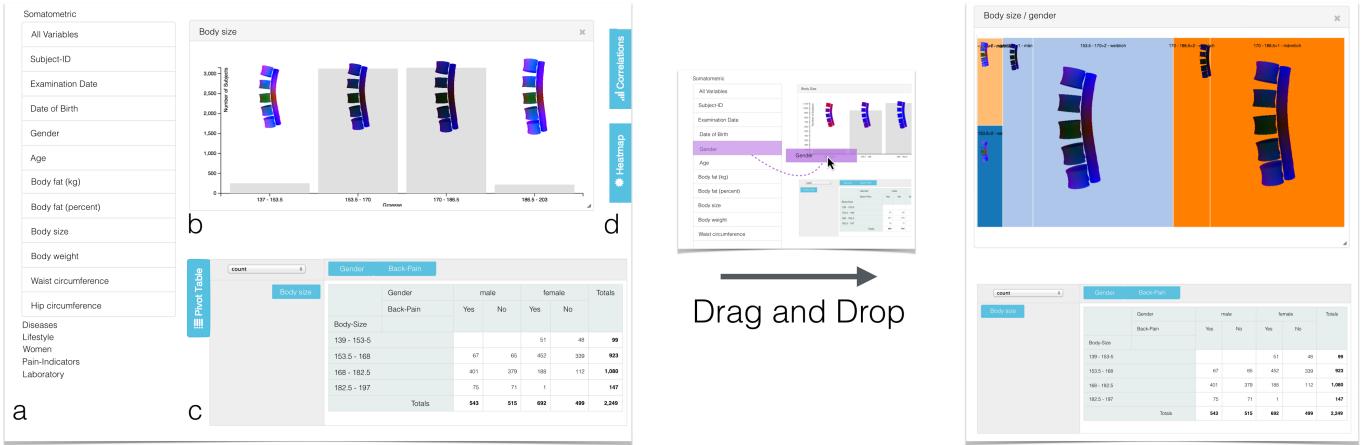


Fig. 3. **Add:** Explanation of the color coding, **ToDo:** Selektion in Abbildung 3 einbauen und daran Selektion erkennen (Left) Screenshot from the front-end, which is divided as follows: (a) The sidebar containing all features as well as the groups defined in the analysis process; (b) the canvas area where features can be added via drag and drop and the visualization is chosen automatically according to the data type; (c) the interactive pivot table showing the exact numbers for each displayed feature combination. The data displayed is used to analyze the lumbar spine. Features can be added freely on the canvas via drag and drop. Dropping the *gender* parameter on the already plotted *body size* container creates a mosaic plot combining both features (right). **3d-Koordinatensystem Kodierung durch Farbe (grün, rot, nuancen)**

- The interactive *pivot table* gives detailed numerical information of the features in the canvas view. This view on the data is familiar to epidemiologists (Fig. 3 (c)).
- The *contingency view* depicts relations for features in the canvas in an adjacency matrix (Fig. 4) and a *contingency list*.

System Layout We experimented with several layouts. The initial idea was to make all components freely arrange- and resizable on a large *canvas* area. This idea was soon dropped since domain experts reported a cluttered workspace, which required a lot of scrolling. A well working configuration was the introduction of separate panes for the adjacency matrix, pivot table and sidebar, displayed with a mouse click on the corresponding button and sliding on top of the *canvas* (Figure 3 shows the system with reeled-out pivot table pane). All user-generated visualizations follow are part of the *canvas* and can be arranged freely.

Sidebar. The *sidebar* is the only visible element when the system is opened. An overview of all features is presented in a sidebar where they are categorized into different types, such as somatometric (measurements of the dimensions of the human body), disease- or lifestyle-related, pain indicators and laboratory data. It also contains subject groups defined by automated shape clustering. Groups are treated like features since they act the same by dividing the subject space into labeled categories. Variables can be dragged from the sidebar into the canvas area for a *feature localization* which works as follows.

Adaptive Feature Visualization. The visualization type is, inspired by GPLOMS [21], dynamically chosen based on the feature types and number. Categorical data are either mapped to bar charts (single features) or mosaic plots (multiple features). Figure 3 describes this dynamic adjustment. Continuous data can be visualized using scatterplots (two features) or parallel coordinates (multiple features), but in epidemiology, this data type is usually categorized into ordinal groups of *equal size*. Since the number of categories often depends on the hypothesis, the discretization steps can be adapted dynamically. Too many groups potentially generate sparse bins not suited for statistical evaluation. Not enough groups overgeneralize information. Adaptive discretization is an option, but imposes possible overfitting to the data. Conclusions based on statistical relationships derived from groups already biased by feature distribution are heavily influenced by the used discretization. Therefore we follow the convention to use bins of equal size.

Following Tufte's concept of *small multiples* [39], information derived from the medical image data are directly incorporated into the plot by including color-coded mean shapes for each manifestation (Figure 3 (b)). The 3D plots can be navigated using standard mouse input, the camera is synchronized between all views to enable direct comparison. The distance from a group mean shape is mapped to the global mean using color allows to assess local shape changes (Figure 3). This is an important information to the epidemiologist, because until now they were not able rapidly validate shape differences based on non-image features. If a feature is dropped on an existing plot, the visualization changes dynamically to make them comparable (Figure 3 (right)).

