Biomedical Informatics in Chronic Infectious and Inflammatory Disease Research: Periodontitis as a Case Study

Bruno G. Loos\textsuperscript{a,1}, Marja L. Laine\textsuperscript{a}, Carol van der Palen\textsuperscript{b}, Francis Lessmann\textsuperscript{b}, Arie-Jan van Winkelhoff\textsuperscript{a}, Ubele van der Velden\textsuperscript{a}, J. Bart A. Crusius\textsuperscript{c}, Servaas A. Morré\textsuperscript{c}, A. Salvador Peña\textsuperscript{a}, Anthony J. Brookes\textsuperscript{d}, Antonio Sousa Pereira\textsuperscript{e}, George de Moor\textsuperscript{f}, Victor Maojog\textsuperscript{g}, Fernando Martin Sanchez\textsuperscript{h}, and the INFOBIOMED consortium

\textsuperscript{a}Dept. of Periodontology, Academic Centre for Dentistry Amsterdam, The Netherlands
\textsuperscript{b}Dept. of Informatics, VU University Medical Center, Amsterdam, The Netherlands
\textsuperscript{c}Dept. of Pathology, VU University Medical Center, Amsterdam, The Netherlands
\textsuperscript{d}Dept. of Genetics, University of Leicester, UK and Center for Genomics and Bioinformatics, Karolinska Institute, Sweden
\textsuperscript{e}University of Aveiro, DET/IEETA, Aveiro, Portugal
\textsuperscript{f}Custodix N.V., Merelbeke, Belgium,
\textsuperscript{g}Polytechnic University of Madrid, Spain
\textsuperscript{h}Carlos III Health Institute, Madrid, Spain

Abstract. The management and control of chronic inflammatory diseases are a major challenge for the health systems of all the European countries. Data collection using new approaches in the field of Biomedical Informatics (BMI) gives the opportunity to build large datasets that can be analyzed. The development and implication of new data warehouse tools that become available for clinicians and researchers, will allow use of newly developed techniques in data mining and visualization. New tools in these fields will help in meta-learning. This opens up new steps towards understanding the (relative) roles of various proposed etiologic factors. The current case study, i.e. periodontitis, will serve as a model for complex chronic inflammatory diseases. A periodontitis data warehouse (PDW) has been structured with the INFOBIOMED network and allows an integrative research approach, where the new field of BMI will contribute to new insights in chronic inflammatory and infectious diseases.

Keywords: Biomedical Informatics, data warehouse, chronic inflammatory diseases, periodontitis, INFOBIOMED.

1. Introduction

Common diseases, such as cardiovascular disease, cancer, metabolic imbalances, and chronic inflammatory illnesses impose a major drain on society, both in terms of financial and human costs. They are thus priority targets for current biomedical research. Chronic inflammatory diseases have a complex pathogenesis and a

\textsuperscript{1}Corresponding Author: Bruno G. Loos, Department of Periodontology, ACTA, Louwesweg 1, 1066 EA Amsterdam, The Netherlands, e-mail: B.G.Loos@acta.nl, Fax: +31 20 518 8512.
multifactorial etiology, involving complex interactions between multiple genetic loci, infectious agents and environmental (behavioral) factors such as diet, smoking habits and physical exercise [1]. The role of infectious agents (bacteria and viruses) in particular, is gaining increased interest and recognition [2-4]. The general paradigm is that certain individuals are genetically more susceptible than others to the environmental risk factors [1], and it is these persons that are more likely to succumb to the illness. Within the European network INFOBIOMED, the disease “periodontitis” is designated as case study for the study into integration of bioinformatics and medical informatics into the new field of bioinformatics (BMI).

2. Periodontitis

Periodontitis is a chronic inflammatory disease of the supporting tissues of the teeth. If left untreated, teeth may show exposed root surfaces, in conjunction with red, swollen gums that easily bleed. Dental radiographs reveal periodontal (alveolar) bone loss around the teeth due to a inflammation process (Figure 1); teeth will become mobile, migrate, and will eventually exfoliate. Patients with periodontitis experience problems with chewing due to tooth mobility and loss of teeth; they have bad breath and suffer from important subjective and objective esthetic problems. Dental practitioners provide labor-intensive diagnostic and treatment sessions to periodontitis patients, including periodontal surgery.