Subject groups can be brushed via a double-click on it's representative in the visualizations. Holding down the shift key allows to select multiple manifestations at once. Brushed groups act as reference for the shape-visualization, calculating distances based on the mean-shape of the brushed selection. The share of subjects of this sub group is linked to all other views (Figure 3 (left)). If the user for example selects all female subjects in a visualization of gender distribution, all other displayed meshes are color coded with their distance to the female mean and the share of female subjects is highlighted in the information visualization.

Pivot Tables. Pivot tables are frequently used to present the data in epidemiological publications. Epidemiologists are used to analyze groups based on table representations. Thus, we decided to introduce an interactive pivot table. These tables clearly convey the subject count in each group (see Figure 3 (c)). However, they quickly get confusing and cluttered when they are divided into many subgroups. We tackled this problem by making the order and number of displayed variables adaptable. This also applies to the assignment of row or column variables. Another way to avoid clutter is the user-driven selection of displayed variables. To allow better comparison with respect to features, the values of each cell can also be displayed as percentage of the feature represented of either the row or column.

Automated Feature Suggestion using a Contingency Matrix. As previously discussed, highlighting potentially interesting values in the data set is one major benefit of the IVA-powered approach and belongs to the *multivariate analysis* pattern, analyzing features outside the shape space. Turkay et al. [40] used the approach to calculate various key figures based on the distribution functions of each feature derived from the image data. Since the majority of our data are categorical features, we have to employ different solutions. The *Cramér's V* contingency coefficient can be used to calculate coherences between

categorical variables [9]. It is based on *Pearson's X²* distribution test [31], which uses contingency tables holding the counts of subjects for all possible manifestations of two variables. *Cramér's V* is defined as:

$$V = \sqrt{\frac{X^2}{N(k-1)}}, \quad (1)$$

where X^2 equals *Pearson's chi squared*, N is the total number of observations and k is either the row or column count, depending on which one is lower. V assumes values between 0, meaning that two variables are completely independent, and 1 indicating they are the same. *Cramér's V* is always positive and does therefore not allow statements about dependency direction.

It shares the same restrictions as *Pearson's X²*. The expected counts in the contingency table have to be larger than 5 for 80% of the entries and no expected value must be smaller than one [8]. Some manifestations and feature combinations, which are only exposed by small subject groups, cannot be assessed with this technique. They cannot be included into the epidemiological analysis, since statistical validation needs a minimum count to be seen as valid. The contingency matrix highlights correlations between all features. This aims to highlight features possibly associated with the focused hypothesis as well as trigger generation of new hypotheses. Contingency is visualized using an interactive adjacency matrix with association power mapped to color brightness. The distinction whether an association is a confounder or an effect, depends on the context defined by the hypothesis and is a decision to be made by the domain expert. The contingency matrix visualization is an appropriate *overview* visualization, something the epidemiological community lacks and is in great need of.

Contingency Pane Dropping a feature into the canvas area adds an entry for each manifestation of it to the *contingency matrix*. Testing sessions revealed that it was very tedious to open the matrix every time a new feature is added. The result was the *contingency pane*, a table containing correlating features for the last added visualizations in descending order of the *Cramér's V* value. *Contingency pane* entries can be dragged and dropped into the *canvas area* just like features in the *sidebar*.

Initialization and Clustering Using feature suggestion allows to initialize the system with a set of potentially interesting visualizations. After testing and domain expert feedback we dropped this idea. Reasons for this are twofold. Very often are high correlations obvious, such as gender with menstrual status. Also, we observed that the variables of interest is dependent on the specialization of the domain expert (explained in detail in Section 6).

Subject clustering is triggered automatically for each variable after it was added to the canvas by the user. A status indicator at the bottom of the screen keeps the user informed about the pending clustering result, since the process can take up to ten seconds. Clustering result are listed in their own category in the *sidebar*.

5.2 Implementation

In this section we discuss how we implemented the presented methods using free, open web standards. To provide a fast communication loop between method development and expert input, we decided to rely on modern web technologies. In addition to the obvious advantages of web technologies, the following aspects are crucial for our work:

- The client-server structure allows for employing heavy computation on a capable server machine and transferring results to the client.
- Since image data for several thousand subjects requires hundreds of gigabytes disk space, it can remain safely on the server and elements can be transferred on demand. High confidentiality standards of the data can be met by restricting access via an account system.



Fig. 5. The front-end solution (left) uses state-of-the-art web technologies such as *HTML5/CSS3*, *WebGL* and *SVG* to display the data. The *NodeJS* based back-end (right) stores all image and non-image data and transfers it to connected clients. All computation-heavy operations, such as calculation of mean shapes or distances, as well as statistical processing are done by the server to keep hardware requirements of client systems low. Client-server communication is accomplished via the *Websocket* protocol.

- Recent developments in *WebGL* applications running in browsers with near-native performance push the development further into the web resulting in many open source libraries, which are well documented, rich in examples and driven by active communities. We use *WebGL* for rendering shape information.

These advantages do not come without drawbacks. Sophisticated libraries/languages, such as the *Visualization Toolkit* (VTK)⁴ or *R*⁵ for statistics, are either not available at all or only accessible through complex client-server systems. Therefore, many standard methods had to be written from scratch. The user needs a fully updated modern web browser, which supports these new technologies, such as *WebGL*.

The back-end is written using *NodeJS*⁶, which is based on the Google V8 Javascript runtime environment. Due to its event-driven non-blocking I/O model it is fast and does not freeze in case of heavy workload like mesh calculation.

Non-image data for all subjects including the data dictionary is stored in a JSON file on the server. Image data are available as raw DICOM files as well as meshes representing segmentation masks of anatomical structures, which can be used to compare subjects. The requested data is transmitted when a client connects. The server performs heavy statistical tasks, such as calculation of *Cramér's V* values for all feature combinations in order to keep the computation time on the client as low as possible.

The front-end is created using *Bootstrap*⁷ as foundation for the layout and basic UI elements using *HTML5*, *CSS3* and *Javascript*. Information visualizations such as scatterplots and bar charts are created using the popular *Data-Driven Documents* (*D3.js*) library [3], which works well for attaching data to visible elements like vector graphics and provides powerful transformation and mapping tools. The pivot table implementation was adapted using *PivotTable.js*⁸. *Three.js*⁹ allows GPU-accelerated data rendering using *WebGL*. Communication between client and server is enabled via the *WebSockets* protocol. Since our clustering algorithms are written in *MatLab*¹⁰, we had to access them using the *NodeJS* server. We accomplished this by converting it to a parameterized standalone console application that is spawned by *NodeJS* on client request and then reads the result from the console standard out and returns it in a proper format to the client. All parameter-steered console applications can be incorporated in this context.

⁴Developed by Kitware Inc; <http://vtk.org>

⁵Open Source; <http://r-project.org>

⁶Developed by Joyent Inc, <http://nodejs.org>

⁷Developed by Twitter, <http://getbootstrap.com>

⁸Developed by Nicolas Kruchten, <http://nicolas.kruchten.com/pivottable>

⁹Originally developed by Ricardo Cabello, <http://threejs.org>

¹⁰Owned by The MathWorks, <http://mathworks.com>

6 APPLICATION

This section describes how the presented *IVA* workflow is used in the epidemiological application. We conducted a qualitative case study with two domain experts on a data set compiled to analyze lower back pain. This is one of the most common reasons for an adult to see a physician in the Western civilization [41]. Epidemiological analysis of lumbar back pain, such as the work of Harreby et al. [17], is largely focused on non-image information. In comparable studies, only a few shape-related features are included [27]. Determining risk factors in this area can lead to [12]:

- a better understanding of effects of preventive measures such as occupational health and safety regulations,
- prognostic features for diagnosis and treatment of lumbar back pain, and
- determination of particularly affected risk groups.

Characterizing the healthy aging process of the spine is a long-term goal for determining age-normalized probabilities for spine-related diseases by incorporating individual risk factors.

6.1 The Lumbar Spine Data Set

We divide the data set into image and non-image data. There are 127 features describing diagnosed diseases, lifestyle factors, women-specific factors, pain indicators, laboratory values and somatometric features for 6,753 subjects (4,420 from SHIP-Trend-0 and 2,333 from SHIP-2). Since data acquisition protocols between these two cohorts are identical, the features between the two cohorts are comparable. The data contains 30 metric, 7 nominal, 29 ordinal and 62 dichotomous epidemiological features. Somatometric features include measures of the human body, such as body height, weight and body fat percentage as well as gender. These measures are reliable and complete. Other features, such as pain indicators or lifestyle indicators such as physical activity are more subjective and less reliable. There are also features missing for each subject, such as features building upon each other (e.g. Do you have high blood pressure? Which medication is prescribed against it?). Therefore some manifestations are sparsely populated, which makes statistical evaluation challenging.

The MRI image data was acquired for each subject on a 1.5 Tesla scanner (Magnetom Avanto; Siemens Medical Solutions, Erlangen, Germany) by four trained technicians in a standardized way. The spine protocol consisted of a sagittal T1-weighted turbo-spin-echo sequence (676 / 12 [repetition time msec / echo time msec]; 150° flip angle; 500 mm field of view; 1.1 × 1.1 × 4.0 mm voxels) and a sagittal T2-weighted turbo-spin-echo sequence (3760 / 106 [repetition time msec / echo time msec]; 180° flip angle; 500 mm field of view; 1.1 × 1.1 × 4.0 mm voxels) [19].

6.1.1 Data Preprocessing

The data are preprocessed as described in Section 4.1.

Non-Image Data. To ensure a fast and easy data access outside of statistical processors like SPSS, the data was exported to the JSON file format. Since it lacks export methods for data dictionaries, we used SPSS to export our data to the 'SAS v9+' format, which writes out the data labels and saved the data values as non-labeled CSV. A short script combined both data sources to a JSON file. The data types had to be transferred manually. Each feature is stored as an object containing:

- the data as array of values—categorical data and error codes are stored using IDs,
- the data type (continuous, nominal, ordinal, dichotomous),

- a detailed description of the feature, and
- the data dictionary translating value or error IDs to the actual values.

Continuous variables are discretized to allow for Cramér's *V* contingency coefficient assessment. In consultation with the collaborating epidemiologists, we set the number of groups to five, to allow for contingency assessment.