![Figure 1. Example of a dental radiograph of tooth 31 and surrounding hard tissue structures showing important landmarks: the cemento-enamel junction (CEJ), the apex and the periodontal bone level (BL) on the mesial site of the tooth. Periodontal bone levels in the healthy situation are located 1-2 mm apical from the CEJ; in this example there is about 55% periodontal bone lost on the mesial surface of tooth 31.](image)

It is accepted that the etiology of periodontitis is multifactorial [5] involving:

a) Infectious component. Of the several hundred oral bacterial species known, a limited number of species are recognized as periodontal pathogens and they have been identified as important markers of progressive disease. These include: *Porphyromonas gingivalis*, *Prevotella intermedia*, *Tannerella forsythensis*, *Fusobacterium nucleatum*,...
Actinobacillus actinomycetemcomitans and the spirochaetal species Treponema denticola [6]. The complexity of the subgingival biofilm as well as the interplay of various species associated with periodontitis is also exemplified by the proposal that infective and protective bacterial species are organized in microbial complexes [7, 8]. In this way a red complex includes the most infective species P. gingivalis, T. denticola and T. forsythensis, while on the other end of the spectrum the yellow complex is consisting of Streptococcal species.

b) Genetic susceptibility. Genetic susceptibility for chronic periodontitis is deduced from family studies and studies in twins. From twin studies it has been estimated that 38% to 82% of population variance in periodontal disease expression, may be attributed to genetic factors [9]. Chronic adult periodontitis was estimated to have 50% heritability, which was unaltered following adjustments for behavioral variables including smoking [10]. The search for genetic markers and candidate disease-modifying genes in periodontitis has recently been receiving considerable attention. In particular, single nucleotide polymorphisms (SNP’s) in genes encoding molecules of the host defense system have been targeted (immunogenetics) [11]. Analogous to other complex inflammatory diseases, it is clear that periodontitis is a polygenic disorder.

c) Environmental factors. Smoking is currently accepted as the most important environmental risk factor in periodontitis [12-14]. Periodontitis susceptible subjects suffer from a more progressive form of periodontitis and smokers have more severe periodontal breakdown than nonsmoker patients [15, 16]. Furthermore, smokers with periodontitis show a less favorable response to periodontal treatment, both for non-surgical and surgical approaches [17]. Other factors that have been proposed as environmental risk factors for periodontitis include diet and stress [18-22]. However these factors are not yet firmly established; more research is needed into their role in the etiology and pathophysiology of periodontitis.

3. Challenges for Biomedical Informatics in the study of periodontitis

Informatics is increasingly pertinent in medicine. First, a discipline called Medical Informatics (MI) has been traditionally involved in the research and development of informatics-based methods and tools for medical and epidemiological research and patient care and management. Bio-Informatics (BI) on the other hand, is a discipline developed around the Human Genome Project, which handles genomic, proteomic and any other biological research data. MI and BI can play an important role in our understanding of complex diseases. Both MI and BI will contribute to gain new insights in complex diseases, will help to better understand the pathophysiology of inflammatory processes, and will help to develop new classification schemes, design new treatment strategies and ultimately preventive and pro-active measures. Biomedical Informatics (BMI) is the emerging discipline that aims to put these two worlds together so that the discovery and creation of novel diagnostic and therapeutic methods is fostered [23].
4. Biomedical Informatics needs in periodontitis

Periodontitis is a complex disease that requires the integration and analysis of multiple sources of data and therefore offers a remarkable set of challenges to BMI (Figure 2). The main challenges for biomedical informatics within the development of a periodontitis data warehouse (PDW) are encountered in the following areas.

Figure 2. Technological aspects for BMI in periodontitis

a) Data Collection. Supporting the whole structure we have the enabling technologies oriented towards facilitating a solution for a PDW. The PDW includes parameters that are collected in daily clinical practice that respond to five distinct levels: genetic, environmental (behavioral), infectious, disease phenotype and intermediate phenotype (individual risk factors).

b) Annotating and standardizing data. Research data, including protocols, primary data collected from the phenotype and intermediate phenotype, information coming from external sources and procedures of reduction and analysis, are the essential components of scientific progress. All the mentioned information is being stored in a PDW. Attention should be given to annotating and documenting computerized information to facilitate detailed review of data that should be treated comparably.

c) Assessing the phenotype data. We developed an ad-hoc Dental Image Analysis (DIA) tool to collect and analyze phenotypic information from dental x-rays. Although there is no standard applied, it is recommended storing radiographic data according to
the DICOM standard. An advantage for this standard is that patient demography data can be coupled to the images and the chances for mislabeling are diminished. Radiographic data in DICOM can be accessed by any computer system, which may be an important advantage in the studies of complex diseases, where multiple databases exist.