Image Data. The lumbar spine was detected in the image data using a hierarchical finite element method by Rak et al. [33]. This semi-automatic method requires the user to initialize the tetrahedron-based finite element models (FEM) with a click on the L3 vertebra. Two user-defined landmarks on the top and bottom of the L3 vertebra are used to obtain an initial height estimation of the model. It uses a weighted sum of T1- and T2-weighted MR images to detect the lumbar spine shape. The registered models capture resilient information about the shape of the lumbar spine canal as well as the position of the L1-L5 vertebrae [24]. Due to incorrect initialization, strongly deformed spines, contrast differences and artifacts, the model was not able to detect lumbar spines for all subjects. We obtained and worked with 2,540 tetrahedron models of the lumbar spine. For clustering purposes, we extracted the centerline of the lumbar spine canal, which captures information about lordosis and scoliosis (the medical terms for spine curvature [24]).

6.1.2 Shape Visualization and Clustering

The tetrahedron-based detection model described in Section 6.1.1 consists of corresponding grid points for each structure instance. This allows for calculation of shape distance and similarity. This information is used to calculate mean shapes as described in Section 5.

Shape distance is mapped onto color. For dichotomous variables, the color codes distances between mean shapes of the two groups, for variables with more than two manifestations it encodes the distance to the global mean shape of all subjects (see Figure ??).

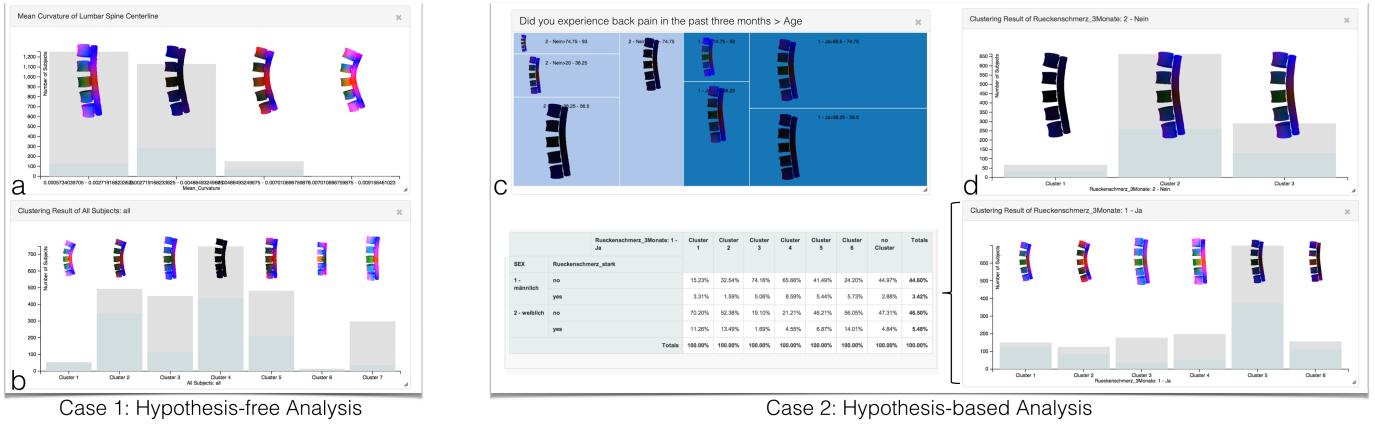
Shape-based clustering is carried out via agglomerative hierarchical clustering of the spine canal centerlines (recall Section 6.1.1 and [24]). Since the “correct” number of clusters in a given group is unknown, an estimate is computed by means of the knee/elbow method [34]. The method has proven to produce sound results on a preliminary data set and was able to reproduce textbook knowledge [24].

6.2 Participants, Setup and Procedure

Guided by the work of Lam et al [26], we conducted an investigation of *Visual Data Analysis and Reasoning* (VDAR). This approach aims to characterize the ability of a system to explore data, discover knowledge, generate hypotheses and help formulating decisions. Since it is hard to quantify these outcomes, Lam et al. suggest case studies for the VDAR analysis by applying the think aloud protocol to understand the domain experts observations, inferences and conclusions when using the system.

Our participants are two epidemiological domain experts who also co-authored this publication. Both of them are practicing physicians with focus on epidemiological research. One is a specialist in internal medicine and also the head of the SHIP, the second a radiologist and responsible for the MRI data acquisition of the study. Years of experience

Setup. To bypass the large geographical distance between us and the domain experts and to take their busy schedule into account, the case study was done completely web-based. The experts accessed the prototype by simply entering the website link into their browser of choice, no exchange of software was required. To make the user input comprehensible, we observed their monitor via screen sharing. Communication was enabled via webcam supported voice over ip. The total setup time including installing the screen sharing application was about five minutes. Video-recordings of the sessions allowed a detailed evaluation afterwards.



Case 1: Hypothesis-free Analysis

Case 2: Hypothesis-based Analysis

Fig. 6. Various case study results. (a) Mean curvature of lumbar spine canal plotted against the mean shape of 58-74 years old female subjects (light-blue bars). Note the high amount of this subject group relative to the total count in the third group compared. The last group only contains four outlier-subjects. (b) Clustering of all subjects yields seven groups, where Cluster 4 assembles the mean. The light blue bars indicate the share of females in the group. (c) A mosaic plot mapping age against dichotomous questionnaire answer to “Did you experience back pain in the past three months?”. (d) Clustering result of “Did you experience back pain in the past three months?” Yes/no with female share in each group. The pivot table shows how many subjects with strong back pain are in each cluster for answer “Yes”. Subjects in Cluster 1, 2 and 6 report strong back pain more often than subjects in other clusters.

Procedure. At first, we controlled mouse and keyboard of the participants PC and demonstrated the basic functionalities of the system. As they understood the concepts, we handed over the mouse and keyboard control and only observed from this point on. The epidemiologists were given two tasks: one hypothesis-free analysis of the data and one starting with an assumption. For each case we conducted one analysis with each expert.

6.3 Case 1: Hypothesis-free Analysis

Analyzing the data set without prior hypothesis requires a starting point giving an overview over the data first [35]. With our tool, there are two ways to achieve this. Performing a *local investigation* by clustering all subjects or use the contingency matrix for a *multivariate analysis*. The latter was chosen first by both domain experts. Before, they were not able to look at all variables in the context of each other. To cite one expert, the contingency matrix “illuminates the data black box”, making it possible to look at the data unbiased from assumptions.

Analysis 1. The radiologist was looking for correlations with shape-related features in the data, finding that they correlate with *leg pain*, *age*, *body height* and *hormone replacement therapy status*. Due to the dense mapping of information in the contingency matrix it was suggested to make this visualization full screen.

After this initial overview, the expert performed a *multivariate analysis* by introducing features, such as *age*, *waist circumference*, *weight* or *lumbar spine canal curvature* as bar charts views into the canvas area and selected subgroups to see how they are distributed and if they could observe unusual behavior in the mean-shapes. This pointed out problems with the used categorization method splitting numerical variables into equally-sized ordinal bins. If a feature contains outliers, such as *waist circumference* (e.g. by subjects with morbid obesity) this approach leads to sparse categories, making it hard to calculate associations. The proposed expert solution for this is categorization using quantiles/quartiles and is described in detail in Section 6.5.

A *multivariate analysis* using the *Cramér’s V* contingency values for subjects with strong lumbar spine curvature showed, that these subjects are primarily females between 58-74 years who also report pain radiating from their back into other body regions Figure 6(a).

Analysis 2. The second expert started also with a *multivariate analysis* using the contingency matrix to analyze non-image features, such as age-associated parameters like *income*, *blood fat values* or

number of born children, but found no associations of interest. Therefore, he applied the *local investigation* pattern by a shape grouping step using shape-based clustering via dragging the *All subjects* from the sidebar into the canvas area, triggering the shape clustering. The results for this step can be seen in Figure 6 (b).

Cluster 4 represents subjects with average shape. Other shapes differ with respect to size, such as cluster 2, 3, 7, where the last one and cluster 5 also represent a more straight spine, which is usual for subjects with larger body size. Cluster 1 and 6 contain outliers, characterized by their unusual shape and small number. To get an overview over the suggested features, the user opened the contingency pane to perform a *multivariate analysis* by looking at *Cramér’s V* contingency values of all clusters, revealing a strong correlation with *gender* and *body size*. Therefore another *multivariate analysis* was carried out by the feature *gender* to the canvas from the contingency pane and selecting all female subjects (Figure 6 (b)). Cluster 1 contained almost only female subjects. Contingency values for this cluster revealed correlations with *leg fatigue*, *physically heavy work*, *body weight*, *dyspnoea* and *headache intensity*. Since it is a pain indicator, headache was of special interest and was further investigated by incorporating a pivot table setting *headache intensity* in relation to cluster affiliation. It was found, that cluster 1 subjects report heavy headaches more frequently than other subjects.

The experts emphasized the importance of methods providing an overview over the data for hypothesis generation. With the presented IVA approaches they were quickly able to confirm medical knowledge as well as elaborate new hypothesis. We observed, that even though the visualization may treat every feature equally, the domain experts are more likely interested in features they are familiar with and have personal clinical experience with.

6.4 Case 2: Hypothesis-based Analysis

If the user proposes a hypothesis about a relation between a non-image feature regarding shape, the workflow slightly differs from the hypothesis-free analysis. The starting point follows the *feature investigation* pattern, where a feature of interest is selected by dragging it into the canvas area and viewing the subject’s distribution as well as their shape differences.

Analysis 1. Hypothesis: “*Back pain is associated with age and lumbar spine shape*”. To validate this hypothesis, a *feature investigation* was performed by introducing the dichotomous feature “*Did you*

experience back pain in the last three month?" together with the age as mosaic plot by dropping both features on the canvas area. The expert was not able to observe the expected effect in the visualization. Reasons for this are twofold. For one, age influences the lumbar spine shape, as seen in Figure 6 (c), while the differences between subjects with and without back pain are small. The major difference seen in the visualization are therefore related to the age feature, masking differences via the back pain parameter. The second explanation is the commonality of back pain in our society. As seen in Figure 6 (c), subjects reporting back pain are even the majority, which makes it difficult to extract parameters which reliably describe back pain. A *multivariate analysis* using the contingency table showed a strong association between *back pain* with *gender* and *body height*. *Body height* was explained to be a confounder for *gender*, since female subjects are smaller on average than male subjects. The analysis solely based on shape accentuated differences in *gender*, not *back pain*. The next step is to use a variable describing the strength of back pain and only analyzing strong back pain.