d) Ensuring privacy. For the studies into periodontitis, patient data are generated. These include disease and intermediate phenotype, genetic data, data on the infectious component and environmental factors. From different sides an integrated data set must be generated and is to be used by dentists and dental researchers, as well as researchers in the field of informatics. To ensure the privacy of the patient an accepted coding and anonymization and/or pseudonymization which may be reversible is applied. In the interactive (reversible) pseudonymization model from Custodix [24], a (transparent) intermediary privacy protection engine is put between the users and the database web server.

e) Mining and visualization data. Data mining is one of the steps in the iterative process of knowledge discovery and it is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data and it consist of applying data analysis and discovery (learning) algorithms that produce a particular enumeration of patterns (or models) over the data [25]. The PDW has large quantities of heterogeneous data collected from diverse sources mentioned above and this information was stored by specific categories so it can be more easily retrieved, interpreted, and sorted by users. Data mining tools and knowledge discovery techniques are now applied to the PDW.

5. Data mining and visualization in the pilot study on periodontitis

   a) Clustering. Clustering is often one of the first steps in medical areas. It identifies groups of related records that can be used as a starting point for exploring further relationships. Clustering techniques are frequently used to discover structure or similarities in data [26, 27]. The classification of periodontitis is still a matter of debate due to its complexity and the subjective criteria used by clinicians; the heterogeneity in diagnostics prevents an objective and general evaluation of the disease.

   b) Feature subjects selection (FSS). Feature subjects selection (FSS) reduces the number of statistical features in high-dimensional classification problems [28]. In the study of this complex disease that involves multiple variables, FSS could help clinicians by finding subsets of features that are most relevant. Models constructed in the subspace of the selected features tend to generalize new data better than classifiers trained on the entire feature set.

   c) Classification. For good classification, there is a need to find relationships that can partition the data into disjoint groups. Classification, perhaps the most commonly applied data mining technique, employs a set of labeled examples to develop a model that can classify the unlabeled population [29, 30]. Example-based data mining methods such as nearest neighbor classification, regression algorithms and case-based reasoning are also examples of solutions to data classification problems. For periodontitis health care professionals, this type of data mining technique would be important in diagnostic decision making since by these probabilistic methods, the clinicians might obtain the probabilities that a given patient is suffering from a certain state of periodontitis.
Visualization. In the data mining process, visualization tools help to explore data before modeling and to verify the results of other data mining techniques. Visualization tools are particularly useful for detecting patterns found in only small areas of the overall data.

6. Conclusions

The management and control of chronic inflammatory diseases are a major challenge for the health systems of all the European countries. Since the etiology is multifactorial and polygenic, a classification of these diseases based on molecular biology, genetic information, phenotype information and modern imaging systems, are bound to provide the clinicians with new tools to identify the different subgroups with different prognosis and help in the development of individualized medicine. Based on these classifications different strategies can be developed for additional diagnosis, therapy and patient management. In this new scenario, old and new treatment strategies can be tested in randomized genomic-based clinical trials using the newly proposed classifications (diagnostic systems) and patient subgroups.

The PDW has used new approaches in the field of BMI. The development and implication of new data warehouse tools that become available for clinicians and researchers, will allow them to perform newly developed techniques in data mining and visualization. New tools in these fields will help in meta-learning. This opens up new steps towards understanding the (relative) roles of various proposed etiologic factors. The current case study, i.e. periodontitis, will serve as a model for other chronic inflammatory diseases. Within the INFOBIOMED consortium, periodontitis is a pilot project and was selected because it provides an example of a chronic, complex disease, where microbiological and genetic factors interact with environmental factors in determining the disease susceptibility, the severity of inflammation and the evolution of the process. Focus on a limited number of etiological factors is likely to lead to the discovery of new clinico-genetic associations and scientific hypotheses, which can be tested. Similarly, new advances in the pathophysiology of periodontitis may provide experience and directions for many other multifactorial/complex diseases.

Acknowledgement

This work was supported by the 6th R&D Framework Programme of the European Commission via the project entitled “Structuring European Biomedical Informatics & Support Individualized Healthcare”, the INFOBIOMED Network of Excellence (IST 2002 507585). The information presented does not necessarily reflect the policy of the CEU.

References