The epidemiologists pointed out that they would like to see a more intuitive and fast way to select subgroups based on different features to make full use of the analysis capabilities, as discussed in Section 6.5.

Analysis 2. Hypothesis: “*Back pain is related to lumbar spine deformation*”. The previously discussed analysis puts the suitability of the lumbar spine segmentation for analyzing back pain into question, leading to this analysis. Therefore the feature “*Did you experience back pain in the past three months?*” is dropped into the canvas area. Then, the features derived from the clustering results for both subject groups are dropped into the canvas area as well (Figure 6 (d)). While the clustering algorithm finds only three homogenous clusters close to the mean shape for subjects reporting no back pain. The cluster analysis for back pain yields diverse clusters with various pathological shape classes. Cluster 5 represents most of the subjects and is very similar to the global mean shape. Cluster 1 and 2 present a *hyperlordosis*, a strong curvature of the lumbar spine, while Cluster 3 and 4 present a more straight shape. A *multivariate analysis* using the pivot tables put gender and strong back pain in context to cluster affiliation (Figure 6 (d)). It shows, that subjects in Cluster 1, 2 and 6 reported strong back pain, while at the same time they also have a considerably higher share of females. To check for unusual correlations, the expert used the *Cramér's V* contingency table. It depicted strong association with *body fat*, *body weight* and *blood pressure* (Cluster 1) *alcohol consumption* and *attentiveness disorder* (Cluster 2), and *amount of sleep* (Cluster 6). For the experts, these observations are a starting point for a number of new hypothesis about possible relationships, for example association between overweight and Cluster 1.

In summary it can be stated that hypothesis-driven analysis leads to hypothesis generation by design of the tool. It is not suited and intended to statistically validate hypothesis like statistical processors. It rather triggers the analysis of potentially associated features with a pathology of interest.

6.5 Further Feedback and Lessons Learned

Both domain experts concluded positively over the benefits of the *IVA* approach when analyzing image-based cohort study data. One expert emphasized the way the image data is included into information visualizations which comes much more natural to her due to her background in radiology. Great potential is also seen in communicating insights efficiently using the presented visualizations (1.08).

Multivariate analysis is most important for hypothesis generation. Both experts emphasized the potential of the *multivariate analysis* capabilities of the adjacency matrix for getting insight into a large amount of features simultaneously. It is also useful to verify established but still controversial risk factors, such as the metabolic syndrome for coronary heart disease and whether the data set provides even more suitable risk factors (39.32). Creating adjacency matrices

for sub groups, such as different age bins can help to characterize the aging process by deriving age-specific risk factors.

Multivariate analysis can also be improved by more ways of brushing the data as well as creating sub groups for comparison as a result of the hypothesis-driven analysis case. To small feature ranges resulting in very sparse groups could potentially hinder the ability to calculate statistical resilient measures, since they require a minimum amount of subjects exhibiting the selected feature ranges.

Segmentation quality is crucial. The radiologist pointed out the unusual strong similarity of the L3 vertebrae throughout the population. The medical explanation is that it represents an angular point of curvature of the lumbar spine. A seconds explanation is the use of the L3 vertebra is initialization point of the lumbar spine model (00.54).

The experts also emphasized that associations related to shape strongly depend on the segmentation quality. The lumbar spine model used in this case study captures deformation of the spine canal well, but lacks precise definition in vertebrae height and shape. Since deformation of the spine canal is the last stage of pathological lumbar spine deformation and is preceded by vertebrae deformation, the system would strongly benefit from more precise segmentation results capturing these prior changes. For the visual comparison itself, the radiologist proposed an abstraction the representation them into representative landmarks, such as centers of the vertebrae and cardinal points of the lumbar spine canal.

Usage of different categorizations depending on expected outcome. When categorizing numerical variables into equal groups it is possible to create very sparse categories due to outliers, for example when analyzing body weight. These outliers are only of high interest for finding pathological subjects. The experts therefore suggested two modes of the tool. The outlier mode still creates categories of equal size, producing sparse categories for outliers. Balanced categories are created on the second mode, which uses quartiles or quintiles to set borders between categories.

Web technologies are well suited for rapid feedback. The web-based approach for both implementing the prototype as well as getting feedback via voice over ip conference calls worked very well. Since the software does not need to be compiled, small changes can even be made on the fly during a testing session. The large data base associated with image-based epidemiological data remains on the server machine and has not to be moved tediously using external hard disks. This approach is well suited for the *VDAR* approach to assess user thought processes using the think aloud technique.

Since it is an entirely different approach to their usual workflows which are heavily influenced by statistical tools, it is to highlight that they had fun browsing through the data set in the presented manner and to dig out new information about the data. This underlines the accessibility of *IVA* tools, which amplifies the natural curiosity of the domain experts in complex medical feature associations.

7 SUMMARY AND CONCLUSION

In this paper, we presented an *IVA* framework for the analysis of complex image-centric epidemiological data. Hence, the framework allows for both, hypothesis-driven analysis and hypothesis generation. The visualization of multivariate data using connected views and different views allows to get fast visual feedback about subject groups. Brushing and linking makes the data tangible and adaptable to formulated hypotheses. The use of pivot tables is familiar to epidemiologists while embracing the power of interactive adjustment of the shown features. The automatic suggestion of correlations using contingency methods like *Cramér's V* triggers *hypothesis generation* by highlighting features potentially overlooked by the experts. Shape-based clustering assesses the variability of an anatomical structure in the context of non-image features such as disease indicators or lifestyle factors.

Our clinical partners from the *SHIP* are for the first time able to assess shape information of the lumbar spine and its influence to diseases. This needs to be adjusted to fit the new Application section

Findings from analyzing lumbar back pain using the *IVA* approach range from deriving shape-based groups of subjects to detailed description of features potentially associated with the disease, such as waist circumference, thyroid nodules and blood fat values. Other somatometric features, such as BMI, are not as influential as expected.

A number of improvements is left open for future work, e.g.,

- shape brushing methods to intuitively query subjects using image data,
- the inclusion of more statistical methods and views that are familiar to the epidemiologists (odds ratios, box plots), or
- adapt the shape visualization to explore other organ data with different variance type (such as texture of liver or white/gray matter distribution in the brain).

As the number of image-centric cohort studies, participating subjects, gathered features and imaging modalities rises, and advances towards comparability between cohort studies are made, the gap between data complexity and analyzability increases. Our work focuses on closing this gap, allowing the domain experts to dig deep into the data and potentially obtain unexpected findings. We believe that web technologies pave the way to analyze this data in a convenient way. They allow a fast exchange between users and developers and employ many different devices. Visual analysis shows to be a promising way to clear the view on complex epidemiological data to uncover its secrets.

ACKNOWLEDGMENTS

SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grant no. 03ZIK012), the Ministry of Cultural Affairs as well as the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Whole-body MR imaging was supported by a joint grant from Siemens Healthcare, Erlangen, Germany and the Federal State of Mecklenburg-Vorpommern. The University of Greifswald is a member of the Centre of Knowledge Interchange program of the Siemens AG. This work was supported by the DFG Priority Program 1335: Scalable Visual Analytics.

REFERENCES

- [1] R. S. Ahima and M. A. Lazar. The health risk of obesity—better metrics imperative. *Science*, 341(6148):856–858, 2013.
- [2] J. Blaas, C. Botha, and F. Post. Interactive visualization of multi-field medical data using linked physical and feature-space views. *Proceedings of EuroVis' 07*, pages 123–130, 2007.
- [3] M. Bostock, V. Ogievetsky, and J. Heer. D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2301–2309, 2011.
- [4] A. Buja, J. McDonald, J. Michalak, and W. Stuetzle. Interactive data visualization using focusing and linking. In *Proceedings of IEEE Visualization*, pages 156–163, 419, 1991.
- [5] S. Busking, C. Botha, and F. Post. Dynamic Multi-View Exploration of Shape Spaces. *Computer Graphics Forum*, 29(3):973–982, 2010.
- [6] J. J. Caban, P. Rheingans, and T. Yoo. An Evaluation of Visualization Techniques to Illustrate Statistical Deformation Models. *Computer Graphics Forum*, 30(3):821–830, 2011.
- [7] K. K. Chui, J. B. Wenger, S. A. Cohen, and E. N. Naumova. Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. *PloS one*, 6(2), 2011.
- [8] W. G. Cochran. The χ^2 test of goodness of fit. *The Annals of Mathematical Statistics*, pages 315–345, 1952.
- [9] H. Cramér. *Mathematical methods of statistics*, volume 9. Princeton university press, 1946.
- [10] X. Dai and M. Gahegan. Visualization based approach for exploration of health data and risk factors. In *Proc. of the International Conference on GeoComputation. University of Michigan, USA*, volume 31, 2005.
- [11] L. Ferrarini, H. Olofsen, W. M. Palm, M. A. Van Buchem, J. H. Reiber, and F. Admiraal-Behloul. Games: growing and adaptive meshes for fully automatic shape modeling and analysis. *Medical image analysis*, 11(3):302–314, 2007.
- [12] R. H. Fletcher, S. W. Fletcher, and G. S. Fletcher. *Clinical epidemiology: the essentials*. Lippincott Williams & Wilkins, 2012.
- [13] O. Gloger, J. Kühn, A. Stanski, H. Völzke, and R. Puls. A fully automatic three-step liver segmentation method on lda-based probability maps for multiple contrast mr images. *Magnetic Resonance Imaging*, 28(6):882–897, 2010.
- [14] O. Gloger, K. D. Tönnies, V. Liebscher, B. Kugelmann, R. Laqua, and H. Völzke. Prior shape level set segmentation on multistep generated probability maps of mr datasets for fully automatic kidney parenchyma volumetry. *IEEE Transactions on Medical Imaging*, 31(2):312–325, 2012.
- [15] D. Greig, B. Porteous, and A. H. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 271–279, 1989.
- [16] D. L. Gresh, B. E. Rogowitz, R. L. Winslow, D. F. Scollan, and C. K. Yung. WEAVE: a system for visually linking 3-D and statistical visualizations applied to cardiac simulation and measurement data. In *Proc. of IEEE Visualization*, pages 489–492, 2000.
- [17] M. Harreby, J. Kjer, G. Hesselsøe, and K. Neergaard. Epidemiological aspects and risk factors for low back pain in 38-year-old men and women: a 25-year prospective cohort study of 640 school children. *European Spine Journal*, 5(5):312–318, 1996.
- [18] K. Hegenscheid, J. Kuhn, H. Völzke, R. Biffar, N. Hosten, and R. Puls. Whole-Body Magnetic Resonance Imaging of Healthy Volunteers: Pilot Study Results from the Population-Based SHIP Study. *Proc. of RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*, 181(08):748–759, 2009.
- [19] K. Hegenscheid, R. Seipel, C. O. Schmidt, H. Völzke, J.-P. Kühn, R. Biffar, H. K. Kroemer, N. Hosten, and R. Puls. Potentially relevant incidental findings on research whole-body MRI in the general adult population: frequencies and management. *European Radiology*, 23(3):816–826, 2013.
- [20] M. Hermann, A. C. Schunke, T. Schultz, and R. Klein. A visual analytics approach to study anatomic covariation. In *IEEE PacificVis 2014*, Mar. 2014.
- [21] J.-F. Im, M. J. McGuffin, and R. Leung. Gplom: The generalized plot matrix for visualizing multidimensional multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2606–2614, 2013.
- [22] S. A. Julius and M. A. Mullee. Confounding and simpson's paradox. *BMJ*, 309(6967):1480–1481, 12 1994.
- [23] P. Klemm, L. Frauenstein, D. Perlich, K. Hegenscheid, H. Völzke, and B. Preim. Clustering Socio-demographic and Medical Attribute Data in Cohort Studies. In *Bildverarbeitung für die Medizin (BVM)*, pages 180–185, 2014.
- [24] P. Klemm, K. Lawonn, M. Rak, B. Preim, K. Tönnies, K. Hegenscheid, H. Völzke, and S. Oeltze. Visualization and Analysis of Lumbar Spine Canal Variability in Cohort Study Data. In J. F. Michael Bronstein and K. Hormann, editors, *VMV 2013 - Vision, Modeling, Visualization*, pages 121–128, 2013.
- [25] Z. Konyha, K. Matkovic, and H. Hauser. Interactive visual analysis in engineering: A survey, Apr. 2009.
- [26] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *Transactions on Visualization and Computer Graphics*, 18(9):1520–1536, 2012.
- [27] M. Lang-Tapia, V. España-Romero, J. Anelo, and M. J. Castillo. Differences on spinal curvature in standing position by gender, age and weight status using a noninvasive method. *Journal of applied biomechanics*, 27(2), 2011.
- [28] U. Niemann, H. Völzke, J.-P. Khn, and M. Spiliopoulou. Learning and inspecting classification rules from longitudinal epidemiological data to identify predictive features on hepatic steatosis. *Expert Systems with Applications*, 2014.
- [29] S. Oeltze, H. Doleisch, H. Hauser, P. Muigg, and B. Preim. Interactive Visual Analysis of Perfusion Data. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 13(6):1392–1399, 2007.
- [30] S. Oeltze, H. Hauser, and J. Kehrer. Interactive visual analysis of scientific data, 2013. Half Day Tutorial at IEEE VIS, Seattle, WA, U.S.
- [31] K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.
- [32] B. Preim, P. Klemm, H. Hauser, K. Hegenscheid, S. Oeltze, K. Toennies, and H. Völzke. *Visualization in Medicine and Life Sciences III*, chapter Visual Analytics of Image-Centric Cohort Studies in Epidemiology. Springer, 2014.
- [33] M. Rak, K. Engel, and K. Toennies. Closed-form hierarchical finite element models for part-based object detection. In *VMV 2013 - Vision, Modeling, Visualization*, pages 137–144, 2013.
- [34] S. Salvador and P. Chan. Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. In *Proc. of Tools with Artificial Intelligence. ICTAI*, pages 576 – 584, 2004.
- [35] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc of Visual Languages*, pages 336–343. IEEE, 1996.
- [36] M. Steenwijk, J. Milles, M. van Buchem, J. H. C. Reiber, and C. Botha. Integrated Visual Analysis for Heterogeneous Datasets in Cohort Studies. *Proc. of IEEE VisWeek Workshop on Visual Analytics in Health Care*, 2010.
- [37] S. Thew, A. Sutcliffe, R. Procter, O. de Brujin, J. McNaught, C. C. Venters, and I. Buchan. Requirements Engineering for e-Science: Experiences in Epidemiology. *Software, IEEE*, 26(1):80–87, 2009.
- [38] J. J. Thomas and K. A. Cook. *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005.
- [39] E. R. Tufte and P. Graves-Morris. *The visual display of quantitative information*, volume 2. CT: Graphics Press, 1983.
- [40] C. Turkay, A. Lundervold, A. J. Lundervold, and H. Hauser. Hypothesis generation by interactive visual exploration of heterogeneous medical data. In *Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data*, pages 1–12. Springer, 2013.
- [41] M. van Tulder, B. Koes, and C. Bombardier. Low back pain. *Best Practice & Research Clinical Rheumatology*, 16(5):761 – 775, 2002.
- [42] H. Völzke, D. Alte, C. Schmidt, et al. Cohort Profile: The Study of Health in Pomerania. *International Journal of Epidemiology*, 40(2):294–307, Mar. 2011.
- [43] H. Wiley. Hypothesis-free? no such thing. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2008.
- [44] Z. Zhang, D. Gotz, and A. Perer. Interactive visual patient cohort analysis. In *Proc. of IEEE VisWeek Workshop on Visual Analytics in Health Care*, 2012.